Consensus Pays: Examining Cross-Platform Rating Divergence

Student Name: Jente-Floor Olthof

Student Number: 666364

Supervisor: Dr. Christian Handke

Master Cultural Economics and Entrepreneurship

Erasmus School of History, Culture and Communication

Erasmus University Rotterdam

Master Thesis

June 13, 2025

Table of Contents

Abstract

Prior studies have examined the effects of UGC on consumer behavior (Chevalier & Mayzlin, 2006), film success in terms of revenue (Duan et al., 2008), or model bias in reputation systems (Wu et al., 2018). Yet, none have isolated the comparative roles of national versus international platforms in this context. Against this backdrop, the present study compares the French platform Allociné with the global benchmark IMDb and sets out to: (i) compare user-generated reviews between a national (Allociné) and international (IMDb) platform, (ii) trace the economic stakes of any divergence, and (iii) the cultural conditions under which those gaps widen or shrink.

To conduct the research, a matched, cross-sectional dataset of 2,296 French and US feature films released between 2010 and 2020 was gathered. After harmonizing all rating scales, I carried out paired t-tests, Bland-Altman plots, ordinary least squares and quantile regressions, and a film-level linear mixed-effects model with random intercepts.

Results indicate that platform context matters. The same films often receive different ratings across the two platforms. This gap between platform ratings narrows only for the very best-reviewed films, and critics magnify these platform differences further. Higher intrinsic quality and larger grosses foster convergence, halving the gap in the upper quartile, whereas every additional point of spectator disagreement is associated with a 36 % drop in worldwide gross for lower-earning films. Yet, against the theoretical premise of cultural proximity theory, there is virtually no domestic uplift on Allociné according to the spectator ratings for French movies (an uplift of 0.09). Rather, according to the mixed-effects results, U.S. films receive a 0.58-point lift on Allociné compared to a -0.64-point more negative rating it gets on IMDb. This means that there is virtually no home bias on Allociné, but rather a preference for U.S. films.

Taken together, the findings portray rating sites as distinct attention markets whose biases can erase or create commercial value and contrast cultural proximity theory. Shrinking cross-platform gaps may therefore offer distributors a route to higher revenues and provide scholars with preliminary findings on how platform ratings diverge.

KEYWORDS: user-generated content; platform effects; cultural proximity; film ratings; box-office revenue

Word count: 13,097

1. Introduction

"Who the hell wants to hear actors talk? They're silent the way they should be!" is a famous statement by H. M. Warner (one of the Warner Bros. founders) in 1927. Within two years later, *The Jazz Singer* (1927), a synchronized film, had exceeded all box-office records. The silent era was over and an entire production infrastructure had to be rebuilt around synchronized sound (Filmsite, n.d.). To this day, the film industry is an industry in which technological advancements have played a pivotal role (Eliashberg et al., 2006).

A comparable inflection point, though more gradual, has been the digital revolution. Unlike the leap from silents to sound, the digital revolution is a continuous process of changes. The way in which films are made, released, and watched has changed as a result of the internet and digital platforms. Especially, since films are a type of good that can create a lot of "buzz" even before they are released. This attribute is mainly caused by the experience good characteristic of films. Consumers face quality uncertainty and therefore tend to frequently consider reviews, affecting whether they will choose to see the film. This is a core component of the dynamic between film and consumer since films face a rather brief theatrical run of weeks, not months. The first week's box office is therefore vital in indicating both the film's eventual success as well as the distribution support it will receive in the weeks that follow (Chakravarty et al., 2010).

One of the main strategies to sway audiences nowadays is to employ viral marketing, which is rooted in encouraging consumers to spread a positive or persuasive marketing message that they receive through the internet. At the core of this strategy, we find electronic word-of-mouth (eWOM) (Moore, 2015). eWOM is best defined as "any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet" (Hennig-Thurau et al., 2004, p.39). This statement is usually communicated through any written or numerical form and can be shared considerably more quickly and effectively than traditional word-of-mouth since the communication does not need to take place in the same location and does not necessitate any type of social connection, it is completely anonymous (Cramer & Kunz, 2025).

Yet, it is not only of interest what consumers say but also where they say it, as these potential intermediaries hold a lot of power over the information that is spread. User-generated content (UGC) for films is spread across platforms such as the Internet Movie

Database (IMDb), Allociné, Letterboxd, and Rotten Tomatoes. These platforms have become certifiers by ranking titles, recommendations, and most of all, monetizing any traffic across the platform through ads. Not only do they affect consumer's decisions, but they also affect the success and visibility of films. Where once the critics and distributors were the main certifiers of the industry, now the consumers largely contribute to the narrative surrounding a film. In platform-economic terms: they operate as multi-sided markets. Within this multi-sided market, filmmakers, advertisers, and viewers each play a role in increasing the platform's value for everyone involved. So-called network effects ensure that the user base keeps on growing, increasing the value even further (Rochet & Tirole, 2003). However, it must be considered that each platform operates in a distinct ecosystem of users. Up until now, no research has been conducted on these differences.

Against this backdrop, I will focus on IMDb and Allociné as they offer a particularly interesting comparison. IMDb began as a small hobby project by film enthusiasts in 1989. After rapidly growing as a company, IMDb was sold to Amazon in 1998. Amazon had a particular interest in IMDb as they wanted to break into the video and DVD market. IMDb already had a large user pool, which could potentially turn into customers. To attract those potential customers Amazon placed matching links to their own site under each movie (Alnefelt et al., 2007). This makes IMDb a key player in the digital platforms that distribute film-related user-generated content. Since it is also the first film platform to emerge on such a scale (and up until today the biggest) it is interesting to consider IMDb as a global player. Allociné, by contrast, was launched in 1988 as a French telephone kiosk service for specific show times by The Box Office Company – Webedia Group (Zhao, 2022). The company switched to an online environment in 1997 and has since then been the dominant national reference for film-related information (Frater, 2001). Both display average ratings and textual reviews but what largely differs is the user ecosystem. Because they overlap in functionality but diverge in audience and ownership, the two sites offer an ideal laboratory for testing whether the "same" movie carries the same reputation once it crosses a digital border.

Prior studies have examined the effects of UGC on consumer behavior (Chevalier & Mayzlin, 2006), film success in terms of revenue (Duan et al., 2008), or model bias in reputation systems (Wu et al., 2018). Yet, none have isolated the comparative roles of national versus international platforms in this context, which makes this research unique. Addressing this omission directly contributes to debates surrounding UGC, platform governance, and cultural proximity.

Therefore, the purpose of this study is threefold: (i) to compare user-generated reviews between a national (Allociné) and international (IMDb) platform, (ii) to trace the economic stakes of any divergence, and (iii) the cultural conditions under which those gaps widen or shrink. Accordingly, this study pursues three closely linked research questions:

How large and in what direction do IMDb and Allociné user ratings diverge for the same films?

To what extent do those cross-platform rating gaps translate into differences in worldwide theatrical revenue?

Is the size or direction of that rating gap influenced by cultural proximity?

As a long-time user of film platforms that distribute UGC, I started to notice that I would not consider a cinema release if the average ratings were below a certain threshold. The power of the information on the platform, therefore, seems enormous, at least to me as a user. This sparked a particular interest in whether these scores would differ systematically by geography or platform design since this would impact entire market segments.

From a scientific standpoint, the study fills a clear gap: no prior work has compared how national versus global rating ecosystems reshape film reputations. By doing so, I contribute to platform economics theory and cultural proximity theory. The social relevance can be identified from the perspective of the consumer. Knowing that the same film can be perceived very differently on two mainstream sites is critical for informed consumption. The remainder of this study proceeds as follows: Chapter 2 will provide the theoretical framework in which the research is grounded; Chapter 3 describes the dataset and methodology; Chapter 4 reports on the empirical results; Chapter 5 will discuss these results in an overarching manner; and lastly, Chapter 6 summarizes the contributions of the study. By probing the rating differences across platforms, the study offers new perspectives on the role of digital platforms in the film economy.

2. Theoretical Framework

The way consumers discover and evaluate films has fundamentally changed due to the rise of digital platforms, or multi-sided markets, such as Allociné and IMDb. These platforms function as intermediaries by collecting and distributing User-Generated Content (UGC) that consists of likes, reviews, and comments (Rochet & Tirole, 2003). The collection and distribution of this UGC would traditionally play an important role in mitigating information asymmetries, where one agent has more or better information than the other (Akerlof,1970), for consumers of films. Akerlof's (1970) lemon problem illustrates that markets with asymmetrical information often lead to inefficiencies where low-quality goods, "lemons", drive high-quality goods out of the market because buyers tend to pay a price that reflects the average quality of the market. However, I would argue this is not the case in the film industry since both consumers and producers face fundamental uncertainties and do not know the exact quality of a particular film. Neither side knows audience reception ex-ante; hence the problem shifts from a one-sided to a joint uncertainty.

This problem is reinforced by the 'experience good' characteristic of films. Consumers can only determine the quality of the product after consumption (Nelson, 1970). However, this analysis constrains itself from traditional markets since it was written in 1970. The film industry has transformed due to digitalization and is nothing like a traditional market. The internet has become a powerful source for distributing information among agents. The new internet economy came with the rise of online reputation systems that now range from restaurants to travel destinations.

However, the idea that perfect information sharing through the internet promotes the optimal outcome is erroneous. This is not the case since the feedback effect, in which users are both consumers and producers of information, is a key feature of this digital market. Information that is produced today affects the choices of future agents, which in turn influences new information that will be generated (Kremer et al., 2014). In addition, ratings can significantly influence buyers' behavior and have a great influence on the success or failure of a product (Chevalier & Mayzlin, 2006).

2.2 Symmetrical Ignorance in the Film Industry

As argued before, Akerlof's (1970) information asymmetry theory does not hold for the film industry because quality is not only unknown to the customers but also to the producer. The producer of a film can hardly predict the audience's perception of the film's quality, which aligns with Caves' fundamental 'nobody knows' principle in the creative industries (Caves, 2003). The concept is quite self-explanatory as it refers to the uncertainty both producers and consumers face. The problem is therefore better described as symmetrical ignorance than one-sided asymmetry. This is especially relevant in the film industry as it deals with huge sunk costs since almost none of the input can be reused and the total production is expensive in itself. Out of 10 major films, 6 to 7 are unprofitable, signaling the uncertainties producers face (Liu, 2006). In addition, the production of a film is dependent on a technological chain of projects. For example, if the shooting of the film is not finished, the post-production team cannot finalize the movie. Each contributor in this order, therefore, sinks money and effort before the next input supplier takes over (Caves, 2003).

