# The Impact of Online Word of Mouth on Consumer Decision-Making
—
# The Case of Cinema Audiences

Richard P. D. Wallace



A Thesis Submitted for the Degree of Doctor of Philosophy

London Metropolitan Business School

London Metropolitan University

December 2012

This thesis contributes to the knowledge on the impact of word of mouth on consumer decisions and the manner by which consumers learn from it. Using a sample of 132 motion pictures released between April and September 2010 and more than 38,000 online ratings from a social network, word of mouth is clustered into two dimensions: volume, representing the amount of online posts, and valence, representing the aggregate opinion of consumers on a particular film. A novel approach to calculate the valence measure is developed in order to account for different ways, in which ratings may be interpreted. Mixed-Effects Methods are used to create a parsimonious model accounting for the systematic variation of clusters of films within the data around the population mean.

The results show that the volume of word of mouth positively affects consumer decisions, indicating that they engage in observational learning. On the other hand, the valence of word of mouth is insignificant, meaning that the qualitative evaluation of motion pictures by consumers does not influence subsequent audience decisions. These findings are attributed to the peculiar nature of motion pictures, as they are unique experience goods commonly only consumed once, have a very short life cycle, and are heavily marketing-driven, leading to a rapid decline in revenues after their opening.

*Consumer Learning, Decision-Making, Experience Goods, Mixed-Effects Models, Online Word of Mouth*

# Acknowledgements

First and foremost, I want to thank my supervisor, John Sedgwick, for consistently providing me with inspiration and valuable feedback over the past three years. Both his dedication to research and his empathy can only be described as exceptional, and I greatly value the relationship we have developed. Vlasios Voudouris both pushed me enough to get to grips with the statistical methods and tools I used and provided me with helpful explanations on technical and statistical issues. As the third person on my supervisory team, Thierry Rayna shared various valuable thoughts on theoretical ideas with me. I am further grateful to Dimitrios Stasinopoulos for introducing me to mixed-effects models, thus showing me a way to analyse my data after initially struggling to apply GAMLSS.

I clearly owe debts to the VC scholarship funding, without which this research would not have been possible. No less important though was the collaborative and friendly atmosphere amongst the PhD fellows; I have met some great people at London Metropolitan University, and I very much enjoyed every single coffee or beer shared while discussing the downsides of life as a PhD student. Thanks to Abdul Kader Aljandali, Daniel Opajemi Amund, Hélène Draux, Özlem Edizel, Thomas Goda, Maryam Hamedi, Cristina Maxim, Mohammad Khaleq Newaz, Anna Nurzynska, and Nikki Zanti.

Finally, I am very thankful to my family for their continuous and unquestioning support on whatever journey I embark upon and to Barbara Müller for her love, compassion and patience with me over the last three years.

# Contents

# List of Figures

# List of Tables

# 1. Introduction

> "The expectancy level of entertainment is a term describing the moviegoer's degree of anticipated enjoyment of the motion picture [...] he is planning to see. This expectancy level may have been contrived in the subject's mind by considering the stars, title, and story type, but more often it will also be affected by advertising, professional critics, and word-of-mouth publicity." (Handel, 1953)

Consumers form expectations about motion pictures in very distinct ways; emotionally from their own past experiences of pleasure and arousal during the consumption of a film, and cognitively by assessing various sources of information. They anticipate how much pleasure they will experience or how scared they will be during the consumption of a film. These expectations, however, are often disconfirmed; they can be disappointed, met or exceeded. A film may be funnier than expected or not as scary as hoped for. The extent to which consumer expectation is disconfirmed and the amount of experienced pleasure and arousal determine consumer satisfaction. Over time, people memorise and learn from these experiences and adjust their expectations of new movies accordingly.

Not only do film consumers learn from their own emotive experience, they also learn cognitively from various sources of information, usually before they go to see a particular film. This information consists of film attributes, such as the cast, the director or the genre of the motion picture, marketing information released by the production studio and the distributor, and word of mouth from individuals who have already seen the film. Using this information, individuals form their expectations and make decisions on whether to see a particular film or not.

Yet, the expectations that consumers form can never be accurate due to the nature of motion pictures. Every film is a unique product differing in some characteristics from any other film. The remake of an old film, for example, would very likely contain a different cast and new technological features giving the remake a different look and feel. Further, the marginal utility of films decreases rapidly; the value of consuming a particular motion picture a second time will generally be much lower than watching an unknown film. Therefore, new films are released onto the market every week, thereby continually providing consumers with new choice options.

For these reasons, motion pictures are experience goods; their utility can only be fully assessed after watching them. This is in stark contrast to search goods, which Nelson (1970) defines as goods whose quality can be assessed precisely and cheaply prior to purchase. Consumers can compare different brands and select the one with the highest expected utility. The quality of experience goods, on the other hand, is either very costly or impossible to assess through search; consumers have to rely on their personal experience and are only able to fully evaluate a product after purchase.

Due to motion pictures' uniqueness, consumers take risks every time they decide to go to the cinema. This risk could be reduced by repeat-watching a film of high quality; however, due to the rapid decline of marginal utility consumers rarely engage in this behaviour and rather search for novelty. Therefore, motion pictures can be classified as a special kind of experience good, since they are usually only consumed once.

The search for novelty paired with the steady release of new motion pictures to the market is apt to lead to a constant change in popular films. Indeed, this is what can be found when looking at the most successful films in terms of revenues on a weekly basis. A particular film only rarely is in the top spot for more than two successive weeks and it usually drops out of the top ranks quickly after its release. Figure 1.1 depicts this behaviour using a data set comprised of 185 motion pictures and their respective weekly rank according to their weekend revenues between April 1998 and March 1999. Most films face a steep decline in ranks; *Saving Private Ryan* ranks at the top position for four successive weeks, longer than any other film in the chosen period. This emphasises another characteristic of motion pictures, namely that they commonly have a very short life cycle of only a few weeks.

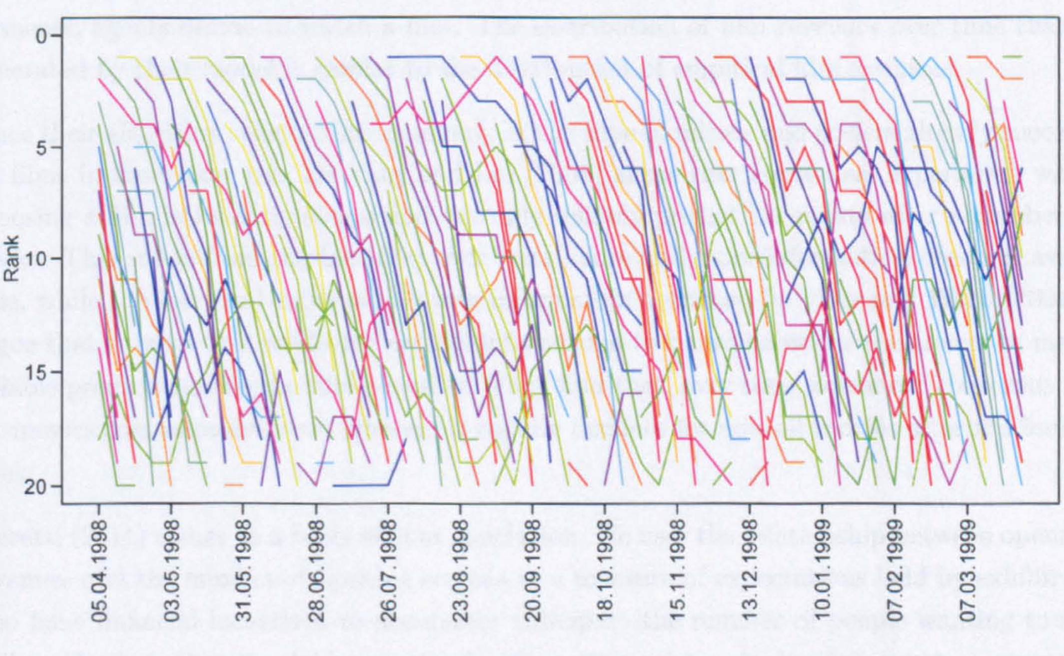



Figure 1.1.: Rank of different films over time

In order to reduce their risk prior to consumption, consumers try to gather reliable information from other sources. A strand of research undertaken especially by De Vany and Walls

(1996, 1999, 2004b), De Vany and Lee (2001), Moretti (2011), and Moul (2007) suggests that word of mouth is such a source. Individuals who have already seen a particular film can give their peers a detailed and fairly reliable assessment of its quality. These peers, in turn, can use feedback from multiple sources to form their expectations and decide whether to watch the film or not. Thus, over time, expectations of later viewers change accordingly to the quality attributed to films by early filmgoers. A film receiving positive reviews in its early release period is likely to attract more filmgoers in subsequent weeks. Moul (2007, p. 859) states that this dynamic can explain "10% of the variation in consumer expectations".

To find out whether this is the typical behaviour of film consumers, De Vany and Walls (1996, 1999) model information transmission between consumers. They conclude that the empirical distribution of motion picture revenues over time can be caused by dynamics, in which consumers discover what they like by choosing a film according to the proportion of consumers in previous weeks. These dynamics are further able to produce information cascades and so-called "superstars" (i.e. successful blockbusters, Rosen, 1981), phenomena which can be found at the box office. However, an unresolved question in these studies is whether individuals actually learn from a qualitative feedback or whether they watch a film because many others have already seen it, thereby simply following the crowd.

De Vany and Lee (2001) argue that film consumers learn from qualitative feedback. In order to account for information on film quality, they create a model, in which agents, besides having a personal opinion or taste, receive either positive or negative feedback on a particular film. Additionally, they place trust in each of their peers' assessments. Depending on these three elements, agents decide to watch a film. The distribution of film revenues over time that is generated by their model is similar to the distribution of empirical film revenues.

Since their algorithm selects films randomly at the time of release and prefers already successful films in their later run, De Vany and Lee (2001) argue that consumers experiment when choosing a film in its early stages and can only reliably predict its quality after a number of weeks. This stems from the fact that over time, more and more information becomes available, while the marginal value of this new information declines. De Vany and Walls (2004b) argue that it takes four weeks for enough information to be available for consumers to make reliable predictions about a film's quality. They find that, over time, consumer consensus for hit movies continuously grows whereas it rapidly declines for non-hit movies after the fourth week.

Moretti (2011) comes to a fairly similar conclusion. He uses the relationship between opening revenues and the number of opening screens as a measure of expectations held by exhibitors who have financial incentives to accurately anticipate the number of people wanting to see a film. He finds that "positive surprises" – films that achieve higher income than expected in their opening week – face a slower decline in weekly revenues than "negative surprises". Moretti (2011) assumes that the former type of films cause positive word of mouth to spread due to their "underlying quality", whereas the latter type of films generate negative word of mouth. The distinct sales patterns are attributed to a process, in which consumers learn from others about the quality of a film and subsequently adapt their expectations and purchases. Individuals that hear about the poor quality of a film are less likely to watch it in any of the subsequent weeks of its run.

However, Moretti (2011) differs markedly from De Vany and Lee (2001) and De Vany and Walls (2004b) in two points; he suggests that consumers somehow already 'know' prior to release whether a particular film will please them or not, thereby causing positive and negative surprises to occur in the opening week. Further, the positive or negative word of mouth arising from these surprises determines a film's subsequent sales trajectory. This supports the often-used quote of producer Robert Evans who stated that if a film doesn't open, it is dead. In contrast, De Vany and Lee (2001) and De Vany and Walls (2004b) believe that consumers experiment in the early weeks and only learn about the true quality of a film over time. According to them, it takes four weeks until sufficient information is disseminated and consumers are able to make precise and well-informed decisions.

Nevertheless, despite these differences, they agree on the importance of word of mouth and stress its impact on consumer learning at the box office. A major issue regarding this literature arises when looking at the empirical revenue statistics for most films; the largest number of consumers decides to see a film in its opening week. Figure 1.2 uses the same data set of motion pictures released between April 1998 and March 1999 as Figure 1.1 and depicts their average weekly revenue over the first 15 weeks of their release. It shows that the diffusion process among the population declines exponentially after a film's initial release. While the 'average' film generates more than \$10 million in its opening week, income drops to \$6.8 million in the second week and further to \$3.6 million by the fourth week. It therefore remains questionable to what extent consumers actually learn about the quality of a film; there seems to be very limited time for it to play a significant role in the diffusion process of motion pictures.

The classical models of word of mouth and product adoption assume that sales build up over time meaning that the rate of adoption increases until a point of satiation is reached when roughly half of the potential population have consumed a product. After this point, the rate of adoption decreases again. Rogers' (1983) S-shaped adoption curve is the classical example of this phenomenon. However, looking at the 'typical' sales trend of motion pictures, they clearly do not build up over time (with the exception of the occasional 'sleeper'). Instead, they taper off very quickly and rather behave like a waterfall.

As stated, De Vany claims that consumers can only make precise predictions about a film's quality after four weeks. If this is the case, then there are only very few potential consumers left to make 'perfectly informed' decisions. The majority guesses what it is that they like or experiments during the early weeks of a film. However, if consumers in these cases choose motion pictures at random as hypothesised in the model by De Vany and Lee (2001), then it

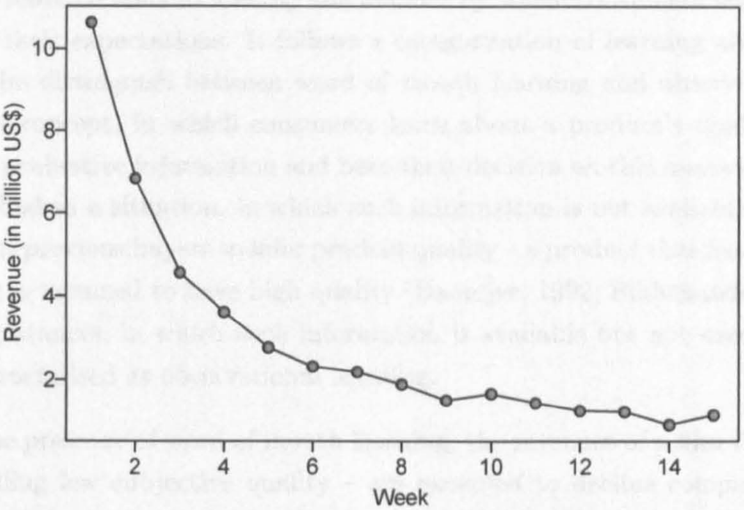



Figure 1.2.: Revenue distribution of the 'average' film over time

remains questionable how hits can arise in the early weeks of a film's release. They should be choosing each film with equal probability.

Moretti, on the other hand, reasons that a motion picture's fate is decided after the opening week. Once a film has generated a positive or a negative surprise (and thus, positive or negative word of mouth), its subsequent sales decline slower or faster, respectively. In this case, consumers already know prior to the release of a film whether it resonates with their personal taste and decide to watch it. A film that manages to appeal to a large audience is therefore likely to become a box office hit. It then becomes debatable whether consumers in subsequent weeks actually learn from peers' qualitative assessments of a film or whether they simply follow the large crowds.

The purpose of this thesis is to investigate consumer learning at the box office and to analyse the impact of word of mouth on consumer decisions. Whereas previous studies of consumer behaviour at the cinema have usually used box office revenue data and made inferences from its statistical distribution, this study follows a more recent strand of research and uses actual word of mouth data from an online social network focussing on motion pictures expressed through film ratings.

The source was chosen, because individuals interacting with a social network website are likely to be film-savvy consumers who are both more interested in the release of information about films and more likely to spread the word about films they have seen. Therefore, learning from others can be expected to be present among this population. Consumers seeing a particular film later in its run are likely to have read and reacted to ratings and reviews posted by earlier filmgoers. Consequently, the rating data provide both a direct measure of the buzz around a film and a direct measure of subjective film quality.

Especially, this research aims to identify the manner by which consumers learn from word of mouth to form their expectations. It follows a categorisation of learning also used by Chen et al. (2011) who distinguish between word of mouth learning and observational learning. The former is a concept, in which consumers learn about a product's quality through the transmission of evaluative information and base their decision on this assessment. The latter has been described as a situation, in which such information is not available and consumers use the actions of previous buyers to infer product quality – a product that has been purchased by many people is assumed to have high quality (Banerjee, 1992; Bikhchandani et al., 1992). However, circumstances, in which such information is available but not used by consumers, can also be characterised as observational learning.

Therefore, in the presence of word of mouth learning, the revenues of a film that receives bad ratings – signalling low subjective quality – are expected to decline comparatively quickly in contrast to a film that receives good ratings. This will also be reflected in the rating data itself. Films receiving low ratings in the opening week are expected to receive a rapidly declining volume of ratings, since fewer consumers go to see them in subsequent weeks. If, however, qualitative assessments of motion pictures by peers do not play a significant role and consumers learn through observational learning, then the 'buzz' surrounding a film will prove to be a valid indicator of film success, since films that appeal to a large audience can be expected to generate a comparatively larger volume of ratings and reviews.

This study contributes to the research on the impact of word of mouth on film consumers. While previous research interested in consumer learning has mostly used statistical methods to analyse the distribution of film revenues and inferred from these that word of mouth has an impact on individuals' decisions, this study uses actual word of mouth data submitted by film consumers. It is thus able to determine the manner by which individuals learn from others.

It further attempts to solve a conflict in the literature regarding the amount of time it takes for consumers to make 'well-informed' decisions. Whereas Moretti (2011) assumes that they are able to assess a film's quality prior to release, De Vany and Walls (2004b) argue that it takes about four weeks until enough information is disseminated. This is an important issue considering the short life cycle of motion pictures.

The thesis contrasts previous research on film consumers employing online word of mouth in two ways; first, prior authors have used the arithmetic mean as a qualitative measure of film ratings (Chintagunta et al., 2010; Duan et al., 2008b). However, ratings are 'fuzzy' data meaning that a rating of four stars does not necessarily indicate that a particular film is perceived to be twice as good as another film that received a rating of two stars. Thus, using the arithmetic mean can lead to arbitrary measures of film quality. This study accounts for the fuzziness of rating data by creating an alternative consensus measure of consumer opinion termed Movie Preference Index (MPI).

Second, prior research has been more interested in using word of mouth to forecast film revenues than in investigating how film consumers integrate word of mouth into their decision-making (Asur and Huberman, 2010; Rui et al., 2011). By distinguishing whether consumers rather tend to engage in observational learning or word of mouth learning, this study is able to make inferences about individuals' behaviour in risky consumption situations. This issue is of interest to film marketers as well, as they can determine whether it is more profitable to invest in film quality or in the creation of buzz.

The remainder of the thesis is structured as follows. Chapter 2 first reviews the literature on the different kinds of consumer learning. Following, the different sources of expectation formation – (online) word of mouth and motion picture characteristics, such as genre, film stars and advertising – are discussed.

Chapter 3 discusses the methodology used for the research based on empirical data. First, the data set is introduced and its features highlighted. Drawing on these, a new measure of aggregate consumer opinion, the MPI, is developed. Finally, the rationale for using mixed-effects methods and the method of model selection are explained.

Chapter 4 presents the results of the data analysis. The selection of the final model is explained step by step and its model fit analysed. Subsequently, the results of the model coefficients are interpreted. This leads on to Chapter 5 discussing the significance of the results and further directions for research emerging from these. Chapter 6 concludes.

# 2. Literature Review

> "The movie industry is still the only major business in the United States which has never made a serious attempt to study its potential market" (Handel, 1953)

It can be argued whether Handel was sincerely concerned about the research undertaken by the film industry or whether he made this statement out of self-interest to further promote his Motion Picture Research Bureau, as Maltby (1999) suggests. In fact, quite some research had been carried out before the 1950s to find out more about film audiences. The earliest forms of research were usually conducted by the cinema-owners themselves by investigating their sales. In the 1920s and 1930s, film producers would analyse fanmail they received or ask readers of magazines to send letters on who their favourite stars were (Bakker, 2003).

When national distributors became the link between film producers and exhibitors, they let the former know what kind of motion pictures were currently in demand. The film producers, in turn, then tried to create strategies for success by marketing sequels or stars, but largely these efforts were still intuitive (Bakker, 2003).

This changed after the huge success of *Gone With The Wind* in 1939, when the first two independent market research firms, the Audience Research Institute and the Motion Picture Research Bureau, were set up in the United States. They represented the first efforts to empirically study Hollywood audiences. Audience Research was the first to introduce systematic sampling methods, and it segmented audiences into age, gender, location and income (Ohmer, 1999).

Their research mainly focussed on attendance frequencies and the composition of the audience with regards to gender, age and social class (Sedgwick and Pokorny, 2010) in order to classify it into different 'taste publics' (Maltby, 1999). It was found that audiences are generally younger, that a huge proportion is teenagers and that a large part of cinema-goers came from lower social classes. Men and women attended about equally often (Lazarsfeld, 1947), but they preferred different kinds of films (Ohmer, 1999).

As film production became more expensive and took more time, market research firms became more and more influential throughout the 1940s. They advised film producers on the kind of films to make, so that they would appeal to an audience as wide as possible. This included the kind of stories they should tell, which title to use for the film, which stars to employ and how to advertise their films (Bakker, 2003).

Nevertheless, Lazarsfeld (1947) argued that there was not much knowledge with regards to the preferences of different groups of people. Generally, people wanted "to hear about themselves" (ibid, p. 166) in the stories a film told, but the extent to which this helped film producers to make films that appealed to different audiences was rather limited.

Since the 1950s, different disciplines have looked at film audiences, from mass communication studies to film studies and economics. Especially film studies have produced a large volume of literature, to a large extent concerned with seeing film as a text set within a wider social context and the ways, in which individual viewers interpret this text. They have often used the audience of a particular film or a particular 'type' of film for their analysis. Examples of these are Barker and Brooks' (1998) investigation of audiences of *Judge Dredd* or Cherry's (1999) analysis of female audiences of horror films.

Economists' involvement and interest in film audiences has only developed recently. Mostly, they were concerned with the risk environment, in which film producers have to operate due to the enormous amount of sunk costs, and the ways, in which they account for these risks. For example, Sedgwick and Pokorny (1998) discuss how production companies started building portfolios, and De Vany and Walls (1996) show how exhibitor contracts dynamically adapt to the demand for particular films. Due to the growing costs of producing motion pictures, many economists have tried to predict revenue outcomes depending on the inputs of a film. Hadida (2009) provides a good overview of the various factors that have been investigated.

The earliest economists concerned with film audiences were De Vany and Walls (1996) who find that films with a large early attendance have a higher chance of being successful conjecturing that "customers choose movies in proportion to the previous film goers who selected that movie" (ibid, p. 1509). This is due to consumers sharing information with each other and discovering the utility a film is likely to have. This research has been followed up by various other studies (e.g. De Vany and Lee, 2001; Hidalgo R et al., 2006; Moretti, 2011) showing how audiences use different kinds of information to make decisions on film consumption. More recently, the growth in online technologies and the emergence of social networks has led researchers to use online word of mouth to investigate the same phenomenon. They have analysed different sources such as Yahoo! Movies (e.g. Chintagunta et al., 2010; Liu, 2006) and Twitter (e.g. Asur and Huberman, 2010; Rui et al., 2011) to forecast box office revenues from early consumer opinions on particular films.

Alongside this, findings from studies in consumer science and psychology show how consumers learn both internally through experience and externally by assessing various sources of information. For example, it has been shown that individuals weigh losses higher than gains and may therefore adapt different strategies of risk aversion and risk taking (Kahneman and Tversky, 1979). Therefore, negative information received prior to the consumption of a motion picture can significantly reduce its appreciation while the same amount of positive information does not increase pleasure to the same extent (Burzynski and Bayer, 1977).

Thus, even if one may argue that Handel was right in 1953, the following years have produced a significant amount of literature on film consumers, so that at least some attempt on studying and understanding Hollywood's audiences has been made. The following chapters review this literature.

## 2.1. Consumer Learning and Choice under Uncertainty

The purchase of any kind of consumer good involves information asymmetry meaning that one party involved in the transaction has more information about the good than the other (Akerlof, 1970). In most cases, the seller knows more about the product than the buyer such as when selling/buying a car or a computer. This can lead to opportunistic behaviour, for example the producer misrepresenting product quality via advertising (Mishra et al., 1998). In order to reduce this asymmetry, consumers generally try to find out as much as possible about a product prior to purchase with the intention of assessing product quality and reducing the risk of making an unsatisfactory purchase.

Consumers can adopt two basic strategies to find out more about a product: search and experience (Nelson, 1970). They will adopt a search strategy, if product characteristics can be searched for in a costless manner or as long as the cost of searching is lower than its utility. For example, information about a pullover can be relatively cheaply gathered by trying it on in a store, looking at the price tag and the label of the producer. Thus, uncertainty about product quality is considerably reduced. Products that predominantly possess search attributes are generally referred to as search goods.

In contrast, when the price of a product is generally very low, consumers are likely to adopt an experience strategy. For example, a consumer is unlikely to spend major effort attempting to assess the quality of canned vegetables through a search strategy; she will rather buy them and 'experience' whether they meet her expectations.

An experience strategy will also be adopted if search is very costly. In order to assess the quality of a car service, for example, and effectively compare it with its competitors, an effective search would need to be substantial. A consumer may therefore rely on her own or other people's experience in order to make a decision. In cases when even the most expensive and detailed search cannot lead to a reliable judgment of product quality prior to purchase, an experience strategy will have to be adopted. Motion pictures and books, for example, belong to the kind of products that often require consumers to use their experience in making a decision and are therefore referred to by economists as 'experience goods'.[1]

[1] It needs to be noted that the distinction between search and experience goods is not always clear-cut. Appliances such as dish-washers possess both search and experience attributes leading the consumer to adopt a mixed strategy in order to reduce information asymmetry prior to purchase.

Thus, purchasing experience goods is generally risky; since consumers cannot assess product quality reliably before buying, a disappointment is not unlikely. However, this risk can be reduced if an experience good is repeat-purchased. This is the case for most utility products, such as canned vegetables or washing powder. A consumer using an experience strategy will test a number of different brands until she has found the one with the highest subjective benefit and subsequently keep on buying it. This kind of learning has been labelled one of the most important processes in reducing risk (Sheth and Venkatesan, 1968).

This situation changes, though, in the case of most hedonic goods, which have a quickly declining marginal utility and are therefore unlikely to be repeat-purchased. Motion pictures are a prime example of this kind of product. It is impossible for a consumer to fully assess the quality and subjective utility of a film prior to consumption. Further, watching a particular film for the second time provides her with a much lower utility than watching an unknown film for the first time (Sedgwick and Pokorny, 2005). And because watching a film is "multisensory", consumers of hedonic products generally look for novelty (cf. Hirschman and Holbrook, 1982; Holbrook and Hirschman, 1982). This is the reason why people only rarely go to see a motion picture at the cinema twice, and a similar argument can be applied to novels.

However, there are some exceptions. An entertainment park, for example, is a hedonic product, which is consumed for enjoyment. Similarly to books or motion pictures, it is difficult to judge prior to a visit how much personal utility will be gained using a search strategy. This is only known after the experience has been made. Yet, in contrast to motion pictures or novels, a consumer may go and visit an entertainment park more than once. While it may not provide a new experience on subsequent visits – the features of the park are now known – she is still likely to enjoy it. Table 2.1 provides an overview of this classification of goods.

| Product Type | Product Attribute | Purchase Frequency | Example |
|---|---|---|---|
| Search | Utility | Repeat | Clothes |
| Experience | Utility | Repeat | Canned Vegetables |
| Experience | Hedonic | Repeat | Entertainment Park |
| Experience | Hedonic | Single | Motion Pictures |

Table 2.1.: A simplified classification of goods

Hedonic goods are generally innovations with distinctive features, which further contributes to the difficulty of assessing their quality prior to purchase (Clement et al., 2006). Every motion picture is unique; although it may share characteristics with other films, some features will not have been shown in the same way before. A sequel may feature the same cast and the same director, but it will tell a different story than the first film. A remake will tell a very similar story, but a different director and cast may adopt an altered view on which parts of the story to emphasise. Technology may have changed and thus visual effects may be stressed. Again, a similar argument can be made for books or music, where some constituent parts of the product are known, but the combination is novel. Therefore, two hedonic products are never the same leading consumers to make new experiences every time they make a purchase decision. It also means that the knowledge gained about a particular product is not necessarily a reliable source of information to base future consumption decisions on.This makes single-purchase hedonic products very risky purchases.

To further complicate decision-making, unlike most goods, motion pictures do not distinguish themselves from each other through price variation, a signal that is often used to indicate the quality of a particular product (Sedgwick and Pokorny, 2005). At any given cinema, prices do not vary between different films.[2] Therefore, consumers cannot use price information to assess the quality of a particular film. They face a risky decision between various unknown products again and again.

With more than 400 films released every year over the past decade in the North American market (MPAA, 2012), consumers face an overwhelming supply. It would be reasonable to assume that it is difficult to make decisions in a situation like this. Generally though, people seem to either 'know' before a film is released whether it will appeal to their personal taste or learn very quickly whether they want to watch it or not. Evidence for this is twofold: most films have a fairly short life cycle and do not run in cinemas for more than eight weeks. Further, most films generate the largest share of their revenues in the first week of their release. After that, revenues generally decline and flatten out over subsequent weeks. Only about ten per cent of the motion pictures released manage to build up an audience over time and increase their revenues in later weeks.

Hence, motion pictures commonly escape the classic inverse u-shaped adoption curve that applies to many innovations and new products (Rogers, 1983). Their adoption curve rather describes a waterfall that drops off a cliff, with the height of the cliff and the steepness of the drop defining the commercial success of the film. This can lead to very different revenue distribution curves, where 'blockbusters' generate large revenues in their opening week, which only decline comparatively slowly over time, and 'bombs' have a large audience in the first week, which quickly disappears in subsequent weeks. Yet, there are a few exceptions, so-called 'sleepers', that manage to build up an audience over time.

Economists have tried to model the underlying learning and decision processes in order to attain these observed outcomes. Building upon the pioneering work of Bandura (1977), both Banerjee (1992) and Bikhchandani et al. (1992) provide simple models of observational learning. In the simplest case, agents face a binary decision problem, e.g. adopting or rejecting an innovation. They hold a private signal whether it is advantageous for them to adopt or reject, which may or may not be correct. Agents make their choice sequentially and cannot delay their decision. Agents can further observe the decisions made by earlier decision-makers, so that every agent except the first receives two signals, her own prior and

[2] However, cinemas may charge different prices for 'premium seats' or similar.

the choices made by others. However, agents cannot observe the private signal of previous agents, and they do not get to know about the payoff of their decision.

Making the assumption that agents weigh their own prior against the decisions made by preceding agents, and further assuming that if there is a 'tie' between options, an agent will always follow her private signal, the models show that, over time, herding will occur, meaning that agents flock to a single option. In the simplest example featuring a decision between two alternatives, the first agent will always follow her private signal. The second agent now receives two signals, her own private signal and the decision made by the first agent. If her signal corresponds with the decision, she will choose the same option; if it differs, she will follow her private signal. The third agent will face one of two options: either the preceding agents have chosen two different options, a situation similar to the one faced by the first agent, in which case she will follow her private signal. Alternatively, the preceding agents have both chosen the same option, in which case she will choose the same option regardless of her private signal; the two signals from preceding agents override her own private signal. Thus, an information cascade is started, and all subsequent agents will choose the same option.

Banerjee (1992) develops the model to show that herding occurs even if agents face a choice amongst a number of options or if some agents do not hold a private signal. In order to test the robustness of their model, Bikhchandani et al. (1992) relax the assumptions made about the agents' payoffs as well as the accuracy of their private signal, still to find that herding nevertheless arises over time.