However, the rise of digital platforms has reshaped the symmetrical ignorance in the film industry. Traditional certifiers (critics and distributors) are enhanced or replaced by new certifiers, such as reviews, ratings, and recommendation algorithms. These signals can help potential customers receive more varied and comprehensive information. Which in theory would allow them to make better-informed decisions. This shift has, however, introduced complexity in how consumers interact with this information. For example, consumers now often engage in a Bayesian learning process when exposed to digital reviews and ratings. This process assumes that consumers are not certain about the quality of a product and therefore update their quality expectations according to past experience and other factors such as marketing communication. In plain terms, consumers start with initial thoughts about a film's quality that are continuously altered by each new rating they perceive. As consumers accumulate more and more ratings from sources they trust, the running average they hold in their minds shifts accordingly (Zhao et al., 2012). During the Bayesian updating process, consumers assign varying weights to reviews. This depends on the credibility of the source, which is the measurement of the level of truth and validity of received messages (Wilson & Sherrell, 1993; Hsieh & Li, 2020). In addition, this spill-over effect to the next film applies most strongly to films that share attributes like the same genre, language, franchise, director, or lead actors because consumers treat these as belonging to a common quality distribution (Liu, 2006).

The signals that update the Bayesian process in each consumer may be significantly different between IMDb and Allociné since there is a difference in timing, volume, and perceived reliability. What may look like the same 7.4/10 score could carry a completely different informational weight depending on where it is posted. This difference produces the very gap this research addresses. By tracking identical films across IMDb and Allociné, the study can look deeply into whether platform-specific differences amplify symmetrical ignorance into persistent rating gaps. This explains why identical ratings can sway audiences differently across platforms.

The importance of how much information flows, rather than how positive that information is, is underscored by the work of Duan et al. (2008). Their research considered this endogenous nature of online user reviews in their study and looked at the interdependent relationship between online word-of-mouth and movie sales. The study analyzed weekly U.S. box-office data for major blockbusters in a 30-week timeframe between 2002 and early 2003. Volume and valence (the mean rating of a film) of Yahoo! Movies posts served as the main variables. They found that higher ratings do not lead to higher sales, rather, the number of posts on the film was significantly associated with movie sales. This highlights that the awareness effect of word-of-mouth is more critical than the persuasive content of a review. Digital platforms make use of this by amplifying this feedback loop, further reinforcing these cycles of informational cascades.

The debate, however, on the influence of viewer ratings on box-office revenue has been divided. A common measure throughout has been the mean rating as a quality indicator for box-office revenues, which this study will also employ. Some research has shown valence to be a predictor of sales, whilst others have found no such relationship (Duan et al., 2008; Chintagunta et al., 2010). The inconsistencies in results may occur because of the following reasons. The most prominent issue might be the fact that user ratings are not collected in a controlled setting, possibly causing random measurement errors and inconsistency between database sources. These inconsistencies are amplified further by the real-time characteristics of the ratings, every database is a snapshot in time (Baugher & Ramos, 2017). Yet, most of these studies remain platform- or language-bound. Baugher & Ramos (2017) conducted the only study that investigated cross-platform consistency. To do so, the study analyzed rating differences between Netflix and IMDb. This research chose to include Netflix as it is one of the most popular benchmarks used in Recommender Systems literature, mainly because rating volumes are often in the millions. This convenience masked three conceptual mismatches. First, the Netflix star score system was never a public, stand-alone quality signal. These ratings were input into the closed-loop recommendation engine and completely disappeared when Netflix replaced the star rating with a thumbs-up metric. This thumbs-up metric is designed to optimize engagement, not peer comparison, revealing Netflix's motives behind the recommendation system (Sims, 2017). Second, only paying subscribers can rate on Netflix which narrows the total voting pool, unlike IMDb. Third, both sites operate in English and therefore cater to a global-scale audience, erasing any local differences. These design features blur the interpretation of where the cross-platform divergence stems from.

A sharper analysis would, therefore, be to compare IMDb against a national, lower-volume platform such as Allociné. This does raise the measurement-error concern flagged in the literature but also allows for a closer look at these locational differences. Trading Netflix for Allociné allows for a smaller but transparent N that lets us disentangle platform differences.

The literature shows that there is a lot of incentive for digital platforms to interfere with what the potential customer sees. The feedback loop can be stimulated by giving certain films priority to what customers see. Distributors that want to promote their film can use relatively few advertisements to stimulate this feedback loop and thus generate a lot of attention for the film with relatively low costs. In addition, it focuses on the number of reviews and not the content. This is beneficial for companies that want to promote their films, as the impact of negative reviews has a larger magnitude than positive reviews (Chevalier & Mayzlin, 2006).

To conclude, the core of systematic ignorance persists, but it evolves in different ways. Stakeholders are still interested in the absolute quality but are also interested in whether and where cascades will ignite.

2.3 User-Generated Content

User-Generated Content (UGC) refers to media material where users contribute to media content and interact with other users. Due to the rise of UGC, platforms, too, have progressed to a central position. Digital platforms are multi-sided markets that monetize the created attention through traffic between all players involved. To determine what is UGC and what it is not, I will define it throughout the study according to the following criteria set by Naab and Sehl (2016):

1. *A certain amount of personal involvement characterizes UGC.* Users must contribute the content themselves. This means that activities of receiving or forwarding content are ineligible.
2. *UGC must be published.* It must be available to the public to facilitate a general conversation throughout society or within a specific group.

3. *UGC is created by non-professionals.* To be able to contrast everyday users with institutional producers.

2.3.1 Impact of UGC on Consumer Behavior

Movie rating platforms like Allociné and IMDb are examples where UGC plays a crucial role in distributing information on films, an industry where the role of critics is most prominent (Basuroy et al., 2003). Consumers perceive UGC as more reliable and authentic. Especially the volume of online posts affects consumers' decisions because this largely affects awareness (Duan et al., 2008). Interestingly, Duan et al. (2008) and Dellarocas et al. (2004) found that higher ratings do not necessarily lead to higher sales and that they are simply a measure of underlying word-of-mouth communication or electronic word-of-mouth (eWOM). When it comes to searching for information on the quality of the film, experts' opinions seem to matter more than eWOM for the revenue of movies (Basuroy et al., 2019). This is consistent with the findings of Rao et al. (2017) who found that the most important element a studio can include in an ad is a review by a trusted critic.

2.3.2 UGC on Allociné vs. IMDb

The cultural and geographical origins of UGC are additional factors that can influence consumer preference and behavior. Hecht & Gergle (2010) conducted research that showed that language barriers and cultural contexts lead to fragmentation in UGC. Out of 25 language versions of the same page on Wikipedia, they discovered vast variations in descriptors and general scope. This demonstrates that UGC on the same topic can vary significantly when compared locally and internationally.

This finding can be related to the differences between Allociné and IMDb. Allociné is an exclusively French platform, which means that non-French speakers are excluded and cannot contribute to the UGC. The information that is distributed on the platform is rooted in French culture and linguistics. In addition to that, French culture has a long history in the film industry. The so-called Nouvelle Vague movement in the 50s and 60s is a characteristic of French cinematography. This movement revolutionized cinematography and introduced new narrative and stylistic techniques that are still being used to this day (Sellier, 2010).

2.4 Cultural Proximity

The term cultural proximity, coined by Straubhaar (1991), refers to the prediction that audiences will prefer national media content that is culturally familiar to them. Factors determining this are language, religion, values, and ethnicity. These factors make the content more relatable to the specific audience. Straubhaar's (2003) follow-up research found that the proportion of people whose identity is deeply globalized is quite small, and traditional layers of identity (local, national, and international) are still the strongest. Therefore, audiences favored cultural-linguistic regional and national content. Of course, this does not hold for the present day. With the rise of digitalization, borders have become vaguer when it comes to the consumption of goods and services. Nonetheless, it is interesting to consider the fact that consumers prefer media content that is culturally familiar to them in the context of the research question.

2.4.1 Cultural Proximity and UGC

Platforms can create connected digital spaces. Within these digital spaces, actors can co-exist and work together. The platforms enable interactions, information flows, and network formation through digital proximity (Panori, 2024). However, the UGC is still generated by users from a certain geographical location. This means that their interpretation and production of content are affected by cultural proximity.

Additionally, research by Forman, Ghose, and Wiesenfeld (2008) examines how the disclosure of reviewer identity influences the perceived helpfulness of online reviews. The study finds that reviews accompanied by identity-descriptive information, such as real names or geographic locations, are perceived as more helpful by consumers. This suggests that consumers place greater trust in reviews from sources they can identify and relate to, thereby assigning more weight to such reviews in their decision-making processes.

2.5 Intrinsic Film Quality

An important problem that still needs to be addressed, referring back to Caves (2003), is the following question: How can we talk about quality when nobody knows? Most empirical work falls back on two observable proxies. The first proxy is based on recognition from critics. Research has linked quality to whether a film has won any major award such as an Oscar or César Award, and whether critics have positively rated the film (Basuroy et al., 2003; Simonton, 2004). The second proxy refers to crowd behavior signals where quality is treated as revealed utility. If diverse audiences consistently attribute high scores to a film, it must be delivering some form of value. This proxy will be used throughout the research as the key indicator of quality as it documents the stated preference of the users and can therefore serve as an indicator of whether higher ratings dampen or amplify the cross-platform rating gap. In addition, worldwide revenue will also be used as an indicator of intrinsic film quality as it documents the revealed preference of audiences.

2.6 Homogenization of Demand

Digital distribution was expected to unleash Anderson's "long tail" theory. The theory revolves around the conception that, because of the internet, consumers will drift away from homogenized hits. Rather, consumers focus on more niche products that are better tailored to their individual tastes. The internet creates this possibility as 'online' stores do not have to deal with limiting shelf space and can reach a bigger customer segment. In other words: the cost of distribution has decreased and access to niche content has increased. Anderson argues that the companies that will prosper, will therefore be those that address the niche products (Anderson, 2007). Even though research demonstrates that the long tail is lengthening, it is more likely that it is in fact extremely flat. Elberse (2013) argues that the products that populate this long tail are mostly a diversion for consumers and that big hits will persist to dominate revenues. Algorithmic recommenders have even further amplified this effect since they are designed to maximize engagement. These systems therefore try to herd users back to content that is already popular, reinforcing the concentration on hits.

This dynamic is relevant to the present study since it suggests that where a rating appears may matter most for films that are not included in the algorithm loop. If the recommender systems already mostly contribute to the popularity of blockbusters, small differences between IMDb and Allociné scores will not hugely impact their outcomes. By contrast, mid-tier and niche films rely more on the average rating to be able to surface in search results (Fleder & Hosanagar, 2009).

3. Research Methodology

This study employs a quantitative comparative research design to examine discrepancies in film ratings across two online platforms: IMDb and Allociné. The approach is cross-sectional, a matched sample of films at a single point in time will be analyzed. A comparison across platforms is needed because previous research has mainly focused on single-platform effects when looking at online movie ratings. This research will therefore look at whether ratings are consistent across platforms and whether they are consistent across a more globally oriented platform (IMDb) vs. a more nationally oriented platform (Allociné). When comparing the origin of users on each site there is a clear global vs. local distinction. I retrieved the user traffic analytics for each site through SimilarWeb, a common and appropriate source for establishing user origins in academia even though user traffic analytics are never completely accurate estimates. SimilarWeb uses three information sources to establish the traffic: people-panel data from volunteers that have a browser plug-in that records site visits, direct site feeds from some websites that share real-time numbers, and SimilarWeb's own scraping of data flows (Prantl & Prantl, 2018; Jansen et al., 2022). The user traffic share on IMDb is 34.73% United States users. This is followed by 10.34% from India, 8.61% from the United Kingdom, 4.85% from Canada, and 2.73% from Australia. For Allociné the share is 87.90% France, 3.79% Belgium, 2.21% Canada, and 1.48% Switzerland. These traffic analytics that were collected on Similarweb show that IMDb's visits are more dispersed across varying countries than Allociné where traffic is based primarily on France. Global versus local is a matter of degree and is not dichotomous. Within this research, I will treat 'global' as a multi-national reach and 'national' as a platform that is dominated by one language and one country. Though some spillover between the platforms will exist, like critic reviews, the origin of users is largely different, justifying the separate analysis. Previous research precisely misses the comparability between the platforms, limiting external validity. Baugher and Ramos (2017) note: "Online movie ratings are often used in studies of Internet word-of-mouth, but the cross-platform consistency of such ratings has not been well established".

Buagher and Ramos' (2017) analysis of Netflix vs. IMDb ratings demonstrated that mean user scores differ significantly and were moderately correlated ($r \approx 0.60$). This correlation implies that platform context matters because each platform has a different user base with different tastes, reinforcing the value of a cross-platform research study.

To test the cross-platform differences, the research adopts a quantitative cross-sectional comparative research design. The study is comparative in the sense that it contrasts two different platforms and cross-sectional in that all data represent a “snapshot” of ratings. This method ensures that time lags or movie reputation changes do not cause differences in ratings. All the ratings were collected in the same period of time, making sure that the total dataset was contemporaneous. A quantitative design is appropriate since measuring and comparing rating scores across platforms is the main interest of the study. To do so, the research has developed hypotheses informed by theory. Statistical tests will either support or refute these hypotheses. There are however some limitations to this design. Being cross-sectional, the study cannot capture evolving trends throughout time. In addition, it limits the study to not being able to determine why differences exist since this would require longitudinal data. This will be considered when presenting the results. Nonetheless, a cross-sectional comparative approach is most feasible and appropriate in the scope of this research. It will allow for a critical examination of cross-platform effects on film ratings.

3.1 Data Collection

The target population of this research is films that have been rated by users and press on both IMDb and Allociné. The data consists of film information and user ratings collected from two sources: the Internet Movie Database (IMDb) and Allociné. IMDb is a globally used platform where users rate (on a scale of 1 to 10, with single-point increments) and review films. Allociné is a French film platform where users rate on a scale of 0 to 5 stars, often with half-star increments, and reviews. In addition, both platforms provide a press rating score. Within IMDb this is called the metascore and this score ranges from 0-100. Allociné refers to a press score which is also set on a 0 to 5-star scale.

Data was collected on French and American productions for which all rating scores were available between 1st of January 2010 and 31st of December 2020 through APIs or by web scraping if necessary. The frame of 2010-2020 was chosen because it allows for the total ratings to develop at a minimum of 5 years. If the years 2020-2025 had been included this would have the possibility of skewing the data since these movies did not have the chance to ‘ripen’ on the platforms. By choosing for 2010-2020 I ensure contemporary relevance and the possibility for films to have accumulated a stable number of votes. The targeted population can therefore be described as follows: all theatre-released feature films that were produced in either France or the United States during the years 2010-2020 that appear on both the IMDb and Allociné databases will be included in the analysis. All the data are aggregated and nonpersonal ratings, meaning that no individual user privacy issues are involved. This resulted in a sample of n = 2.296 films (of which 517 were French productions, and 1.779 were U.S. productions) that were matched across IMDb and Allociné. The distribution across the years was remarkably even with around 200 titles per year. Every observation contained IMDb spectator rating (1-10), Allociné spectator rating (0-5 stars, converted to 1-10 scale), IMDb press rating (1-100, converted to 1-10 scale), Allociné press rating (0-5 stars, converted to 1-10 scale), number of votes, genre, budget, worldwide gross, year of release, runtime, language.

To ensure accuracy and consistency in the dataset I verified the collected film data by manually checking around 5% of the film ratings to see if the entries matched against the websites. This is especially important because the IMDb and Allociné datasets had to be merged together. Due to language differences many titles did not exactly match (e.g. *The Hangover* vs *Very Bad Trip*) in French. To address this problem, I applied fuzzy string-matching techniques because there is an absence of a common unique identifier for a film. The use of this algorithmic system allows for a matching of two items through different identifiers that might otherwise overlap with multiple items. For each IMDb title, the script searched for the best matching title according to the following identifiers: release year, duration, and director. If the match exceeded the threshold of 70% the two platform scores were linked. Films that could not be matched (e.g. niche movies that were only present on one of the two platforms) were dropped from the analysis. After the merging was complete each film had four key rating values: IMDb spectator and press rating, and Allociné spectator and press rating.

Another important consideration in sampling is the difference in voting population between the two platforms. Since IMDb is a globally used platform, and Allociné a more local one, voting numbers could differ largely. Therefore, vote counts were added to the dataset as a variable to ensure the ability to later assess possible disparities. If, for example, a film has far fewer votes on Allociné than IMDb, its average on Allociné might gravitate toward a few extreme ratings (that consists of fan ratings, which can happen with a film that gains a cult status after some time). In addition, it must be acknowledged that IMDb employs a weighted average algorithm to combat ballot stuffing (IMDb, n.d.). The calculation method of Allociné is not clearly stated but is presumed to be the straightforward mean of all user ratings. This means that there is a difference in the rating metrics between the two platforms. However, in this analysis, we treat both as comparable measures since IMDb's weighting measures are likely to be consistent amongst most of the films. This means that the difference in rating metrics would only be a problem if the weighted average disproportionately impacts certain types of films.

3.2 Data Cleaning

After finalizing the data collection, the combined set was cleaned and normalized to prepare for further analysis. The cleaning of the data involved checking for inconsistencies, errors, or missing values in the merged data. First, I checked for any duplicates in the dataset, which was not the case. 615 movies (26.8%) lacked the spectator rating score on either platform (for example the title: *The Evil Within*). I deliberately chose to still include these titles as the main focus of the research is based on spectator ratings. When using the press rating score I account for the smaller sample and the possible selection bias that accompanies it.

3.2.1 Rating scale alignment

A key normalization of the data was to align all rating scales. As stated before in section 3.1, the platforms applied different rating scores to different types of ratings. Therefore, the ratings needed to be transformed to a comparable scale. All ratings were converted to a 10-point scale. This was done with simple linear transformation: all Allociné ratings were multiplied by two and the IMDb press rating scores were divided by ten.

3.2.2 Budget and Worldwide Gross Normalization

Another important normalization included the budget variable. All budget information was collected through IMDb which meant that it was based on the dollar currency consistently. Because budget figures are highly skewed (a small number of blockbusters have very large budgets) I applied a logarithmic transformation for use in parametric analysis. This transformation helps normalize the distribution of the budget variable and reduces heteroscedasticity. This is especially helpful when running analyses like correlation and regressions because these assume a normal distribution.

3.3 Operationalization of the Variables

The operationalization of all variables was, wherever possible, aligned with how other scholars have defined similar constructs.

<table>
<tr><th>Variable</th><th>Key concepts</th><th>Notes</th></tr>
<tr><td>Platform-specific film rating – main dependent variable</td><td>eWOM and crowd assessment of quality (Chevalier & Mayzlin, 2006). This variable serves as the main quality indicator since it documents the stated preference of the audience.</td><td>All scores were converted to a 10-point rating scale, which is a common practice in platform studies (Baugher & Ramos, 2017).</td></tr>
<tr><td>Cultural Origin – independent variable</td><td>Cultural proximity theory – audiences prefer content that is already familiar to their culture (Straubhaar, 1991).</td><td></td></tr>
<tr><td>Rating gap – independent variable</td><td>This variable was computed by taking the absolute difference between the two rating scores. The absolute difference quantifies disagreement between raters instead of the mean signed difference which diagnoses systematic bias by repeatedly substracting one platform from the other (Bartok & Burzler, 2020; Moss, 2024).</td><td>| IMDb – Allociné |</td></tr>
<tr><td>Genre – control variable</td><td>This variable contains the primary genre listed on IMDb (e.g. drama, comedy, etc.). IMDb assigns multiple genres to a single film, to ensure that the first genre listed is the primary genre I checked multiple films to see if this ranking was not alphabetical. For example, for some films comedy was listed as the primary genre and action as the second. This confirmed that the ranking of genres was not done alphabetically but in subsequent order of importance. After assigning one primary genre to each film I created a list with all genres</td><td>Collapsing the genres may mask the differences between mainstream genres such as Action or Comedy and niche genres (e.g., Musical or Western)</td></tr>
</table>

	noted. This list contained twenty different genres with some only having one or two cases. Since the number of categories and for some of those the number of cases would not provide sound statistical analyses, I clustered the genres into 5 main categories. Following established practice in box-office research (e.g., De Vany, 2004; Litman, 1989; Redfern, 2011), twenty IMDb/Allociné genre labels were recoded into four macro-genres: Action/Adventure, Comedy, Drama/Romance, Thriller/Horror.	
Budget – control variable	Films that have a higher budget often have more resources available for production and marketing, both of these can inflate the financial outcome of the film.	ln(budget) to correct heavy right skew (min \$0.1 m; max \$350 m).
Worldwide Gross – dependent variable	Commercial performance both affects and is affected by perceived quality (word-of-mouth loop). Worldwide Gross serves as a quality indicator since it reveals the preference of the audience.	Worldwide theatrical box-office (USD), log-transformed for regression.
Year of Release – control variable	Accounts for temporal drift in rating norms and platform demographics.	

3.5 Limitations

Despite the care taken in data collection, some constraints narrow the overall generalisability of the findings. Firstly, the composition of the audience is uncertain and unknown. When looking at the national differences, this can only be observed at the film level, not at the rater level because the data can only adjust for the country of origin for the film and not the country of origin of each voter. This indicates that when different scores are assigned to the same film, the gap might be caused by the platforms themselves (different interfaces, languages, algorithms, etc.) or simply by the fact that the people who use the different platforms are very distinct. Part of the platform effect could therefore be a compositional effect because the research is not a controlled experiment in which identical audiences use both platforms. The possibility of either platform effect or different people effect can therefore hardly fully be untangled. However, despite this caveat, comparing IMDb and Allociné is still highly informative because moviegoers consult whichever platform they already use. In practice, these users do not see a demographically corrected rating. In that sense, the study captures the real information that users would encounter, which is valuable in itself. By flagging this issue, the study sets the agenda for possible follow-up work that could look into interface experiments. In sum, the absence of voter demographics prevents the perfect separation of the platform effect from the people effect. However, the chosen design provides the best available, real-world portrait of how ratings differ across a national and international platform on which future research can directly build.

Secondly, the rating scales differ across both platforms. Because IMDb uses a 10-point rating scale and Allociné a 5-star rating scale, the psychological quality signal might differ. These different interfaces can anchor votes on different spots on the scale. The same goes for the IMDb Metascore which is translated from 100-point scale to 10. A translated 6.5 on Allociné might therefore not mean the same 6.5 on IMDb. This makes the 10-point rescaling of all other scales an approximation rather than a perfect translation.

Thirdly, the vote reliability is not stable. Throughout the analysis, mean ratings are treated as if measured with equal precision, however, vote counts range largely. This is due to both the difference in size of the user base, but also the difference for each film. Niche movies get viewer ratings and are therefore less reliable to represent a solid mean.