However, these cascades are very fragile. Once a cascade has started, the information that an agent provides to her follower through her choice is uninformative, since the agent's private signal was overridden by the choices of earlier agents. Even if the agent had a private signal to make a particular choice, the fact that one option had been picked by multiple other agents led her to the rational conclusion that this option is likely to have a higher payoff. Therefore, if new information favouring an alternative choice becomes available at this point, for example through public release, this can lead to the breakdown of the cascade and potentially start a new one at a different option.

This behaviour of consumers has also been shown in laboratory experiments. Anderson and Holt (1997) create a setting, in which subjects have to choose between two different events, $A$ and $B$, represented by two urns; urn $A$ contains two balls $a$ and one ball $b$ while urn $B$ contains one ball $a$ and two balls $b$. The initial probability of drawing balls $a$ and $b$ is therefore equal at $\frac{1}{2}$, since there are three balls $a$ and three balls $b$. First, a group of subjects is randomly assigned an urn, from which one ball is drawn, shown to the first subject, and replaced. The subject thus receives an informative private signal about the urn, which is correct with a probability of $\frac{2}{3}$. She then makes a decision between $A$ and $B$, which is announced to all other subjects. This procedure is sequentially repeated for the remaining group. Consequently, every subject but the first receives two signals - her private signal drawn from the assigned urn, and a public signal about the decision of previous subjects. Using six decision-makers in each round of this procedure, Anderson and Holt's (1997) results are fairly consistent with the prediction of herding models; information cascades occurred in more than 73 per cent of the experiments.

In this experiment, subjects only have to choose between two alternatives and there is an obvious optimum choice. In real-life decisions, however, there are often a multitude of options, and an option with a maximum payoff for everyone may not exist. In the case of hedonic products, for example, different tastes or preferences affect both the decision and the utility gained from it. The difficulty of making a choice also increases with the number of options available, if product or service attributes are difficult to process and if there is a large amount of uncertainty regarding the quality of attributes (Bettman et al., 1991). This has been shown to be the case for products such as books or motion pictures. Consumers may therefore use a different learning strategy.

Consumers may decide to use their personal experience from previous purchases in order to assess the quality of a product. It has been shown that this strategy is especially valuable for durable consumer goods or goods that are repeatedly purchased (Arndt and May, 1981). However, McFadden and Train (1996) argue that in the case of single-purchase products such as motion pictures or books, the experience gained cannot be used for future purchases. It is debatable, though, whether the personal experience made with a particular 'kind of film' cannot be used for future film-going decisions.

However, another option for gathering information about a product is learning from others. Especially word of mouth from peers has been attributed to be valuable and have an influence on the consumption decision. Investigating the consumption of motion pictures, Faber and O'Guinn (1984) ask students about the usefulness, importance, frequency of consultation, credibility and impact on the decision-making of various information sources. While film previews are perceived as the most important and the most useful sources, comments from peers are more credible and more frequently consulted. The largest impact on the decision-making process though is accredited to film previews.

Similarly, Mahajan et al. (1984) question students about their information sources and find that, over a time period of ten weeks, consistently more than one third stated that they ask their friends for information about a film indicating that word of mouth indeed is a very important source of information for hedonic products. These results have to be taken with care though, since self-reported influences cannot always be trusted.

Therefore, researchers have used different methods to investigate consumer learning in the case of hedonic products. De Vany and Walls (1996), for example, attempt to model the underlying processes and dynamics of film revenue distribution curves, as they cannot be described by the simple models of observational learning. They assume that the quality of the product is unknown prior to consumption and that consumer evaluations differ. The utility that a film-goer gains consists of a 'common film quality' and an individual deviation from this quality depending on the individual's taste. The cost of seeing a film is equal for each individual, and consumers decide to watch it if their expected utility is higher than this cost. Further, it is assumed that film-goers update their initial evaluation of a film using information from various sources, such as film reviews, advertising and word of mouth in a "sequential Bayesian decision process" (ibid, p. 1498). Essentially, consumers rely on decision-makers ahead of them to update their estimate of the 'common film quality'. This decision process leads to the reinforcement of prior outcomes; if a film was successful in its opening week, it will be successful in later weeks, because it reveals positive information about the film to later consumers.

This process also reflects the Bose-Einstein process leading to a Pareto distribution, meaning that for every film each revenue outcome is, initially, equally likely. However, the choice probabilities of consumers develop in a reinforcing manner, implying that once differences in the quality of films emerge, they can grow at exponential speed leading to the huge successes observed in the motion picture market. This is due to the information feedback from early consumers to later ones. With regards to the release strategy of motion pictures, this means that a wide release can lead to a huge success, because early positive information leads consumers to watch a particular film. At the same time, though, spread of negative information can make a film disappear quickly from cinema screens.

In order to illustrate how this process compares to observational learning, De Vany and Lee (2001) build upon the model created by Bikhchandani et al. (1992). In their model, every agent holds a private signal about the quality of the product, which may or may not be correct. They make their decisions sequentially and observe the choices made by previous agents. However, they also receive a quality signal from previous agents, which is either positive or negative. This quality signal can either be trusted or disapproved of; the trust that an agent places in a previous agent's quality evaluation depends on the market share of the product. De Vany and Lee (2001) argue that an individual will trust an evaluation more if it is confirmed by many further choices. However, this leads to a situation, in which information is only significant when the product has a high market share.

This is similar to a study carried out by Chen et al. (2011), who use a natural experiment setting of sales of digital cameras on Amazon, and distinguish between observational learning and word of mouth learning. They define that, on the website of each product, the section displaying the top-five products in a category and stating "X% out of Y people bought this product" influences observational learning in a similar manner to the models of Banerjee (1992) and Bikhchandani et al. (1992), where agents can observe the actions taken by previous agents. They further define that word of mouth learning is apparent if product ratings displayed on the same website influence product sales. Since Amazon removed the section on observational learning for a couple of months, Chen et al. (2011) are able to study the different effects of observational learning and word of mouth learning.

They show that positive observational learning information, that is a large percentage of people purchasing a particular product, positively influences product sales. However, negative observational learning information, that is a low percentage of people purchasing a particular product, has no influence on future product sales. It is argued that "observational learning information contains the discrete signals expressed by the actions of other consumers but not the reasons behind their actions" (Chen et al., 2011, p. 240). Therefore, negative observational learning information is not informative, since a product adopted by only few people may simply be a niche product of high quality instead of a bad product. Positive observational learning information, on the contrary, shows that a product appeals to a broad population and therefore has at least decent quality. Word of mouth information, in contrast, has the opposite effect: negative product ratings hurt sales much more than positive ratings help them.

To show how this kind of learning behaviour affects the market share of different products, De Vany and Lee (2001) use their model to run multiple simulations with only five products and varying product quality. In a market with one bad product and four good products, the bad product will generally fail, because negative information spread will lead consumers to choose another product. However, if its initial market share is high enough, observational learning can start a cascade that will let the bad product maintain its position. In contrast, in a market with four bad products and one good product, the good product can capture nearly the full market if its initial market share is high enough. However, if its market share is low and agents thus do not trust earlier agent's product evaluations, observational learning can also lead agents to remain loyal to one of the bad products. This is in line with De Vany and Walls' (1996) finding of a Bose-Einstein information process, where individuals choose a product according to its previous success.

Another experimental study by Hanson and Putler (1996) confirms this hypothesis. On a website hosting software for free download, they identify various pairs of similar software and manipulate the number of downloads for one of the two files. Following this treatment, they show that the file with substantially higher download numbers expands its lead over the other file in relative terms; its download rate increases by up to 65 per cent compared to the untreated file. This indicates that consumers are drawn to popular products if there is little other information available that distinguishes available options noticeably.

Similar results are obtained in a more sophisticated two-stage laboratory experiment conducted by Narayan et al. (2011). In the first stage, they provide seventy MBA students with information on six product attributes of four different electronic book readers. Subjects are then asked to use this information and their personal preferences to choose one amongst the four products. Subjects are also asked to identify individuals within the group that may influence their purchase decision and the extent to which each of these individuals may influence it.

In the second stage, subjects get to know whom of their self-reported 'influencers' chose which product and are asked to make a choice based on product attributes, personal preferences, and the knowledge of their influencers' decisions. Narayan et al. (2011) are able to show that subjects change their attribute preferences due to peer influence and display conforming behaviour. The least popular brand is chosen less frequently while the other three brands are chosen more often.[3]

These results show that herding or cascades do not only arise when consumers use simple observational learning to make decisions; they may also arise when learning occurs due to information spread. Yet, all of these studies have used either agent-based models or laboratory experiments to obtain their results. Employing empirical data to investigate consumer learning is a more recent phenomenon.

De Vany and Walls (2004b) apply their finding of a Pareto distribution of motion picture revenues (De Vany and Walls, 1996, 1997, 1999) in order to analyse the extent to which the box office revenues earned up to a certain week after release influence future revenues. Their sample consists of films that ran for at least ten weeks in cinemas and distinguishes between 'hits' earning more than $50 million and 'non-hits' earning less over their theatrical run. Because the Paretian distribution implies that future revenue is proportional to past revenues, they estimate a linear model, in which future revenues depend upon past revenues.

They find that revenues in a particular week can best be predicted by using the cumulative past revenues. For example, the revenues of week eight cannot be precisely forecasted using only the revenues of weeks one and two, but they can be more precisely forecasted using the revenues up until week seven. In the case of hits, weekly revenues cannot be accurately predicted in the early weeks of a film's release, but can be very well predicted in later weeks. Accordingly, the model fit improves every week, especially between weeks three and four. In the case of non-hits, weekly revenues can be reasonably well predicted in the early weeks of a film's release, but prediction accuracy does not increase significantly in later weeks. Accordingly, the model fit remains moderate from the early until the late weeks of theatrical release.

The Paretian distribution is further characterised by a tail weight, which indicates how much mass the tail of the distribution carries, that is the extent to which the mean of the distribution is dominated by a few extreme values. De Vany and Walls (2004b) use this tail weight in order to investigate the differences between 'hits' and 'non-hits'. They find that the tail weight is generally lower for 'hits' indicating more mass in the tail. This is consistent with the idea that the motion picture industry is driven by - relatively few - box office hits. The

[3]Further models on consumer learning from word of mouth are created by McFadden and Train (1996) and Ellison and Fudenberg (1995). While agents display conforming behaviour in the former model, word of mouth leads to diversity in the latter. These differing results occur due to the different ways in which agents 'calculate' their expected utility to make decisions.

tail weight declines over time for both 'hits' and 'non-hits', which they state is consistent with models of consumers sharing information.

They further interpret the coefficient of variation of weekly revenues as a measure of consumer consensus on the film's utility. It is shown that this coefficient continuously rises over time for hits indicating a tendency for consumer tastes to converge on a few hit products, whereas it approximates an inverted U-shape curve for non-hits, signifying that consumers do not agree on the quality of non-hits. They also look at the week-to-week correlations of film revenues and find that these are close to zero or even negative for hits during the early weeks and tend to rise during the later weeks. For non-hits, correlations are high throughout the weeks of their theatrical release.

They interpret these data as indications of "turbulent dynamics" (ibid, p. 60), which is consistent with information sharing. The finding that revenues generated early in films' life cycles are a poor predictor of final box-office revenues implies that a large opening is not necessary to create a hit. Consumers learn about the true quality of a film over time, and after about the fourth week a bifurcation sets in, where the differences between hits and non-hits grow at exponential rate; revenues for hits develop expansively whereas revenues for non-hits develop contractively.

Empirical market-level data are also used by Moretti (2011). Specifically, he uses data of all films released in the North American market between 1982 and 2000. He builds upon the theory of disconfirmation (cf. Sedgwick, 2007) stating that consumers hold expectations prior to experiencing a product, which may be positively or negatively disconfirmed. Employing weekly sales data, he measures the extent of this positive and negative surprise that consumers experience in the opening week, and tests its impact on future sales. He defines film-specific surprise as the "residual from a regression of first week log sales on log number of screens" (Moretti, 2011, p. 368).

More explicitly, he assumes that the number of screens a motion picture opens on is an indicator of the sales expectations of profit-maximising exhibitors who have an incentive to predict consumer expectations as precisely as possible prior to a film's release. He further states that the number of screens therefore reflects all the information that is available to consumers prior to a film's release, such as the cast, director, budget, critical reviews etc., a hypothesis for which he finds some support: first, via a regression of the number of screens on box office revenues that yields a good model fit and second, via interviews with theatre owners. Thus, if more people went to see the film than expected in the opening week, this indicates that they somehow 'knew' that it would be a good film and experienced a positive surprise. On the contrary, if less people went to see the film than expected, they experienced a negative surprise.

Subsequently, Moretti (2011) shows that motion pictures with a positive surprise in the opening week face a comparatively slower decline in revenues over subsequent weeks, whereas motion pictures with a negative surprise in the opening week face a comparatively faster decline in revenues. Specifically, the former's logarithm revenue curve is concave whereas the latter's logarithm revenue curve is convex. This difference is statistically significant, as is proven in a surprise model in which the logarithm of weekly revenues depends on the week, an interaction effect between week and positive surprise, and a fixed effect for the film. The decay rate for positive-surprise films is significantly lower than for negative-surprise films. He believes that word of mouth is responsible for these distinct sales trends. Once consumers learn that a particular film is of high quality, they will raise their expectations and, subsequently, a comparatively larger share will attend it in subsequent weeks.

However, since word of mouth is not directly measurable, he conducts various tests that indicate word of mouth learning. In a first step, he adds advertising expenditures to his surprise model. However, his results are not sensitive to advertising expenditures and their impact on the different sales patterns of positive- and negative-surprise films is marginal. This means that advertising has little or no effect on the sales trajectories once a film is released. It does not rule out the possibility of advertising having an influence on the opening week results; however, this effect is assumed to be captured by the number of opening screens. Similarly, adding critical reviews to the surprise model also does not affect the different rates of revenue decline.

Next, Moretti (2011) hypothesises that word of mouth will have a smaller impact in situations where consumers have a fairly precise prior and face comparatively smaller uncertainty. He characterises sequels as products, of which consumers have an idea whether they will like them or not. Adding an interaction term for sequels and week as well as an interaction term for sequels, week and positive surprise to the surprise model, he finds support for this assumption. The triple interaction term has a negative coefficient indicating that a positive surprise in the opening week – and thus, positive word of mouth has a smaller effect on revenues in subsequent weeks for sequels than for other films.

He hypothesises that word of mouth effects are more pronounced when consumers have a large social network through which they can spread. He further assumes that teenagers generally have stronger and larger networks and thus conjectures that surprises will have a larger effect on films targeted at teenagers. Adding an interaction term for teenage films and week as well as an interaction term for teenage films, week and positive surprise to the surprise model, he finds confirmation for this conjecture; the triple interaction term is positive.

Subsequently, he assumes that the marginal amount of learning declines over time, because the value of new information in week two is much higher than the value of new information in week three; consumers will already have had access to information to update their expectations of a particular film. Testing for the concavity of positive-surprise film revenues and the convexity of negative-surprise film revenues over time he states to find support for this.

Finally, in an attempt to test for externalities as an alternative explanation to word of mouth learning, he tests for weather shocks in six big cities. He hypothesises that weather shocks should not drive people away from films of good quality if word of mouth learning is present. Instrumenting weather surprises instead of film quality surprises in the surprise model, he finds that the coefficients point in the wrong direction and are partially not significant depending on the exact specification of weather shocks. Overall, though, weather shocks cannot explain the different sales patterns of motion pictures.

Moretti (2011) concludes that the sum of his findings strongly indicate that social learning takes place through the spread of information that consumers receive from their peers. This kind of learning makes successful motion pictures more successful, and unsuccessful motion pictures more unsuccessful, over time.

Moretti (2011), De Vany and Walls (2004b) and De Vany and Lee (2001) all agree on the importance of quality feedback and its effect on the future of a motion picture's sales. However, they differ on a crucial point. In Moretti's (2011) case, the fate of a film is more or less decided after the opening week. A positive surprise at this early point in time leads to higher subsequent sales due to the positive feedback loop that sets in. De Vany and Lee (2001), on the other hand, find support for the 'unpredictability' of final outcomes under the condition that consumers pass on quality feedback to one another. According to them, a 'good' film with a high initial market share can still die fairly quickly due to the "complex dynamics of the motion picture market" (ibid, p. 611). It is these dynamics letting De Vany and Walls (2004b) conclude that a big opening does not necessarily make a hit and that it takes about four weeks until a film's fate is decided. According to Moretti (2011), a 'good' film would have a positive surprise in the opening week and thus both face a slower decline and generate larger revenues.

Moretti's (2011) results indicate a conforming outcome, where consumers quickly learn through word of mouth which product has the highest expected utility and choose the same option. This is consistent with other studies (e.g. De Vany and Walls, 2004b; McFadden and Train, 1996). However, when looking more closely at the research conducted by De Vany and Lee (2001) and Chen et al. (2011), there is something crucial nevertheless about the amount of people that have previously chosen a particular option. Consumers seem to trust the decisions more, if many others have already made the same choice. And if consumers listen to their peers' opinions, negative reviews have a larger effect than positive reviews. Thus, it seems clear that products, which have been chosen by many people and received positive reviews, will continue to do well in the future; similarly, products chosen by only few people receiving bad reviews will not fare well and exit the market.

Less clear though is whether niche products with good reviews have a chance of becoming more popular (and becoming so-called 'sleepers') – according to De Vany and Lee (2001) this information will not be trusted, but according to Chen et al. (2011), positive reviews do have some effect, although smaller than for negative reviews. And similarly, it remains uncertain what happens to popular products with negative reviews. In this case, positive observational learning and negative word of mouth seem to operate in opposing directions; while the former leads to the reinforcement of large market shares (De Vany and Walls, 1996), the latter should 'kill' the product. This paper aims to gain further insight about the interaction between observational learning and word of mouth learning. To this end, the next section reviews the literature on online word of mouth.

## 2.2. Online Word of Mouth

Over the past two decades, the online sphere has experienced exponential growth. More and more households are connected to the internet and, more recently, with the emergence of smartphones and other handheld devices alongside faster mobile connections, many people are online even when they are not at home or in the office. This development has happened alongside a growth in online social networks and other types of platforms such as forums and blogs, in which people post their comments on various issues and products. This multi-user data has become a rich source of word of mouth information for researchers, since it provides a direct measure of consumer opinions that can be easily and costlessly collected. This is especially true of rating systems, where consumers appraise a product on a given scale.

Research using online word of mouth has largely focussed on two different dimensions – volume and valence. The volume of word of mouth is said to increase consumer awareness of a product, because the more publicity a product gets, the more likely people are to know about it and include it in their choice set (Berger et al., 2010; Liu, 2006). Transferred to the market for motion pictures, however, it can be assumed that the number of individuals talking about a particular film is influenced by the number of people who have seen it. Thus, a large volume of online word of mouth represents a large audience that has already consumed the film.

Consumer learning from the volume of word of mouth can therefore be compared to the models of herding behaviour (Banerjee, 1992; Bikhchandani et al., 1992) – individuals observe the actions of previous consumers and subsequently decide to neglect their private signal in order to follow the crowd – or observational learning (Chen et al., 2011). Learning from the volume of online word of mouth is therefore comparatively 'easy' for consumers, since it is reflected in the current charts of the most successful films, which are easily accessible and may even be advertised through other channels such as magazines, radio or television.

The valence of online word of mouth, in contrast, represents a quality feedback signal. Consumers who have seen a particular film evaluate their personal utility derived from watching the film and attach a rating value to it. These quality assessments may influence individuals who have not made their decision on seeing a particular film or help them make a decision between various alternatives.

Learning from the valence of online word of mouth requires more effort on behalf of consumers. They actually need to search for this kind of information. They may need to read the review alongside the numerical rating in order to see how the user arrived at this particular rating. They may need to compute the 'average' rating out of a number of ratings, though many websites readily offer this information. Table 2.2 summarises the literature on online word of mouth.

One of the first studies to use online word of mouth was conducted by Liu (2006). He collects weekly box office and online word of mouth data on forty motion pictures released between May and September 2002. From the Yahoo! Movies message board, he collects more than 12,000 ratings, which he decomposes into volume and valence measures. The latter is coded into five different categories – positive, negative, mixed, neutral and irrelevant. From various public sources he gathers data on box office revenues, the number of screens the film is shown on, estimates on the production budget, critical reviews, the MPAA rating, genre and the star power of a film. The data is collected from the date of the earliest post – which is usually in the pre-release phase – until the eighth week of a film's run.

In his analysis of the dynamics of these word of mouth measures, Liu (2006) finds that the volume of ratings is greatest in the opening week, closely followed by the pre-release period. Looking at the valence, he finds that it is mostly positive in the pre-release period, but the percentage of positive word of mouth drops significantly for half of the films in his data set once they are released while the percentage of negative word of mouth increases. This indicates that audiences hold comparatively high expectations prior to a film's release, but that these expectations are often not met.

In order to estimate the effect of word of mouth on consumer decisions he uses double log specifications in a regression-type model for each week from the opening until week eight of a film's release, where the natural logarithm of weekly revenues depends upon the number of screens, the volume of word of mouth, the number of critical reviews, the number of new releases among the top twenty films, the average age of the top twenty films (all in natural logarithms), the percentage of positive word of mouth messages and the percentage of positive critical reviews. The number of new releases and the average age of films among the top twenty are used to measure the degree of competition.

Liu (2006) further estimates a model, in which the aggregate box office revenue depends upon pre-release word of mouth using the same characteristics as in the previous model. In both of these models, he finds that including film-specific variables such as star power or genre does not improve his model as these variables are insignificant. The volume of word of mouth provides explanatory power on box office revenues whereas its valence does not, both for weekly box office and gross revenues. However, he finds that this effect disappears after six weeks. This indicates that once the majority of the population has become aware of a film's existence, an additional message cannot increase this awareness further.

| Author | Dimensions of WOM | Valence Coding | Source of WOM | Type of Model | Number of Products |
|---|---|---|---|---|---|
| Asur and Huberman 2010 | Volume Valence | Positive Neutral Negative | 2.89 million tweets from Twitter | OLS | 24 films released between Nov 2009 and Feb 2010 |
| Chakravarty et al. 2010 | Valence | Textual | Yahoo! Movies | 3 experimental studies | 1 film in each study |
| Chintagunta et al. 2010 | Volume Valence Precision | Arithmetic Mean (3-13) | Yahoo! Movies | Log-linear OLS with film and market fixed effects | 148 films released between Nov 2003 and Feb 2005 |
| Dellarocas et al. 2007 | Volume Valence | Arithmetic Mean | 55,156 user reviews from Yahoo! Movies | Hazard rate formulation Bass diffusion model | 80 films released in 2002 |
| Duan et al. 2008a | Volume Valence | Arithmetic Mean (3-13) | 95,867 user posts from Yahoo! Movies | Linear three-stage least square (3SLS) with film fixed effects | 71 films released between Jul 2003 and May 2004 |
| Duan et al. 2008b | Volume Valence | Arithmetic Mean (3-13) | 95,867 user posts from Yahoo! Movies | Log-linear three-stage least square (3SLS) with film fixed effects | 71 films released between Jul 2003 and May 2004 |
| Godes and Mayzlin 2004 | Volume Dispersion (Valence) | Positive Negative Mixed | Usenet message boards | OLS | 41 TV shows premiered between 1999 and 2000 |
| Joeckel 2007 | Volume Valence | Arithmetic Mean (1-10) | Ratings from Gamespot | OLS | 201 video games |

Table 2.2.: Literature on Online Word of Mouth

| **Author** | **Dimensions of WOM** | **Valence Coding** | **Source of WOM** | **Type of Model** | **Number of Products** |
|---|---|---|---|---|---|
| Li and Hitt 2008 | Volume<br>Valence | Arithmetic Mean (1-5) | 82,131 user reviews from Amazon | Linear negative exponential OLS with non-linear fixed effects | 2,203 books published between Jan 2000 and Feb 2004 |
| Liu 2006 | Volume<br>Valence | Positive percentage<br>Negative percentage | 12,136 user posts from Yahoo! Movies | Log-linear OLS | 40 films released between May and Sep 2002 |
| Rui et al. 2011 | Volume<br>Valence<br>Intention (to see the film) | Positive<br>Neutral<br>Negative | 4.17 million tweets from Twitter | OLS with film-specific fixed effects | 63 films released between Jun 2009 and Feb 2010 |
| Wu et al. 2011 | Valence<br>Readability<br>Length | Arithmetic Mean (1-5) | 7,659 user reviews from Amazon | OLS | 776 books |

Table 2.2.: (continued)

Moreover, he finds that the size of the effect of the volume of pre-release word of mouth is larger than the effect of the number of screens for first-week revenues. He therefore analyses whether this early word of mouth basically consists of marketing information by including the advertising budget in his regression, finding that the impact of the volume of word of mouth remains significant.

He concludes that the volume of word of mouth can serve as a reliable additional variable in revenue forecasting for film production firms. It represents a measure of consumer awareness, which directly translates into sales. He further links the fact that the valence of word of mouth is not a significant explanatory variable to a common finding of behavioural research, that attitude is not always a reliable predictor of behaviour.

Liu (2006) assumes word of mouth to be an exogenous variable that influences audiences in the following week. Although it changes on a weekly basis, it is set by external circumstances similarly to the number of screens a film is shown on. Yet, the volume of word of mouth may be heavily influenced by the number of past viewers (Godes and Mayzlin, 2004) and therefore be an endogenous variable. The more people who go and see a film, the more it will be talked about. It may in fact be ticket sales and film attendance that lead to a rise in the volume of word of mouth.

Duan et al. (2008a,b) investigate this issue and analyse the relationship between these two variables. They collect data on 71 motion pictures released in the North American market between July 2003 and May 2004. They collect daily revenue data, the number of screens and other film-specific variables such as critical reviews, production budget, MPAA rating and stars from Variety and Box Office Mojo. Their online word of mouth data is collected from Yahoo! Movies and coded into measures of volume and valence. Daily posts for each film are counted to generate the volume. To calculate the valence, numerical values (ranging from 13 to 3) are assigned to each letter grade of the reviews (ranging from A+ to D). Subsequently, the arithmetic mean is calculated for each film both on a daily as well as on a cumulative basis. This data is collected for a period of six weeks after a film's release.

Like Liu (2006), they find that the volume of word of mouth is highest directly after the release of a film and then declines rapidly, with small surges on the weekend. This pattern is very similar to the pattern of sales – in fact, both volume of word of mouth and box office revenues decline simultaneously – and therefore a first indicator that the volume of word of mouth represents previous viewers. In contrast, the average rating does not change considerably over time.

In order to investigate the interrelationship between box office revenues and word of mouth, they use a dynamic simultaneous equation system, a three-stage least square procedure consisting of two equations, one in which revenue is specified as the dependant variable and one in which the volume of word of mouth takes this role. Duan et al. (2008a) use a linear model with one-day lagged variables. Daily revenues depend upon the daily number of posts, the one-day lagged cumulative average rating, daily revenues of the previous day, an indicator of whether the current day is a weekend and film-specific fixed effects. The daily number of posts, in turn, depends upon the daily revenue, the number of posts on the previous day, the cumulative number of posts up to the previous day, a weekend indicator and a film-specific fixed effect.

Duan et al. (2008b) specify a log-linear model with multi-lag terms. The log-linear terms are meant to reflect a "multistage consumer decision process" (ibid, p. 238), where, for example, the number of consumers seeing the film is dependent on the consumer base, the probability of a consumer being aware of the film, and a consumer actually deciding to see the film given his awareness. Multi-lag terms are incorporated assuming that word of mouth affects decisions not just on the same day, but also on subsequent days. The daily revenues depend on the multi-lagged daily volume of posts, the cumulative average rating on the same day, the average rating on the same day, the number of screens, the age of the film, an indicator of whether the day is a weekend and film-specific fixed effects. The daily volume of posts depends on the multi-lagged daily revenues, the cumulative average rating, the daily average rating, the age of the film, a weekend indicator, and film-specific fixed effects. All of these coefficients are presented in logarithms except the weekend indicator. In both models, the weekend variable is meant to capture the different behaviour of consumers on a weekend. Generally, less people go to the cinema during the week.

The findings of Duan et al. (2008a) and Duan et al. (2008b) are qualitatively similar. They both state that there is a positive interaction between the volume of word of mouth and box office revenues, meaning that more online posts lead to higher sales and higher sales lead to more word of mouth. Thus, word of mouth is endogenous and has a dual role, one as a "precursor" and one as an "outcome" of film sales.

Duan et al. (2008b) provide some additional findings. The positive effect mentioned above diminishes quickly over time. They show that the coefficients of the multi-lagged variables are either negligible or insignificant after three days and conclude that online word of mouth rapidly spreads through a community. Further, because they include the average rating in their word of mouth equation, they are able to show that the valence does not directly influence box office revenues, but it does impact on the volume of posts. Since the volume, in turn, positively affects sales, the valence has an indirect effect on box office revenues. Seemingly, consumers are more likely to talk about films which they have had a positive experience with. This leads to a higher awareness among the general population, which subsequently translates into higher sales. The fact that the awareness effect seems to be driving film sales once again highlights the importance of spreading awareness through the population as fast as possible.

Another study sheds more light on the different effects of volume and valence on consumer decisions. Interestingly, it comes to a different conclusion. Chintagunta et al. (2010) collect daily box office data on 148 motion pictures released between November 2003 and February

2005 and the according word of mouth data from Yahoo! Movies. Similar to previous studies, they use volume, valence, and variance as their dimensions of online word of mouth. However, they account for the sequential rollout of films by employing "Designated Market Area (DMA) level local geographic box-office performance" data (ibid, p. 944). They argue that even a film on wide release will only be shown after the national release date in some locations. The decisions of consumers in such locations may be influenced by the word of mouth of consumers in earlier release-date locations. Therefore, they collect theatre-level box office data from 3,830 geographical locations.

In their model, they control for the number of screens, pre-release advertising, and competition and incorporate measures for seasonal variations and holidays as well as fixed effects for the film and the geographical market. They account for endogeneity of user ratings in subsequent markets and for correlation in the error term of films released across different markets.

Using this model, they find that only the valence of word of mouth significantly drives audience decisions. In contrast to previous studies, neither the volume nor the variance is significant. They hypothesise that there is an aggregation bias in previous studies meaning that by using national-level data the marginal effect of user ratings is to some extent masked. It does not account for the release of films across geographical markets and the different ways, in which consumers are influenced depending on the time of release. Therefore, they re-estimate their model using national-level data and indeed find that this leads to different results – the volume of ratings is now the only word of mouth variable that has a significant impact on box office revenues.

Another approach estimating the different effects of volume and valence is taken by Dellarocas et al. (2007). They collect online rating data from Yahoo! Movies, weekly box office and marketing data from Box Office Mojo and estimates of star power from Hollywood Reporter. They create four different models based upon a hazard rate formulation in order to forecast total box office revenues from early sales, other film-specific variables, and word of mouth data. They include an external force – including marketing and publicity, critical reviews and unobservable attributes such as attractiveness of plot or quality of trailer – and an internal force that relates to word of mouth. They further incorporate a discount factor for a decreasing external factor and a "time-discounted integral of past adopters" (ibid, p. 31) to account for a diminishing effect of word of mouth over time.

Model A includes all available data – marketing expenses, number of screens, critical reviews, early box office revenues and word of mouth data. Model B accounts for the fact that early revenues may not be available to forecast revenues and uses the volume of early online reviews as a proxy for sales. Model C stands as the benchmark model and includes neither early revenues nor word of mouth data. Finally, Model D only uses online rating data.

They test their models' goodness of fit by using the mean absolute percentage error (MAPE). As hypothesised, Model A provides the best fit. It is followed by Model B, Model D and Model C. The significant finding is that removing marketing expenses, the number of screens and critical reviews from the full model (Model D) does not increase the MAPE by as much as the removal of word of mouth data (Model C).[4] This shows the importance of word of mouth when forecasting consumer decisions.

They further find that marketing expenses are a significant predictor of the external force only in Model C, where no word of mouth data is present. This may indicate that early word of mouth is simply a reflection of a film's marketing and contains the same information. It is intuitive to assume that early word of mouth is strongly influenced by advertising. However, they do not test this hypothesis.