Finally, it is important to consider the limitations of a cross-sectional snapshot. All the ratings were collected at one single moment in time. Temporal nuances, which could consist of algorithmic changes, award surges, or any other impactful effect, therefore fall outside of the data frame. The study can not perform any panel regression or dynamic feedback loops because of this and will refrain from framing any result to be highly causal. Instead, I will focus on associative relations. Future research, which is granted more time, could employ monthly snapshots to investigate how the cross-platform gaps operate across time.

In conclusion, these constraints reduce the study's contribution without diminishing it. By analyzing more than 2,000 movies the research provides the first systematic, side-by-side perspective of how film ratings across a more international and a more local platform. Even though rater demographics are hidden, scales fluctuate, vote counts vary, and the snapshot is cross-sectional, those very limitations mirror the way real-world platform users encounter ratings when deciding what to watch. The findings therefore make clear if a film's public reputation can differ when compared between platforms – even if the gap can never be fully attributed to the platform alone. By documenting these limitations, the study provides a clear agenda for the next wave of research.

4. Results

The central question to be addressed in the analysis is whether identical films receive comparable ratings on a more global (IMDb) versus a more national (Allociné) platform, and how any divergences relate to film attributes such as budget, country of origin, and genre. To start the analysis, IMDb's press rating, Allociné's press rating, and Allociné's spectator rating have been harmonized to a 10-point scale. Scale minima are therefore 1 for all spectator and Allociné press scores, and 0 for the IMDb press score. This is the least-distorting compromise because upscaling Allociné's ratings by 20 would create 10-point jumps erasing nuances. The results section will be structured in the same order as the research questions and will start with looking at rating divergence.

4.1 How large and in what direction do IMDb and Allociné user ratings diverge for the same films?

Two univariate tables are provided as a starting point. Table 1 displays all continuous variables and Table 2 displays all discrete variables.

Table 1 Descriptive Statistics – Continuous Variables

	n	Min	Max	Mean	SD
Spectator rating IMDb	2270	1.50	9.20	5.35	1.33
Spectator rating Allociné	2240	2.00	9.00	5.77	1.51
Press rating IMDb	793	.10	10.00	5.42	1.73
Press rating Allociné	1379	2.00	10.00	6.12	1.34
Spectator Gap	2296	0.00	6.20	2.45	1.25
Press Gap	1681	0.00	6.20	0.78	1.41
Budget (Log)	697	0.00	8.55	6.43	1.23
Worldwide Gross (Log)	1071	2.64	9.45	6.15	1.42

Table 2 Descriptive Statistics – Discrete Variables

	Categories	*n*	%
Country of origin	France	514	22.39
	USA	1782	77.61
Macro-genre	Action/Adventure	431	18.77
	Comedy	675	29.40
	Drama/Romance	682	29.70
	Thriller/Horror	508	22.13
Year	2010	172	7.49
	2011	199	8.67
	2012	203	8.84
	2013	224	9.76
	2014	217	9.45
	2015	222	9.67
	2016	246	10.71
	2017	222	9.67
	2018	260	11.32
	2019	224	9.76
	2020	107	4.66

Allociné spectators award films with a mean score of 5.8 (SD=1.5), compared with 5.3 (SD=1.3) on IMDb. A larger variance is noticeable between the press rating score on Allociné 6.1 (SD=1.3) and IMDb 5.4 (SD=1.7). The gap variable translates those raw means into absolute differences. Spectator gaps average 2.45 points, whereas press gaps average 0.78. This hints toward a first understanding of where differences between platforms are most pronounced. However, in order to make any worthwhile statements this needs to be further analyzed in detail. The average budget in the sample is e6.43 ≈ €620,000 and the worldwide gross e6.15 ≈ 470,000. Turning to the categorical composition, U.S. productions dominate the sample with 78%, however, the French subset is with 22% still large enough to create meaningful within-country analysis. The composition of genres is fairly balanced with each genre taking approximately 20% of the share. Finally, the release-year distribution is flat from 2010-2019 with a predictable dip in 2020 (the first pandemic year).

To conduct an initial test, of whether various types of ratings are statistically different from each other, a set of two-sample *t*-tests were conducted and displayed in Table 3. The IMDb score was subtracted from the Allociné score, therefore, positive values indicate higher ratings on Allociné compared to IMDb. To ensure robustness, I complement the *t*-based confidence interval with a bias-corrected and accelerated (BCa) bootstrap interval based on 5000 resamples (Efron, 1987). The method adjusts for both bias and skewness in the bootstrap distribution by resampling the data. This produces a second-order accurate 95% interval without having to rely on normality. For all two-sample *t*-tests, the interval was almost identical. The test confirms that the difference between spectator ratings on IMDb and Allociné is statistically significant, $t(2{,}213)= 11{,}51$, $p= <.001$. The midpoint of this interval is the mean gap, which is about 0.42, meaning that Allociné scores run 0.42 points higher than IMDb. Cohen's d translates the raw point difference into expressing the mean gap between Allociné and IMDb in units of the films' own variability. The effect size is small (Cohen's $d = 0.25$), but on a one-digit rating scale, the 0.25-point uplift reflects a consistent boost that Allociné spectators give their scores supporting. When further breaking down the sample, the gap widens and narrows in different ways. Looking at the different genres, the gap is trivial for Comedy, 0.14 ($p = .013$) but increases to 1.00 ($p < .001$) for Thriller/Horror. This indicates that Allociné horror watchers reward titles with one point above IMDb's. The contrast within comparison groups is sharpest for country of origin. French films show no meaningful gap (–0.09, $p = .042$) whereas U.S. titles receive 0.57 higher scores ($p < .001$). The lower quartiles of the budget show a 0.96 ($p < .001$) to 1.08 ($p < .001$) boost in ratings on Allociné. For the highest quartile, this was -0.25 ($p = .002$). These initial indications need to be further tested with more sophisticated and reliable multivariate tests.

Table 3 Two-samples *t*-tests - Spectator

Comparison	*n*	*t*(df)	95% Cl (param.)	95% Cl (BCa 5000)	Cohen's *d*
Spectators	2,214	11.51***(2,213)	0.35-0.49	0.35-0.49	0.25
Genre: Action/Adventure	413	6.35***(412)	0.39-0.74	0.39-0.73	0.31
Genre: Comedy	656	2.49*(655)	0.029-0.25	0.033-0.25	0.10
Genre: Drama/Romance	659	2.91**(658)	0.057-0.29	0.057-0.29	0.11
Genre: Thriller/Horror	486	11.00***(485)	0.83-1.16	0.83-1.16	0.70
Country: France	510	-2.04*(509)	-0.17- -0.003	-0.17 - -0.005	-0.09
Country: USA	1,704	12.70***(1,703)	0.48-0.66	0.48-0.66	0.44
Budget: Q1 (lowest)	176	6.77***(175)	0.77-1.39	0.75-1.38	0.51
Budget: Q2	173	7.32***(172)	0.71-1.23	0.71-1.22	0.56
Budget: Q3	161	0.21(160)	-0.18-0.22	-0.16-0.22	0.02
Budget: Q4 (highest)	169	-3.20**(168)	-0.40- -0.096	-0.41- -0.11	-0.25

Note. BCa= bias-corrected accelerated bootstrap (5000 resamples). Significance codes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4 applies the same logic to press' scores. Across the full set of press ratings, Allociné reviews are on average 0.55 points above IMDb ($p < .001$). because the BCa interval is virtually identical to the parametric one, the result is not a distributional artifact. When I zoom in further on different comparisons for press rating scores, a few findings are of interest. In Action/Adventure, Allociné critics award on average 0.43 points more than their IMDb peers, $p = .002$; $d = 0.29$, virtually the same standardized gap that spectators show ($d = 0.31$), so the press neither amplifies nor dampens audience bias in this genre. Similarly to spectator ratings, press ratings for Thriller/Horror films see a 1.06 higher rating on Allociné than on IMDb, $p < .001$. The effect size is quite moderate with 0.54, but lower than the effect size for the spectator ratings ($d = 0.70$), meaning that on average Allociné thriller/horror spectator ratings are 0.70 SD above the IMDb average for the same films and the Allociné press mean is 0.54 SD higher than the IMDb press mean on those films. Therefore, audiences diverge more strongly than critics in this genre even though both raw gaps are around 1. National origin sharpens these contrasts further, for French titles, press scores on Allociné are 0.67 ($p < .001$) points higher compared to the converging spectator scores across both platforms (–0.09; $p = .042$). This might be an initial indication of the absence of home bias for spectators, which will be further investigated. By contrast, the press gap (0,53, $p < .001$) is slightly smaller than the spectator gap (0.57, $p < .001$) for U.S. films, so spectator ratings are even a bit higher and diverge more for U.S. movies on Allociné compared to press ratings on Allociné. For the lower quartiles (Q1 + Q2) spectators double the critic gap. For Q1 and Q2 the press ratings are uplifted by 0.53 ($p = .025$), and 0,51 ($p = .019$) respectively. The upper quartile results were not significant.

Table 4 Two-samples *t*-tests - Press

Comparison	*n*	*t*(df)	95% Cl (param.)	95% Cl (BCa 5000)	Cohen's *d*
Press	491	7.41***(490)	0.40-0.70	0.41-0.70	0.33
Genre: Action/Adventure	123	3.18**(122)	0.16-0.70	019-0.72	0.29
Genre: Comedy	133	4.20***(132)	0.31-0.87	0.32-0.87	0.36
Genre: Drama/Romance	169	3.31**168)	0.17-0.66	0.18-0.67	0.25
Genre: Thriller/Horror	66	4.39***(65)	0.58-1.55	0.62-1.54	0.54
Country: France	83	4.76***(82)	0.39-0.95	0.42-0.97	0.52
Country: USA	408	6.22***(407)	0.36-0.70	0.36-0.70	0.31
Budget: Q1 (lowest)	65	2.30*(64)	0.070-0.99	0.11-0.99	0.29
Budget: Q2	65	2.41*(64)	0.088-0.94	0.13-0.95	0.30
Budget: Q3	64	1.11(63)	-0.12-0.41	-0.093-0.40	0.14
Budget: Q4 (highest)	65	0.95(64)	-0.15-0.41	-0.15-0.40	0.12

Note. BCa= bias-corrected accelerated bootstrap (5000 resamples). Significance codes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 5 displays the paired *t*-test of 1681 films that carry both gap measures and confirms that the difference is of significance ($t(1680)=50.14$, $p<.001$; $d= 1.22$). The

Bootstrapped Bca confidence interval (1.762-1.905) mimics the parametric one, indicating that the result is not a distributional artifact. This means that audiences diverge far more sharply than critics and drive cross-platform volatility. The results display that platform rating gaps are not significantly larger for the press than for spectators on both IMDb and Allociné in the sample.

Table 5 Two-sample *t*-test for Rating Gaps

	t(df)	*p*	95% Cl (param.)	95% Cl (BCa 5000)	Cohen's *d*
Spectator/Press Gap	50.14(1680)	<.001	1.764-1.907	1.762-1.905	1.22

Having mapped where and how much the scores differ, the next step is to investigate how tightly the two rating scales move together film by film. I, therefore, compute paired Pearson correlations for each rater group. The resulting coefficients (Table 6) range from 0.29 to 0.42, demonstrating a moderate, highly significant (p <.001) relationship. This modest relation confirms that Allociné and IMDb are related yet clearly distinct.

Table 6 Paired Pearson correlations

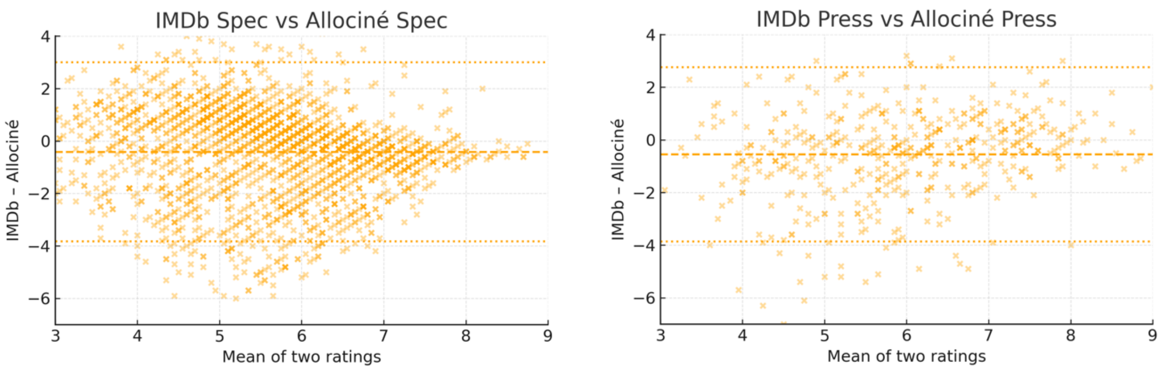
	n	Pearson *r*	95% Cl (param.)
Spectator ratings	2,214	0.29***	0.25-0.32
Press ratings	491	0.42***	0.35-0.49

Note. BCa= bias-corrected accelerated bootstrap (5000 resamples). Significance codes: * $p < .05$, ** $p < .01$, *** $p < .001$.

These correlations, however, only summarise a single linear link. It does not document where along the rating scale agreement is strongest or weakest. To initially investigate this I turn to Bland-Altman plots, which graph the difference (Allociné-IMDb) against the mean of the two scores of each film. This gives valuable insight into possible systematic bias, and heteroscedasticity (Giavarina, 2015). The fact that platforms are related but distinct slightly points towards the conception that ratings do not inherently converge to a single value. This conception is confirmed by the Bland-Altman plots (Figure 1). In the plots, the center line displays the average bias between the two sites. The upper line is the upper limit of agreement under which 95% of all differences should fall, the lower line is the lower limit of agreement of which 95% of all differences should fall above. The spectator graph is funnel-shaped and this funnel starts to tighten around a rating of 6-7. What is interesting is that the press plot has roughly the same funnel, but even wider. This shows that convergence is conditional. For the middle-category films, the gap is quite small but there is clear polarisation in the higher and lower rating segments.

Figure 1

Bland-Altman Plots



To further investigate the gap variable I first compute an ordinary least-squares (OLS) regression that will serve as the baseline. This model estimates the average relationship between the amount of disagreement (the gap) and quality indicators (the average of all four ratings, budget, and worldwide gross). I will perform the model for both the spectator and press gap. In addition, I will run a Breusch-Pagan and White test to check for heteroscedasticity. The diagnostic function of these tests is to determine the reliability of the OLS assumption of constant error variance. The Breusch-Pagan test regresses the squared OLS residuals on the original regressors. White's test adds all cross-products of regressors, capturing any nonlinear form of heteroskedasticity. The model is as follows:

$$\mid gap \mid i = \beta 0\tau + \beta 1\tau + \beta 2\tau + \beta 3\tau + \beta 4\tau + \Sigma g + \beta 5\tau + \beta 6\tau + \varepsilon i\tau, \qquad Q\tau(\varepsilon i\tau) = 0,$$

- $\mid gap \mid i$ Absolute spectator gap (IMDb vs. Allociné) for film *i*
- $\beta 0\tau$ Average of the four normalised ratings for film *i*
- $\beta 1\tau$ Release year

- $\beta 3\tau + \beta 4\tau$ Country dummies, USA i + France i
- Σg Dummy for genre g (baseline = Action/Adventure)
- $\beta 5\tau$ Natural log of production budget
- $\beta 6\tau$ Natural log of worldwide gross
- $\varepsilon i\tau$ Asymmetric error term

Table 7 OLS Regression with spectator gap as the dependent variable

	B	SE	*t*
Average quality	-0.11*	0.036	-3.12
Release year	-0.005	0.014	-0.33
USA (vs France)	0.016	0.11	0.15
Genre: Comedy	-0.19	0.11	-1.65
Genre: Drama/Romance	-0.039	0.11	-0.34
Genre: Thriller/Horror	-0.18	0.13	-1.38
Budget	-0.11	0.086	-1.32
Worldwide Gross	-0.14*	0.043	-3.34
Constant	8.53	18.70	0.46

Note. n= 389 films. Genre baseline = Action/Adventure. $R^2 = 0.16$. $F(8, 380) = 10.49$***. Significance codes: *** $p < .001$; ** $p < .01$; * $p < .05$. Standard errors are conventional OLS; heteroskedasticity tests rejected constant variance (Breusch-Pagan $\chi^2 = 17.4$, $p = .026$; White $\chi^2 = 64.0$, $p = .003$).

The OLS benchmark (Table 7) reveals that, on average, higher-quality films tend to have more audience agreement. A one-point increase in the mean of the four normalized ratings is linked to a 0.11 decrease in the absolute gap ($p < .05$). A larger Worldwide Gross also seems to trim the gap on average, reducing the difference by roughly 0.14 points ($p = .002$), indicating that popular films receive higher overall consensus. Both the increase of stated, as well as revealed preferences seem to narrow the gap. On the other hand, the OLS documents that the control variables genre, country of origin, and production budget do not really alter the mean gap as these results are also insignificant. However, this mean-centered perspective lacks important nuances. The heteroskedasticity tests imply that OLS coefficient estimates remain unbiased but the standard errors are unreliable and therefore the average effect could potentially mask different important patterns (Breusch-Pagan $\chi^2 = 17.4$, $p = .026$; White $\chi^2 = 64.0$, $p = .003$). Since the research is particularly interested in the differences in ratings and how these are linked to quality indicators, these alternative patterns are of great importance. Quantile regression (Koenker & Bassett, 1978) models the *τ-th* percentile of the distribution of gaps. Unlike OLS, this model requires no constant-variance assumption because standard errors are bootstrapped. Therefore, this model will demonstrate whether the quality-convergence premise holds for the most extreme disagreements as well as for the typical ones. This is especially important because the OLS model did not reveal variance in heterogeneity and that is exactly what is of interest to uncover. Table 8 displays the Quantile regression results.

Table 8 Quantile regression results for the spectator rating gap model

	q25		q50		q75	
	B	*t*	*B*	*t*	*B*	*t*
Average quality	-0.038	-1.53	-0.096*	-2.96	-0.16*	-3.84
Release year	0.008	0.81	0.004	0.33	0.002	0.15
USA	-0.005	-0.79	-0.002	-0.26	-0.028	0.00
France	-0.048	-0.75	-0.019	-0.23	0.028	0.00
Genre: Comedy	0.006	0.07	0.049	0.49	0.003	0.03
Genre: Drama/Romance	0.058	0.74	0.084	0.83	0.11	0.82
Genre: Thriller/Horror	0.046	0.50	-0.022	-0.19	-0.062	-0.40
Budget	0.051	0.82	0.021	0.27	-0.11	-1.15
Worldwide Gross	-0.062*	-2.10	-0.16*	-4.09	-0.23*	-4.88

Notes. BCa= bias-corrected accelerated bootstrap (5000 resamples). Dependent variable = |IMDb − Allociné| spectator scores (n = 389). Genre baseline = Action/Adventure. Country baseline= France. Significance code: *p < .01; Pseudo-R^2: .10 (.10), .07 (.25), .05 (.50), .06 (.75), .05 (.90).

Similarly to the OLS estimate, a one-point increase in overall quality reduces disagreement by around 0.10 points at the median of the gap distribution (q50, p < .01). This effect nearly doubles for the upper tail where the reduction is approximately 0.16 for the most extreme 25% of gaps (q75, p < .01), confirming that films with higher rating scores attract sharper consensus where disagreement is largest. Conversely, quality is unrelated to the smallest gap (q25) as the result was insignificant. The effect of Worldwide Gross documents the same pattern. Moving up towards the upper tail, the negative effect on rating gaps becomes more pronounced until it reaches a 0.23 decrease for q75 ($p <.01$). The other control variables remain negligible.

The findings confirm that IMDb and Allociné do not converge to a single crowd verdict; instead, the divergence is to some degree dependent on both stated and revealed preference (being the mean rating of all four ratings and the worldwide gross). Consistent with symmetrical ignorance theory, quality remains uncertain ex-ante, however, higher average ratings and higher grosses do result in more alignment of audience opinions on both platforms. Especially for the titles that would otherwise have the largest gaps. It can be concluded that scores on IMDb and Allociné often disagree, but the amount of disagreement is affected by its worldwide gross and total mean rating and this effect is most pronounced for the movies that receive high gross or high ratings on average.

4.2 To what extent do those cross-platform rating gaps translate into differences in worldwide theatrical revenue?

The results reveal that ratings not only depend on stated preference (average ratings) but also on revealed preferences which are measured by the worldwide gross. It would therefore be of interest to rerun the OLS regression with the gross variable as the dependent variable. The model will use the same variables as the model in which the spectator gap is the dependent variable, only the mean ratings will be swapped with the spectator gap. The press gap could not be used because this would even further the already lower N of 394.

Figure 2 – Distribution among ratings across Gross

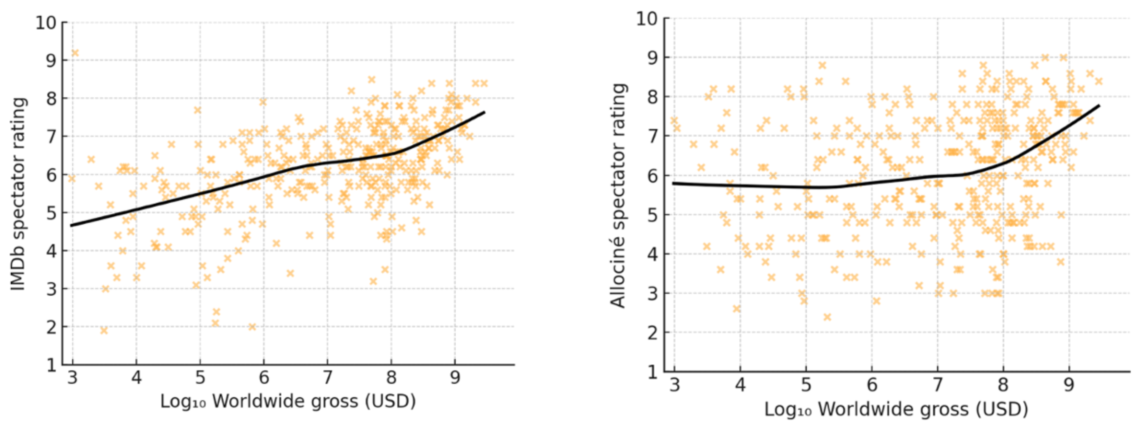


Table 9 OLS regression with dependent variable gross

Predictor	*B*	SE	*t*
Spectator gap	-0.26***	0.06	-4.32
Year	-0.03	0.02	-1.61
USA (vs France)	0.32*	0.09	3.56
Genre: Comedy	0.20	0.14	1.47
Genre: Drama/Romance	0.12	0.14	0.91
Genre:Thriller/Horror	0.14	0.16	0.87
Budget	1.43*	0.07	19.11
Constant	-2.06	0.41	-5.06

Note. $R^2 = 0.59$. $F(7, 381) = 91.6$***. *Heteroskedasticity diagnostics:* Breusch–Pagan $\chi^2 = 49.30$, $p < .001$; White $\chi^2 = 80.19$, $p < .001$. Significance codes: *** $p < .001$; ** $p < .01$; * $p < .05$.