They distinguish between blockbuster-type films and sleepers by assuming that the latter are released on less than 600 screens. They continue to test their models for these two categories and find that Model A remains to have the best fit for blockbusters, but Model D using only word of mouth data has the best fit for sleepers. This indicates that blockbusters are driven more by first-week revenues whereas word of mouth plays an important role in determining the success of sleepers.

With regards to the volume and the valence of online ratings, they find that the valence significantly influences the internal force in Model A, B and D. This shows that individuals seem to take the evaluation of other consumers into consideration when making a decision. The volume of ratings is significant in Model B and D, where it is used as a proxy for weekly revenues. This is in line with Duan et al.'s (2008a,b) finding that the volume of word of mouth is an outcome of film sales. Unfortunately, Dellarocas et al. (2007) do not include the volume in the internal force equation. Therefore, they do not test for the awareness effect of word of mouth.

Some studies have used different sources of online word of mouth in order to analyse its effects on consumption behaviour. Both Asur and Huberman (2010) and Rui et al. (2011) use Twitter in order to count the number of messages or 'tweets' for a specific motion picture and employ sentiment analyses to code the messages into positive, neutral and negative comments.

Asur and Huberman (2010) analyse 24 motion pictures and collect 2.89 million tweets, which they code into valence measures using linguistic software. They define a critical period – the period during which a film's fate is decided – as the period from one week prior to a film's release until two weeks after release. They find that the amount of tweets is highest just before a film's release, but that these are mostly neutral and anticipatory, whereas their subjectivity increases once a film is released. They interpret that consumers are likely to make

[4] It needs to be noted that the lower the MAPE the better the model fit.

their positive or negative opinion heard only after they have seen the film. This contrasts with the findings by Liu (2006) who states that pre-release word of mouth is mostly positive.

Asur and Huberman (2010) define tweets containing marketing information as messages with web addresses (URLs) or retweets, which means that a message was forwarded without alteration. These kind of marketing messages are also highest prior to the release of a film. They run a simple regression using marketing tweets to forecast film revenues, but find them to be a poor predictor. They interpret this as a failure of promotional material to increase film revenues.

They further define a tweet-rate as the number of tweets mentioning a particular film per hour and use this as a measure of the volume of word of mouth a film receives. This tweet-rate has a high correlation with box office revenues, and a linear regression using only the average tweet-rate of the tweets prior to release to forecast gross box office revenues provides a good fit. They improve this model by using the average tweet-rate for each of the seven days prior to release and adding the number of screens a film is released on as an additional variable.

Defining the ratio between positive tweets and negative tweets as their valence measure leads to another variable, which they add to their model. Using second-week revenues as their dependent variable and both the average tweet-rate as well as the ratio as independent variables, they further improve their model. However, they find that the valence measure is "not as important as the rate of tweets themselves" (ibid, p. 7).

More sophisticated in the way online word of mouth data is classified is the method adopted by Rui et al. (2011). In a first step they cluster tweets into intention messages expressing that the user will go and see a film and sentiment messages expressing the user's opinion about a film. Thus, they create measures of pre-consumption and post-consumption word of mouth, which is different from pre-release and post-release word of mouth used in other studies (e.g. Asur and Huberman, 2010; Liu, 2006), since pre-consumption word of mouth can still occur after the film is released.

In a second step, sentiment messages are further categorised into positive, neutral and negative tweets using linguistic software. Further, for each Twitter user the number of followers, that is the number of people who will automatically receive every tweet from this user, is recorded. They use this measure as a proxy for social influence and opinion leader status; users with more than 400 followers are characterised as users with a high social influence. Finally, advertising messages are identified by filtering tweets containing URLs and subsequently removed from the data set.

Rui et al. (2011) create a regression-type model, in which the weekly revenues depend on the revenues of the previous week, the total number of tweets, the ratio of tweets from users with a large audience, the ratio of intention tweets, the ratio of positive and negative tweets and film-specific fixed effects.

Intention tweets have the highest coefficient indicating one of two things. Either this kind of information contains high credibility and people follow its suggestion. Alternatively, there is a substantial direct effect of such pre-consumption word of mouth, because not only does it have an awareness and persuasive effect on other consumers, but the user who posted the message will still go to see the film herself and may be doing so with a group of people. It remains unclear though which of the two effects is dominant.

Both positive and negative word of mouth affect the decisions of audiences, and both in the expected direction – positive tweets help film sales while negative tweets hurt them. However, the effect of the latter is larger than the effect of the former. Accounting for the fact that less than five per cent of all tweets are negative messages, this shows that consumers react very sensitively to negative information.

The effect of users with a large audience is also significant, though the coefficient is not as large as the previous three word of mouth dimensions. However, the model does not explain why this is the case. This kind of users may simply raise awareness in a larger part of the population and therefore positively affect motion picture revenues. It may also be the case that users with a large audience are more influential and therefore the persuasive effect is dominant.

In contrast to these dimensions of word of mouth, the volume of tweets, despite being significant, carries by far the lowest coefficient indicating that it is not as important as the previously mentioned dimensions. However, the volume of word of mouth may also be largely captured by the other variables.

While Rui et al. (2011) analyse the effect of social influence of a person, Chakravarty et al. (2010) are more interested in a different characteristic of consumers, namely their cinema-going frequency. They argue that frequent filmgoers will react differently to certain types of information than infrequent filmgoers. They distinguish between two types of information: (online) affective word of mouth and professional critical reviews concentrating more on artistic and technical aspects of the film.

In order to test their hypotheses, they run three different experiments each time simulating a Yahoo! Movies message board. They distribute questionnaires amongst their participants and define frequent filmgoers as either having watched three or more films in the past two months or having watched ten or more films in a year.

In the first experiment, 157 students are presented with a website on the film *National Treasure*. Two groups of subjects are shown either a mix of positive and neutral reviews or a mix of negative and neutral reviews. The results of two questionnaires, completed respectively before and after the experiment, show that negative reviews are read more carefully than positive reviews. Further, infrequent filmgoers are more prone to change their opinion about the film, and the change is greater for negative word of mouth than for positive word of mouth.

In the second experiment, Chakravarty et al. (2010) want to find out how word of mouth is moderated in the presence of contradictory critical reviews. Using 128 students and the film *Sahara*, they add an average numerical rating from professional critics, which differs from the word of mouth posted by consumers, to the situation in the first experiment. Thus, they create a situation, in which subjects are either exposed to positive word of mouth and a negative critical rating or to negative word of mouth and a positive critical rating. Subjects in the control group do not receive any critical ratings.

Frequent cinemagoers significantly attenuated their evaluation of the film in the direction of the critical rating, in both cases almost neutralising the effect of word of mouth. In contrast, the effect is smaller for infrequent cinemagoers, and they did not alter their assessment of a film in the case of negative word of mouth and a positive critical rating. This shows that they place their trust rather in the opinion of other consumers representing the mass appeal of a particular film.

In the third experiment, they account for the fact that critical reviews are only presented as a summary rating whereas word of mouth referrals are presented in full text format and provide subjects with textual comments from different critics. 119 students are shown two websites on the film *Déjà Vu*, one group is shown the critical reviews first, and the other gets to see the Yahoo! Movies message board first. Again, positive word of mouth reviews are combined with negative critical reviews and vice versa.

The results are similar to those of their second study. Frequent filmgoers adapt their assessment towards the critical review whereas infrequent filmgoers are influenced in the direction of word of mouth. This is especially the case for frequent filmgoers and negative word of mouth contrasted by positive critical reviews. The mean effect is close to zero, meaning that the two kinds of information nearly cancel each other out. In contrast, infrequent filmgoers presented with the same situation significantly lowered their opinion about the film in the direction of the negative word of mouth referrals.

Chakravarty et al. (2010) conclude that frequent filmgoers, due to their experience and familiarity with the product, develop a more 'elite' taste similar to that of professional film critics and are able to assess the quality of a motion picture based on its cast, director and other attributes more easily. Therefore, they tend to hold very strong opinions about films and are not as easily affected by exogenous influences such as word of mouth.

Researchers have not only examined the effect of online word of mouth on films, they have also looked at other entertainment and experience goods, such as books, video games, and TV shows. One of the first studies to use online word of mouth was conducted by Godes and Mayzlin (2004) who used conversations from Usenet to measure their influence on the viewing habits of TV shows. Usenet was a collection of newsgroups, in which users could post their opinion on various topics ranging from religion to science, from news to literature. Godes and Mayzlin (2004) collect more than 20,000 posts from newsgroups beginning with either alt.tv or rec.arts.tv, indicating that they are discussing television, and search for posts relating to 44 TV shows that were launched between 1999 and 2000.

They compute the volume of online word of mouth for each of these TV shows by counting the number of posts they received between two episodes. They further calculate the dispersion of these posts across different newsgroups between two episodes. The theoretical underpinning for using this dimension of word of mouth leads back to the work of Granovetter (1973) who shows that information travels quickly among communities with strong ties, but slowly across communities with weak ties. In order for a product or service to be positively influenced by word of mouth, it is crucial that as many people as possible know about it. Therefore, Granovetter (1973) argues, these weak ties are essential in order to spread information widely.

Building upon this, Godes and Mayzlin (2004) argue that a newsgroup is like a community. Information spreads quickly within a newsgroup notifying its users about the existence of a new TV show, but it only spreads slowly across different newsgroups. They further hypothesise that the more newsgroups contain posts about a specific TV show, the more people will know about it and, subsequently, the more successful the show will be.

In order to measure the success of the TV shows, they collect viewership data from Nielsen ratings. They are able to control for a time trend, since they collect data on every single episode for both the viewership and the word of mouth data.

They use a regression model with a fixed effect for each TV show, in which the viewership of the current episode depends upon the viewership of the last episode, the number of posts and the dispersion of posts between the last and the current episode, and the time variable, that is the episode of the TV show.

They present their results for the time trend between episodes four and seven. They show that the dispersion has a significant effect on viewership in the early period of a new TV show. It is significant in weeks four, five and six, and the coefficient declines between week four and six. This indicates that the spread of early word of mouth is important to make people aware of the existence of a TV show; however, once awareness has been raised, this effect declines. The volume of word of mouth, on the other hand, is insignificant during the early weeks of a new TV show. Only in week seven does it reach marginal significance at the 10-percent level. The fact that the dispersion is more important than the volume further confirms Granovetter's (1973) findings – awareness is quickly raised within a community; however, to make a product successful, awareness needs to spread across communities.

Finally, Godes and Mayzlin (2004) analyse the effect of the valence of their posts. To do this, they sample ten per cent of all posts for each TV show and have assistants categorise each of them into one of six categories - positive, negative, neutral, mixed, irrelevant and not sure. They subsequently include positive, negative and mixed valence ratings in their regression and find none of these measures to be significant. Instead, the volume remains insignificant, while the dispersion becomes even more significant. They attribute this to the fact that irrelevant posts have been left out of the equation.

One important thing to note is that Godes and Mayzlin (2004) investigate the effect of word of mouth on TV shows. These are generally cheap to consume, so once an individual is aware of the existence of a particular show – and finds that people are talking about it – she may watch it without spending much time on decision-making and without closely considering the quality other viewers have attributed to it. There is only little risk involved in watching a TV show – the potential loss being limited to the time spent watching the show –, a situation different from other experience goods with higher costs.

One such good are video games. Joeckel (2007) analyses how the 201 top-selling games between 1995 and 2005 were affected by online word of mouth and press ratings. He uses the penetration rate – the percentage of installed consoles reached by a particular game – to account for a game's success and consumer popularity. For the rating data, he uses average online press ratings, which are collected and integrated by GameSpot and average user ratings posted on the same website.

He finds that not all games that received a very high rating also sold very well. This is indicated by the moderate correlations between the penetration rate and both average user and average press ratings (.341 and .359, respectively). When controlling for the number of ratings, these correlations decrease considerably for user ratings to .202 and less significantly for press ratings to .305. He suggests that this is due to a lower variation in the number of press reviews, a first indication that the volume of ratings plays an important role. He subsequently uses the volume of ratings; the correlation increases to .444 for user ratings and decreases to .212 for press reviews.

Joeckel (2007) builds three different models, the first one using the penetration rate as a dependent variable and both average user rating and average online press rating as independent variables, accounting for the correlation between the two. He finds that this model can explain 13 per cent of the variation in the penetration rate.

Arguing that the two kinds of ratings measure the same phenomenon due to their high correlation, he constructs a latent variable labelled "Perceived Quality". He assumes that perceived quality actually influences the penetration rate, and ratings are merely a manifestation of this variable. Using perceived quality as the independent variable, a second model can explain 14 per cent of the variation in the penetration rate.

In his final model he also includes the volume of user ratings as a third manifestation of the latent variable perceived quality. He finds that the perceived quality is influenced strongly by both the average user rating and the average press rating, but only to a much weaker degree by the volume of ratings. Overall, this model can explain 15 per cent of the variation in the penetration rate.

Another experience good that falls into the category of not being costless are books. These are generally much less heavily advertised than motion pictures or video games, and therefore freely available information is limited. Thus, consumers may rely more on the experiences of other users and how they evaluate a particular book. Within the online sphere, Amazon.com provides a rich data source, which most researchers investigating the influence of word of mouth on books have tapped into.

Li and Hitt (2008) investigate the self-selection bias within online reviews meaning that early buyers of a product hold different preferences and expectations than later buyers. They assume that this has an effect on long-term market outcomes. They investigate a durable experience good without repeat purchase, namely books, which, in contrast to films at the cinema, have a longer life cycle. They state that books have both search attributes, which are easily accessible prior to consumption such as the author of the book, and experience attributes, which cannot be accessed until the book has been read. Once individuals have consumed the book they voice their opinion giving late buyers the option of including these evaluations in their assessment and updating their expectations accordingly. In this case, the self-selection bias presents itself through the likelihood of positive ratings from early buyers who may be fans of the author, which stand in contrast to the majority of the population. Thus, late buyers reading these reviews may hold pre-purchase expectations that are higher than the utility they receive from the book leading to disappointment and subsequent negative reviews.

Analysing user reviews from Amazon over a long period of time and using a linear negative exponential regression to account for the time trend of the reviews, Li and Hitt (2008) find that for 70.81 per cent of the 2,203 books in their sample the average rating declines over time, for 18.20 per cent the average rating increases over time, and the remaining 10.99 per cent do not show a clear trend. Among the books that have a declining average rating, 27.37 per cent show an undershooting effect, meaning that after high initial ratings, ratings decline dramatically and, for a period of time, are even below the long-term average rating.

Li and Hitt (2008) also look at the long-term effects of the self-selection bias by testing whether consumers do correct for an early review bias. They create a model including the number of reviews and both the long-term average rating and a time-varying rating at a particular point in time for each book and test their effect on sales. All three variables prove to be significant indicating that the volume of reviews positively affects sales and further that consumers do not fully account for the early review bias. Thus, higher early ratings can also lead to higher sales.

However, the overall findings on the influence of product review valence are conflicting. Some studies find a significant negativity bias indicating that negative ratings hurt product sales more than positive ratings benefit them. Other studies find the valence to be insignificant. This conflict interests Wu et al. (2011) and they examine whether the helpfulness of a review may have a moderating effect on how consumers use a review to shape their decision. They collect data on the helpfulness of 7,659 customer reviews on 776 books from Amazon. Consumers can always select whether a particular review on Amazon is helpful or not. Besides this variable, Wu et al. (2011) use the length and the readability of a review as additional variables influencing the helpfulness. The readability is measured using the Flesch Reading Ease, "a popular readability index designed to measure the easiness of comprehension on a piece of text of standard English" (ibid, p. 5). Finally, they operationalise the arithmetic mean of the five-star consumer ratings as the rating valence.

They find that negative reviews are generally perceived as more helpful, although one-star ratings are perceived as less helpful than two- or three-star ratings. However, this negativity effect disappears once they account for the length and the readability of reviews. They conclude that the trust or helpfulness of a review is actually more important than its valence alone. Since satisfied customers provide lengthier and more helpful reviews, the negativity bias disappears when accounting for the helpfulness of a review.

This chapter has shown that a significant amount of research has been done directly measuring online word of mouth through social websites or forums. Yet, there are a number of conflicts emerging from the literature.

Most studies agree that the volume of word of mouth positively influences product sales (Chintagunta et al., 2010, being the sole exception). However, volume has a dual role. It raises awareness of a film among the population that has not seen the film and thereby influences future consumer decisions. On the other side, it is an outcome or a representation of these decisions. This is due to the intuitive fact that the more people go and see a film at the cinema the more people are likely to post their opinion on the internet.

Generally, the volume of word of mouth is highest around the release of a film and declines quickly afterwards, similar to the pattern of box office revenues. This further supports the hypothesis that volume not only influences (future) consumer decisions, but is also influenced by their (past) decisions. Therefore, the effect of volume of word of mouth needs to be interpreted carefully.

The findings on the valence of word of mouth are much more contrary. Some studies find it to have a significant impact on revenue outcomes, while others find it to be insignificant. In some cases the valence is less significant than the volume (Asur and Huberman, 2010; Joeckel, 2007), in other cases it only has an indirect effect on box office revenues (Duan et al., 2008b), while in the case of accounting for the sequential release of motion pictures across different geographical markets, valence has the highest impact on sales (Chintagunta et al., 2010). Li and Hitt (2008) show that high early ratings can lead to higher product sales in the case of books, because consumers do not correct for a self-selection bias. It is questionable, though, how likely this is to happen in the case of motion pictures, since they generally have the highest attendance in the early period of their release.

Due to this ambiguity it also remains unclear whether film consumers exhibit a negativity bias with regards to film reviews, meaning that negative product evaluations hurt sales more than positive evaluations help them. Some studies find support for this effect (Chakravarty et al., 2010; Rui et al., 2011). However, it disappears if the helpfulness of a product review is accounted for (Wu et al., 2011).

While the volume of word of mouth declines quickly after the film is released, it is not clear whether this happens for the valence. Li and Hitt (2008) find a self-selection bias in the reviews of books, leading to high early ratings and a declining trend over time. Liu (2006) provides some indication of the trend of rating valence in the case of motion pictures. He shows that positive ratings are highest prior to the release of a film, but that the trends of positive and negative ratings after release are fairly random. The positive ratings prior to a film's release may therefore reflect consumer anticipation, individuals who are looking forward to see the film (cf. Asur and Huberman, 2010).

Noteworthy is also the difference between pre-release and post-release word of mouth. Prior to the release of a motion picture, the valence of a rating cannot contain any experienced utility, but only an indication of how much the user is looking forward to seeing the film. It can therefore be hypothesised that it is heavily influenced by advertising information (Dellarocas et al., 2007). On the other hand, both Asur and Huberman (2010) and Rui et al. (2011) provide some indication that even when omitting reviews that contain marketing information, pre-release word of mouth still significantly influences consumer decisions. However, they do not analyse whether reviews that do not contain obvious marketing information were nevertheless influenced by heavy advertising.

Finally, some of the differences emerging in the findings on online word of mouth may be due to the different products used for analysis. TV shows are surely cheaper to consume than motion pictures and individuals may not spend as much time on deciding whether to watch them once they are aware of their existence. Books are generally not as heavily advertised as films and only have the author as a visible cue (unless it is a series of books, such as Harry Potter or The Dark Tower series). Hence, consumers may value word of mouth higher. Video games are bought by a younger demographic than motion pictures, which again may affect the ways in which word of mouth is both spread and used to inform decisions. Yet, despite these differences in product characteristics, there are some commonalities: in all cases, word of mouth raises the awareness of the existence of a product and can influence sales. It remains ambiguous, though, exactly how this happens.

## 2.3. Motion Picture Characteristics

Although motion pictures are difficult to assess prior to consumption, it is not only word of mouth from peers or other sources that consumers use in order to make their decision. They have learned from their own experience what kind of films they generally prefer over others, for example whether they usually enjoy comedies more than action films or prefer historical costume films over science fiction. Thus, they may be able to look at publicly available cues such as the genre, the cast or the director of a film and estimate whether a particular motion picture reflects their cineastic interests.

It is therefore unsurprising that a number of studies has looked at different film-specific characteristics that are available before a film's release in order to analyse whether they carry a positive signal that consumers look for. These studies are mostly forecasting exercises using empirical quantitative data intended to help film producers create films carrying signals that consumers are interested in. In order to determine the success of a given film, they generally use box office revenues, which are modelled as a function of a number of explanatory variables such as production budget, advertising, cast (or 'stars'), director, genre, critical reviews or the number of screens. However, despite the similarities in their approaches, the outcomes of these studies are multi-faceted (see Table 2.3 for a summary of the variables employed and their significance [highlighted in bold format]).

Some of the earliest studies were conducted by Litman (1983) and Litman and Kohl (1989), both using the same model to respectively analyse factors driving motion picture success during the 1970s and 1980s. In the more recent study, the rentals of 697 films that were released between 1981 and 1986 and earned more than $1 million are modelled as a function of a number of explanatory variables in an ordinary least-squares (OLS) regression. Variables for genre, MPAA rating, production budget, critical reviews, number of opening screens, annual admissions and competition are included as well as binary variables for sequel, prequel, remake, adaptation, North American origin, star actor, top director, major distributor, Christmas, Easter, summer, award nomination and award win.

They find the MPAA rating and most of the genres to be insignificant explanatory variables. Only science-fiction/fantasy and drama are significant. In contrast to the 1970s, the Christmas release period is not as important anymore in the 1980s and only the summer remains significant. While being nominated for an award still increases earnings, winning an award has lost its importance. The production budget has also become less important, though it remains significant. Further, the number of opening screens and critical reviews remain significant predictors of box office success. Finally, only the binary variables sequel, North American origin and competition are positively correlated with film revenues.

Two more recent studies using a similar approach are Chang and Ki (2005) and Terry et al. (2005) who both employ a number of film-specific variables to analyse their respective influence on cinema attendance. Chang and Ki (2005) sample 463 North American motion pictures that were released between 2000 and 2002 and earned at least $1 million from IMDb and use three dependent variables for their analysis: total domestic box office revenues, first-week revenues and length of run.

| **Author** | **Number of Films** | **Variables tested** |
|---|---|---|
| Ainslie et al. 2005 | 404 films relased in 1995-1998 | **Advertising**, **competition**, critical reviews, distributor*, film type (art house vs. block-buster), star actor, **director**, **opening screens**, **sequel**<br>*some distributors are significant |
| Bagella and Becchetti 1999 | Italian films released in 1985-1996 | Age restriction (similar to MPAA rating R), foreign co-producer, genre (**comic**, **socio-political**, ...), distributor, producer (**Filmauro**, ...), **star actor**, **director**, state subsidies |
| Basuroy et al. 2003 | 175 films released in 1991-1993 | **Budget**, critical reviews (total number, **positive**, **negative**), MPAA rating (G, PG, PG-13, R), seasonality, star power, director, sequel*, **screens**<br>*significant in the first week only |
| Basuroy et al. 2006 | 175 films released in 1991-1993 | **Advertising**, competition, critical reviews (**positive**, **consensus**), distributor, **screens**, seasonality, **sequel**, **star power** (award wins and nominations), **week**, **word of mouth** (cumulative number of screens) |
| Chang and Ki 2005 | 463 films released in 2000-2002 | **Audience rating**, **budget**, critical reviews, distributor, genre (**drama**), MPAA rating (**PG**, **R**), season (**Easter**, **summer**, Christmas, other), **sequel**, **star actor**, director, **opening screens** |
| De Vany and Walls 1996 | 300 films released in 1985-1986 | |
| De Vany and Walls 1997 | 350 films released in 1985-1986 | First-run bookings, weekly revenue, number of weeks already in top 50, rank in top 50 and opening screens |
| De Vany and Walls 1999 | 2,015 films released in 1985-1996 | Budget, star, director, sequel, genre, MPAA rating and release year |
| De Vany and Walls 2002 | 2,015 films released in 1985-1996 | MPAA rating (G, PG, PG-13, R) |
| Elberse and Anand 2007 | 280 films released in 03/2001-05/2003 | Advertising, critical reviews, expected revenues |

Table 2.3.: Literature on cues available to consumers

| Author | Number of Films | Variables tested |
|---|---|---|
| Elberse and Eliashberg 2003 | 164 films released in 1999 | **Advertising**, **competition**, **critical reviews**, seasonality, **screens**, **star actor**, director, word of mouth |
| Eliashberg and Shugan 1997 | 56 films released in 1991-1992 | Critical reviews (total number, **positive**, **negative**, mixed), **screens** |
| Gemser et al. 2007 | 84 films released in 1998-2003 in the Netherlands | Budget, critic reviews (**total number**, rating 1-5, **size** [in a newspaper]), distributor, **film type** (art house vs. mainstream), seasonality, sequel/adaptation, **screens**, star power |
| Jedidi et al. 1998 | 102 films released in 1990-1992 | **Award win**, competition, **consumer rating**, MPAA rating (**G/PG**, PG13, R), **screens**, **seasonality**, **sequel**, star actor |
| Litman and Kohl 1989 | 697 films released in 1981-1986 | **American origin**, annual admissions, **award nomination**, award win, **budget**, **competition**, **critical reviews**, **major distributor**, genre (**science-fiction/fantasy**, **drama**), MPAA rating, season (Easter, **summer**, Christmas), **opening screens**, prequel, **sequel**, adaptation, **star actor**, **director** |
| Neelamegham and Chintagunta 1999 | 35 films released in 1994-1996 | **Age** (in weeks), **cumulative viewers**, distributor, genre (action, comedy, drama, romance, thriller), **screens**, **star actor**, country-specific intercept |
| Prag and Casavant 1994 | 652 films | **Advertising** (for 195 films), award wins*, budget*, **critical reviews**, genre (romantic/family, comedy, drama*, action), MPAA rating (G*, PG*, PG13*, R*), **sequel**, star* *significant only when advertising is excluded |
| Ravid 1999 | 180 films released in 1991-1993 | **Budget**, critical reviews (**total number**, percentage of good reviews, percentage of good and mixed reviews), MPAA rating (**G**, **PG**, PG13, R), seasonality, **sequel**, star actor (won an award, participated in top-10 grossing film in the previous year, unknown cast), director (won an award) |

Table 2.3.: (continued)

| **Author** | **Number of Films** | **Variables tested** |
|---|---|---|
| Reinstein and Snyder 2005 | 609 films | Critical reviews (predictive effect*, influence effect*), **film quality**, **genre**, **producer**, **seasonality**, **screens**<br>*mixed results |
| Sawhney and Eliashberg 1996 | 111 films released in 1992 | **Critical reviews**, genre (**action**, children, comedy, drama, horror, sci-fi), MPAA rating, **sequel**, **sexual content**, special effects, **star actor** |
| Sharda and Delen 2006 | | Competition, genre (action, cartoon, comedy, documentary, horror, historic epic drama, modern drama, politically related, sci-fi, thriller), MPAA rating (G, PG, PG13, R, NR), **screens**, sequel, **star actor**, **technical effects** |
| Smith and Smith 1986 | 600 films released before 1980 | Award (**total number**, best picture*, best actor, best actress*, best director*), **year**<br>*mixed results |
| Sochay 1994 | 263 films released in 1987-1989 | Award (**nomination**, **win**), competition, **critical reviews**, genre (action/adventure, children, **comedy**, drama, horror, sci-fi), distributor, MPAA rating (G, PG, PG-13, **R**), seasonality (**Christmas**, Easter, **Summer**), **screens**, **star actor** |
| Terry et al. 2005 | 505 films released in 2001-2003 | **Award nomination**, **budget**, **critical reviews**, genre (action, children), MPAA rating (**R**), season (Memorial Day, Independence Day, Thanksgiving, Christmas, New Years), **sequel**, **screens** |
| Zufryden 1996 | 63 films released in France | **Advertising**, **age** (in weeks), genre (**comedy**, **action**), **screens** |

Table 2.3.: (continued)

Numerous explanatory variables are modelled. The appeal of the cast is represented both by the box office performance of the last film the lead actor appeared in as well as the total number of films he has appeared in. The appeal of the director is measured similarly. Dummy variables for whether a film is a sequel or an adaptation are included as well as the production budget. Seven categories are used for genre, four categories for MPAA rating and four categories for important seasons. Further, critical reviews are calculated by averaging three different sources, and average audience ratings are taken from IMDb. The market power of the distributor is calculated by taking the number of films from the distributor in the top 100 in the previous year. Finally, the number of opening screens is also included.

The model fits best to explain first-week revenues, where budget, audience rating, sequel, number of screens and summer releases positively influence sales, while the lead actor has a negative impact. The model also provides good fit to explain total revenues, where, in addition to the previously mentioned variables, PG rating, critical reviews and the Easter release period positively influence sales, while R rating and drama have a negative impact. The length of run is not as well-explained, however, the number of opening screens lose their significance whereas the market power of the distributor becomes more important.

Terry et al. (2005) use a different data set consisting of 505 motion pictures released between 2001 and 2003 in the North American market that reached at least one hundred screens during their release time. Critical reviews are sourced from the Rotten Tomatoes website, which calculates an average rating from a number of critics. Seasonality is characterised by binary variables for Memorial Day, Independence Day, Thanksgiving, Christmas and New Year while MPAA ratings are represented solely by R-rated films. Genres are represented by action and children, and the number of award nominations is counted. Finally, the number of screens and the production budget are also included in the model.

Both a linear and a semi-log model are defined, achieving substantively similar results. Budget, critical reviews, sequel, award nominations and opening screens positively influence cinema attendance while R ratings have a negative impact.

In all of these studies the production budget has a positive impact on box office revenues. A high budget generally signals a motion picture of high quality, and it can be argued that consumers are drawn to this higher quality. Nevertheless, a high production budget does not guarantee a box office hit and flops are still possible. However, researchers agree on a generally positive direct effect of the budget on box office revenues (Litman, 1983; Litman and Kohl, 1989; Terry et al., 2005). The budget also indirectly influences revenues by having a positive impact on the length of run. More expensive motion pictures are shown longer at cinemas, thus increasing the opportunities to generate revenues (Chang and Ki, 2005).

Pokorny and Sedgwick (2010) go one step further and look at the profits of films. Using a data set of 2,116 films released between 1988 and 1999, they estimate profits by making a number of assumptions. They hypothesise that promotion and distribution costs are initially set as a proportion of production costs and that distribution costs are further related to the revenues generated by the film. Parameter estimates are calculated through regression analysis using motion pictures, for which all of the data is available, as a sample and subsequently compared to industry data provided by Vogel (2001).

The production costs are adjusted for the proportion of revenues generated in the domestic market as well as the proportion of revenues generated through theatrical release. Using the resulting estimates for film profit, they find that high-budget productions have higher variability in revenues, but at the same time are responsible for the majority of profits during the 1990s. This serves as further proof for the generally positive relationship between the production budget and the success of a film.

Although the production budget is considered one of Hollywood's best kept secrets, there are ways to estimate it. Numerous websites such as the-numbers.com and boxofficemojo.com provide publicly available estimates of the production budget, and they rarely differ significantly from each other. Yet, it seems unlikely that the average consumer conducts extensive research on the production budget. There are, however, other cues signalling a high budget. Advertising expenditures, for example, are usually highly correlated with the production budget. Estimates are that the advertising budget amounts to about 30 to 50 per cent of the production budget (Elberse and Anand, 2007; Prag and Casavant, 1994).

Prag and Casavant (1994) focus their attention on the influence of advertising when analysing a number of variables that impact box office revenues. From Variety, data on the production costs and the costs on 'prints and advertising' are sampled. They further include dummy variables on MPAA ratings, award wins, sequels and four different genres: romance/family, comedy, action and drama. Star power is measured by a self-constructed system varying between 0 for no stars and 2 for more than one established star featuring in a film. Critical reviews are summarised by calculating an average rating on a ten-point scale from two different sources.