The OLS model, as documented in Table 9 explains around 60% of the variance in worldwide gross. Three clear signals can be derived from the model. First of all, budget is one of the key drivers of revenue, which is not a surprise. A 1% rise in budget predicts around a 1.4% rise in gross (p <.05). Secondly, the spectator gap predictor shows that each one-point widening of the IMDb and Allociné gap corresponds to a 0.26 drop for the log of gross, which roughly translated means a 23% lower revenue (p <.001). With France functioning as the baseline, U.S. films outperform French by 37%, which makes sense since U.S. films have a bigger reach. The outcomes of the heteroskedasticity diagnostics confirm the need for another quantile analysis. This is especially the case since Figure 2 visualizes the distribution of gross compared to the different ratings. By including a locally estimated scatter-plot smoother (LOESS) the figures are able to show the average direction of the data without having to form a straight line. In the case of budget and gross, it shows non-linearity, justifying the quantile regression.

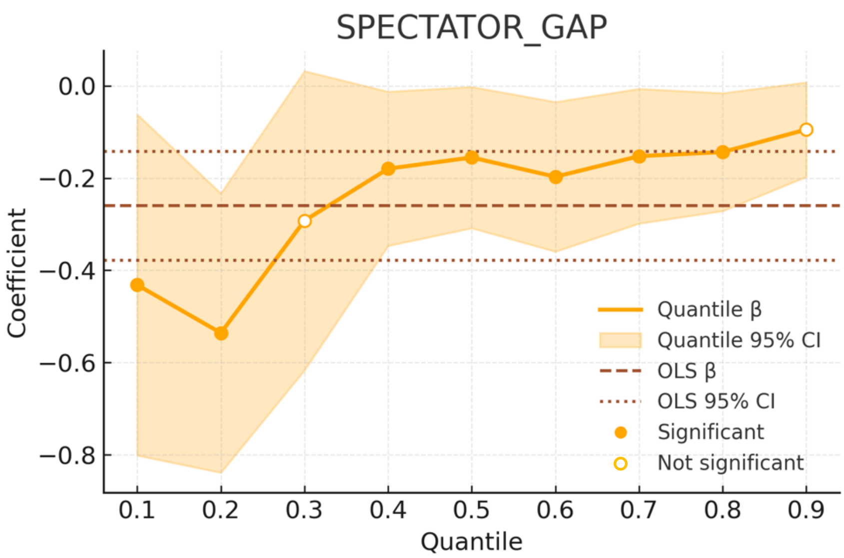
Table 10 Quantile regression results for the worldwide gross model

	q25		q50		q75	
	B	*t*	*B*	*t*	*B*	*t*
Spectator gap	-0.44*	-6.91	-0.16*	-2.93	-0.14*	-2.61
Release year	0.01	-0.46	0.01	-0.71	0.00	-0.03
USA (vs France)	0.18	0.53	0.25	0.94	0.43	1.78
Genre: Comedy	0.24	1.61	0.19	1.55	0.16	1.44
Genre: Drama/Romance	0.18	1.16	-0.04	-0.35	0.11	1.01
Genre: Thriller/Horror	0.09	0.53	-0.02	-0.19	-0.062	-0.40
Budget	1.42*	17.00	1.44*	18.70	1.41*	14.90

Notes. BCa= bias-corrected accelerated bootstrap (5000 resamples). *n* = 389. Genre baseline = Action/Adventure. Country baseline= France. Significance code: *$p < .01$; Pseudo-R^2: .10 (.10), .08 (.25), .05 (.50), .06 (.75), .05 (.90).

The quantile regression nuances the outcome of the OLS further. Spectator disagreement is largest in the bottom quartile (-0.44, $p < .01$) and slowly fades when moving upwards, indicating that movies that receive higher worldwide gross will display smaller gaps. Figure 3 provides a visualization of the spectator gap across the full quantile. The budget remains a positive driver across the full quantile, confirming that the available budget scales proportionally with revenues, even when accounting for different controls. Comparing this to the OLS findings it can be concluded that cross-platform is not merely a correlate of success on average but is especially critical in the lower quantile of the revenue.

Figure 3 Visualisation of the OLS and quantile regression with dependent variable Worldwide Gross



However, some limitations must be reported. To be able to include all variables the usable sample dropped from 2,296 to 389 films. This means that there is the possibility of sampling selection bias. In addition, the quantile regression had to be based on <100 cases per quantile, which is not ideal. The impact of disagreement on revenue must therefore be viewed as associative rather than strictly causal. The coefficients do, however, display a clear pattern: even after adjusting for budget and genre, consensus between platforms is an important factor in determining the commercial success of lower-grossing films. A 1-point gap can cost a small film roughly a third of its gross.

4.3 Is the size or direction of that rating gap influenced by cultural proximity?

In addition to the platform effect, the geographical location of raters can shape evaluation patterns. Allociné is dominated by a domestic audience, whereas IMDb reflects a more international user base. The cultural proximity thesis states that viewers tend to favor works that are linguistically and culturally related to their own experiences (Straubhaar, 1991). It is therefore expected that French films should receive higher ratings on Allociné compared to IMDb. To test this, I restrict the sample to French productions (n = 514). To be able to single out title-specific heterogeneity in the cross-market comparison, I treat every French production as its own control by matching the IMDb score with the Allociné score.

Table 11 Descriptive Statistics of French Productions

	n	Min	Max	Mean	SD
Spectator rating IMDb	514	1.70	7.90	5.80	.87
Spectator rating Allociné	513	2.0	8.40	5.72	1.40
Press rating IMDb	88	.10	8.50	6.03	1.48
Press rating Allociné	479	2.00	9.20	6.10	1.24

Table 11 showcases that spectator ratings on IMDb have a mean score of 5.80 (SD = .87), whilst Allociné spectators give out an average score of 5.72 (SD = 1.40), hinting at more polarised domestic ratings. This difference is, however, more marginal than the difference displayed in Table 1 (5.35 on IMDb compared to 5.77 on Allociné). Press ratings have a mean of 6 (IMDb) and 6.1 (Allociné). However, $n = 88$ for press ratings on IMDb, which shows a severe asymmetry compared to $n = 479$ for press ratings on Allociné. This small overlap, which would contain less than 20% of the Allociné press sample would create the risk of selection bias and would have little statistical power when incorporated in the paired test because any visible difference could simply reflect which films happened to attract English-language reviewers on IMDb. For completeness of the analysis, however, press ratings will be included even though the interpretation of these results should be carefully considered.

Table 12 Two-samples *t*-tests – French films

	t(df)	*p*	95% Cl (param.)	95% Cl (BCa 5000)	Cohen's *d*
Spectators	2.04(509)	.042	0.003-0.17	0.002-0.17	0.09
Press	4.76(82)	< .001	0.39- 0.95	0.38- 0.97	0.52

Note. BCa= bias-corrected accelerated bootstrap (5000 resamples).

The paired-sample analysis (Table 12) of the 510 French productions that were rated by spectators on both platforms shows that the difference is statistically significant, $t(509)$ =

2.04, p = .042. The 95% confidence interval of 0.003-0.17 is almost identical to the BCa-bootstrapped interval of 0.002-0.17. Nevertheless, Cohen's d = 0.09, which is very small and not likely to alter any ratings that are rounded off with one decimal. For Press, Cohen's d is significantly higher (0.52, p < .001). However, since the sample for press is quite small, the result should be taken with caution

French productions do not have significantly higher spectator ratings on Allociné than on IMDb in the tested sample. In addition, it must be taken into account that the sample includes only French productions that are listed on both platforms. Niche domestic releases that never reach IMDb are underpopulated in the sample, the same goes for big-budget productions that are marketed at an international level, which are likely over-represented. Moreover, IMDb users who watched and rated French production might have more cultural capital since they have to put in more effort to find these films and watch them, hence they reflect a different audience than the more general French public on Allociné. Thus the observed gap would not explain too much about the identity of the platform, but more on the identity of the viewer.

To look further into the influence cultural proximity has on the rating of films, I will employ a film-level linear mixed-effects model (LMM). I will do so because each film is rated on two different platforms by two different rating groups. Ratings within the same film are therefore not independent. The LLM with a random intercept for film will allow for an analysis that treats coefficients like platform and press reviews and their interactions into within-film contrasts. This is exactly what is needed in order to further determine the effect of cultural proximity. By employing this approach I follow previous works that have also nested different coefficients inside films to test the interactions (Chen et al., 2020; Nalabandian & Ireland, 2018; Nalabandian & Ireland, 2022). This leads to the following model:

$$\begin{aligned} Rating = \beta 0 + \beta 1 Platform + \beta 2 USA + \beta 3(Platform \times USA) + \beta 4 IMDbspec \\ + \beta 8 Year + \beta 9 Gross + \gamma \top Zj + u0j + u1jPlatform + \varepsilon ij, \end{aligned}$$

- $\beta 0$ grand mean (baseline: French film x IMDb x spectator, at average year & gross)
- $\beta 1$ Allociné – IMDb gap for French films
- $\beta 2$ USA – France gap on IMDb

- $\beta 3$ *cultural-proximity term*: extra gap Allociné gives to USA films
- $\beta 4$ press – spectator shift (platform-invariant)
- $\beta 5, \beta 6$ Year and Gross controls (mean-centred)

The intercept coefficient serves as the reference category. This baseline was set to the dummy codes French film x IMDb x spectator because this combination is present for every film in the dataset. In addition, it is the most logical starting point as I want to compare this base against both the domestic platform (Allociné) as well as foreign content (USA films). The outcome of the intercept will indicate what score is expected for a French film rated on IMDb by spectators. Every other coefficient represents a shift from that starting point.

The Platform x Nationality two-way interaction describes the cultural proximity test; it shows whether the gap between IMDb and Allociné will decrease or grow when switching between U.S. and French titles. Furthermore, the random intercept is of interest because it determines how much of the rating variance is due to intrinsic quality indicators such as Worldwide Gross and average rating.