This data is sampled for 652 motion pictures from a variety of years, the oldest film being released 77 years prior to the data analysis. However, since the data on advertising expenditure is available for only 192 films, two models are created. In the first model, revenues are dependent on production budget, critical reviews, star power and the dummy variables. In the second model including only 192 films, revenues are dependent on the same variables and the advertising budget.

In the first model, production budget, critical reviews, star power, sequel, award win, drama and the four MPAA ratings are all significant variables. In the second model, only advertising budget, critical reviews, sequel and romance/family are significant explanatory variables. Thus, when including the advertising budget, production costs as well as star power and award wins become insignificant. Prag and Casavant (1994) conclude that advertising heavily influences consumers in their decision making.

However, production budget, stars and awards are not unimportant in determining a film's appeal to consumers. The advertising budget is highly correlated with both production costs and star power and moderately correlated with award wins. This indicates that production studios are likely to invest more money into advertising, if their film features stars, has won an award or had high production costs. Creating a model with advertising costs as the dependent variable, Prag and Casavant (1994) indeed find that these variables influence the advertising expenditure.

In an effort to look at the effect of advertising on consumer decisions more closely, Zufryden (1996) proposes a three-stage model based on a behavioural framework, in which advertising is initially linked to awareness, because it alerts consumers to the existence of a particular film. In the second stage, this awareness as well as film characteristics influence the intention to see a film, which in the final stage leads to actual ticket sales.

Consumer awareness in a particular week is modelled as a function of the awareness in the previous week, a fraction for the previously unaware that have now been captured as a result of advertising expenditures, a word-of-mouth effect between previously aware and unaware consumers and a loss of awareness due to memory loss.

The intention to see a film in a particular week is modelled as a function of the current awareness level, the number of screens a film is shown on, advertising expenditures and two binomial variables for the genres comedy and action. Finally, ticket sales in a particular week are modelled in a log-linear response model as a function of the intention to see a film, the number of screens and the number of weeks since release.

Overall, data from 63 motion pictures released over a six-month period in the French market are employed for the analysis. Further, data on the number of theatre tickets sold and screens displaying a particular film are provided by a marketing research department of a film studio. The studio also provided data on the aggregate weekly advertising expenditure of their own films. In order to test the model's predictions against empirical data, consumer awareness of these 63 films and intention to see them are sampled through telephone-based interviews.

Using this data, the model can explain 97 per cent of the variability in awareness, with all of the variables being significant. It can further explain 75.8 per cent of the variability in the intention to see a film. Awareness is the best predictor of the intention to see a film; the number of screens, the amount spent on advertising as well as the genres action and comedy are also significant. Since the genres drama, children, horror and science fiction are not significant, they are not included in the final model. Finally, the model explains 89.9 per cent of the variability in ticket sales, with both the number of screens and the age of the film being significant explanatory variables.

Thus, film advertising is a key variable to influence consumer awareness of a film and thus subsequent decision-making. However, the temporal pattern of this effect is not clear. Finding that the majority of advertising expenditures is spent prior to a film's release, Elberse and

Anand (2007) adopt a slightly different approach in that they do not use box office revenues as their dependent variable, but the pre-release market expectations as held by the Hollywood Stock Exchange (HSX). In this virtual market, individuals can trade 'film stocks'. The higher the value of a particular film's share, the higher the market expectations. The HSX has proven to be a reasonable predictor of final revenue outcomes.

Sampling both HSX data as well as weekly television advertising expenditures for 280 motion pictures released between 2001 and 2003 that are widely released on 650 screens or more, they collect data for the 12 weeks prior to the release of a film. Additionally, the quality of the film is assessed by collecting the average critical review from Metacritic.

By using a dynamic model specification, in which the time-series data is first-differenced in order to get rid of "movie-specific time-invariant unobserved heterogeneity" (Elberse and Anand, 2007, p. 327), their model focusses solely on the two variables mentioned above. They argue that while film-specific factors may influence the level of advertising expenditures, it is unlikely to impact on its weekly changes. Their final model estimates that the weekly change in expectations depends on the weekly change in the level of advertising, the change of expectations in the previous week (due to changes in advertising) and a moderating effect of the quality of the film, indicating that the effect of advertising may be altered for films of different quality.

Elberse and Anand (2007) find positive results for both the direct effect of advertising changes and the carryover effect of advertising from previous weeks. This means that an increase in advertising expenditure leads to higher expectations not only in the following week, but also in subsequent weeks. Estimates show that an increase of $1 in advertising leads to an increase of $0.65 in expected revenues. They further find support for the moderating effect of film quality reflected by critical reviews. For a given rise in advertising expenditures, market expectations grow more for films with good critical reviews than for films with bad critical reviews. These results to some extent justify the large amounts spent on advertising in order to influence consumer attendance. However, the returns to advertising are negative indicating that many distributors spend too much money on advertising.

It can be concluded that advertising has positive effects on the awareness of a motion picture, making consumers conscious about a film's existence and creating interest. Advertising subsequently has an impact on the intention to see a film and thus influences ticket sales (Zufryden, 1996). This effect is especially noticeable in the opening week (Elberse and Eliashberg, 2003); however, raised awareness prior to release also has been found to have a long-term effect on sales, not only raising opening week sales, but also revenues later on (Ainslie et al., 2005; Elberse and Anand, 2007).

Another element of the production budget that consumers may look for – and one that is very visible – is the cast of a film. Prag and Casavant (1994) show that the appearance of stars is highly and positively correlated with the production budget. People may be a fan of a particular star and watch almost any film she appears in.

Some authors argue that film stars are one of the least noisy signals that a film has towards consumers. Albert (1998) therefore clusters film types according to their leading actor meaning that every film Clint Eastwood stars in is of the same film type. He assumes that consumers choose films in a stochastic process based on their previous experience with films of a similar type. Therefore, the success of a film type is dependent upon the success of similar film types in the past. He further hypothesises that the likelihood of the next successful film being of a new type – indicating a new star – is rather small. These assumptions lead to a steady-state distribution of successful films. For example, the proportion of film types that generate one successful film is 0.50 and the proportion of film types that generate two successful films is 0.167. Sedgwick (2002) largely confirms this hypothesis looking at film data from 1946 to 1965, although he finds that the number of stars with a large number of successful films is overpredicted.

Albert (1998) composes a data set consisting of the top 20 films according to rentals of each year between 1940 to 1955 and 1960 to 1995, resulting in a data set of 960 films, which are each marked by a starring actor. Subsequently, the predicted number of successful film types is compared with their actual number. Calculating the Chi-square statistic the results show that the empirical distribution is not significantly different from the predicted one. Albert (1998) concludes that films can be clustered according to leading actors and that these actors function as markers for specific film types, both drawing in audiences and providing a signal for success.

This is further investigated by Ravid (1999). He formulates two hypothesis to test whether stars have an impact on revenues and especially on the profit of a motion picture. The "rent capture hypothesis" assumes that a star essentially earns the money he adds to a film's revenues. The "marking hypothesis" states that a hired star functions as a positive signal to audiences, thereby increasing consumer interest in a film and subsequently its revenues and profits.

In order to characterise stars, he determines whether a cast member won an Oscar as Best Actor, Best Actress or Best Director, whether she "participated in a top-ten-grossing movie in the previous year" (ibid, p. 469) or whether the cast consists of unknown people. Using these measures of star power he compares star-studded films with other films and finds that the former have a significantly higher budget, higher revenues and a larger number of critical reviews. However, the rate of return is not significantly different.

Ravid (1999) develops four different models with domestic revenues, international revenues, video revenues and total revenues as the respective dependent variable. Explanatory variables in each case are budget, MPAA rating, percentage of non-negative ratings, number of critical reviews, seasonality, sequel and binary variables for a cast member or director winning an award, for the cast consisting of unknown members only and for a cast member participating in a top-ten grossing film in the previous year.

He finds that budgets are the most significant variable indicating that expensive films lead to higher revenues. However, none of the measures of star power are significant in any of the four models. Additionally, sequel, number of critical reviews, and the MPAA ratings G and PG are significant variables. He finally models the rate of return as a function of the same set of independent variables. Neither the budget nor the star measures are significant; only the MPAA ratings G and PG are.

He concludes that stars indeed increase box office revenues, but so generally does a higher budget. However, stars (or higher budgets) do not necessarily increase profits. These results lead to a non-rejection of the rent-capture hypothesis. Nevertheless, star-studded films are reviewed more often and "attention by reviewers seems to be important to success" (ibid, p. 488).

Hiring stars seems to be somewhat of a two-edged sword for film studios. On the one hand, they seem to provide a positive signalling function, as a cast including stars leads to higher box office revenues (Albert, 1998; Neelamegham and Chintagunta, 1999; Sharda and Delen, 2006). They especially seem to draw audiences to a motion picture early in its run (Ainslie et al., 2005) thus potentially starting a positive cascade: once a film has attracted a large number of people, it will keep on attracting more.

However, stars are generally able to capture fully the additional cost of hiring them – hence the strong correlation between star power and production budget – and therefore may not lead to additional profit. De Vany and Walls (1999) show that films featuring stars have both a higher budget and a wider release, two factors that have a higher impact on revenues than stars themselves. Yet, they state that stars affect the "staying power" of a film, strengthening its survival and lenghtening its life cycle in the cinema. They go on to analyse the effect of individual actors and actresses, but find that each of them has a sizeable standard error, indicating that there is no such thing as a "bankable" star. This means that consumers either each have their individual disposition towards actors and actresses, or that stars alone do not provide a reliable cue for consumers to make their decision. They are, in fact, not able to 'mark' a film as one of high quality.

The director of a motion picture provides a similar signal to consumers. Some directors have managed to achieve a 'star' status over time and new releases from these directors can be long-awaited products. Numerous studies have included directors in their models, but few have concentrated on the effect directors have on box office success. Bagella and Becchetti (1999) hypothethise that there is a positive interaction effect between star actors and directors. They build different models, in which the number of total admissions depend on star actors and director, state subsidies, producer, foreign co-producer, distributor, genre and age restriction (similar to an MPAA rating of R) and use data from Italian motion pictures released between 1985 and 1996.

Using a GMM-HAC (Generalised Method of Moments Heteroskedasticity and Autocorrelation) approach in order to account for the non-normal distribution of the data and to set the correlations between instruments as close to zero as possible, they achieve the best model fit when they include a quadratic interaction term between actors and director as an additional explanatory variable.

The quadratic interaction term and the individual effects of actors and director have the strongest correlation with total admissions providing a first indication that they influence consumers in their decision-making. In their model, only the quadratic interaction term is significant, which they attribute to the correlation between the interaction term and the individual effects. It is therefore difficult to separate the influence of actors from the influence of directors. However, the results indicate that a film with both popular actors and a popular director provides a signal of much higher quality to audiences than the sum of two films in which one includes the same cast but a different director and the other a different cast but the same director.

Again, though, research shows that the impact of directors on film choice is not straightforward. While some studies (Bagella and Becchetti, 1999; Litman and Kohl, 1989) find that well-known directors contribute positively to film success, Ainslie et al. (2005) state that the director's influence is more indirect, contributing rather to a longer staying power at the box office than directly to box office revenues. Further, a number of studies that include the director as an explanatory variable in their analysis find his influence on film success to be insignificant (e.g. Chang and Ki, 2005; Liu, 2006; Ravid, 1999).

Audiences may therefore not only look at the creative hands of a film, but rather at the kind of film they can expect, which can be represented by its genre. It is not unlikely that individuals prefer a particular genre such as horror or romantic comedy over others. They can combine this information with the cast to estimate their liking of the film, for example judging Tom Hanks to fit well into a comedy, but not to be a good actor for a crime story.

Numerous studies forecasting box office success have incorporated film genres into their analyses. Yet, there is little congruency between both the number and the different kind of genres these studies use, and different results have been achieved on the effect of varying genres. Comedy, for example, was found to either have a significantly positive impact (Bagella and Becchetti, 1999; Sochay, 1994; Zufryden, 1996) or to be insignificant (Chang and Ki, 2005). Drama, on the other hand, was mostly estimated to have a negative impact on revenues (Chang and Ki, 2005; Litman and Kohl, 1989). The action genre was found to influence revenues both negatively (Zufryden, 1996) and positively (Sawhney and Eliashberg, 1996).

In order to somewhat resolve this inconsistency, Neelamegham and Chintagunta (1999) test for marked differences in taste across markets. They conduct a cross-cultural study and predict the number of viewers for sequential releases of motion pictures in both domestic and international markets. They use a Poisson regression model, in which the parameters are modelled as a country-specific function of the number of screens, the number of previous viewers, a time trend, film genre, stars and the distributor.

Using five different genres – action, comedy, drama, romance, and thriller – they find that thriller is the most popular genre overall, but mostly preferred in Japan and Mexico. While action is a less popular genre overall, it is a significant variable in Canada, the UK and the US. Cultural differences are largest for the romance genre, but it is favoured in Germany, Sweden, South Africa and the US. They further find that the number of screens is a significant explanatory variable in all of their fourteen countries, and that stars have a positive effect on viewership while the time since release and the past audience have a negative effect.

Overall, there is a large ambiguity regarding the influence of film genres on consumer decisions. This may be due to the different kinds of genres employed in past research. It may also be due to genres generally being broad classifiers that do not say very much about a particular film. To account for this fact, films are now often classified into multiple genres (for example, IMDb classifies the film *Iron Man* into action, adventure and sci-fi). Nevertheless, consumer tastes differ across populations and it may be difficult to say that dramas are generally received negatively while action films are received positively.

However, the genre is not the only cue that tells consumers something about the kind of film they can expect to see. The Motion Picture Association of America (MPAA) rates films according to their violence and sexual content in order to assign age restrictions to them. Films are classified into suitability for the general audience (G), for parental guidance (PG), for a strong caution of parents (PG-13), and into a restricted (R) category for films containing "some adult material" (MPAA, 2011). Therefore, these MPAA ratings communicate the suitability of a particular film for children and adults.

Litman (1983) hypothesises that motion pictures rated G or PG should generate higher revenues, since a larger proportion of the audience is able to watch them. However, he finds all of the ratings to be insignificant, both in the 1970s as well as in the 1980s (Litman, 1983; Litman and Kohl, 1989).

While Basuroy et al. (2003) achieve similar results, other authors find significant differences between different MPAA ratings. Some indeed find that a PG rating contributes positively to domestic film revenues, though this is not true for a G rating (Ravid, 1999; Chang and Ki, 2005). This effect seems to be negligible, though, once the influence of advertising is accounted for (Prag and Casavant, 1994). The strongest findings concern the R rating; a number of studies find this variable to deter people from going to see a film (Sochay, 1994; Chang and Ki, 2005).

De Vany and Walls (2002) investigate this issue further. They use a sample of 2,015 motion pictures released between 1985 and 1996. During this decade, the majority of films released were R-rated, namely 1,057. Similarly, the majority of box office hits, which they characterise as films generating more than $50 million in revenues, have an R rating, namely 68, while only eight G-rated films earned that much money. However, the rate of success for R-rated films is only 6 per cent, whereas it is 13 per cent for G- and PG-rated films. They continue to look at "returns hits", which they classify as films earning more than three times their production budget. Again, only 11 per cent of R-rated films fall into this category, whereas it is 20 per cent for G-rated and 16 per cent for PG-rated films.

They show that the stable Paretian model can represent the distribution of box office revenues well, especially its heavy tails and its extreme skewness. Using this distribution, they compare the probabilities of outcomes for different MPAA ratings and find that G- and PG-rated films' revenue and profit outcome probabilities stochastically dominate R-rated films. This means that for any given outcome, the probability that a G- or PG-rated film earns more than the outcome is higher than the probability for an R-rated film.

Thus, similar to film genres, the MPAA rating, while providing consumers with an indication of the content with regards to violence, language and sexuality, is a rather imprecise signal. There is some support for R-rated motion pictures deterring people from seeing them, but there is ambiguity about whether any of the other ratings is able to draw audiences towards a film.

Therefore, people may turn to 'experts' who can tell them more about a film's content and quality. These experts are professional film critics who are often allowed an advance screening of a motion picture, usually because production studios hope that positive criticism will induce people to watch the film. However, it has been argued that the judgments of experts and the tastes of the mass audience differ drastically from one another (Bourdieu, 1984, 1993).

Holbrook (1999) tests two claims empirically: whether expert judgments differ demonstrably from mass consumer taste and whether there is a negative correlation between experts and ordinary consumers. He assumes that consumers generally prefer to have the world portrayed in a realistic fashion, whereas experts prefer abstract representations. From this he formulates six hypotheses: consumers prefer MPAA ratings for a wide audience; consumers prefer particular genres (e.g. family entertainment) over others (e.g. sci-fi); consumers prefer motion pictures originating from the U.S. over foreign productions; newer, longer films in colour resonate more with ordinary consumers; some stars or directors appeal more to mainstream audiences than others; award wins are more strongly related to expert judgments.

He samples 1,000 motion pictures that were released in the U.S. before 1986, won an academy award and were among the critical favourites. Subsequently, two dependent variables are constructed. For the taste of ordinary consumers, ratings on a six-item scale from HBO consumers are employed for each of the films. Similarly, two comprehensive film guides are used for expert judgments. As independent variables, measures for violence – on a six-item scale – and sexuality – on a seven-item index – are included as well as six dummy variables for the genres family, comedy, drama, musical, sci-fi, and western. Further, the country of origin, colour, length, year, the number of awards a film won, and dummy variables for the presence of 11 directors and 39 actors are also included.

Successively, Holbrook (1999) runs two regressions and finds that ordinary consumers respond significantly more positively to the family genre, domestic origin, colour features, longer duration, recency and star power than experts. They are influenced significantly more negatively by violence, sexual content and science-fiction and respond less positively to award wins than experts do. However, despite these seeming differences in taste, the popular appeal as reflected in consumer ratings and the expert judgments from film guides are positively correlated. This indicates that the two groups actually share somewhat similar tastes. Holbrook (1999) attributes these findings to unobserved variables such as the storyline, themes or motifs or special effects.

Maybe inspired by these findings, further research on the influence of professional film critics on consumer choice was conducted. Eliashberg and Shugan (1997) state that film production studios often use the reviews from film critics in order to influence consumers at the beginning of a film's release when no other information is available. According to this industry wisdom, they hypothesise that critics are influencers who have an impact on audiences in the early weeks of a film's release. However, this effect is assumed to die out as more word of mouth from peers becomes available, overriding critical reviews.

They sample 56 motion pictures released between 1991 and early 1992, which ran for at least eight weeks at the cinema. Weekly box office revenues are collected from Baseline Inc. and Entertainment Data Incorporated. They further obtain 2,104 reviews on these films from 181 critics from Variety, which are classified into three categories – positive, negative and mixed.

Accounting for the fact that the percentage of positive and negative reviews are highly correlated, they separately model both weekly and cumulative revenues as a function of either the percentage of positive or negative reviews, the number of screens to account for the studio's marketing efforts and the total number of critical reviews.

They find that both the percentage of positive reviews as well as the percentage of negative reviews are insignificant explanatory variables during the first four weeks of a film's release, but become significant in later weeks as well as for cumulative box office revenues. Further, the number of reviews is significant only in the opening week, whereas the number of screens is positively correlated and significant throughout a film's release as well as for its cumulative revenues.

This leads to a rejection of the original hypothesis of film critics being influencers. Instead Eliashberg and Shugan (1997, p. 76) characterise them as predictors meaning that critics' reviews are "predictive of performance" although they do not necessarily cause it. This may be the case because film critics are able to express consumers' feelings towards a film or stand as representatives for the mass audience. Alternatively, it may simply be the high quality of the film that leads experts to write positive reviews and causes people to watch it.

Basuroy et al. (2003) further investigate the dual role of critics. Similar to Eliashberg and Shugan (1997), they hypothesise that if critics have the greatest effect on early box office revenues but not on late or cumulative revenues – meaning their impact declines over time –, they can be characterised as influencers. If critics affect late or cumulative revenues, they are predictors and in case they have an effect on both, they can be considered being both influencers and predictors.

They employ a random sample of 175 motion pictures released between 1991 and 1993 from Baseline and Variety. First, they replicate Eliashberg and Shugan's (1997) model by modelling weekly revenues as a function of the total number of reviews, the percentage of either positive or negative reviews and the number of screens for each of the first eight weeks. In contrast to Eliashberg and Shugan (1997), they find positive and negative reviews to be significant over the whole time period. This may be due to the larger sample size, the slightly longer time period they employ or the random selection of motion pictures.

In order to strengthen their findings, they include a number of dummy control variables in their model: star power – measured by whether a cast member or director won an award prior to the release of the film –, production budget, sequel, MPAA rating and seasonality. The results essentially remain the same; positive and negative reviews are significant for each of the eight weeks as well as the number of screens. The budget is significant for the first four weeks, but becomes insignificant thereafter, and sequel is significant only in the opening week.

Next, they run a cross-section time-series analysis adding a variable for the week of release as well as an interaction variable between either positive or negative reviews and week in order to test for the declining effect of reviews over time. Their results again confirm the previous outcomes; both positive and negative ratings are significant explanatory variables as well as the budget, the number of screens and the week. However, while the interaction term between positive reviews and week is insignificant, the interaction term between negative reviews and week is significant indicating that the effect of negative reviews declines over time. This result partially supports the role of critics as influencers.

In order to test for a negativity bias the authors use the number of positive and negative reviews (instead of the percentage), on the grounds that the reviews are only weakly correlated and therefore can be legitimately included in the same regression model. They find both negative and positive reviews to be significant and to affect revenues in the expected direction, respectively negatively and positively. However, the difference between the two coefficients is insignificant indicating that positive reviews influence consumers as much as negative reviews. They attribute this to the diminishing effect of negative reviews over time, and subsequently limit their analysis to the first week. Now positive reviews become insignificant, but the difference between positive and negative reviews becomes significant. This points to the existence of a negativity bias, where negative reviews lead consumers to shun a film in the opening week.

They conclude that film critics play a role as both influencers and predictors. They are able to 'correctly' predict both weekly and cumulative success of a motion picture, while the existence of a negativity bias in the early weeks of a film's release suggests that critics can convince individuals to stay away from a particular film.

It can be argued, though, that the correlation between critical reviews and consumer demand is spurious due to the underlying quality of a film. A good film will achieve both good reviews and a large audience, but these may happen independently from each other. In order to investigate this, Reinstein and Snyder (2005) account for the timing of a review. They argue that a review published after a film's opening weekend cannot influence the revenues of that weekend, and can therefore only have a prediction effect for later weeks. In contrast, a review published before a film's opening can have both an influence and a prediction effect. Differencing between these two, they propose to isolate the influence effect.

They sample 609 motion pictures and use both the opening weekend revenue as well as the total revenue as dependent variables. Critical reviews are collected from two popular film critics, Siskel and Ebert who each assign a 'thumb up' or a 'thumb down' for each motion picture. Therefore, a particular film can either have zero, one or two thumbs up.

Opening weekend revenues are modelled as a function of the timing of the review – before or after the opening weekend –, dummy variables for receiving one or two thumbs up, an interaction effect between the timing and these two dummy variables, control variables for the number of opening screens as well as dummy variables for long weekends, year, month, genre and production company, and an additional variable of unobserved film quality on a four-point scale by employing the film rating by critic Leonard Maltin.

They find that all of their critics' variables are insignificant in this model. Merely the number of screens, film quality and the dummy variables for year, month, genre and producer are significant.

Subsequently, they divide their sample into wide releases shown on more than the median number of screens and narrow releases as well as into the genres drama, action and comedy. For wide releases and comedies, both variables for receiving one and two thumbs up are significant, whereas both interaction terms are insignificant, indicating that film critics are predictors for these kinds of motion pictures. In contrast, both one and two thumbs up are insignificant variables for narrow releases and dramas, whereas the interaction between the timing and two thumbs up is marginally significant and the interaction between the timing and one thumb up is significant for dramas, pointing to an influence effect. None of the critics' variables are significant for action films.

They conclude that critics are more influential for 'art' films than for big 'event' films. This may be the case because consumers have enough information for the latter kind already, since these films are usually heavily advertised. In order to get more information about the former kind, they turn to experts, since other information is not as easily accessible. Alternatively, consumers do not share the experts' opinions for blockbuster-type films, whereas they rather trust their opinion when it comes to an 'elitist' taste on art films.

However, the precise effect that film critics have on consumer choice remains inconclusive. Numerous studies have found reviews to positively affect box office revenues (Chang and Ki, 2005; Elberse and Eliashberg, 2003; Terry et al., 2005). It is less clear whether this happens due to film critics directly influencing consumer choice by starting a positive or negative cascade or whether expert judgments and popular appeal are correlated due to the underlying quality of a film. While Eliashberg and Shugan (1997) find support for the former, both Basuroy et al. (2003) as well as Reinstein and Snyder (2005) get mixed results, supporting both roles to some extent.

Of a similar influence, one may argue, is the recognition that particular motion pictures get from the industry via the award of prizes. These are generally awarded by film experts as well and say something about the artistic recognition of film. The most famous of these are the Academy Awards or Oscars. Different authors have included either Oscar nominations, Oscar wins or both into their analysis.

Smith and Smith (1986, p. 502) sample 600 "box office champion films" released before 1980 and model domestic rental income as a function of the year of release, the total number of Oscar wins, and dummy variables for Best Picture, Best Actor, Best Actress and Best Director separately for the 1950s and before, the 1960s and the 1970s. They find that the total number of Oscars significantly influences consumers to go and see a film, and that this effect grew from the 1950s to the 1970s. However, the results of the individual awards are less conclusive. An Oscar for Best Picture influences film revenues negatively during the 1960s, but positively during the 1970s. An Oscar for Best Actor is insignificant, while an award for Best Actress positively influenced consumers during the 1950s and negatively during the 1970s. Similarly, an Oscar for Best Director contributes positively to film success during the 1960s, but negatively during the 1970s.

Other authors have usually not looked at the different awards separately, but rather created one variable for whether a film won or got nominated for an award. Terry et al. (2005) limit their analysis to Oscar nominations in only a few categories, namely Best Picture, Best Director, Best Leading and Supporting Actor and Actress, and find that they positively influence consumers to go and see a film. Litman and Kohl (1989), in contrast, who concentrate only on Best Picture, Best Actor and Best Actress, but include both award nominations as well as award wins, find that award nominations lead to higher attendance, but that an additional award win has no further significant effect. Prag and Casavant (1994) state that award wins contribute indirectly to a film's success, namely through increased advertising expenditure.

It is not unlikely, though, that even a film only being nominated for an Oscar may use this information for an additional or increased advertising campaign, thus leading to Litman and Kohl's (1989) results. However, Ravid (1999) who uses award wins as a measure for star power finds this variable to be insignificant.

The problem of relating award nominations to consumer decisions or film success when using a data set that stretches over a number of months or years is that the effect of an award nomination or win will be different depending on the stage of a film's life cycle. Nelson et al. (2001) show that a film released just before nominations are announced has a higher probability of gaining from it.

In order to provide consumers with a more precise signal of what to expect from a film, production studios have often invested in brand extensions by creating sequels or prequels to an already existing - and usually successful - motion picture. The idea is that if a large enough audience enjoyed a particular film, they will also enjoy a very similar film involving the same characters and a continuation of the story being told. On the other hand, sequels may reach a point of saturation, because film audiences are to some extent looking for novelty, which is why they rarely go to see a film twice at the cinema. Therefore, if a sequel is too similar to the original idea, people may form low expectations about it and not watch it.

Basuroy et al. (2006) assume that both sequels and advertising provide a positive signal to consumers, since a production studio would not invest money into a product if it was not of high quality. Thus, sequels and advertising serve as signals of quality in the absence of other cues. However, once more signals become available, such as critical reviews and word of mouth, sequels and advertising become less important to consumers. Especially if film critics have a uniform opinion about a particular film, this reduces uncertainty about its quality and therefore reduces the signalling quality of sequels and advertising.

In order to test their hypotheses, they randomly sample 175 films released between late 1991 and early 1993, in which sequels represent 6.3 per cent of the sample. They create a variable for critics' consensus by summing up the squares of proportions of positive, negative and mixed reviews from Variety. They further operationalise word of mouth as the cumulative number of screens since a film's release implying that the more people are able to go and see a film the more people are likely to talk about it. And they measure star power by the total number of awards that any crew member either won or was nominated for.

They develop a dynamic simultaneous equations model with three independent equations, wherein weekly revenue, weekly advertising spending and weekly screens are the dependent variables to account for the "interrelationships of behaviors among movie audiences, studios, and exhibitors" (Basuroy et al., 2006, p. 291). Opening revenues are modelled as a function of the number of screens, star power, competition, seasonality, advertising expenditure, percentage of positive reviews, dummy variables for major distributor and sequel as well as interaction terms for advertising and critics' consensus, sequel and critics' consensus, and sequel and advertising. Revenues of later weeks are modelled as a function of the number of screens, competition, seasonality, advertising expenditure, week of release, word of mouth, and interaction terms for advertising and word of mouth as well as sequel and word of mouth.

They find that sequels, advertising expenditure, positive reviews, star power and opening screens are positive and significant variables, whereas seasonality is insignificant. Competition is marginally significant and negative. Further, the interaction term between sequel and critics' consensus is negative and significant while the interaction term between advertising and critics' consensus is negative and marginally significant. This supports their hypothesis that the more critics share an opinion about a film's quality prior to its release, the less valuable is the signal provided by a sequel or advertising, because uncertainty is significantly reduced.

For the later weeks of a film's release, they find the number of screens, advertising expenditure and word of mouth (the cumulative number of screens) to be positive and significant. The week of release is negative and significant, while competition and seasonality are insignificant. Similarly to the opening week, they find both the interaction terms for advertising and word of mouth as well as for sequel and word of mouth to be negative and significant. Again, this supports their hypothesis that the existence of word of mouth reduces the signalling effect of both sequels and advertising to audiences.

Nevertheless, for the opening week, the interaction term for sequel and advertising is positive and significant, indicating that consumers not only use both of these signals, but that they positively influence each other. An explanation would be that consumers assume that a film studio would not independently invest money in both if the film was not of high quality.

Many other studies are roughly in line with these findings. Mostly, though, they have simply included sequel as a dummy variable into their regression analysis and found it to be a significant explanatory variable (e.g. Chang and Ki, 2005; Litman and Kohl, 1989; Prag and Casavant, 1994; Ravid, 1999). However, other studies have made less consistent findings with regards to the impact of sequels on consumer decisions. Basuroy et al. (2003) find it to be significant only in the opening week, but not in subsequent weeks, and two studies find it to be insignificant (Gemser et al., 2007; Sharda and Delen, 2006). Yet, the fact that out of the top ten grossing films in the North American market, five were sequels in 2009, four in 2010 and a staggering nine in 2011 (Box Office Mojo, 2012), lends some support to the assumption that consumers are attracted to stories and characters that they are already familiar with.

A final cue that many studies have identified as being very significant when determining box office revenues is the number of screens that a particular film is shown on. It is reasonable to assume that consumers do not actually count the number of screens or that it significantly sways their opinion about a film. However, the availability of a film in a cinema nearby drives the decision on whether it will actually be watched (Elberse and Eliashberg, 2003; Zufryden, 1996). The number of screens is usually set in negotiations between the distributor and exhibitors and often dependent on the release strategy. A wide release is often accompanied by a large advertising budget, while narrow releases, in contrast, generally advertise much less but rather hope to build up word of mouth over time thus increasing demand for the film (Eliashberg et al., 2000).

Elberse and Eliashberg (2003) assume that screens are allocated by exhibitors depending on their expectations of demand. Subsequently, audiences consume films depending on their availability, and this consumption in turn updates exhibitor decisions regarding screen allocation in later weeks.

In order to test and account for this interrelationship between consumers and exhibitors, they develop a dynamic simultaneous-equations model. They further employ a three-stage least-squares (3SLS) procedure to account for the endogeneity of the number of screens. They create a data set consisting of 164 motion pictures that were produced or co-produced in the U.S., released in the U.S. in 1999, and reached the top-25 U.S. box office at least once during their cinematic release.