Table 13 LLM model to test for cultural proximity

Effect	*B*	SE	*z*
Intercept	5.81***	0.06	95.87
Platform: Allociné	0.07	0.06	1.15
Nationality: USA	-0.64***	0.07	-9.44
Press vs Spectator	0.09**	0.03	2.63
Platform x Nationality	0.58***	0.07	8.31
Year	0.016*	0.008	2.04
Gross	0.17***	0.02	7.42

Note. Effect baseline = IMDb x French x spectator. ICC = 0.35. Significance codes: *** $p < .001$; ** $p < .01$; * $p < .05$.

The results (Table 13) indicate a significant interaction between Allociné and USA ($B = 0.58$, $p < .001$). This supports the early two-sample *t*-tests by indicating that Allociné raises the rating of U.S. films by 0.58 points relative to IMDb, providing a robust angle to the cultural proximity debate. French audiences seem to rate U.S. films substantially higher on the domestic platform instead of displaying home bias for French productions. This result echoes the effect of Nationality: USA, where U.S. films start 0.64 points ($p < .001$) lower on IMDb than French films. In addition, the press is 0.09 points ($p < .01$) more positive than the spectators, even after being controlled for platform and nationality. Year and Gross have small but significant ($p < .05$ and $p < .01$) positive impacts on scores. The intraclass correlation documents that one-third of all rating differences are due to films being inherently different, and 35% of all rating variation occurs between films. The remaining 65% of the films is the portion the fixed effects (platform, nationality, and revenue) try to explain. This means that intrinsic film quality, which is related to anything that makes a film good or bad before any audience bias, platform culture, or marketing comes into play, still dominates a fair amount of the total films. Together, these patterns illustrate that the reputation of a film travels rather through platform-specific lenses than home vs. foreign binaries. It therefore matters more where and by whom the film is rated, not where it was made.

5. Discussion

The results of this study reveal a picture of how a national platform (Allociné) and an international platform (IMDb) shape symmetrical ignorance in film ratings. Firstly, the study reveals that the platform context matters in this. The same films often receive different ratings across the two platforms. This was demonstrated by computing paired t-tests that showed systematic rating gaps between the two platforms. On average, Allociné users rated films higher than IMDb users, this gap was even more pronounced for the press. In other words, the reputation a film has is slightly monolithic; it matters where you look for information on the quality of the film. These findings are in line with the work of Baugher and Ramos (2017) who state that cross-platform consistency of ratings cannot be assumed especially because different platforms can attract different users, have different rating goals, and be influenced by factors unrelated to the core subject being rated (like movie quality), all of which can impact the consistency and validity of the ratings across platforms. This study extends this insight by not only comparing different platforms but also considering the difference between a local vs. global platform. Suggesting that symmetrical ignorance exists not only between film and audience but also between different audience communities. This supports the notion of each platform creating its own market for information and therefore users on Allociné and IMDb operate in two very distinct information ecosystems.

Secondly, the results nudged in the direction of the intrinsic quality of a film fostering consensus across platforms. If a movie has high intrinsic quality audiences will converge, whereas films with medium and low intrinsic quality will generate more disagreement. This was evidenced from two different angles. Firstly, the Bland-Altman plots that showed funnel shapes indicated that for ratings of 7 and upward, there is increasing consensus. The quantile regression affirmed this by indicating that higher average film quality is associated with smaller inter-platform. At the median level of the gap, a film's overall rating had a slightly reducing effect on the gap but when looking at the upper quantiles (.75 and .90), the effect of high quality reducing the gap was much stronger. This indicates that universally acclaimed movies with a rating of 8+ from all four rating audiences rarely received big differences in ratings. This echoes the findings of other research. For instance, the research of Burmester et al. (2024) found that higher-quality evaluations reduce outcome heterogeneity. Translating that to this research, quality reduces the heterogeneity of opinions across different audiences and, therefore, I argue that it is not sheer luck whether a film will receive high ratings.

Thirdly, the study's results expose a consistent yet uneven divide between the press and the spectator. Across the 481 films for which press reviews exist on both platforms, the press score films about 0.55 points higher on Allociné than on IMDb, while the average spectator disparity is 0.40 points. The linear-mixed model that was performed confirms this pattern. After adding the film-level controls, press ratings remain 0.09 points more generous than spectator ratings, regardless of platform or nationality. This has two implications. First, critics amplify platform differences. The scores they attribute to films move in the same direction as the spectators but are more widely dispersed. Second, press influence is asymmetric because both platforms curate different pools of critics. To what extent the sourcing of the press ratings overlap is unknown, but since Allociné uses mostly French reviews it is most likely there is a cirtical differentiation between the two. Especially so, because French critics will tend to emphasize what is domestically relevant in France. Because the press rating gap is larger and more stable, due to smaller sampling error and lower relative dispersion, it matters for a film's average rating whether a press review is posted on IMDb or Allociné.

Fourthly, the analysis of revenue revealed that cross-platform consensus is economically consequential. The OLS regression indicated that every 1-point widening of the gap roughly decreases worldwide gross by 23% after controlling for budget, genre, and release year. This result is further sharpened by the quantile regression. The same 1-point widening of the gap results in a 44% impact on gross for the bottom quartile of the revenue distribution. This percentage drops to roughly 14% in the upper segment, indicating that small-budget films are most heavily impacted by negative disputes. It therefore seems that consensus pays off in terms of revenue. For example, a typical lower-tier movie roughly shows a gap of 0.30 points. If the gap is closed by 0.20 points, 7% of revenue would be recovered (0.20 x 36% = 7%, because e^-0.44 ≈ 0.64 means that revenue falls to 64% of what it would have been, which is equal to a 36% reduction). For studios and distributors, this means that, especially for the lower-tier budget movies, shrinking cross-platform disagreement is beneficial for revenue outcomes. However, these results only track theatrical grosses, leaving out streaming revenues, and focus on a cross-sectional dataset. As a result, the outcomes are unable to determine if earnings are driven by ratings or vice versa. In addition, it would be of value to capture the dynamics of word-of-mouth after a film has been released. Future research should look into this to further strengthen causality.

Finally, the results contradict the home-country-bias expectation which stems from cultural-proximity theory. There is virtually no domestic uplift on Allociné according to the spectator ratings for French movies (an uplift of 0.09). Rather, according to the mixed-effects results, U.S. films receive a 0.58-point lift on Allociné compared to a -0.64-point more negative rating it gets on IMDb. This means that there is virtually no home bias on Allociné, but rather a preference for U.S. films, making U.S. imports look better than they do on the international benchmark. This reversal could be caused by the hybridized media consumption climate we now live in. The application of cultural proximity theory might be challenged by the platform-audience interaction.

Having considered these angles, the results present a complex yet coherent picture. Depending on whether a user consults IMDb or Allociné, the same movie may appear better (or worse); platform context matters. The gap narrows only for the very best-reviewed films, and critics magnify these platform differences further. Notably, disagreement also has an influence on revenues. Every extra rating point of cross-platform discord can eliminate up to one-third of the potential box office revenue. Yet, against the theoretical premise of cultural proximity theory, Allociné does not promote French titles more positively than IMDb; instead, it is more positive on U.S. movies. These findings collectively demonstrate that the platform-specific reputation of a film is not a neutral mirror of its intrinsic quality but rather an economically powerful construct.

6. Conclusion

This thesis set out to answer the following questions: How large and in what direction do IMDb and Allociné user ratings diverge for the same films? To what extent do those cross-platform rating gaps translate into differences in worldwide theatrical revenue? Is the size or direction of that rating gap influenced by cultural proximity?

By comparing the national site Allociné with the international benchmark IMDb across more than 2,000 releases, I provide the first systematic evidence of platform-specific rating asymmetries and their financial stakes. The following key takeaways were identified. First, platform context matters. On a 10-point scale, paired tests reveal an average spectator gap of around 0.40 points and a press gap of about 0.55 points, confirming that a movie can carry varying reputation depending on where audiences look for the information. Second, higher ratings moderate disagreement. The quantile regressions and Bland-Altman funnels show that the best-reviewed films, with a score of 8 and above, attract convergent scores. Critics amplify these gaps even further. Even after film-level controls, press scores continue to be 0.09 points higher than spectator scores. Third, consensus is profitable. Each additional rating point of cross-platform disagreement reduces worldwide revenues by about 36% in the bottom quartile of the revenue distribution. Films in the higher-grossing segment can more easily withstand the reputational shock, whereas low and mid-tier films suffer the most. Fourth, the home bias premise is reversed because Allociné does not provide French films with a significant benefit. Instead, Allociné displays a preference for U.S. films. Platform-audience interactions, therefore, seem to override the straightforward local allegiance.

Together, these findings portray rating platforms as distinct information markets. Even though the exact same movies are rated, each site attracts a different crowd of voters, pools reviews from different critics, and presents the scores in unique ways on the platform interface. These differences indicate that an intrinsically good film will not be equally perceived by the different platforms. This contingency is more than just an identified difference: it can add or subtract the gross outcome of a film and overturn the conception of cultural proximity. In summary, where a movie is rated can affect its perceived value and financial success.

For studios and distributors, the results give a quantified indication of the effect of cross-platform divergence on financial success. For cultural economics scholars, the study demonstrates that platform effects can neutralize or even reverse expected cultural proximity patterns. This calls for future research on the role of proximity theory in an era of hybridized media consumption.

These findings are subject to some limitations. The audience composition is unknown, which means that platform disparities may be inflated by variations in users' gender, age, or language. Vote counts vary for each film but are unknown in this study, resulting in the fact that the primary estimates treat each mean identically. In addition, the variation in vote counts makes niche as well as blockbuster films' averages less reliable since it is likely they will vary the most. The causal direction of the revenue effects cannot be established as the study is limited to cross-sectional data. Finally, rescaling the rating measures to a 10-point metric is the best approximation rather than full certainty.

Future research should investigate and track ratings over time. This timeline would give valuable insights into whether ratings rise first and sales follow, or vice-versa. In addition, sentiment analysis of textual reviews could reveal more detailed differences than a 10-scale rating score. Interface experiments in which users are shown cross-site averages would clarify how platform design cues affect the rating behavior.

Through the integration of spectator and press ratings, worldwide revenues, and mixed-effects modeling, this thesis shows that platform-specific rating scores are both a strategic asset and a potential liability.