Subsequently, a number of explanatory variables are constructed. Data from the Hollywood Stock Exchange is used as a measure of expected first-week revenue. Word of mouth is operationalised as the revenues per screen in the previous week. Star power and director power are measured according to a publication from the Hollywood Reporter. Finally, three measures of competition are created, one in which the number of new releases in a given week are weighted by their production budget, one in which the age (in weeks) of the top-25 films in the previous week is averaged, and one in which the number of films in the top-25 that are of the same genre or MPAA rating are counted and weighted by their age (in weeks). The former two measures account for the competition of screen space, whereas the latter represents the competition for audiences.

Their estimation strategy consists of four equations, with weekly revenues and weekly screens as dependent variables, modelled seperately for the opening week and later weeks. They model opening revenues as a function of the number of screens, word of mouth, competition from films with a similar target audience, seasonality, star actor, director, advertising expenditure and critical reviews. Later weeks' revenues are modelled as a function of the number of screens, competition, seasonality, word of mouth, and time-variant dummy variables. Opening screens are modelled as a function of expected revenues, word of mouth, competition from both new and on-going films, budget, star actor, director, advertising expenditure, critical reviews, and distributor. And later weeks' screens are modelled as a function of expected revenues, competition from both new and on-going films, word of mouth and time-variant dummy variables.

For opening week revenues, screens, star power, advertising, critical reviews and competition are significant, with competition being the only variable having a negative influence. Later weeks' revenues are significantly influenced by screens, competition and word of mouth. The number of screens is a very important explanatory variable both in the opening week as well as later in a film's life cycle, exceeded only by word of mouth in later weeks – however, the word of mouth measure is also related to the number of screens a film is shown on.

For opening week screens, expected revenues, advertising and critical reviews are significant. Later weeks' screens are significantly influenced by expected revenues, competition from new films and word of mouth. While critical reviews contribute positively to opening revenues, they contribute negatively to screen allocation in the first week. Elberse and Eliashberg (2003) attribute this to the negotiation power of distributors to put low-rated films on more screens in order to draw as many people as possible to a film in its early weeks.

When comparing the coefficients of their 3SLS model with the coefficients of a similar ordinary least-squares (OLS) model, they find that certain variables ore over-emphasised in the OLS model. For example, the influence of advertising on opening revenues is much smaller in the 3SLS estimation, whereas its influence remains the same on opening screens. This leads them to the conclusion that many film-specific variables influence box office revenues indirectly via the allocation of screens. In other words, the number of screens captures to a considerable extent the film characteristics such as budget, advertising and stars in regression-type analyses, an approach that is taken e.g. by Moretti (2011).

In conclusion, there has been a vast amount of research on signals that may influence consumer decisions regarding film consumption. Regardless of this, there is little congruency between the different results that have been reached using different kinds of regression methods. There seems little reason to believe that higher advertising budgets will generally lead to higher rates of return or that action films will generally fare better than dramas.

However, despite the different approaches taken, the different variables employed and the different time periods covered, the listed studies all achieve good levels of fit for their models. And although the extent to which each cue contributed to consumer decisions remains ambiguous, overall, these studies show that the signals that are available to consumers prior to the release of a film can to a large extent explain their decisions.

Finally, there is some indication that the film-specific characteristics can to an extent be captured in the number of screens a film is released on. This is due to the fact that more expensive films featuring stars or well-known directors and a large advertising budget are generally released on a larger number of screens.

## 2.4. Conclusion

Motion pictures are a very particular kind of product. They are experience goods whose quality cannot be assessed prior to consumption. Every film is unique in its own sense, and therefore the knowledge gained about the quality of a particular film cannot be directly used for future consumption decisions. Further, if it were possible to gather full information about a film prior to consumption, this may reduce the individual pleasure derived from watching the film - because motion pictures are hedonic goods, consumers look for some kind of novelty.

The reviewed literature has highlighted a couple of further issues. Motion pictures are high-risk products – for consumers, since they do not know whether they will like a film prior to consumption, but especially for film producers, since nearly all of the expenses are sunk costs and there is no guarantee of recouping them. This is mainly due to the market being dominated by only very few products and it being difficult to land one of the few box office 'hits'.

It is therefore unsurprising that a large amount of research has been conducted to predict whether certain characteristics such as the cast, the genre or the marketing strategy can enhance the probability of landing a hit. The results are, overall, not very conclusive though. While there seem to be some intuitive factors driving success - the number of screens or the amount of advertising are positively related to box office revenues - the impact of other factors is not as clear-cut.

Thus, it is not clear how consumers form their expectation of a film and make their decisions. It has been argued that word of mouth providing reliable information is one of the key drivers of product adoption. A strand of research has inferred this effect from the distribution of film revenues, yet it remains ambiguous in these studies how consumers learn from word of mouth. This thesis aims to investigate whether cinema-goers tend to engage in observational learning, inferring a film's quality from the number of people who have seen the film before them, or word of mouth learning, using the evaluation of previous consumers to update their expectations.

Another strand of research has used online word of mouth to empirically assess its impact on consumer decisions. Usually separating word of mouth into volume and valence measures, these studies have achieved conflicting results. While the volume of word of mouth seems to positively affect consumer decisions, the effect of the valence is not clear. Some studies have found it to positively influence consumer decisions, while others have found it to be insignificant.

Arguing that the way these studies have measured the valence of word of mouth may have contributed to these conflicting results, this thesis develops a new method of aggregating consumer opinions and analyses its impact on consumption decisions at the cinema. This measure further allows to test whether consumers are more sensitive to positive or negative word of mouth, thus assessing the existence of a negativity effect in the case of motion pictures.

As the distribution strategy or the number of screens a motion picture is shown on has proven to be an important determinant of success, the thesis examines whether the effect of word of mouth differs depending on the release strategy of the film. Specifically, it investigates whether so-called 'sleepers' are driven to success by positive word of mouth.

# 3. Methodology

> "The necessity of introducing 'error terms' in economic relations is not merely a result of statistical errors of measurement. It is as much a result of the very nature of economic behavior, its dependence upon an enormous number of factors, as compared with those which we can account for, explicitly, in our theories." (Haavelmo, 1943)

This study conceptualises the decision-making of consumers as having three main influences: the individual's previous experience with a product or a product class, which is generally unobserved; the characteristics of the product that can be easily accessed by consumers, i.e. the search attributes of a product; and the word of mouth that consumers receive about a product both from their peers or the online sphere. The amount of previous experience is likely to moderate the influence of both word of mouth and product characteristics on the decision. This relationship is illustrated in Figure 3.1 and is very similar to conceptualisations used in previous research (e.g. De Vany and Lee, 2001; Moretti, 2011).

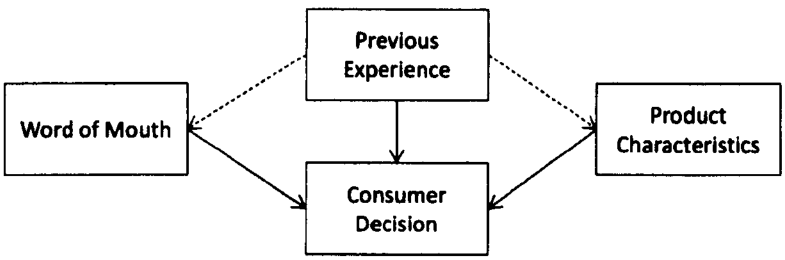



Figure 3.1.: Conceptualisation of the influences on consumer decision-making

It is further assumed that prior to the release of a new product a consumer can only use her previous experience and the known product characteristics to make a purchase decision. Only after some consumers have bought the product are they able to communicate to other people their experience with it. The effect of word of mouth on consumer decisions may therefore vary over time.

It is acknowledged that word of mouth is existent prior to the release of many products, yet it cannot contain any reliable qualitative information about the product. In the case of motion pictures, a considerable amount of word of mouth is published prior to the release of a film in the online sphere. Yet, this information consists largely of anticipatory comments pointing out consumer expectations (Asur and Huberman, 2010). It does not contain reliable information regarding the quality of a film and thus provides little information that a consumer can learn from in order to make his decision. Therefore, this study focusses on post-release word of mouth, since this is likely to have been communicated by consumers who have already seen the film.

In order to address the research aim stated in Chapter 2.4, this conceptualisation is adopted in a two-stage methodology within the framework of a sensitivity analysis as shown in Figure 3.2. The methodological design is based upon the process illustrated in Figure 3.1 and therefore closely mirrors it. Similar to previous studies employing empirical data (e.g. Liu, 2006; Moretti, 2011), a regression-type method is used to analyse the effect of different variables on consumer learning and decision-making.

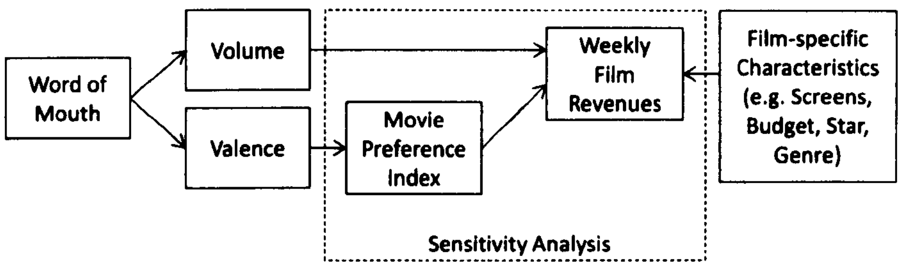



Figure 3.2.: Research design

In this model, film revenues are used as the dependent variable, since they are a quantifiable expression of consumer decisions. Of particular interest is the effect that post-release (online) word of mouth has on shaping consumer decisions. In order to quantify this influence, the data on word of mouth is clustered into two different dimensions, namely volume and valence, to test their respective effects and infer the way in which consumers learn about single-purchase experience products.

The volume is a simple number specifying how many people talk about a particular product. It represents the amount of 'buzz' that surrounds it. The volume of word of mouth has been shown to influence consumer awareness about a product (Berger et al., 2010; Liu, 2006). However, the volume of word of mouth is also a representation of past audience behaviour, since plausibly it follows that the more people purchased a product, the more likely they are to communicate their views (Godes and Mayzlin, 2004). Volume may therefore serve as an indicator of the extent to which observational learning takes place. Consumers may infer from a large volume that a product is not only popular, but also of high quality otherwise not as many people would have purchased it.

In contrast to the volume, the valence contains evaluative information about the experience that a previous consumer made with a product. It says something about the product's quality. Valence therefore serves as an indicator of the extent to which consumers learn from qualitative word of mouth information.

Finally, product characteristics are represented by the observable film-specific characteristics that are available to consumers prior to the release of a motion picture, such as the production budget – an indication of which is visible to consumers via marketing expenses , star presence, number of screens, or the film genre. Because the previous experience of consumers is not observable, it is not directly included as an independent variable in the regression analysis. However, since every regression model contains an error term or a residual capturing what the independent variables cannot explain, the intrinsic influences form part of the residual in this model.

In order to collect the necessary data, publicly available secondary sources are used. Box office statistics are aggregated from a provider of entertainment news. Information on word of mouth is collected from an online social network, in which consumers can both rate and review motion pictures publicly, and data on film characteristics are gathered from various websites.

In this study, the valence is computed by means of a Movie Preference Index (MPI) to represent the aggregate consumer opinion on particular motion pictures. A number of MPIs are tested in order to account for different weightings that consumers may apply to film ratings. Individuals may focus on positive reviews and go to see a film if it received high ratings; they may also be influenced more by negative reviews, in which case a film would quickly lose its audience.

Each of these different MPIs is employed to run separate regression-type analyses, in which film revenues are the dependent variable representing consumer decisions and both film-specific characteristics and word of mouth are the independent variables representing the measurable influences on consumer decisions (see Figure 3.2). Subsequently, a sensitivity analysis is conducted to determine whether different calculations of the MPI lead to significantly different outcomes. Thus, the way in which consumers handle product ratings can be evaluated.

The following section describes the method of data collection and the characteristics of the data set. Subsequently, the reasons for using the MPI as a representation of word of mouth valence are explained and its calculation demonstrated. Finally, mixed-effects models are introduced, the chosen type of regression analysis for this research.

## 3.1. Data Collection

In line with much of the previous research undertaken on the film industry, this study concentrates on the North American market. Due to the existence and dominance of Hollywood in the industry, it is the largest and most important geographical market in terms of revenues, and it shapes the overall financial outcome of a film to a great extent (Elberse and Eliashberg, 2003; Neelamegham and Chintagunta, 1999). Another reason to select the North American market is its abundance of available data.

This study uses data solely from publicly available sources. The general approach of the data collection is to follow new releases through their life cycle and collect their revenues as a representation of consumer decisions as well as their film-specific characteristics and the word of mouth data that each of these films generates.

Data collection took place over a period of twenty weeks, from April to September 2010. During this time, every new release among the top 100 weekly domestic box office list published by Variety (2012), a renowned provider of entertainment news, was added to the data set. For each of these films, the title, distributor, weekly rank, weekly box office revenue, percentage change in weekly revenues, number of screens, and the week of release are recorded from Variety as well.

Further, data on whether the film contains a major star is recorded using the comprehensive top-ten money making stars list by Quigley Publishing (2011). A film is defined as containing a star if any member of its cast was featured in this list in any of the previous five years. Data on the production budget, the genre, the MPAA rating and whether the film is a sequel is collected from the Internet Movie Database (IMDb, 2012) and backed up by data from Box Office Mojo (2012) and The Numbers (2012), especially in the case of production budgets, since these are generally estimates. Finally, data on the opinion of film critics is collected by using the Metascore, a weighted average score of numerous respected film critics calculated by Metacritic (2012).

The final data set consists of 132 motion pictures. As the descriptive statistics in Table 3.1 show, films generally have a short life cycle, with the average film running for less than eight weeks in cinemas. The median film earns about $180,000 over its life time, but both the mean and the skewness statistic indicate heavy tails, meaning that the revenue distribution is dominated by very few films generating extremely large revenues. This is a defining characteristic of the motion picture industry.

| | Weeks on Release | Cumulative Revenues | Opening Revenues | Opening Screens | Budget | Critical Rating | Rating Volume | Week 1 Rating Volume | Average Rating |
|---|---|---|---|---|---|---|---|---|---|
| Mean | 7.74 | 25,573,798 | 11,448,626 | 765 | 40,451,483 | 57.58 | 335 | 174 | 68.45 |
| Median | 7 | 179,031 | 42,884 | 4 | 10,000,000 | 59 | 12.5 | 12 | 70 |
| Std Error | 0.42 | 5,967,493 | 2,629,646 | 126 | 6,834,251 | 1.6 | 74 | 40 | 1.34 |
| Std Deviation | 4.86 | 68,561,272 | 29,282,500 | 1,405 | 57,586,423 | 16.44 | 789 | 352 | 14.35 |
| Kurtosis | -0.85 | 12.89 | 12.48 | 0.39 | 1.81 | -0.52 | 14.99 | 12.17 | -0.07 |
| Skewness | 0.44 | 3.51 | 3.38 | 1.47 | 1.7 | -0.28 | 3.56 | 3.18 | -0.09 |
| Minimum | 1 | 67 | 828 | 1 | 100,000 | 21 | 1 | 1 | 31.7 |
| Maximum | 20 | 410,010,779 | 167,551,682 | 4468 | 200,000,000 | 92 | 4958 | 2014 | 100 |
| N | 132 | 132 | 124 | 124 | 71 | 105 | 114 | 77 | 114 |

Table 3.1.: Descriptive Statistics of the Motion Picture Data Set

Generally, films earn the largest share of their revenues in the opening week, about 45 per cent on average. The distribution of first-week revenues shows the same characteristics as cumulative revenues: a very high skewness and thus heavy tails indicating the dominance of a small number of motion pictures.

The median film opens on four screens, but again this data is positively skewed, signifying that a small fraction of films opened on an extremely large number of screens and that the majority of films opened on fewer than the arithmetic mean number of screens. The budget is only available for 71 films, since reliable estimates do not exist for a significant proportion of films. Nevertheless, the production budget also shows a positively skewed distribution meaning that only few films are produced on large budgets – the standard deviation is larger than the mean –, increasing the mean production budget to $40.5 million.

Similarly, not all of the 132 motion pictures are reviewed by a large enough number of film critics for Metacritic to calculate its Metascore, a score that ranges from 0 to 100. The films in the sample achieved a mean critical rating of 57.58. The lowest score is 21 and the highest 92, showing that critics voice a wide array of opinions.

With regards to film genres, the data set contains 18 action films, two adventures, three animations, 35 comedies, six crime films, 35 documentaries, 43 dramas, seven horror films, and two musicals. Using the convention explained above, ten of the motion pictures contain a major film star. Finally, the Motion Picture Association of America awarded two films a G rating, 22 were rated PG, further 19 rated PG-13, and 45 rated R. The remaining motion pictures were not rated at the time of data collection.

For each of the films, individual-level rating data is collected from Rotten Tomatoes, an online community focussing on motion pictures. Rotten Tomatoes was chosen amongst a number of alternative sources for various reasons. First, it is a very popular online film review hub (Alexa Internet, 2010). Second, due to its community character, it is more likely that users interact with each other than on more forum-like websites such as Yahoo! Movies. Finally, most people using Rotten Tomatoes are from North America, which does not apply to other film review hubs (Alexa Internet, 2010). Thus, the film ratings submitted on Rotten Tomatoes are most likely to have been published by North American film-goers and consequently provide the most valid and reliable data source for online word of mouth.

The data collected from Rotten Tomatoes consists of film title, date of submission, rating – on a scale from 0 (lowest) to 100 (highest) in decimal steps –, and name of the user. The final data set consists of 38,235 individual ratings, for which descriptive statistics are displayed in Table 3.1. The mean film received 335 ratings over its life cycle, yet, similar to revenues and budgets, these ratings are positively skewed both over films and time. Very few films receive an extremely large number of ratings while the majority of films get comparatively few ratings. And films generally receive most of their ratings in the opening week of their release – approximately half of their ratings are submitted then. It is noteworthy, though, that some films do not receive any ratings during their opening week.

Over its time of theatrical release, the average motion picture receives an arithmetic mean rating of 68.45, indicating that consumers generally rate films positively. There is some deviation around the mean, but the relatively small value of the standard deviation and the fact that the distribution of rating valences is barely skewed further hint at the fact that consumers rarely rate a film very negatively. In fact, the lowest mean rating that a film received is 31.7. The arithmetic mean of a film rating has its shortcomings as an aggregate measure of consumer opinions, though, as the next section shows. It is therefore not used in this study.

This data set needs to be adjusted to be used in subsequent data analyses so that the results are valid, reliable, and transferable to a wider population. Because the aim of this study is to analyse the impact of word of mouth over time, motion pictures that were not listed in Variety's weekly top 100 list for at least three weeks were removed from the data set. For films that ranked outside the top 100 for not more than one consecutive week and re-entered the following week, the revenues of the missing week could be calculated using the given weekly change rates.

Further, in order to calculate a valid valence measure of consumer opinion on a particular film, it is necessary that this film was rated by at least a certain number of people. Obviously, it is desirable to include as many films as possible in the analysis; however, a minimum number of ten different ratings is considered necessary to express an 'average' consumer evaluation. During the 20-week period, 21 films did not receive any rating, and another 44 films did not receive ten or more ratings. Thus, these films were dropped. The final data set, on which all subsequent analysis is based, consists of 58 motion pictures, which were listed in Variety's top 100 for at least three weeks and received at least ten ratings on Rotten Tomatoes during the time of data collection.

## 3.2. The Movie Preference Index

It is a common approach in empirical research using online word of mouth to separate word of mouth into different dimensions, most commonly volume and valence. In all of these studies, the volume is composed by counting the number of posts over a certain period of time, for example the number of comments during a week, a day, or an hour. However, in order to calculate the valence, authors have used different approaches. They have either coded it into positive and negative sentiments about the product and subsequently used the respective percentages as independent variables within their models (Liu, 2006; Asur and Huberman, 2010) or calculated the arithmetic mean (Duan et al., 2008b; Chintagunta et al., 2010).

However, the 'sentiment approach' has limitations. For instance, it is insensitive to subtleties such as the difference between an 8/10 rating and a 10/10 rating. Both are positive ratings, but they may be interpreted differently by consumers. Therefore, coding both of these ratings as 'positive' is too simplistic. The sentiment approach also generally neglects ratings in the middle range, since these are interpreted as neutral and therefore implicitly assumed to not have an effect on consumer decision-making.

In contrast, the 'arithmetic mean approach' incorporates all of the different ratings and thereby the subtleties between them. It was therefore initially decided to use this approach to aggregate consumer opinions, as it had been used in previous research as well. Yet, preliminary results showed that the arithmetic mean rating does not influence consumer decisions significantly, and, even more confounding, that it seems to have a negative impact. A full discussion of this issue can be found in Appendix A.

It is thus surmised that the arithmetic mean provides an inappropriate measure of aggregated consumer opinions, since it does not properly account for the 'fuzziness' of rating data. Whenever a complex object is evaluated, individuals assign weights to each of its different characteristics. Some features may be of more importance than others. For example, when buying a car, consumers may assess its fuel and cost efficiency, comfort or appearance. Different weights may be assigned to these features deliberately, for instance focussing mainly on fuel and cost efficiency and putting comfort in third place while neglecting appearance, but often consumers weigh the features of a global evaluation unconsciously (Kahneman, 2011).

Applied to the evaluation of film ratings, this suggests that users may deliberately ascribe varying weights to different ratings. A negative rating may be perceived much stronger than a positive rating (cf. Basuroy et al., 2003; Kahneman, 2011) and thus have a much bigger effect on the consumption decision. Similarly, the negative evaluation of a film by one close friend or a trusted expert may override the positive feedback that an individual got from several acquaintances. For these reasons, a motion picture rated with, say, 8/10 is not necessarily twice as good as a motion picture rated 4/10, an assumption that the arithmetic mean implicitly makes.

Finally, using the arithmetic mean to aggregate the qualitative information of film ratings may obscure very different 'rating profiles' that these films received. Table 3.2 provides an example for this. Both Film A and Film B received the same number of ratings. Yet, Film A received very diverse ratings, being rated positively by part of the audience and badly by the other part. Film B, in contrast, received generally moderate ratings. Nevertheless, the mean rating for both films is the same, and a model using the mean as its measure of valence fails to capture differences in the spread of audience opinion.

In order to overcome these limitations and account for the characteristics of rating data, a different method of aggregating ratings and calculating the valence is employed. This measure of consumer opinion is termed the Movie Preference Index (MPI). The MPI is able

| Rating | Film A | Film B |
|---|---|---|
| 5 | 15 | 0 |
| 4 | 10 | 17 |
| 3 | 5 | 24 |
| 2 | 8 | 9 |
| 1 | 12 | 0 |
| Count | 50 | 50 |
| Mean | 3.16 | 3.16 |

Table 3.2.: Two films with different rating profiles

to account for different ways of interpreting film ratings, such that it can be analysed whether consumers are influenced more by positive or negative opinions. It is also able to be 'fine-tuned' so that the differences between 'good' and 'very good' ratings are visible. Finally, it is able to express the aggregate consumer opinion of a particular film as a numerical value. Below, its explicit characteristics and its differences with regards to the arithmetic mean are highlighted alongside its mathematical model.

Generally, in order to calculate the valence of ratings that users give to a specific product $i$, a Z-score of the form of

$$Z_i = \sum_{r=1}^{k} \frac{w_r v_{ir}}{n_i} \tag{3.1}$$

is calculated, where $v_{ir}$ denotes the number of $r$th rank ratings product $i$ received, $w_r$ is the weight assigned to this rank, and $n_i$ is the total number of ratings product $i$ received. In Table 3.2, in order to calculate the arithmetic mean, the weight $w_r$ would be the numerical value of the rating (ranging from 1 to 5), $v_{ir}$ would be the number of ratings it got for each rank (e.g. Film A received 15 five-star ratings), and $n_i$ would be represented by the 50 ratings that each film received.

The essential contribution of the MPI is that it does not assign weights to each rank in a linear fashion. For instance, when calculating the arithmetic mean, the difference between two adjacent weights in Table 3.2 is always one (e.g. 4 minus 3 equals 1). Thus, when using the arithmetic mean, there is a linear relationship between different ratings and it is implicitly assumed that a film rated with four stars is twice as good as a film rated with two stars. Yet, because consumers are likely to interpret ratings differently, this may not be the case.

The MPI, in turn, accounts for the varying weights that consumers attach to different ratings. These weights are determined using Ordered Weighted Averaging (OWA, Yager, 1988). The OWA operator is chosen for this study, because it is the most widely used aggregation model amongst a large number of different methods (Zhou et al., 2008). It is based upon the idea that any decision-making paradigm requires the aggregation of different preferences or criteria into an overall evaluation that appropriately weighs the contribution of these preferences or criteria. When deciding on buying a car, the question is how much more important is fuel efficiency than comfort. Applied to the evaluation of film ratings, it is necessary to estimate how much more important a positive rating is to a negative rating (or vice versa) to consumers.

In general, there are different decision-making paradigms: group decision-making, multi-criteria decision-making, or a mixture of both. A group decision involves different individuals who each may have different preferences and/or have a higher or lower influence in comparison to others. For example, the opinion of an expert or a superior may have a greater impact on the overall decision. In order to determine the overall evaluation of a group decision (the Z-score in Equation (3.1)) it is therefore necessary to assign the appropriate weight to the influence or importance of each individual. The result thus reflects both the different preferences within the group as well as each individual's influence.

In a multi-criteria decision, one option amongst different alternatives is selected based on multiple criteria. These criteria also need to be weighted according to their importance before an overall evaluation of each alternative can be calculated. For example, when purchasing a car, the fuel efficiency may be more important than the cost efficiency, which in turn is more important than the comfort. Only after each criterion has been assigned a weight can an overall evaluation of the car be calculated.

In essence, the difference between group decision processes and multi-criteria decision processes is that in the former weights are assigned to individuals according to their influence on the decision and in the latter weights are assigned to criteria according to their importance.

In OWA applications, these weights usually carry values between zero and one and they add up to one (Triantaphyllou, 2010). Plausibly, the least important criterion is assigned the lowest weight – potentially as low as zero in case it does not play any role in the final decision – whereas the most important criterion is assigned the highest weight – potentially as high as one if it is the only criterion that plays a role in the final decision.

The OWA operator has previously been used in research on management decisions within a multi-criteria environment or in ballot decisions as well as in different applications within computational intelligence. For the purpose of this study, OWA provides a class of aggregation operators or weights, where the different film ratings from Rotten Tomatoes are ranked according to their importance (Yager, 1988). Similar to the arithmetic mean, the aggregated result using OWA operators represents a mean consumer evaluation of the film. Yet, due to the different way in which weights are assigned to each rating, this result reflects the different ways, in which consumers may interpret film ratings, more accurately than the arithmetic mean.

Two measures have been defined to determine the weights of OWA operators: the orness and the dispersion. The orness degree describes the extent to which the weights are distributed towards the 'extreme' ranks. For example, if there are five criteria ranked by importance, from $a_1 \dots a_5$ (this could be the five different film ratings mentioned in Table 3.2), then the two most extreme weight vectors are $[1,0,0,0,0]$ and $[0,0,0,0,1]$, because all of the weight is put on either the most important or the least important criterion. Mathematically, the orness degree $\alpha$ is defined as

$$\alpha = \frac{1}{n-1} \sum_{r=1}^{n} (n-r)\, w_r \tag{3.2}$$

where $n$ denotes the number of criteria, $r$ signifies the rank of a criterion, and $w_r$ is the weight ascribed to the according rank.

The weight vector $[1,0,0,0,0]$ would therefore have an orness degree of 1, whereas the orness degree of the weight vector $[0,0,0,0,1]$ would be 0. In the case of film ratings, a high orness degree (close to one) puts a lot of weight on high or positive film ratings whereas a low orness degree (close to zero) puts a lot of weight on low or negative film ratings (see Table 3.3 for the distribution of weights ascribed to the different ranks of film ratings from Rotten Tomatoes using different orness degrees). Thus, the orness degree can be interpreted as a decision maker's optimism, since it describes how much weight is put on the best or worst outcome (Wang et al., 2007a; Yager, 1988).

Yet, it is possible to create two weight vectors with equal orness degree, but very different characteristics, such as $[0,0,1,0,0]$ and $[\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5},\frac{1}{5}]$. Both weight vectors have the same degree of orness ($\alpha = 0.5$) in that they are balanced between the two extreme distributions of weights, but the weights are spread over the five criteria in the latter case.

Therefore, a measure of dispersion is introduced describing the entropy of the weight distribution or the degree to which all criteria are equally weighted. The dispersion is mathematically defined as

$$dispersion(W) = -\sum_{r} w_r \ln w_r \tag{3.3}$$

where $W$ is the weight vector containing all the weights, $r$ is the rank of the criteria and $w_r$ is the weight ascribed to the according rank. The dispersion is minimum (or zero) if any weight $w_r$ is equal to one and is maximum (or $\ln n$) in case the weights are equally distributed over the number of criteria and each weight $w_r$ is equal to $\frac{1}{n}$.

Different methods have been suggested to determine the weight distribution over a given number of criteria. Most of these have either used the dispersion measure, the orness degree or both to calculate the final weight vectors. In this thesis, the chi square method (CSM) suggested by Wang et al. (2007b, p. 209) is applied to aggregate rating data, because the weights determined by the CSM model "do not follow a regular distribution". This means that neither the difference between two adjacent weights nor the ratio of two adjacent weights remains the same. While the difference between two weights in Table 3.2 is always one, Table

3.4 shows that, in the case of an orness degree of 0.8, the difference between two weights decreases exponentially, thus putting a lot of weight on the highest ranked criterion. Further, the weights change more smoothly with the degree of orness using the CSM method than with other methods suggested for determining OWA weights.[5]

In order to generate the weights $w_r$ using the CSM method, the function

$$Minimise J = \sum_{r=1}^{n-1} \left( \frac{w_r}{w_{r+1}} + \frac{w_{r+1}}{w_r} - 2 \right) \tag{3.4}$$

fulfilling the constraints

$$s.t. orness(W) = \alpha = \frac{1}{n-1} \sum_{r=1}^{n} (n-r)\, w_r, \quad 0 < \alpha < 1$$

$$\sum_{r=1}^{n} w_r = 1, \quad 0 \leq w_r \leq 1, \quad r = 1, \ldots, n$$

is optimised, where $r$ denotes the rank of the criterion, $n$ specifies the number of ranks, $w_r$ specifies the weight attached to the rank and $\alpha$ denotes the orness degree.