References

Akerlof, G. A. (1970). The Market for 'Lemons': Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, *84*(3), 488-500. https://doi.org/10.2307/1879431

Alnefelt, P., Berg, Z., & Wahlström, H. (2007). IMDb – strategic development of a small hobby project to a large IT system. Linköping. Retrieved May 29, 2025, from https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=9d0829a9336b2ad22804c8ccde840f200ba1250f

Anderson, C. (2007). The long tail: why the future of business is selling less of more. *Choice Reviews Online, 44*(05), 44–2783. https://doi.org/10.5860/choice.44-2783

Bartok, L., & Burzler, M. A. (2020). How to assess rater rankings? A theoretical and a simulation approach using the sum of the pairwise absolute row differences (PARDs). *Journal of Statistical Theory and Practice, 14*(3). https://doi.org/10.1007/s42519-020-00103-w

Basuroy, S., Chatterjee, S., & Ravid, S. A. (2003). How Critical are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal Of Marketing*, *67*(4), 103–117. https://doi.org/10.1509/jmkg.67.4.103.18692

Basuroy, S., Ravid, S. A., Gretz, R. T., & Allen, B. J. (2019). Is everybody an expert? An investigation into the impact of professional versus user reviews on movie revenues. *Journal Of Cultural Economics*, *44*(1), 57–96. https://doi.org/10.1007/s10824-019-09350-7

Baugher, D., & Ramos, C. (2017). The Cross-Platform consistency of online user movie ratings. *Atlantic Marketing Journal*, 5(3), 121-136. https://digitalcommons.kennesaw.edu/amj/vol5/iss3/9/

Berg, M. (2017). The importance of cultural proximity in the success of Turkish dramas in Qatar. *International Journal of Communication, 11*, 3415-3430. https://ijoc.org/index.php/ijoc/article/view/6712

Caves, R. E. (2003). Contracts between art and commerce. *The Journal of Economic Perspectives, 17*(2), 73–83. https://doi.org/10.1257/089533003765888430

Chakravarty, A., Liu, Y., & Mazumdar, T. (2010). The differential effects of online Word-of-Mouth and critics' reviews on pre-release movie evaluation. *Journal of Interactive Marketing, 24*(3), 185–197. https://doi.org/10.1016/j.intmar.2010.04.001

Chen, Z., Zhu, S., Niu, Q., & Zuo, T. (2020). Knowledge discovery and recommendation with linear mixed model. *IEEE Access, 8*, 38304–38317. https://doi.org/10.1109/access.2020.2973170

Chevalier, J. A., & Mayzlin, D. (2006). The Effect of Word of Mouth on Sales: Online Book Reviews. *Journal Of Marketing Research*, *43*(3), 345–354. https://doi.org/10.1509/jmkr.43.3.345

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The Effects of Online User Reviews on Movie Box Office Performance: Accounting for Sequential Rollout and Aggregation Across Local Markets. *Marketing Science*, *29*(5), 944–957. https://doi.org/10.1287/mksc.1100.0572

Cramer, S., & Kunz, R. E. (2025). The Role of Electronic Word-of-Mouth as a Quality Signal for Film Consumption. *Jahrestagung der Fachgruppe Medienökonomie der Deutschen Gesellschaft für Publizistik-und Kommunikationswissenschaft*, 91-105. https://doi.org/10.21241/ssoar.101890

Das, P. R., & Chakrabarti, T. (2016). Application of Bayesian credibility theory in movie rankings to reduce financial risk of production houses. Parikalpana *KIIT Journal of Management, 12*(2), 95. https://doi.org/10.23862/kiit-parikalpana/2016/v12/i2/132968

Dellarocas, C., Awad, N. F., & Zhang, X. (2004). EXPLORING THE VALUE OF ONLINE REVIEWS TO ORGANIZATIONS: IMPLICATIONS FOR REVENUE FORECASTING AND PLANNING. *International Conference On Information Systems*, 379–386. http://ccs.mit.edu/dell/papers/movieratings.pdf

De Vany, A. (2004). *Hollywood Economics: How Extreme Uncertainty Shapes the Film Industry*. http://ci.nii.ac.jp/ncid/BA64429117

Duan, W., Gu, B., & Whinston, A. B. (2008). Do Online Reviews Matter? - an Empirical Investigation of Panel Data. *Decision Support Systems*, *45*(4), 1007-

1016. https://papers.ssrn.com/sol3/Delivery.cfm/SSRN_ID652123_code412577.pdf?abstractid=616262&rulid=234955&mirid=1

Efron, B. (1987). Better Bootstrap confidence intervals. *Journal of the American Statistical Association*, 82(397), 171–185. https://doi.org/10.1080/01621459.1987.10478410

Elberse, A. (2013). *Blockbusters: hit-making, risk-taking, and the big business of entertainment.* https://ci.nii.ac.jp/ncid/BB14694962

Eliashberg, J., Elberse, A., & Leenders, M. A. (2006). The Motion Picture Industry: Critical Issues in Practice, Current Research, and New Research Directions. *Marketing Science*, *25*(6), 638–661. https://doi.org/10.1287/mksc.1050.0177

Filmsite. (n.d.). The Jazz Singer (1927). Filmsite. Retrieved May 29, 2025, from https://www.filmsite.org/jazz.html

Fleder, D., & Hosanagar, K. (2009). Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity. *Management Science, 55*(5), 697–712. https://doi.org/10.1287/mnsc.1080.0974

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, *19*(3), 291–313. https://doi.org/10.1287/isre.1080.0193

Frater, P. (2001). Founders quit as AlloCine expands online ticketing. Screen Daily. Retrieved June 11, 2025, from https://www.screendaily.com/founders-quit-as-allocine-expands-online-ticketing/405046.article

Giavarina, D. (2015). Understanding Bland Altman analysis. *Biochemia Medica*, *25*(2), 141–151. https://doi.org/10.11613/bm.2015.015

Hecht, B., & Gergle, D. (2010). The tower of Babel meets web 2.0. *Proceedings Of The SIGCHI Conference On Human Factors in Computing Systems*. https://doi.org/10.1145/1753326.1753370

Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *Journal of Interactive Marketing, 18*(1), 38–52. https://doi.org/10.1002/dir.10073

Hsieh, J., & Li, Y. (2020). Will You Ever Trust the Review Website Again? The Importance of Source Credibility. *International Journal Of Electronic Commerce*, *24*(2), 255–275. https://doi.org/10.1080/10864415.2020.1715528

IMDb. (n.d.). Weighted Average Ratings. Retrieved May 13, 2025, from https://help.imdb.com/article/imdb/track-movies-tv/weighted-average-ratings/GWT2DSBYVT2F25SK#

Jansen, B. J., Jung, S., & Salminen, J. (2022). Measuring user interactions with websites: A comparison of two industry standard analytics approaches using data of 86 websites. *PLoS ONE*, *17*(5), e0268212. https://doi.org/10.1371/journal.pone.0268212

Kremer, I., Mansour, Y., & Perry, M. (2014). Implementing the "Wisdom of the Crowd". *Journal Of Political Economy*, *122*(5), 988–1012. https://doi.org/10.1086/676597

Litman, B. R., & Kohl, L. S. (1989). Predicting financial success of motion pictures: The '80s experience. *Journal Of Media Economics*, *2*(2), 35–50. https://doi.org/10.1080/08997768909358184

Liu, Y. (2006). Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue. *Journal Of Marketing*, *70*(3), 74–89. https://doi.org/10.1509/jmkg.70.3.74

Moore, S. (2015). Film Talk: An investigation into the use of viral videos in film marketing, and the impact on electric word of mouth during pre-release and opening week, *Journal of Promotional Communications*, *3*(3), 380-404. https://promotionalcommunications.org/index.php/pc/article/download/64/81/305

Moss, J. (2024). Measures of Agreement with Multiple Raters: Fréchet Variances and Inference. *Psychometrika, 89*(2), 517–541. https://doi.org/10.1007/s11336-023-09945-2

Naab, T. K., & Sehl, A. (2016). Studies of user-generated content: A systematic review. *Journalism*, *18*(10), 1256–1273. https://doi.org/10.1177/1464884916673557

Nalabandian, T., & Ireland, M. E. (2018). Genre-typical narrative arcs in films are less appealing to lay audiences and professional film critics. *Behavior Research Methods, 51*(4), 1636–1650. https://doi.org/10.3758/s13428-018-1168-7

Nalabandian, T., & Ireland, M. E. (2022). Linguistic gender congruity differentially correlates with film and novel ratings by critics and audiences. *PLoS ONE, 17*(4), 1-30. https://doi.org/10.1371/journal.pone.0248402

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, *78*(2), 311–329. https://doi.org/10.1086/259630

Panori, A. (2024). Platforms Enhancing Proximity in the Digital Era. *Platforms*, *2*(1), 1–14. https://doi.org/10.3390/platforms2010001

Prantl, D., & Prantl, M. (2018). Website traffic measurement and rankings: competitive intelligence tools examination. *International Journal of Web Information Systems*, *14*(4), 423–437. https://doi.org/10.1108/ijwis-01-2018-0001

Rao, V. R., Ravid, S. A., Gretz, R. T., Chen, J., & Basuroy, S. (2017). The impact of advertising content on movie revenues. *Marketing Letters*, *28*(3), 341–355. https://doi.org/10.1007/s11002-017-9418-5

Rochet, J., & Tirole, J. (2003). Platform competition in Two-Sided markets. *Journal of the European Economic Association*, *1*(4), 990–1029. https://doi.org/10.1162/154247603322493212

Sellier, G. (2010). French New Wave Cinema and the Legacy of Male Libertinage. *Cinema Journal*, *49*(4), 152–158. https://doi.org/10.1353/cj.2010.0007

Simonton, D. K. (2004). Film Awards as Indicators of Cinematic Creativity and Achievement: A quantitative comparison of the Oscars and six alternatives. *Creativity Research Journal, 16*(2), 163–172. https://escholarship.org/content/qt6kp6d661/qt6kp6d661.pdf?t=lnqyoa

Sims, D. (2017). Netflix Believes in the Power of Thumbs. Retrieved June 1, 2025, from https://www.theatlantic.com/entertainment/archive/2017/03/netflix-believes-in-the-power-of-thumbs/520242/?utm_source=chatgpt.com

Straubhaar, J. (2003). Choosing National TV: Cultural Capital, Language, and Cultural Proximity in Brazil. In *Routledge eBooks* (pp. 77–110). Https://doi.org/10.4324/9781410607041-6

Straubhaar, J. D. (1991). Beyond media imperialism: Assymetrical interdependence and cultural proximity. *Critical Studies in Mass Communication*, *8*(1), 39–59. https://doi.org/10.1080/15295039109366779

Sunder, S., Kim, K. H., & Yorkston, E. A. (2019). What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions. *Journal of Marketing, 83*(6), 93–112. https://doi.org/10.1177/0022242919875688

Walls, W. D. (2005). Modeling Movie Success When 'Nobody Knows Anything': Conditional Stable-Distribution Analysis of Film Returns. *Journal Of Cultural Economics*, *29*(3), 177–190. https://doi.org/10.1007/s10824-005-1156-5

Wilson, E. J., & Sherrell, D. L. (1993). Source effects in communication and persuasion research: A meta-analysis of effect size. *Journal Of The Academy Of Marketing Science*, *21*(2), 101–112. https://doi.org/10.1007/bf02894421

Wu, L., Ren, Z., Ren, X., Zhang, J., & Lü, L. (2018). Eliminating the effect of rating bias on reputation systems. *Complexity, 2018*(1). https://doi.org/10.1155/2018/4325016

Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2012). Modeling Consumer Learning from Online Product Reviews. *Marketing Science*, *32*(1), 153–169. https://doi.org/10.1287/mksc.1120.0755

Zhao, L. (2022). AlloCiné movies databases : matching data using semi automatic algorithm. Medium. https://webedia.io/allocin%C3%A9-movies-databases-matching-data-using-semi-automatic-algorithm-7606c120670b?gi=758cd8f968aa&utm_source=chatgpt.com