This method generates a different weight $w_r$ for each rank $r$. For the data in Table 3.2, five different weights are needed, summing to one. However, since Rotten Tomatoes, the source chosen in this study for data on word of mouth, allows film ratings to vary between 0 and 100 in decimal steps, eleven weights are calculated, one for each possible rating. The sum of these weights also adds up to one (see Table 3.3).

| $W$ | $Orness(W) = \alpha$ | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | 0.9999 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0.0001 |
| $w_1$ | 0.9914 | 0.4852 | 0.2659 | 0.1787 | 0.1309 | 0.0909 | 0.0588 | 0.0435 | 0.0244 | 0.0075 | 0.0003 |
| $w_2$ | 0.0043 | 0.2378 | 0.2033 | 0.1554 | 0.1229 | 0.0909 | 0.0630 | 0.0461 | 0.0258 | 0.0080 | 0.0003 |
| $w_3$ | 0.0009 | 0.1128 | 0.1452 | 0.1301 | 0.1138 | 0.0909 | 0.0681 | 0.0503 | 0.0289 | 0.0091 | 0.0003 |
| $w_4$ | 0.0007 | 0.0584 | 0.1024 | 0.1081 | 0.1045 | 0.0909 | 0.0739 | 0.0565 | 0.0341 | 0.0112 | 0.0003 |
| $w_5$ | 0.0006 | 0.0337 | 0.0734 | 0.0901 | 0.0958 | 0.0909 | 0.0805 | 0.0648 | 0.0422 | 0.0149 | 0.0004 |
| $w_6$ | 0.0005 | 0.0214 | 0.0545 | 0.0759 | 0.0878 | 0.0909 | 0.0878 | 0.0755 | 0.0545 | 0.0214 | 0.0005 |
| $w_7$ | 0.0004 | 0.0149 | 0.0422 | 0.0649 | 0.0805 | 0.0909 | 0.0958 | 0.0901 | 0.0734 | 0.0337 | 0.0006 |
| $w_8$ | 0.0003 | 0.0112 | 0.0341 | 0.0565 | 0.0739 | 0.0909 | 0.1045 | 0.1082 | 0.1024 | 0.0584 | 0.0007 |
| $w_9$ | 0.0003 | 0.0091 | 0.0289 | 0.0504 | 0.0681 | 0.0909 | 0.1138 | 0.1303 | 0.1452 | 0.1128 | 0.0009 |
| $w_{10}$ | 0.0003 | 0.0080 | 0.0258 | 0.0462 | 0.0630 | 0.0909 | 0.1229 | 0.1557 | 0.2033 | 0.2378 | 0.0043 |
| $w_{11}$ | 0.0003 | 0.0075 | 0.0244 | 0.0436 | 0.0588 | 0.0909 | 0.1309 | 0.1790 | 0.2659 | 0.4852 | 0.9914 |
| Sum | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 | 1.0000 |

Table 3.3.: The OWA operator weights determined by the chi square model (Source: Wang et al., 2007b)[6]

[5]Other methods for calculating OWA weights include the maximum entropy method (O'Hagan, 1988), the minimum variance method (Fullér and Majlender, 2003) or the minimax disparity approach (Wang and Parkan, 2005).

The CSM method also allows calculating different 'weighting schemes' by using different orness degrees, which vary between 0 and 1. Depending on the orness degree, the distribution of the weights over the ranks varies. Since there is no conclusive research on whether consumers are generally optimistic or pessimistic decision-makers, selecting one orness degree and thus one weighting scheme would be an arbitrary choice. In order to overcome this problem, different weighting schemes are calculated for eleven different orness degrees, ranging from 0 to 1 in 0.1-margin steps (see Table 3.3). In this way, both the subtleties of film ratings can be captured and the optimism or pessimism of consumers can be analysed.

Since the orness can be interpreted as a decision-maker's optimism, an orness degree of 0.8, for example, would describe an optimistic, risk-seeking consumer who puts a lot of weight on positive ratings and almost neglects lower ratings associated with a poor experience. The cumulative weight of the top-three ranks is 0.6144 whereas the cumulative weight of the bottom-three ranks is 0.0691. In contrast, an orness degree of 0.1 would describe a risk-averse consumer putting a lot of weight on negative ratings. This kind of individual may be persuaded to stay away from a particular film by negative ratings. In this case, the cumulative weight of the top-three ranks is 0.0246 whereas the cumulative weight of the bottom-three ranks is 0.8358.

In both cases, Film A in Table 3.2 would receive the greater valence measure, because the extreme ratings are weighted much more than the middle-range ratings. In the case of an orness degree of 0.8, the high 5-star ratings are assigned the highest weight, whereas in the case of an orness degree of 0.1, the low 1-star ratings are assigned the highest weight. Since Film A has a larger number of these extreme ratings (positive as well as negative), its final valence measure would be greater than for Film B. Thus, if film consumers are influenced by word of mouth and are optimistic decision-makers, the MPI with an orness degree of 0.8 should have a significant positive impact on film revenues, and Film A should fare better, ceteris paribus. In contrast, if they are pessimistic decision-makers, the MPI with an orness degree of 0.1 should have a significant negative impact, and Film A should fare worse, ceteris paribus.

| Difference between two adjacent weights | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| $w_1$-$w_2$ | $w_2$-$w_3$ | $w_3$-$w_4$ | $w_4$-$w_5$ | $w_5$-$w_6$ | $w_6$-$w_7$ | $w_7$-$w_8$ | $w_8$-$w_9$ | $w_9$-$w_{10}$ | $w_{10}$-$w_{11}$ |
| 0.2474 | 0.125 | 0.0544 | 0.0247 | 0.0123 | 0.0065 | 0.0037 | 0.0021 | 0.0011 | 0.0005 |
| | | | | | | | | | |
| Ratio of two adjacent weights | | | | | | | | | |
| $w_1/w_2$ | $w_2/w_3$ | $w_3/w_4$ | $w_4/w_5$ | $w_5/w_6$ | $w_6/w_7$ | $w_7/w_8$ | $w_8/w_9$ | $w_9/w_{10}$ | $w_{10}/w_{11}$ |
| 1.3079 | 1.4001 | 1.418 | 1.3951 | 1.3468 | 1.2915 | 1.2375 | 1.1799 | 1.1202 | 1.0574 |

Table 3.4.: The distribution of the OWA operator weights determined by the CSM method under $\alpha = 0.8$

Each of the eleven degrees of orness results in a different weighting scheme. Thus, for every film, there are eleven potential MPI values at any given point in time. Each of these va-

[6]The CSM method approximates the orness degrees of 1 and 0 by 0.9999 and 0.0001, respectively, for mathematical reasons.

lence measures is subsequently used in regression-type analyses to examine how consumers interpret different ratings. Higher degrees of orness give greater emphasis to positive ratings, whereas lower degrees of orness give more weight to negative ratings. The size of the respective orness coefficient in the regression analysis will indicate which weighting scheme most realistically reflects film consumers' attitudes to risk. It is anticipated that higher orness degrees will influence consumer decisions positively, whereas lower orness degrees affect consumer decisions negatively.

To illustrate this, Table 3.5 shows two films, *The Last Airbender* and *Toy Story 3*, and their rating profile for the first week of their release. Both films received a similar volume of ratings. However, the valence of the ratings is very different, as is apparent looking at the spread of the rating volume over the different ratings in the top half of the table.

The bottom half of the table presents the MPI calculated by Equation (3.1) for the different orness degrees ranging from 0 to 1 and their according weights $w_r$ displayed in Table 3.3 as well as the arithmetic mean for both films. The arithmetic mean provides further proof for the fact that consumers enjoyed *The Last Airbender* much less than they enjoyed *Toy Story 3*. Yet, it doesnt say very much about the spread of ratings.

| Rating | 100 | 90 | 80 | 70 | 60 | 50 | 40 | 30 | 20 | 10 | 0 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| The Last Airbender | 84 | 57 | 54 | 44 | 45 | 38 | 26 | 42 | 50 | 282 | 73 | 795 |
| Toy Story 3 | 559 | 129 | 58 | 18 | 7 | 2 | 2 | 1 | 1 | 3 | 2 | 782 |
| | | | | | | | | | | | | |
| Orness degree $\alpha$ | 1.0 | 0.9 | 0.8 | 0.7 | 0.6 | 0.5 | 0.4 | 0.3 | 0.2 | 0.1 | 0 | Mean |
| The Last Airbender | 0.11 | 0.09 | 0.08 | 0.08 | 0.08 | 0.09 | 0.10 | 0.11 | 0.13 | 0.14 | 0.09 | 39.81 |
| Toy Story 3 | 0.71 | 0.40 | 0.24 | 0.17 | 0.13 | 0.09 | 0.06 | 0.05 | 0.03 | 0.01 | 0.00 | 94.74 |

Table 3.5.: Two exemplary films, their ratings and their MPI for different orness degrees

Looking at the different MPI measures, more details become apparent. For example, the fact that *Toy Story 3* has an MPI of 0.71 for an orness degree of 1.0 (this will be notated as $\text{MPI}_{\alpha=1.0}$ hereafter) expresses that it received a large percentage of maximum ratings. Similarly, an $\text{MPI}_{\alpha=0.1}$ of 0.01 indicates that only a very small fraction of ratings are extremely negative.

*The Last Airbender*, on the other hand, receives its largest MPIs for orness degrees of 0.1, 0.2 and 1.0 as well as 0.3 in descending order. This shows that most of its ratings are extremely negative, but that a large fraction are nevertheless extremely positive, indicating an overall large spread of ratings.

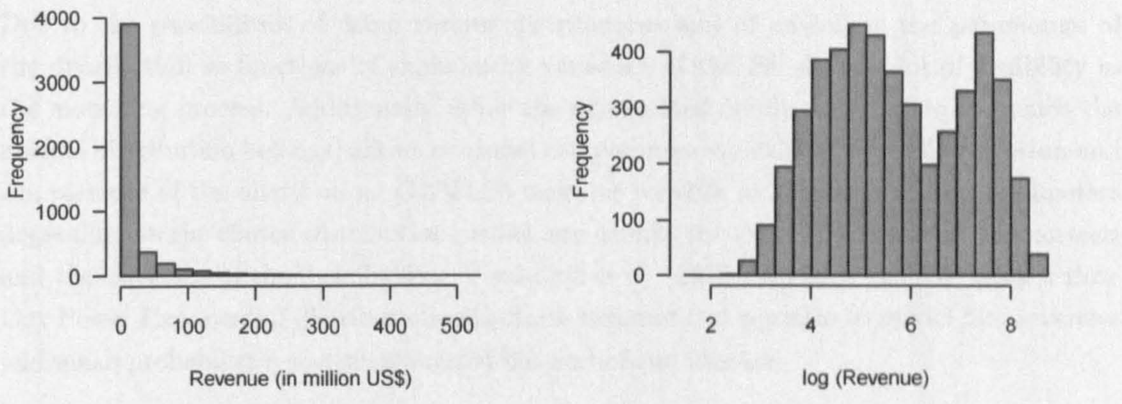
If, in a subsequent analysis of the different MPIs, the $\text{MPI}_{\alpha=0.2}$ had the largest coefficient, this would indicate that consumers are generally more influenced in their decision-making by negative ratings, which cause them to stay away from a particular film, than by positive ratings. If, on the other hand, the $\text{MPI}_{\alpha=0.9}$ had the largest coefficient, this would mean that consumers are generally positive, look whether a film received positive ratings and almost neglect negative ratings in their decision.

In conclusion, using the different MPIs allows on the one hand a more detailed picture of the ratings a film received and on the other hand an analysis of whether consumers are generally rather pessimistic or optimistic in their decision-making with regards to the influence of word of mouth. The latter is tested using mixed-effects models, a regression-type of analysis that the next section explains.

## 3.3. Mixed-Effects Methods

The "great workhorse" of economic and econometric analysis is the least squares or ordinary least squares (OLS) model, which is a common choice when the change in one variable is sought to be explained by the change in one or a number of other variables (Engle, 2001, p. 157). The general approach of empirical research in the motion picture industry, for example, is to use film revenues as the dependent variable that is to be explained by various explanatory variables.

OLS models generally assume the data to be normally distributed, an assumption that does not hold for motion picture revenues (see Figure 3.3 for an example using data from all films released in the American market in 2008). Researchers have often addressed this issue by using the logarithm, thereby flattening the curve of film revenues and assuming that the data is log-normally distributed (e.g. Chintagunta et al., 2010; Liu, 2006).




(a) in absolute numbers (b) in logarithms

Figure 3.3.: Distribution of film revenues

Yet, it has been shown that the distribution described by film revenues, whether on a daily, weekly or cumulative basis, does not belong to the normal family. Essentially, the normal family cannot capture the extreme variance that characterises film revenues and the motion picture industry. It is important, though, to use an appropriate statistical model with appropriate stochastic assumptions to describe the data; the wrong statistical model may lead to a false economic theory (Haavelmo, 1943).

Using a sample of 300 motion pictures, De Vany and Walls (1996) suggest that a stable distribution with a Paretian upper tail describes the distribution of film revenues well. This distribution implies that the variance of film revenues is theoretically infinite and the final outcome of a particular film is therefore virtually unpredictable.

This finding is refined by De Vany and Walls (1999) who use a larger sample of motion pictures and state that film revenues are Pareto-Lévy distributed. These distributions feature heavy upper tails and are extremely skewed: they have a theoretically infinite variance. The implications are very similar to their earlier paper. The mean motion picture revenue is dominated by very few blockbusters, and the chance of landing a blockbuster is very small.

In conclusion, OLS models and the normal-family distributions cannot capture the extreme variance displayed by film revenues, because the range of possible outcomes under these distributions is not wide enough. De Vany and Walls (1996, 1999) state that Pareto-Lévy distributions with infinite variance can capture the distribution, thus reverting to another extreme, where the range of possible outcomes is (at least theoretically) without boundaries.

Voudouris et al. (2012) show that film revenues can be accurately modelled using a less extreme distribution, thus making the final outcome more predictable once certain features of the motion picture are known. Modelling post-opening box office revenues as a function of opening week revenues, they employ generalised additive models for location, scale and shape (GAMLSS) introduced by Rigby and Stasinopoulos (2005) in order to select the distribution that best fits the data.

Due to the possibilities of using various distributions and of modelling the parameters of the distribution as functions of explanatory variables, GAMLSS allows a lot of flexibility in the modelling process. Additionally, while the exponential family distribution (to which the normal distribution belongs) allows to model two parameters, which are usually the mean and the variance of the distribution, GAMLSS makes it possible to model up to four parameters depending on the chosen distribution , which are usually the mean, the variance, the kurtosis and the skewness of the distribution. Voudouris et al. (2012) conclude that by using a Box-Cox Power Exponential distribution with finite variance it is possible to model film revenues and make probabilistic statements about the end-of-run income.

The model that Voudouris et al. (2012) create uses only two measurements over time, namely the opening-week box office revenues and the end-of-run revenues. They are mainly interested in building a stochastic model to predict the final outcome of motion picture revenues and thus provide a tool for film producers, distributors, and exhibitors.

In contrast, this study is interested in both the effect that word of mouth has on consumer decisions over time, and whether it affects them differently for various clusters or 'kinds' of motion pictures. The model employed therefore needs to be able to capture the variation across these different clusters, which is something that GAMLSS cannot easily do in its current implementation. To account for this, mixed-effects methods are used in this research, an approach that has previously been used mainly in the physical, biological and medical sciences. They are called mixed effects, because this method combines both fixed and random effects.

Fixed-effects parameters can be compared to the covariate parameters in an OLS regression. They describe the relationship between the independent variables and the dependent variable for the entire population. However, the results of a study may vary systematically for different subjects or clusters within the sample. For example, if a study wants to make general claims about students, then the results may systematically vary for the sampled classrooms or schools. In the case of motion pictures, the effects may vary for e.g. different genres or 'blockbuster' versus 'sleeper'-type films.

Random effects fulfil two tasks. First, they are classification variables with different levels (e.g. classrooms, schools, or different films), which have been randomly sampled from a population of different levels to be analysed. Random effects are specific to such clusters or units. Second, they provide parameters describing how the results of a study using a random sample vary around the population mean with regards to the specified clusters or units. Thus, mixed-effects models can describe the relationship between a response variable and some explanatory variables in data "grouped according to one or more classification factors", such as clustered data, repeated-measures data, or longitudinal data (Pinheiro and Bates, 2000, p. 3).

For instance, if the aim of a study is to make a statement about (the population of) fifth grade students, the fixed effects describe how the independent variables (e.g. student characteristics) influence the dependent variable (e.g. math grades). If the study samples students from different classes within a school or from different schools, then the dependent variable may systematically vary for these different levels. Random effects can be classified according to classrooms or schools and account for this systematic variation. Similarly, the effect of word of mouth on motion picture consumption may vary systematically for different films, genres etc. In this case, random effects can be classified according to these levels and account for their systematic variation.

Essentially, mixed-effects models extend standard linear models through random effects, which are additional error terms that account for both the correlation among measurements of the same subject at different points in time or under different conditions and the systematic variation in the response variable according to this subject.

The aim of this research is to estimate how word of mouth impacts consumer decisions in the case of motion pictures using repeatedly measured weekly revenues as the dependent variable. Since it is only possible to sample a certain number of motion pictures, the different levels are the different sampled films. Random factors are used to capture the variation in the response variable across the different sampled films, so that the results of the data analysis can be generalised to a greater population (West et al., 2007). The random factors are therefore specific to units or clusters of units within the population of motion pictures.

Following West et al. (2007), Equation (3.5) provides a general specification of a linear mixed-effects model using longitudinal data.

$$Y_{it} = \beta_1 X_{it}^1 + \beta_2 X_{it}^2 + \cdots + \beta_p X_{it}^p + b_{1i} Z_{it}^1 + b_{2i} Z_{it}^2 + \cdots + b_{qi} Z_{it}^q + \epsilon_{it} \tag{3.5}$$

$Y_{it}$ represents a continuously measured response variable, and $i$ represents the subject while $t$ indexes the time period at which measurement took place. The model includes two sets of covariates, $X$ and $Z$. The $p$ $X$-covariates are associated with the fixed effects, $\beta_1 \dots \beta_p$ and the $q$ $Z$-covariates are associated with the random effects, $b_{i1} \dots b_{iq}$ that are specific to subject $i$. Finally, $\epsilon_{it}$ stands for the residual of the $t$th observation on the $i$th subject.

The $p$ covariates can be either time-invariant characteristics of the subject (e.g. the production budget of a film) or time-varying for each measurement (e.g. the number of screens a film is shown on each week). The $\beta$ parameters represent "the fixed effect of a one-unit change in the corresponding covariate on the mean value of the dependent variable assuming that the other covariates remain constant" (West et al., 2007, p. 16). Estimating these parameters therefore allows a direct interpretation whether a covariate influences a response variable and in which direction this influence lies.

This model can be more efficiently expressed in matrix notation, as Equation (3.6) shows:

$$\begin{aligned} Y_i &= X_i\beta + Z_i b_i + \epsilon_i \\ i &= 1 \dots n_i, \quad b_i \sim N(0, D), \quad \epsilon_i \sim N(0, \sigma^2 I) \end{aligned} \tag{3.6}$$

$Y_i$ is the vector of the response variable containing the values for the $i$th subject. $X_i$ is a $n_i \times p$ matrix with the $p$ covariates, $X^1 \dots X^p$, for each of the $n_i$ measurements on the $i$th subject (for example, each film $i$'s number of screens over the $n$ weeks of data collection). $\beta$ is a vector of $p$ fixed-effects parameters associated to each of the $p$ $X$-covariates.

$$Y_i = \begin{pmatrix} Y_{1i} \\ Y_{2i} \\ \vdots \\ Y_{n_i i} \end{pmatrix} \qquad X_i = \begin{pmatrix} X_{1i}^1 & X_{1i}^2 & \cdots & X_{1i}^p \\ X_{2i}^1 & X_{2i}^2 & \cdots & X_{2i}^p \\ \vdots & \vdots & \ddots & \vdots \\ X_{n_i i}^1 & X_{n_i i}^2 & \cdots & X_{n_i i}^p \end{pmatrix} \qquad \beta = \begin{pmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}$$

$Z_i$ is the $q \times n_i$ matrix representing the covariates, $Z^1 \ldots Z^q$, for the $i$th subject. The columns of this matrix contain the values for the $q$ parameters which have an effect on the response variable and are associated to vary randomly for each subject. In the case where only the intercept is assumed to vary randomly across subjects, this vector contains only one column of 1's. The $b_i$ vector contains $q$ random effects associated with the $q$ covariates in the $Z_i$ matrix. Finally, $\epsilon_i$ is a vector containing the $n_i$ residuals for each of the $i$ subjects on the $t$th occasion.

$$Z_i = \begin{pmatrix} Z^1_{1i} & Z^2_{1i} & \cdots & Z^q_{1i} \\ Z^1_{2i} & Z^2_{2i} & \cdots & Z^q_{2i} \\ \vdots & \vdots & \ddots & \vdots \\ Z^1_{n_i i} & Z^2_{n_i i} & \cdots & Z^q_{n_i i} \end{pmatrix} \qquad b_i = \begin{pmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{qi} \end{pmatrix} \qquad \epsilon_i = \begin{pmatrix} \epsilon_{1i} \\ \epsilon_{2i} \\ \vdots \\ \epsilon_{n_i i} \end{pmatrix}$$

It is assumed that the random effects $b_i$ are random variables following a multivariate normal distribution with mean 0 and the variance-covariance matrix $D$. Elements along the diagonal of this matrix represent the variance of each random effect in $b_i$, and the other elements represent the covariances between two random effects. Since there are $q$ random effects, $D$ is a symmetric $q \times q$ matrix.

$$b_i \sim \mathcal{N}(0, D) \qquad D = Var(b_i) = \begin{pmatrix} Var(b_{1i}) & cov(b_{1i}, b_{2i}) & \cdots & cov(b_{1i}, b_{qi}) \\ cov(b_{1i}, b_{2i}) & Var(b_{2i}) & \cdots & cov(b_{2i}, b_{qi}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(b_{1i}, b_{qi}) & cov(b_{1i}, b_{qi}) & \cdots & Var(b_{qi}) \end{pmatrix}$$

The residuals of a mixed-effects model can be correlated, which is in contrast to standard linear models. It is also important for this study, since the residuals of weekly film revenues are likely to be correlated within each film. It is assumed that the $n_i$ residuals in the vector $\epsilon_i$ are random variables following a multivariate normal distribution with mean 0 and the variance-covariance matrix $\sigma^2 I$. It is further assumed that the residuals of different subjects are independent of each other.

$$\epsilon_i \sim \mathcal{N}(0, \sigma^2 I) \qquad \sigma^2 I = Var(\epsilon_i) = \begin{pmatrix} Var(\epsilon_{1i}) & cov(\epsilon_{1i}, \epsilon_{2i}) & \cdots & cov(\epsilon_{1i}, \epsilon_{n_i i}) \\ cov(\epsilon_{1i}, \epsilon_{2i}) & Var(\epsilon_{2i}) & \cdots & cov(\epsilon_{2i}, \epsilon_{n_i i}) \\ \vdots & \vdots & \ddots & \vdots \\ cov(\epsilon_{1i}, \epsilon_{n_i i}) & cov(\epsilon_{1i}, \epsilon_{n_i i}) & \cdots & Var(\epsilon_{n_i i}) \end{pmatrix}$$

These specifications are used in combination with the conceptualisation of consumer decision-making (Figure 3.1) to create the specific model employed in this paper. Similar to previous studies, the weekly film revenues will be used as a representative of consumer decisions (e.g. Liu, 2006; Rui et al., 2011). The independent variables of greatest interest are the valence and the volume of word of mouth. These are complemented by various film-specific characteristics, which can either vary over time, such as the number of screens or the week of release, or be time-invariant, such as the film budget or the presence of a star actor, in order to test their respective impact on consumer decisions. This model is presented in Equation (3.7):

$$
\begin{aligned}
\log(Rev_{it}) = {} & \beta_0 + \beta_1 \log(MPI_{it}) + \beta_2 \log(Vol_{it}) + \beta_3 \log(Screens_{it}) \\
& + \beta_4 \log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i \\
& + \beta_8 Sequel_i + \beta_9 Star_i + \beta_{10} MPAA_i + \beta_{11} Genre_i + b_{0i} \\
& + b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it} \\
i = 1 \ldots n_i, \quad & b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)
\end{aligned}
\tag{3.7}
$$

$\log(Rev_{it})$ is the logarithm of the weekly revenue of film $i$ at time $t$, $\log(MPI_{it})$ is the logarithm of the MPI of film $i$ at time $t$, $\log(Vol_{it})$ is the logarithm of the weekly volume of ratings of film $i$ at time $t$, $\log(Screens_{it})$ is the logarithm of the number of screens film $i$ is shown on at time $t$, $\log(Budget_i)$ is the logarithm of the budget of film $i$, $Week_{it}$ is the number of weeks for which film $i$ has been on release at time $t$, $Week_{it}^2$ is the square of this number, and $Critic_i$ is the rating that film critics attributed to film $i$. $Star_i$ and $Sequel_i$ both are binary variables denoting the presence of a star actor in film $i$ and signifying a sequel respectively. $MPAA_i$ is a categorical variable representing the MPAA rating of film $i$, and $Genre_i$ is a categorical variable representing the genre of film $i$. $\beta_0$ represents the population intercept term, and $\beta_1$ to $\beta_{11}$ represent the different fixed effects. $b_{0i}$, $b_{1i}$ and $b_{2i}$ are the random effects associated with film $i$, and $\epsilon_i$ is the $n_i$-dimensional within-group error vector with a spherical Gaussian distribution.

The logarithm is used for weekly revenues, MPI, rating volume, screens and budget in order to smoothen their distribution so that it approaches a normal distribution. The week of release and the squared week of release are included to allow an exponential influence of time on consumer decisions, something that has been observed for a significant number of motion pictures: revenues decline at exponential speed after a film's release.

Finally, the random effects introduce by-film adjustments both to the intercept ($b_{0i}$) and to the slope with regards to the week of release ($b_{1i}$ and $b_{2i}$). This means that the intercept of each film is allowed to vary randomly around the population mean, and that each film's slope is allowed to vary around the population mean slope over time. In other words, since the trajectories of individual film revenues are likely to be different from each other, the random effects capture the variability between these trajectories (West et al., 2007). Thus, the fixed-effects parameters allow an interpretation of the influence of the covariates on film consumption decisions in general, while the random-effects parameters represent the variation of the selected sample around the population of motion pictures.

## 3.4. Model Selection

The model described in Equation (3.7) contains all the covariates and three random effects associated with each film: a random intercept, a random week effect and a random week-squared effect. It is therefore referred to as the full or the loaded model. This model structure allows each film to have an individual trajectory, with coefficients that vary randomly around the fixed effects defining the mean growth in revenues for each film (West et al., 2007). Yet, it is important to find the 'best' model in the sense that it is both able to explain the variation in the dependent variable and parsimonious with regards to the number of independent variables. In order to test the goodness or the fit of this model, a number of tests are conducted.

First, the random-effects structure is analysed. This is done indirectly by creating new nested models, which means that these models are special cases of the loaded model in the sense that constraints are placed on some of the parameters, and comparing them to the full model. The first nested model, presented in Equation (3.8), omits the random effect for the intercept, $b_{0i}$, while the remaining structure of the full model is kept. This means that a restriction is placed on the intercept of each film's trajectory: it is not allowed to vary around a population mean; it is the same for every film.

$$\begin{aligned} \log(Rev_{it}) = {} & \beta_0 + \beta_1 \log(MPI_{it}) + \beta_2 \log(Vol_{it}) + \beta_3 \log(Screens_{it}) \\ & + \beta_4 \log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i \\ & + \beta_8 Sequel_i + \beta_9 Star_i + \beta_{10} MPAA_i + \beta_{11} Genre_i + b_{1i} Week_{it} \\ & + b_{2i} Week_{it}^2 + \epsilon_{it} \\ & i = 1 \ldots n_i, \quad b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I) \end{aligned} \tag{3.8}$$

Further nested models are fitted omitting either parts of or the whole random effect for the slope, $b_{1i} Week_{it} + b_{2i} Week_{it}^2$, while keeping the remaining structure of the model. One example omitting the week-squared random effect is presented in Equation (3.9). All of these models are nested models of Equation (3.7), because restrictions are placed on the random-effects structure. While in the loaded model both the slope and the intercept are allowed to vary around the population mean, in Equation (3.8) only the slope is allowed to vary, and in Equation (3.9) the slope is only allowed to vary for the week, but not for the week-squared.

$$
\begin{aligned}
\log(Rev_{it}) = {} & \beta_0 + \beta_1 \log(MPI_{it}) + \beta_2 \log(Vol_{it}) + \beta_3 \log(Screens_{it}) \\
& + \beta_4 \log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i \\
& + \beta_8 Sequel_i + \beta_9 Star_i + \beta_{10} MPAA_i + \beta_{11} Genre_i + b_{0i} \\
& + b_{1i} Week_{it} + \epsilon_{it} \\
& i = 1 \dots n_i, \quad b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)
\end{aligned}
\tag{3.9}
$$

Generally, two estimation methods are used to calculate the parameters of a mixed-effects model, maximum likelihood (ML) and restricted maximum likelihood (REML). These are both mathematical methods to generate estimates of unknown parameters through the optimisation of a likelihood function (West et al., 2007). Whereas ML produces estimates that are biased because it does not account for the loss of degrees of freedom from the estimation of the fixed effects, REML takes into account the loss of degrees of freedom and thus corrects for the presence of additional parameters (Laird and Ware, 1982). Both of these maximum likelihood methods estimate the covariance structures in $D$ and $\sigma^2 I$ and subsequently the fixed-effects parameters $\beta$ are calculated.

In order to decide which random-effects structure to keep, the ML or REML estimates are used in likelihood ratio tests (LRTs), a common method of testing hypotheses when one hypothesis "involves a restriction on the values of some of the parameters" (Morrell, 1998, p. 1561). As explained above, some parameters in the nested models are constrained in comparison to the loaded model, thus making LRTs an appropriate method. The loaded model contains the random effect parameter being tested whereas the nested model does not. The LRT statistic is calculated by

$$
-2 \log \left( \frac{L_{nested}}{L_{reference}} \right) \sim \chi^2_{df} \tag{3.10}
$$

where $L_{nested}$ denotes the likelihood value of the nested model and $L_{reference}$ refers to the likelihood value of the reference or loaded model. The LRT statistic asymptotically follows a $\chi^2$ distribution and the degrees of freedom (df) are obtained by calculating the difference between the number of parameters in the nested and the full model (Morrell, 1998; West et al., 2007). If the LRT statistic is large and significant, then the reference model is preferred. In case the LRT statistic is small and insignificant, this indicates that the nested model should be preferred.

After fitting the random effects structure, the second step to find the 'best' model is by looking at the parameters of the fixed effects. It may be possible to remove insignificant variables from the model and thus improve the model fit. This is done by creating a new model without the insignificant variables and comparing it to the full model via an ML-based LRT. In the case where two models have different fixed effects, an ML-based LRT is used, because the

REML estimate accounts for the covariance structure of the fixed effects parameters. Since this covariance structure is different for two models with different fixed effects, a comparison using an REML-based LRT is not meaningful.

Similarly to the LRT for the random effects structure, a fixed effect can be omitted from the model if the LRT statistic is small and insignificant. If the test statistic is large and significant, the fixed effect should be kept. This procedure is repeated removing one insignificant fixed effect at a time, until every fixed effect in the model is significant or the LRT statistic becomes significant.

In addition to the LRT, there are two further test statistics, which are of value when selecting amongst two competing models: the Akaike Information Criterion (AIC, Akaike, 1973; Sakamoto et al., 1986) and the Bayesian Information Criterion (BIC, Schwarz, 1978). These information criteria can be used to compare two models independent of whether they are nested or not. The smaller the value of the information criterion, the better the model fit. The information criteria are respectively calculated as

$$AIC = -2\log(L) + 2n_{par} \tag{3.11}$$

$$BIC = -2\log(L) + n_{par}\log(N) \tag{3.12}$$

where $\log(L)$ denotes the logarithm of the likelihood value, $n_{par}$ the number of variables in the model and $N$ the total number of observations used to fit the model. Both information criteria assess the fit of a model based on its likelihood value, but add a penalty for the number of parameters used. This penalty is larger in the case of the BIC; it therefore puts more weight on the parsimony of a model.

After removing the insignificant independent variables, the final model's fit is assessed by a number of tests. The residuals or error terms, $\epsilon_{it}$, are tested for homoscedasticity, making sure that they are uniformly distributed over the range of the data. The residuals are further assessed over time to test whether there is a discernible pattern in the data that evolves over time. And the random effects terms, $b_i$, are assessed in terms of whether they follow or approximate a normal distribution.

All of these tests are conducted to assess whether the final model satisfies the assumptions of a linear mixed-effects model and whether there are any significant outliers not captured by the model. Finally, in order to assess whether the model is able to explain all of the observations made in the sample, the predicted values are plotted against the observed values.

## 3.5. Conclusion

This study builds upon previous empirical quantitative research on online word of mouth. Similarly to these studies, publicly available data on motion pictures and word of mouth are collected from online sources. However, a novel data set is created aggregating film-specific data on 132 motion pictures together with film ratings from an online social network that has not been used in academic research before.

A new method for the aggregation of consumer opinions is introduced, as the methods previously used in online word of mouth research – the arithmetic mean and the sentiment approach – may be inappropriate. The 'Movie Preference Index' accounts for different weighting schemes that consumers may use when interpreting product ratings. By modifying the 'orness degree', this MPI can be adjusted to put more weight on either positive or negative ratings, thus reflecting whether consumers are more influenced by good or bad ratings.

Finally, this thesis uses mixed-effects methods to analyse the effect of word of mouth on consumer decisions, a method previously employed mainly in physical, biological and medical sciences. This regression-type method is able to account for the systematic variation of the selected sample in comparison to the population with regards to different films or clusters of films. By employing mixed effects, this research is able to analyse the differing effects of word of mouth on consumer decisions for different 'kinds' of motion pictures, such as wide and narrow releases.

# 4. Results

The descriptive statistics presented in Table 3.1 already illustrate some of the features of the data set. In order to examine the characteristics more closely and to see the relationship between different variables, Table 4.1 highlights the correlation between selected variables of the model.

| | Revenue | Week | Screens | Budget | Star | Sequel | Critic | AvgRat | Volume |
|---|---|---|---|---|---|---|---|---|---|
| Revenue | 1 | | | | | | | | |
| Week | -0.4069 | 1 | | | | | | | |
| Screens | 0.6969 | -0.4166 | 1 | | | | | | |
| Budget | 0.4300 | 0.0529 | 0.5909 | 1 | | | | | |
| Star | 0.3288 | 0.0295 | 0.4177 | 0.6683 | 1 | | | | |
| Sequel | 0.3242 | 0.0249 | 0.3086 | 0.3817 | 0.1547 | 1 | | | |
| Critic | 0.1003 | 0.0104 | -0.0858 | -0.1017 | 0.0648 | 0.0952 | 1 | | |
| AvgRat | 0.1673 | -0.0561 | 0.0435 | 0.0524 | 0.1955 | 0.1101 | 0.6513 | 1 | |
| Volume | 0.8208 | -0.3526 | 0.6403 | 0.4835 | 0.3438 | 0.2419 | 0.0585 | 0.1638 | 1 |

Table 4.1.: Correlation between selected variables

As can be expected, the number of weeks a film has been on release is negatively related to weekly motion picture revenues, as most films generate the largest share of their revenues in the opening week. The number of screens, the presence of a star, a sequel, and the budget of a film are all positively related to its revenues and thus may influence consumer decisions positively.

Weekly film revenues are especially strongly correlated with the volume of word of mouth; however, it remains unclear as to whether the volume influences consumer decisions through the awareness effect (Berger et al., 2010; Liu, 2006) or whether the volume is merely a reflection of the number of people who have gone to see a particular film. With regards to evaluative measures of a film's quality, both the arithmetic mean consumer rating and the opinion of critics are weakly, but positively related to film revenues. This indicates that successful motion pictures are more highly rated by both professional film critics and consumers, although the correlations are not very strong.

The number of weeks a film has been on release is negatively related to both the number of screens a film is shown on and the volume of ratings it receives. This makes sense in the light of the fact that most motion pictures earn the largest share of their revenues in the early weeks and exhibitors are thus inclined to replace a particular film with a new release once it does not attract a large enough audience anymore. This also means that, over time, fewer people go and see a particular film who could subsequently submit a rating, thus decreasing the volume of ratings. The idea that the volume may simply be a representation of audience attendance is further supported by the positive relationship between the number of screens and the volume; the more cinemas a particular film is shown in, the more people are able to see it and subsequently post a rating.

The arithmetic mean rating is positively, but weakly, related to the volume of ratings, which may indicate that popular films are generally liked better, irrespective of whether an individual has a taste for the popular or the 'niche' product (cf. McPhee, 1963). The arithmetic mean is further strongly correlated with the opinion of professional film critics, which shows that while consumers and critics may not agree on the quality of all films, they do share a similar opinion on a large fraction of motion pictures.

Film stars are positively correlated with both the number of screens a film is shown on and the volume of word of mouth. This may hint at the fact that stars can make a film more popular and more talked about. Yet, the film budget is also positively correlated with screens and volume. Additionally, the film budget and the presence of a star are strongly positively related indicating that hiring a famous star may absorb a large fraction of the total film budget. It is therefore not clear whether the presence of a star or rather the budget has a stronger influence on the number of screens a film is shown on and the volume of word of mouth it creates.

The reason that the arithmetic mean rating is only weakly related to film revenues may be due to the way of measurement. Section 3.2 has highlighted a number of reasons why this study does not use the arithmetic mean and introduced an alternative measure of aggregate consumer opinion termed the Movie Preference Index in order to analyse the relationsip between rating valence and consumer decisions expressed via film revenues.

In order to investigate the characteristics of the MPI further, Figure 4.1 presents the correlations between the MPI measures and the weekly film revenues. The MPI measures with orness degrees closer to 1 – putting more weight on positive ratings and thus representing optimistic decision-makers – are positively correlated with film revenues, indicating that films achieving many positive ratings tend to do well at the box office.

MPI measures with orness degrees closer to 0 – putting more weight on negative ratings and thus representing pessimistic decision-makers – are, in contrast, negatively related to film revenues, indicating that films, which people do not like, also do not fare well at the box office. This correlation increases with higher orness degrees that do not put weight solely on extremely negative ratings, but also put some weight on positive ratings. The fact that these orness degrees have a stronger correlation with film revenues – and thus, consumer decisions – may be due to the fact that consumers generally do not award very negative ratings to motion pictures (see Table 3.1).

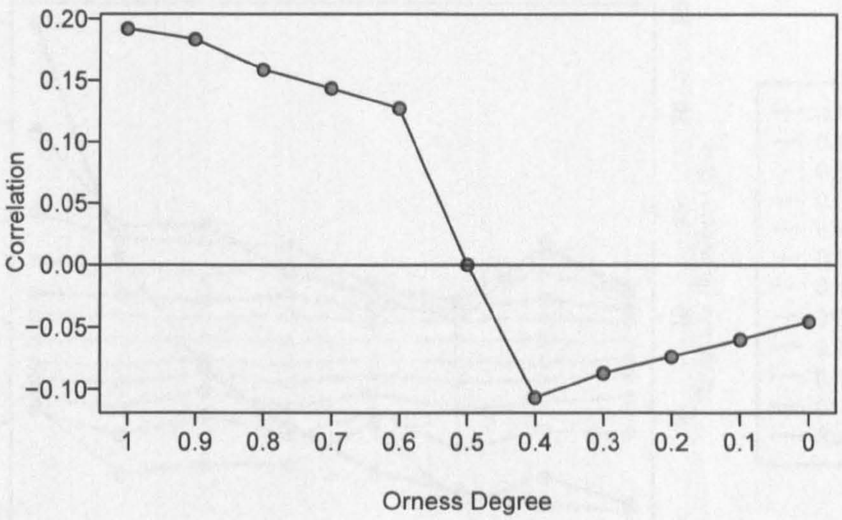



Figure 4.1.: Correlation between the MPIs and revenues

In order to look at the different MPIs more closely, and to examine the manner in which they develop over time. Figure 4.2 charts the the mean MPI of films during the course of the first eight weeks of their release for each degree of orness. This is overlaid with the mean revenue curve demonstrating the 'waterfall' effect of time on revenues; the largest share of overall revenues is generated in the opening week, and weekly revenues decline rapidly thereafter.

In contrast, most of the valence measures do not change significantly over time, indicating that consumers generally agree on the quality of films. It is only in the case of the most extreme orness degrees that a different pattern emerges - a downward sloping correction over time.

An interpretation of the curves for high orness degrees could be that audiences during the early stages of a film's release 'know' that they will enjoy the film – this could be fans of the actor, story etc. – and consequently award high ratings. However, subsequent audiences learn about the high ratings that a film achieved and raise their expectations in later weeks of a film's release, which consequently leads to a less pleasurable consumption experience and the award of comparatively lower ratings. This would be in line with the findings of Li and Hitt (2008) who find a self-selection bias among early reviewers.

At the other end of the spectrum, the decrease in the MPI for low orness degrees may similarly indicate that consumers learn from the negative word of mouth a film received and either downward-correct their expectations or stay away from it; as a consequence, films do not receive a lot of extremely negative ratings in later weeks of their release.

It is also worthwhile to notice that, on average, higher orness degrees achieve higher MPIs as aggregate measures of consumer opinions. This can be explained by the fact that consumers generally award high ratings to motion pictures (see Table 3.1 on page 63).

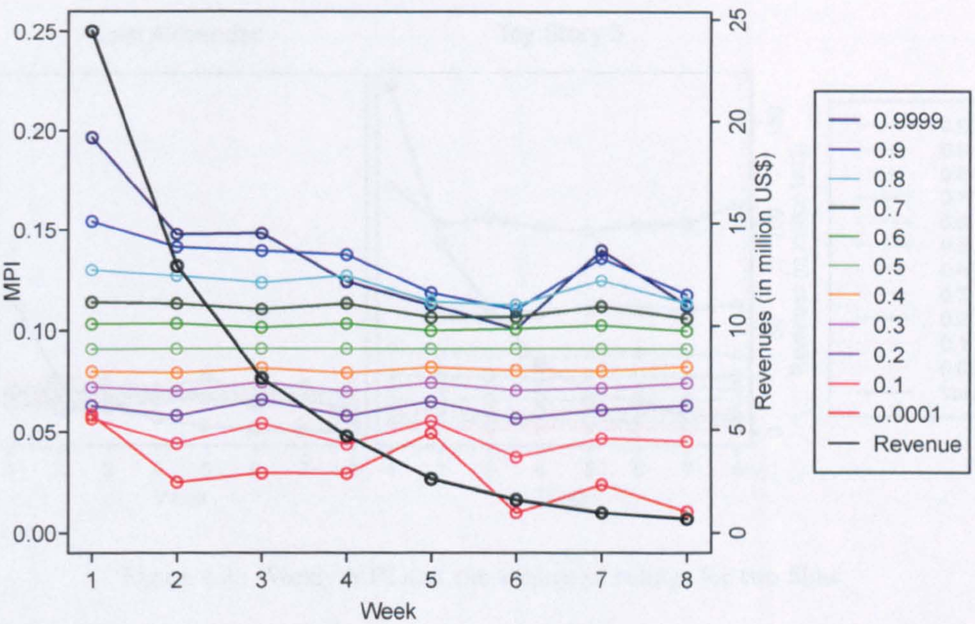



Figure 4.2.: Weekly MPIs and revenues for the 'average' film

Figure 4.2, however, provides little indication whether the valence measures have an impact on consumer decisions. In order to examine this more closely, Figure 4.3 looks at the two films that were introduced in Table 3.5, *The Last Airbender* and *Toy Story 3*, by illustrating the MPI for different orness degrees and the revenues over time.

For both films, the MPI is fairly constant over time, regardless of the chosen orness degree. This again shows that audiences largely share the same opinion on motion pictures. In the case of *Toy Story 3*, the MPIs with higher orness degrees are constantly featured at the top of the diagram signalling that this film consistently received very high ratings during the first eight weeks of its cinematic release.

In contrast, for *The Last Airbender*, MPIs with low orness degrees are featured at the top of the spectrum, though not as high as for *Toy Story 3*. Nevertheless, this shows that *The Last Airbender* mostly received bad ratings. These ratings are fairly consistent for the first three weeks, but in weeks four and five the $\mathrm{MPI}_{\alpha=1.0}$ leaps to the top indicating that the film received very positive ratings during these weeks. It may be a sign that the bad ratings during the opening deterred a lot of people from the film and mostly people who were certain that they would nevertheless enjoy the film went to see it.

Overall, the difference in enjoyment of these two films can be clearly seen. This may be one of the reasons why *Toy Story 3* generated \$410 million at the box office, whereas *The Last Airbender* 'only' generated \$130 million. For both films, however, weekly revenues decline quickly over time, indicating that fewer people go to see either film in later weeks. Yet, *Toy Story 3* generates much higher revenues throughout the first eight weeks of its release,

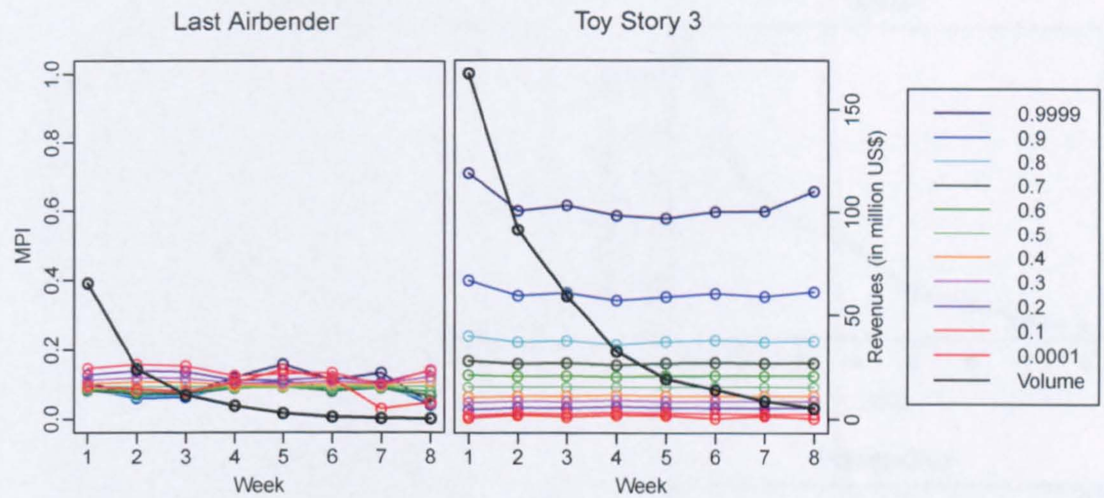



Figure 4.3.: Weekly MPI and the volume of ratings for two films

earning $167 million in its first week alone. Similarly, the volume of online word of mouth is highest in the opening week for both films, declining quickly afterwards. In contrast to the revenues, though, *The Last Airbender* and *Toy Story 3* received a similar number of online posts (see Table 3.5 on page 72).

Although both of these motion pictures display the 'typical' behaviour of motion picture revenues, it is important to notice that not all of the films in the data set develop in the same way. In fact, very different 'revenue profiles' can be identified, as Figure 4.4 shows.

The 'waterfall' behaviour of revenues is typical for blockbuster-type releases, such as *Iron-Man 2*. These films are typically released on a large number of screens (4,380 in the case of *Iron-Man 2*) and generate the largest share of their revenues in the opening week. This distribution is not limited to wide releases, though, as the example of *Babies* shows. Released on comparatively fewer screens (534), it also earns most of its revenues in the opening week, albeit on a much smaller scale than *Iron-Man 2*.

In contrast, there are films which build their revenue over time and grow. Typically, these kinds of revenue distributions are not to be found amongst the most successful films. These films are often narrow releases, testing whether they catch on with their initial target audience and subsequently increasing the number of screens they are shown on. *Winter's Bone*, for example, is shown on only four screens in its opening week. This number increases to 83 screens by week four when it generates its largest weekly revenues. A rather extreme example is the behaviour displayed by *Please Give*, which manages to increase its weekly revenues, albeit on a comparatively small scale, until week eight of its life cycle. Similar to *Winter's Bone*, this film opened on only five screens; by week eight it was shown in 272 cinemas.

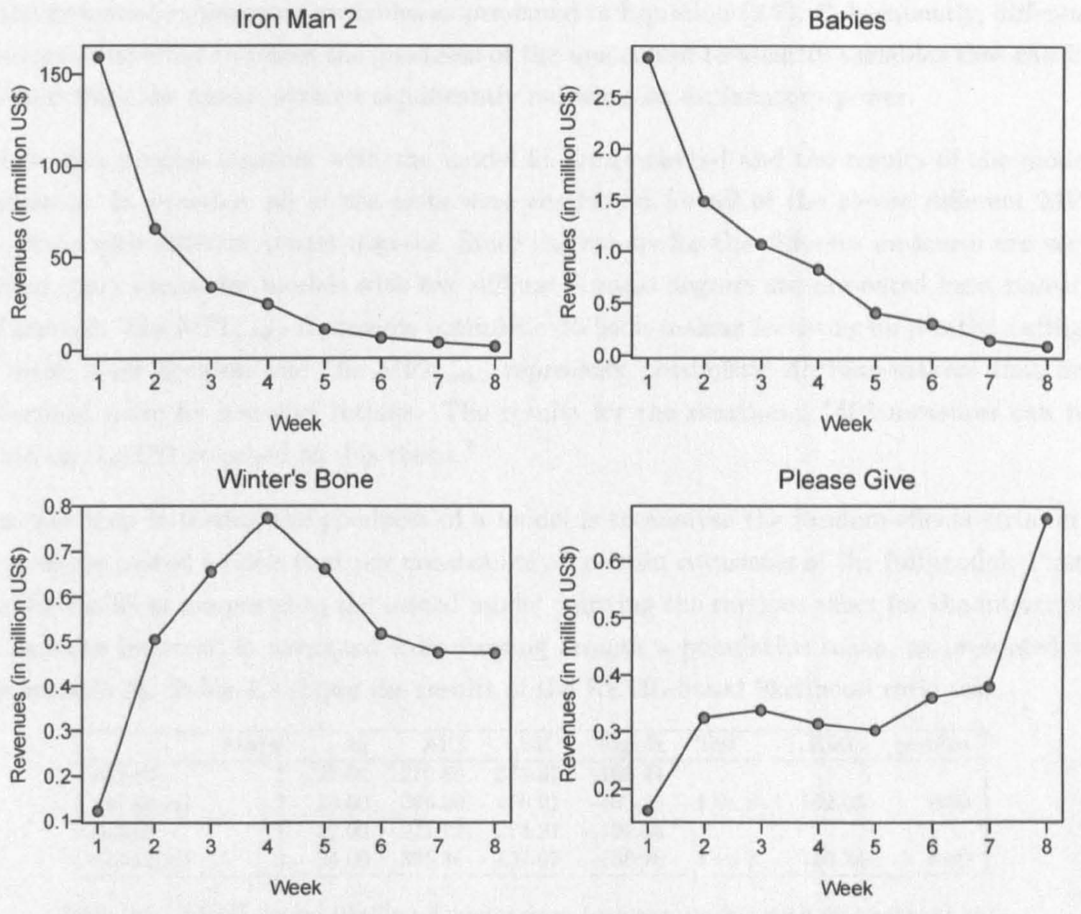



Figure 4.4.: Different 'kinds' of revenue distributions

## 4.1. Model Building and Selection

The previous section has highlighted a number of characteristics of the data set, which need to be taken into account when building a model for the assumptions of the model to correspond to what can be found in the data. As section 3.3 emphasises, mixed-effects methods are chosen for this study, because they are able to capture the random variation in the data according to different groups or clusters.

In order to account for the variation of motion picture revenues and especially the development of revenues over the life cycle of a film, random effects for both the week and the week-squared are introduced seperately for each film. This means that, initially, every film is seen as its own cluster, for which the intercept and the slope are allowed to vary. For example, in Figure 4.4, the slopes of the revenues for *Iron-Man 2* and *Babies* are similar, but the intercepts are very different. In contrast, the intercepts for *Winter's Bone* and *Please Give* are similar, but the slopes are different.

Following West et al. (2007), the model selection process starts with a full model including all of the potential explanatory variables as presented in Equation (3.7). Subsequently, different tests are conducted to assess the goodness of the model and to identify variables that can be omitted from the model without significantly reducing its explanatory power.

Below, this process together with the model fit are presented and the results of the model discussed. In practice, all of the tests were conducted for all of the eleven different MPI measures with different orness degrees. Since the results for the different measures are very similar, only results for models with two different orness degrees are presented here, namely 0.8 and 0.2. The $\text{MPI}_{\alpha=0.8}$ represents optimistic decision-makers focussing on positive ratings to make their decision and the $\text{MPI}_{\alpha=0.2}$ represents pessimistic decision-makers that are influenced more by negative ratings. The results for the remaining MPI measures can be found on the CD attached to this thesis.[7]

The first step in testing the goodness of a model is to analyse the random-effects structure by creating nested models that put constraints on certain covariates of the full model. First, the full model is compared to the nested model omitting the random effect for the intercept, so that the intercept is restricted from varying around a population mean, as presented in Equation (3.8). Table 4.2 shows the results of the REML-based likelihood ratio test.

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.all | 1 | 27.00 | 270.88 | 373.37 | -108.44 | | | |
| m8.all.nri | 2 | 24.00 | 366.90 | 458.01 | -159.45 | 1 vs 2 | 102.03 | 0.00 |
| m2.all | 1 | 27.00 | 271.72 | 374.21 | -108.86 | | | |
| m2.all.nri | 2 | 24.00 | 365.96 | 457.07 | -158.98 | 1 vs 2 | 100.24 | 0.00 |

Table 4.2.: REML-based likelihood comparison between models with and without random intercept

m8.all and m2.all are the full reference models for the $\text{MPI}_{\alpha=0.8}$ and $\text{MPI}_{\alpha=0.2}$, respectively. m8.all.nri and m2.all.nri are the corresponding models without a random effect for the intercept.[8] df stands for the degrees of freedom of the model. AIC and BIC represent the two information criteria, the Akaike Information Criterion and the Bayesian Information Criterion, respectively, and logLik stands for the logarithm of the likelihood value of the model. The most critical measures of this test are represented by L.Ratio and p-value, which denote the likelihood ratio test statistic and the significance of this test.

The fact that the LRT statistic is both large and significant suggests keeping the random effect for the intercept in the model. The smaller values for the AIC, the BIC and the likelihood value of the reference model further confirm this result.

In a second step, the two random effects for the slope are tested by fitting a nested model omitting both random effects (all.s) and another nested model omitting only the random

[7] The CD further contains the data sets and the R scripts used for statistical analysis.

[8] The equation for the .all models is $\log(Rev_{it}) = \beta_0 + \beta_1\log(MPI_{it}) + \beta_2\log(Vol_{it}) + \beta_3\log(Screens_{it}) + \beta_4\log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i + \beta_8 Sequel_i + \beta_9 Star_i + \beta_{10} MPAA_i + \beta_{11} Genre_i + b_{0i} + b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it}$ and the equation for the .all.nri models is $\log(Rev_{it}) = \beta_0 + \beta_1\log(MPI_{it}) + \beta_2\log(Vol_{it}) + \beta_3\log(Screens_{it}) + \beta_4\log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i + \beta_8 Sequel_i + \beta_9 Star_i + \beta_{10} MPAA_i + \beta_{11} Genre_i + b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it}$.

effect for the week-squared (all.nrw) while keeping the same fixed effects and the film-specific random intercept.[9] These models are again compared to the loaded model using an REML-based LRT. The results are presented in Table 4.3.

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.all | 1 | 27.00 | 270.88 | 373.37 | -108.44 | | | |
| m8.all.s | 2 | 22.00 | 371.30 | 454.81 | -163.65 | 1 vs 2 | 110.42 | 0.00 |
| m2.all | 1 | 27.00 | 271.72 | 374.21 | -108.86 | | | |
| m2.all.s | 2 | 22.00 | 374.27 | 457.79 | -165.14 | 1 vs 2 | 112.55 | 0.00 |
| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
| m8.all | 1 | 27.00 | 270.88 | 373.37 | -108.44 | | | |
| m8.all.nrw | 2 | 24.00 | 277.75 | 368.85 | -114.87 | 1 vs 2 | 12.87 | 0.00 |
| m2.all | 1 | 27.00 | 271.72 | 374.21 | -108.86 | | | |
| m2.all.nrw | 2 | 24.00 | 278.46 | 369.57 | -115.23 | 1 vs 2 | 12.74 | 0.01 |

Table 4.3.: REML-based likelihood comparison between models with and without random effects for the slope

The upper half of Table 4.3 indicates that the random-effects structure of the full model should be clearly preferred to the structure of the nested model without any random effects for the slope. The high and significant LRT statistics demonstrate this as well as the two information criteria, which are smaller for the full model.

The comparison between the loaded model and the nested model omitting the random effect of week-squared is less straightforward. The AIC and the likelihood value both prefer the loaded model, whereas the BIC prefers the nested model. The LRT statistic is comparatively small, but it is significant. Based on the significant p-value of this test, it is decided to keep the random effect of week-squared. It is thus also decided to retain the random effect of week, so that the model is well formulated in a hierarchical sense (Morrell et al., 1997).

After determining the random-effects structure of the model, in a second step to test and improve the model fit, fixed effects that are not significant in the full model can be removed. Table 4.4 and Table 4.5 highlight the results of the fixed effects of the full model. The headline shows the fixed effects structure of the model and below, Value denotes the coefficient of the independent variable while Std.Error reports its standard error. DF again stands for the degrees of freedom. The t-value is a common statistic dividing the coefficient value by its standard error and the p-value reports whether a variable is significant.

The results show that log(*Vol*), log(*Screens*), log(*Budget*), $Week^2$, and *Critic* all have a positive and significant effect on consumer decisions at the 95% confidence interval, meaning that the p-value is smaller than 0.05, whereas *Week* has a negative and significant impact on consumers. log(*MPI*), *Sequel*, and *Star* are insignificant variables, because their p-value is greater than 0.05.

[9] The equation for the .all.s models is $\log(Rev_{it}) = \beta_0 + \beta_1\log(MPI_{it}) + \beta_2\log(Vol_{it}) + \beta_3\log(Screens_{it}) + \beta_4\log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i + \beta_8 Sequel_i + \beta_9 Star_i + \beta_{10} MPAA_i + \beta_{11} Genre_i + b_{0i} + \epsilon_{it}$ and the equation for the .all.nrw models is $\log(Rev_{it}) = \beta_0 + \beta_1\log(MPI_{it}) + \beta_2\log(Vol_{it}) + \beta_3\log(Screens_{it}) + \beta_4\log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i + \beta_8 Sequel_i + \beta_9 Star_i + \beta_{10} MPAA_i + \beta_{11} Genre_i + b_{0i} + b_{1i} Week_{it} + \epsilon_{it}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + Sequel + Star + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 5.161533 | 1.7711831 | 295 | 2.914172 | 0.0038 |
| log(MPI) | -0.044863 | 0.0544635 | 295 | -0.823732 | 0.4108 |
| log(Vol) | 0.227032 | 0.0295092 | 295 | 7.693621 | 0.0000 |
| log(Screens) | 0.743964 | 0.0250028 | 295 | 29.755246 | 0.0000 |
| log(Budget) | 0.202889 | 0.0957200 | 34 | 2.119610 | 0.0414 |
| Week | -0.564695 | 0.0378962 | 295 | -14.901073 | 0.0000 |
| Week2 | 0.033477 | 0.0040766 | 295 | 8.211980 | 0.0000 |
| Critic | 0.023564 | 0.0055217 | 34 | 4.267452 | 0.0001 |
| Sequel | 0.042531 | 0.2602283 | 34 | 0.163436 | 0.8711 |
| Star | 0.215213 | 0.2474320 | 34 | 0.869786 | 0.3905 |
| MPAA NR | -0.927444 | 0.5398033 | 34 | -1.718115 | 0.0949 |
| MPAA PG-13 | 0.083116 | 0.4236181 | 34 | 0.196206 | 0.8456 |
| MPAA PG | 0.412519 | 0.4086299 | 34 | 1.009517 | 0.3199 |
| MPAA R | 0.034349 | 0.4028129 | 34 | 0.085273 | 0.9325 |
| GenreAdventure | 0.275125 | 0.3892276 | 34 | 0.706849 | 0.4845 |
| GenreAnimation | 0.055549 | 0.4260392 | 34 | 0.130384 | 0.8970 |
| GenreComedy | 0.603185 | 0.2426311 | 34 | 2.486017 | 0.0180 |
| GenreCrime | 0.811594 | 0.4418945 | 34 | 1.836623 | 0.0750 |
| GenreDrama | 0.437788 | 0.2637549 | 34 | 1.659828 | 0.1061 |
| GenreHorror | -0.271639 | 0.3328482 | 34 | -0.816104 | 0.4201 |

Table 4.4.: Fixed-Effects Results of the 'Loaded' Model (Equation (3.7)) using $MPI_{\alpha=0.8}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + Sequel + Star + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 5.270428 | 1.7879025 | 295 | 2.947828 | 0.0035 |
| log(MPI) | -0.010972 | 0.0484004 | 295 | -0.226701 | 0.8208 |
| log(Vol) | 0.224206 | 0.0294354 | 295 | 7.616873 | 0.0000 |
| log(Screens) | 0.743137 | 0.0249983 | 295 | 29.727447 | 0.0000 |
| log(Budget) | 0.203057 | 0.0964192 | 34 | 2.105982 | 0.0427 |
| Week | -0.563886 | 0.0378452 | 295 | -14.899801 | 0.0000 |
| Week2 | 0.033364 | 0.0040799 | 295 | 8.177645 | 0.0000 |
| Critic | 0.022913 | 0.0055875 | 34 | 4.100792 | 0.0002 |
| Sequel | 0.057055 | 0.2620406 | 34 | 0.217733 | 0.8289 |
| Star | 0.207867 | 0.2491165 | 34 | 0.834418 | 0.4099 |
| MPAA NR | -0.917354 | 0.5448443 | 34 | -1.683699 | 0.1014 |
| MPAA PG-13 | 0.089671 | 0.4269334 | 34 | 0.210036 | 0.8349 |
| MPAA PG | 0.438535 | 0.4110867 | 34 | 1.066770 | 0.2936 |
| MPAA R | 0.044990 | 0.4061051 | 34 | 0.110785 | 0.9124 |
| GenreAdventure | 0.259243 | 0.3918162 | 34 | 0.661644 | 0.5127 |
| GenreAnimation | 0.031275 | 0.4291684 | 34 | 0.072873 | 0.9423 |
| GenreComedy | 0.597907 | 0.2446738 | 34 | 2.443689 | 0.0199 |
| GenreCrime | 0.770451 | 0.4450626 | 34 | 1.731108 | 0.0925 |
| GenreDrama | 0.409772 | 0.2653770 | 34 | 1.544115 | 0.1318 |
| GenreHorror | -0.259885 | 0.3352919 | 34 | -0.775100 | 0.4436 |

Table 4.5.: Fixed-Effects Results of the 'Loaded' Model (Equation (3.7)) using $MPI_{\alpha=0.2}$

The values of the different levels of the dummy variables, *MPAA* and *Genre*, represent comparisons to the respective reference levels, *G* and *Action*. The coefficients in Table 4.4 and Table 4.5 are therefore estimates how the intercepts of the different MPAA ratings and genres vary from the reference level's intercept keeping all other parameters constant, and the p-values signal whether this difference is significant. For example, the population intercept for action films (the reference level for *Genre*) using $\text{MPI}_{\alpha=0.2}$ is 5.27, whereas the intercept for adventure films, keeping all other things constant, lies approximately 0.26 higher at 5.53. However, this difference is not significant, since the p-value of 0.51 is much greater than the required level of significance of 0.05. This signifies that an adventure film does not influence consumer decisions significantly different from an action film. It does not mean that the adventure genre does not have an impact on consumer decisions at all.

Throughout all of the models employing different orness degrees for the MPI measure, *Sequel* and *Star* are insignificant variables. It is therefore decided to test whether they can be removed from the model by comparing the reduced model with the loaded model using an ML-based LRT. In a first step, *Sequel* is removed, because it is highly insignificant throughout the different models. Equation (4.1) presents this reduced model.

$$
\begin{aligned}
\log(Rev_{it}) &= \beta_0 + \beta_1 \log(MPI_{it}) + \beta_2 \log(Vol_{it}) + \beta_3 \log(Screens_{it}) \\
&+ \beta_4 \log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i + \beta_8 Star_i \\
&+ \beta_9 MPAA_i + \beta_{10} Genre_i + b_{0i} + b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it} \\
i &= 1 \dots n_i, \quad b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)
\end{aligned}
\tag{4.1}
$$

Table 4.6 shows the results of the comparison, where red1 indicates the model omitting the variable *Sequel*. ml indicates that the models have been fitted using ML estimates.

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.all.ml | 1 | 27.00 | 209.52 | 313.61 | -77.76 | | | |
| m8.red1.ml | 2 | 26.00 | 207.53 | 307.76 | -77.76 | 1 vs 2 | 0.00 | 0.98 |
| m2.all.ml | 1 | 27.00 | 210.26 | 314.35 | -78.13 | | | |
| m2.red1.ml | 2 | 26.00 | 208.26 | 308.50 | -78.13 | 1 vs 2 | 0.00 | 0.95 |

Table 4.6.: ML-based likelihood comparison between models with and without fixed effects for Sequel

The LRT statistic is very small and insignificant. Additionally, both the AIC and the BIC produce lower values for the reduced model thus indicating that removing the independent variable *Sequel* produces a 'better' model in terms of parsimony while not significantly affecting its explanatory power.

The coefficients of the fixed effects for the reduced model presented in Equation (4.1) are very similar to the coefficients of the full model shown in Table 4.4 and Table 4.5. It is therefore decided to remove the variable *Star* from the model as well, since it is not significant. Equation (4.2) presents this model.

$$\begin{aligned}\log(Rev_{it}) &= \beta_0 + \beta_1 \log(MPI_{it}) + \beta_2 \log(Vol_{it}) + \beta_3 \log(Screens_{it}) \\ &\quad + \beta_4 \log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i \\ &\quad + \beta_8 MPAA_i + \beta_9 Genre_i + b_{0i} + b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it} \\ i &= 1 \ldots n_i, \quad b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)\end{aligned} \tag{4.2}$$

This further reduced model is compared to the previous model using an ML-based LRT. Table 4.7 presents the result of the model comparison, where red4 indicates the model omitting both *Sequel* and *Star*.

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.red1.ml | 1 | 26.00 | 207.53 | 307.76 | -77.76 | | | |
| m8.red4.ml | 2 | 25.00 | 207.03 | 303.41 | -78.52 | 1 vs 2 | 1.51 | 0.22 |
| m2.red1.ml | 1 | 26.00 | 208.26 | 308.50 | -78.13 | | | |
| m2.red4.ml | 2 | 25.00 | 207.71 | 304.09 | -78.86 | 1 vs 2 | 1.45 | 0.23 |

Table 4.7.: ML-based likelihood comparison between models with and without fixed effects for Sequel and Star

The LRT statistic is again very small and insignificant. Further, both the AIC and the BIC produce smaller values for the model dropping *Sequel* and *Star* from the loaded model. The further reduced model, red4, thus represents a better fitted model. Table 4.8 and Table 4.9 present the fixed-effects results for this model.

The fixed-effects coefficients of this model do not majorly differ from the coefficients of the loaded model presented in Table 4.4 and Table 4.5. Most of the variables are significant at the 95% confidence interval; only the MPI remains insignificant.

Further, irrespective of the orness degree for the MPI measure used in the model omitting both *Sequel* and *Star* (red4), none of the levels of the dummy variable *MPAA* is significantly different from the reference level *G*. Only films that are not rated (*NR*) reach a level of marginal significance at the 90% confidence interval. Further, most of the different genres are not significantly different from the reference level *Action* either. Only *Comedy* and *Crime* are significantly different.

This may indicate that both dummy variables, *MPAA* and *Genre*, can be removed from the model without affecting its explanatory power. In order to test this hypothesis, a reduced model omitting both variables is created (red7) and compared to the model omitting *Sequel* and *Star* (red4). This model is presented in Equation (4.3).

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + Sequel + Star + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 4.728972 | 1.6080321 | 295 | 2.940844 | 0.0035 |
| log(MPI) | -0.043977 | 0.0543833 | 295 | -0.808653 | 0.4194 |
| log(Vol) | 0.227352 | 0.0294112 | 295 | 7.730117 | 0.0000 |
| log(Screens) | 0.744846 | 0.0249615 | 295 | 29.839786 | 0.0000 |
| log(Budget) | 0.230642 | 0.0843648 | 36 | 2.733861 | 0.0096 |
| Week | -0.564654 | 0.0378733 | 295 | -14.909017 | 0.0000 |
| Week2 | 0.033499 | 0.0040743 | 295 | 8.222042 | 0.0000 |
| Critic | 0.024782 | 0.0051499 | 36 | 4.812234 | 0.0000 |
| MPAA NR | -1.024618 | 0.5226212 | 36 | -1.960536 | 0.0577 |
| MPAA PG-13 | 0.058997 | 0.4142618 | 36 | 0.142414 | 0.8875 |
| MPAA PG | 0.337966 | 0.3895614 | 36 | 0.867556 | 0.3914 |
| MPAA R | -0.047269 | 0.3858920 | 36 | -0.122494 | 0.9032 |
| GenreAdventure | 0.217356 | 0.3761803 | 36 | 0.577797 | 0.5670 |
| GenreAnimation | 0.017713 | 0.3761186 | 36 | 0.047093 | 0.9627 |
| GenreComedy | 0.590204 | 0.2328028 | 36 | 2.535210 | 0.0157 |
| GenreCrime | 0.84929 | 0.3962882 | 36 | 2.143113 | 0.0389 |
| GenreDrama | 0.421228 | 0.2417714 | 36 | 1.742255 | 0.0900 |
| GenreHorror | -0.257219 | 0.3256874 | 36 | -0.789774 | 0.4348 |

Table 4.8.: Fixed-Effects Results of Equation (4.2) omitting Sequel and Star using $\mathrm{MPI}_{\alpha=0.8}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + Sequel + Star + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 4.83309 | 1.6192233 | 295 | 2.984820 | 0.0031 |
| log(MPI) | -0.010594 | 0.0483427 | 295 | -0.219153 | 0.8267 |
| log(Vol) | 0.224644 | 0.0293411 | 295 | 7.656296 | 0.0000 |
| log(Screens) | 0.744167 | 0.0249567 | 295 | 29.818276 | 0.0000 |
| log(Budget) | 0.231254 | 0.0848158 | 36 | 2.726544 | 0.0098 |
| Week | -0.56389 | 0.0378252 | 295 | -14.907775 | 0.0000 |
| Week2 | 0.033402 | 0.0040786 | 295 | 8.189585 | 0.0000 |
| Critic | 0.024033 | 0.0052052 | 36 | 4.617045 | 0.0000 |
| MPAA NR | -1.01718 | 0.5269210 | 36 | -1.930423 | 0.0615 |
| MPAA PG-13 | 0.061523 | 0.4166552 | 36 | 0.147659 | 0.8834 |
| MPAA PG | 0.358607 | 0.3909685 | 36 | 0.917226 | 0.3651 |
| MPAA R | -0.037293 | 0.3880767 | 36 | -0.096097 | 0.9240 |
| GenreAdventure | 0.201949 | 0.3779326 | 36 | 0.534351 | 0.5964 |
| GenreAnimation | 0.004068 | 0.3781841 | 36 | 0.010757 | 0.9915 |
| GenreComedy | 0.587445 | 0.2343883 | 36 | 2.506291 | 0.0169 |
| GenreCrime | 0.822017 | 0.3983807 | 36 | 2.063396 | 0.0463 |
| GenreDrama | 0.401333 | 0.2429154 | 36 | 1.652153 | 0.1072 |
| GenreHorror | -0.244159 | 0.3275602 | 36 | -0.745386 | 0.4609 |

Table 4.9.: Fixed-Effects Results of Equation (4.2) omitting Sequel and Star using $\mathrm{MPI}_{\alpha=0.2}$

$$
\begin{aligned}
\log(Rev_{it}) &= \beta_0 + \beta_1 \log(MPI_{it}) + \beta_2 \log(Vol_{it}) + \beta_3 \log(Screens_{it}) \\
&+ \beta_4 \log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i \\
&+ b_{0i} + b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it} \\
i &= 1 \dots n_i, \quad b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)
\end{aligned}
\tag{4.3}
$$

Table 4.10 presents the results of the ML-based comparison.

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.red4.ml | 1 | 25.00 | 207.03 | 303.41 | -78.52 | | | |
| m8.red7.ml | 2 | 15.00 | 219.74 | 277.56 | -94.87 | 1 vs 2 | 32.71 | 0.00 |
| m2.red4.ml | 1 | 25.00 | 207.71 | 304.09 | -78.86 | | | |
| m2.red7.ml | 2 | 15.00 | 219.86 | 277.69 | -94.93 | 1 vs 2 | 32.15 | 0.00 |

Table 4.10.: ML-based likelihood comparison between models with and without fixed effects for Sequel, Star, MPAA Rating and Genre

The results of this comparison are not unambiguous. The AIC and the likelihood comparison statistic are in favour of the model keeping the dummy variables *MPAA* and *Genre* (red4). However, the BIC, which penalises the use of additional variables more heavily than the AIC, favours the reduced model without the dummy variables *MPAA* and *Genre* (red7).

These results together with the fixed-effects coefficients presented in Table 4.8 and Table 4.9 indicate that while the various film genres and MPAA ratings do not influence consumers in a significantly differing manner, they do provide some explanatory power for their overall decision. It is therefore decided to retain both dummy variables in the model.[10]

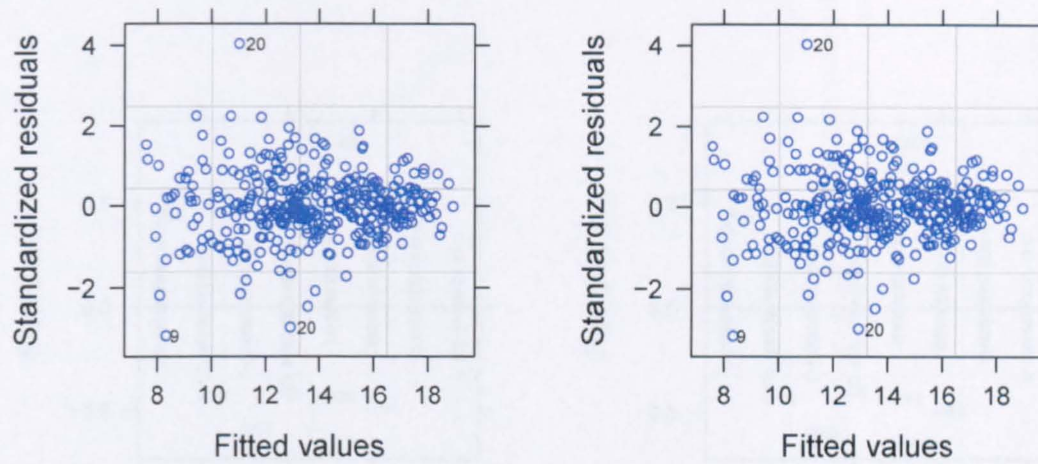
## 4.2. Model Fit Analysis

After deciding on the random-effects structure and the fixed-effects covariates to retain in the model, residual diagnostics are carried out to further assess the goodness of the model. It is tested whether the residuals or the random effects are homoscedastic and approximately normally distributed.

Figure 4.5 plots the fitted values of the model against the standardised residuals. There is no discernible pattern regarding a systematic increase or decrease in the residuals over a particular range of the data, indicating that the residuals are homoscedastic. Further, there seem to be two outliers, namely the films *Survival of the Dead* (9) and *Jonah Hex* (20).

Figure 4.6 plots the model residuals for the first eight weeks of the motion pictures' cinematic release. Again, there is no discernible pattern, no systematic increase or decrease in the

[10] It was also tested whether only removing one dummy variable would improve the overall model fit, achieving negative results (for the full analysis see Appendix B).




(a) $\mathrm{MPI}_{\alpha=0.8}$ (b) $\mathrm{MPI}_{\alpha=0.2}$

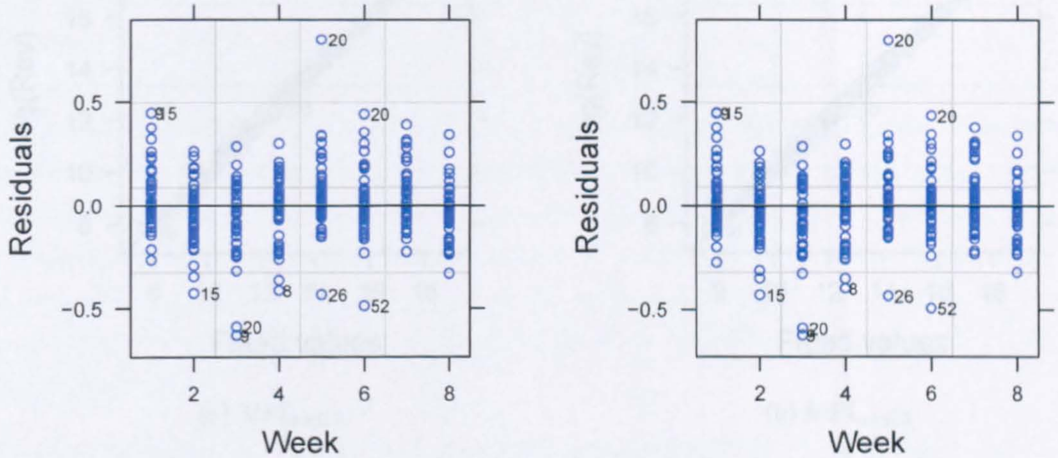
Figure 4.5.: Model residuals versus fitted values for the final model (red4)

variance of the residuals over time, indicating that the proposed model is reasonable. Again, a number of outliers are identified, amongst which both *Survival of the Dead* and *Jonah Hex* can be found.

Figure 4.7 presents a Q-Q plot of the random effects, which compares the distribution of the random effects with a hypothetical distribution, in this case the normal distribution (Thode, 2002). If the random effects were distributed exactly normal, they would follow a diagonal line from the bottom left to the top right corner. As the figure shows, the data approximately follows this distribution. Yet, the week effect has slightly longer tails to the left, because it ranges from -0.4 to 0.2, whereas the week-squared effect has heavier tails on the right ranging up to 0.04. The films *Survival of the Dead*, *The Kids Are All Right* (23), and *Kites* (26) are identified as outliers with sizeable random effects.

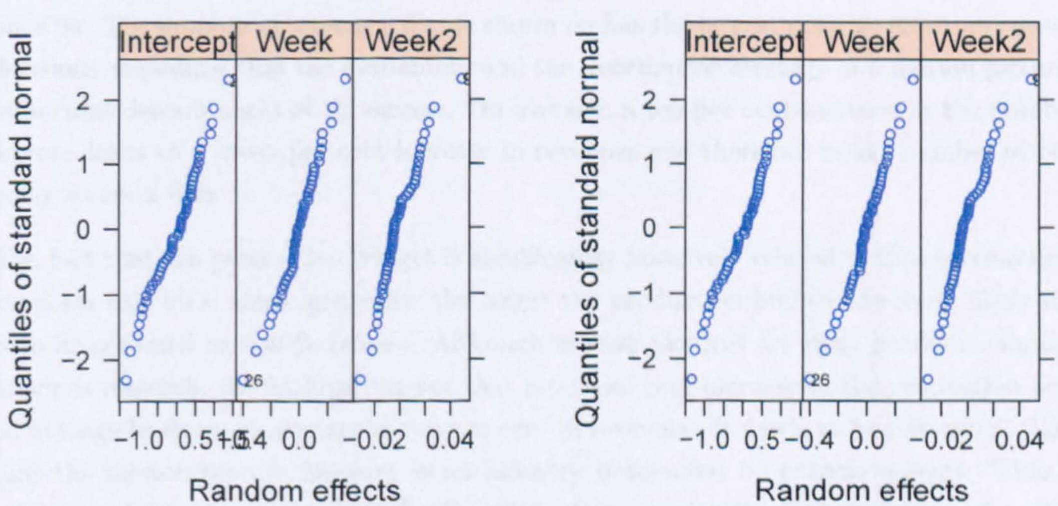
Finally, it is tested whether the model is successful in explaining the observations made in the data. To this end, Figure 4.8 plots the predicted values against the observed values. If the model is able to explain the data, then the data points should follow a 45-degree line from the bottom left to the top right corner. The figure shows that this is almost the case. Only two films, *Survival of the Dead* and *Jonah Hex*, are identified as outliers for the model.

Thus, it can be stated that the final model presented in Equation (4.2) on page 93 is a reasonable model. It provides a good model fit, is parsimonious, and is able to explain the variation in the observed data. Only a minor fraction of outliers have not been fully captured by the model. In order to test the influence of these outliers on the results, the model was re-run excluding the two identified outliers, *Survival of the Dead* and *Jonah Hex*, achieving results that are not significantly different (the full analysis and model results can be found



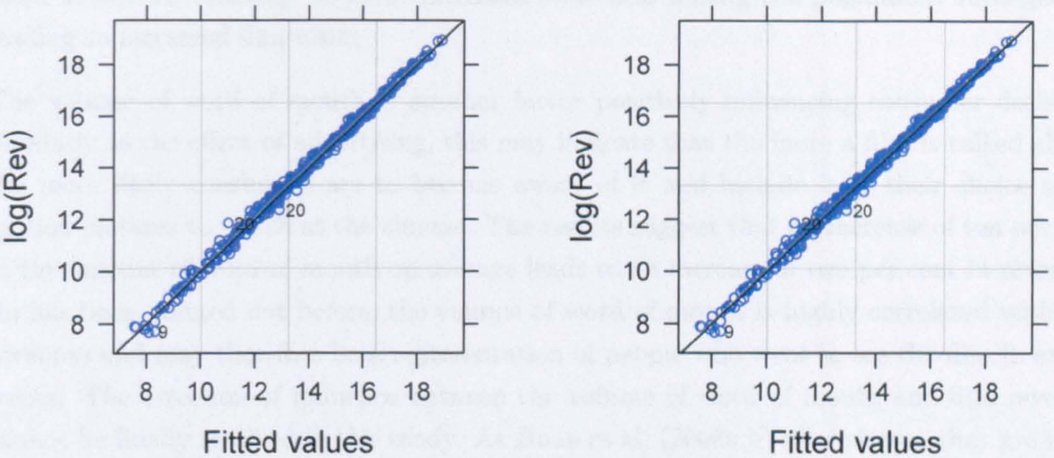

(a) $\mathrm{MPI}_{\alpha=0.8}$ (b) $\mathrm{MPI}_{\alpha=0.2}$

Figure 4.6.: Residual variance over time for the final model (red4)




(a) $\mathrm{MPI}_{\alpha=0.8}$ (b) $\mathrm{MPI}_{\alpha=0.2}$

Figure 4.7.: Q-Q Plot of the random effects for the final model (red4)




(a) $\text{MPI}_{\alpha=0.8}$ (b) $\text{MPI}_{\alpha=0.2}$

Figure 4.8.: Predicted versus observed values for the final model (red4)

in Appendix C). Therefore, it can be said that the final model is able to explain the various influences on consumer decisions at the cinema.

## 4.3. Model Results

The fixed-effects coefficients for this final model are presented in Table 4.8 and Table 4.9 on page 94. The number of screens a film is shown on has the largest positive effect on consumer decisions, signalling that the availability and the distribution strategy of a motion picture are important determinants of its success. On average, a ten per cent-increase in the number of screens leads to a seven per cent-increase in revenues and therefore in the number of people going to see a film.

The fact that the production budget is significantly positively related to film success further supports this idea, since, generally, the larger the production budget the more likely a film is to be screened as a wide release. Although motion pictures are risky products, similar to previous research, the findings suggest that a ten per cent-increase in the production budget on average leads to an increase of two per cent in revenues. It needs to be reiterated, though, that the motion picture industry is an industry dominated by extreme events. Thus, the influence of one extremely successful film with a large production budget (such as *Iron-Man 2* or *Toy Story 3*) may overshadow numerous small failures.

Yet, the production budget is able to provide a positive signal to consumers, most likely via popular film stars, appealing visual effects, or the advertising budget, which is generally estimated to be a fraction of the production budget (Elberse and Anand, 2007; Prag and Casavant, 1994). Advertising, in turn, increases awareness among the population subsequently leading to increased film visits.

The volume of word of mouth is another factor positively influencing consumer decisions. Similarly to the effect of advertising, this may indicate that the more a film is talked about, the more likely consumers are to become aware of it and include it in their choice set of motion pictures to watch at the cinema. The results suggest that an increase of ten per cent in the amount of word of mouth on average leads to an increase of two per cent in revenues. As has been pointed out before, the volume of word of mouth is highly correlated with film revenues and may therefore be a representation of people who went to see the film in earlier weeks. The direction of influence between the volume of word of mouth and film revenues cannot be finally resolved in this study. As Duan et al. (2008a,b) have shown, they are likely to be interrelated.

Finally, critical reviews also affect consumer decisions positively. The fact that the coefficient is comparatively small has to be interpreted with care. Most of the other covariates including the response variable are in logarithms, whereas critical reviews are not. In fact, a one-unit change in critical reviews (expressed via the Metascore) on average leads to a six per cent-change in revenues, indicating that consumers are either influenced by professional film critics or simply share their taste (cf. Basuroy et al., 2003; Chakravarty et al., 2010). The relatively high correlation between critical reviews and consumer ratings suggests that the latter may be the case (see Table 4.1 on page 83).

The number of weeks a film has been on release negatively affects the number of individuals going to see a film. This is likely to happen because consumers enjoy novelty and generally choose to watch a motion picture early in its life cycle. It may also be connected with the fact that the number of screens a film is shown on decreases over time and thus the availability of a film declines. The effect of week-squared is positive, providing another indicator for the fact that film revenues mostly develop in a convex manner.

The coefficient of the MPI is insignificant for each of the different orness degrees apart from one exception. This signifies that evaluative quality information about a motion picture from other consumers does not have an impact on subsequent individuals' decision-making. The only model, in which the MPI measure is significant, is the model using the $\mathrm{MPI}_{\alpha=1.0}$, the measure taking into account only the highest ratings. In this model, however, the coefficient of the MPI is negative, indicating that films receiving a large amount of very positive ratings influence consumer decisions negatively. This result does not make intuitive sense. A possible explanation is that individuals who rate motion pictures with the highest rating are seen as fanatical and therefore not trusted by the majority of consumers. A film receiving a lot of maximum ratings is treated sceptically and subsequently not attended.

The various MPAA ratings signalling the suitability of motion pictures for certain audiences regarding their age do not have a significantly differing impact from each other on consumer decisions. Therefore, in contrast to the findings of some previous research (e.g. Chang and Ki, 2005; Sochay, 1994; Terry et al., 2005), R-rated motion pictures do not deter consumers from watching them. Overall, however, MPAA ratings do seem to provide some valuable information to consumers, since removing them from the model reduced the model fit.

Similarly, the majority of the various genres employed in this study do not have a significantly differing impact from each other on consumption decisions. The results indicate that only comedies and crime films fare better than action films. It can thus be concluded that these genres are liked better by consumers.

Yet, this conclusion needs to be qualified: first, the selection of genres is, to some extent, arbitrary as previous studies have employed very different labels for genre classification and the results have been very ambiguous (e.g. Bagella and Becchetti, 1999; Chang and Ki, 2005; Litman and Kohl, 1989). Second, the sample may contain an overrepresentative number of successful comedies or crime films affecting the results; in fact, the sample of this study contains only three crime films. Thus, it is difficult to conclude on the influence that different genres have on consumer decisions. It can be stated, though, that genres provide a valuable signal to consumers, since removing the dummy variable from the model decreased the model fit.

Lastly, the random-effects coefficients show the extent to which the motion pictures in the selected sample vary around the population mean trend. Table 4.11 and Table 4.12 show the random effects results. The random effect for the intercept is approximately 0.70 across all of the models using different orness degrees. This means that, in the case of $MPI_{\alpha=0.8}$, where the fixed effect for the intercept is 4.73, for every film in the sample the intercept can vary between 4.03 and 5.43. Similarly, the random effect for the week may vary between approximately -0.38 and -0.74, and the random effect for week-squared may vary between approximately 0.01 and 0.05.

| Random Effects: Formula: ~ Week + Week2 \| Film | | | |
|---|---|---|---|
| | StdDev | Corr | |
| (Intercept) | 0.70343586 | (Intr) | Week |
| Week | 0.18325978 | -0.497 | |
| Week2 | 0.01985168 | 0.132 | -0.871 |
| Residual | 0.19935174 | | |

Table 4.11.: Random effects results for the model using $MPI_{\alpha=0.8}$

| Random Effects: Formula: ~ Week + Week2 \| Film | | | |
|---|---|---|---|
| | StdDev | Corr | |
| (Intercept) | 0.69904214 | (Intr) | Week |
| Week | 0.18249892 | -0.473 | |
| Week2 | 0.01989892 | 0.100 | -0.868 |
| Residual | 0.19921704 | | |

Table 4.12.: Random effects results for the model using $MPI_{\alpha=0.2}$

This is illustrated in a very simplified manner in Figure 4.9. It is simplified, because it ignores most of the covariates and only illustrates the effects, both fixed and random, of the week of release on a film's revenues. The middle curve shows the influence of the fixed effects and thus the trajectory of the population mean motion picture. The top curve illustrates the most positive outcome – the most extreme deviation from the population mean – that is theoretically possible according to the model. It shows that it is possible for a motion picture to increase its revenues in later weeks. On the other hand, the bottom curve illustrates the most negative outcome theoretically possible. In summary, the trajectories of the film revenues in the sample can vary between the upper and the lower curve.

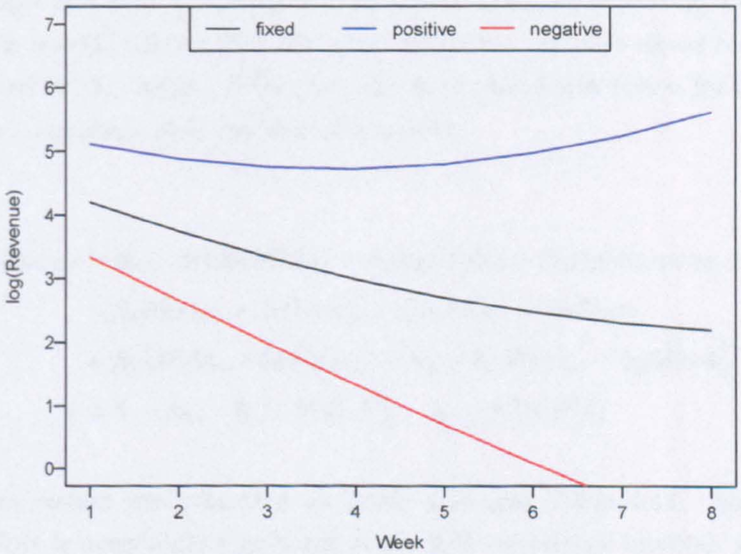



Figure 4.9.: The effect of the random effects on a film's revenue's trajectory

An upward shift in the intercept indicates that a film did better than could have been expected on the opening weekend. In Moretti's (2011) terms, this film is a positive surprise to exhibitors considering the signals that the film provided to consumers, as they did not expect it to fare so well. This may happen when consumers 'know' prior to release that this particular film will provide them with a pleasurable experience. In contrast, a downward shift in the intercept means that a film did worse than could have been expected on the opening weekend. This may be the case because the signals provided to consumers prior to the release of the film were not strong enough to convince them to see the film at the cinema or, alternatively, suggested that they would not enjoy it.

Similarly, a steeper downward slope indicates that the film's weekly revenues decreased faster than the weekly revenues of the population mean film. For example, this is the case for *Iron-Man 2* (however, its intercept was higher that the population mean). In contrast, a flatter downward or an upward slope indicate a film that becomes more popular over time, something that is observed e.g. for the films in the bottom of Figure 4.4.

## 4.4. Further Investigations

Looking at the results achieved this far, two issues call for further investigation. The fixed-effects coefficients of the final model (Table 4.8 and Table 4.9) do not largely differ from the coefficients of the loaded model presented in Table 4.4 and Table 4.5. Two differences are notable though. First, the intercept decreased by approximately 0.44. Second, the coefficient for Budget increased by approximately 0.03 and became more significant, improving from significant at the 95% confidence interval to significant at the 99% confidence interval.

This change occurs only after the removal of the variable *Star* from the model. Since Table 4.1 shows that *Budget* and *Star* are positively correlated, these variables may be substituted for each other. This would indicate that film stars provide a valuable signal to consumers that is strongly related to the budget. Therefore, the final model was tested for a substitution of *Budget* by *Star*. Equation (4.4) presents this model.

$$\begin{aligned}\log(Rev_{it}) &= \beta_0 + \beta_1\log(MPI_{it}) + \beta_2\log(Vol_{it}) + \beta_3\log(Screens_{it}) \\ &+ \beta_4 Week_{it} + \beta_5 Week_{it}^2 + \beta_6 Critic_i + \beta_7 Star_i \\ &+ \beta_8 MPAA_i + \beta_9 Genre_i + b_{0i} + b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it} \\ i &= 1 \dots n_i, \quad b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)\end{aligned} \tag{4.4}$$

The fixed-effects results are presented in Table 4.13 and Table 4.14. Most notably, the coefficient for *Star* is marginally significant at the 90% confidence interval. According to the results, the presence of a film star can be expected to increase the revenues after eight weeks by 47 per cent in comparison to a film without a star. These results prove that *Star* can be used as a substitute for *Budget*, yet it does not provide the model with the same explanatory power.

This is further confirmed by the ML-based LRT in Table 4.15. It shows that the model using the variable *Budget* (red4) is preferred to the model using *Star* (red4.nbud) as an independent variable. The AIC, the BIC, and the likelihood value are smaller for the former model. Therefore, in terms of a model providing a good fit, *Star* should only be used in cases when the production budget is not available as an explanatory variable.

These findings suggest that the possibility that a star has some impact on consumer decision-making cannot be ultimately ruled out. Yet, since the presence of a star is often well-explained by an increase in the production budget, using the budget as an aggregate signal for stars (and advertising expenditures, visual effects etc.) is appropriate for a parsimonious model.

Another issue for investigation emerges from the insignificance of the valence measures in the models. A problem of the model may be that, although allowing the sales trajectories to vary

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + Week + Week2 + Critic + Star + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 8.999511 | 0.5557032 | 295 | 16.194815 | 0.0000 |
| log(MPI) | -0.045435 | 0.0544632 | 295 | -0.834230 | 0.4048 |
| log(Vol) | 0.235498 | 0.0291594 | 295 | 8.076235 | 0.0000 |
| log(Screens) | 0.754081 | 0.0245738 | 295 | 30.686437 | 0.0000 |
| Week | -0.568449 | 0.0381957 | 295 | -14.882539 | 0.0000 |
| Week2 | 0.034703 | 0.0041694 | 295 | 8.323103 | 0.0000 |
| Critic | 0.017911 | 0.0051533 | 36 | 3.475704 | 0.0013 |
| Star | 0.385673 | 0.2259764 | 36 | 1.706698 | 0.0965 |
| MPAA NR | -1.166832 | 0.5396675 | 36 | -2.162131 | 0.0373 |
| MPAA PG-13 | 0.052006 | 0.4091125 | 36 | 0.127120 | 0.8996 |
| MPAA PG | 0.364839 | 0.3898464 | 36 | 0.935852 | 0.3556 |
| MPAA R | -0.078075 | 0.3907827 | 36 | -0.199792 | 0.8428 |
| GenreAdventure | 0.397396 | 0.3665525 | 36 | 1.084145 | 0.2855 |
| GenreAnimation | 0.216996 | 0.3662695 | 36 | 0.592448 | 0.5573 |
| GenreComedy | 0.436344 | 0.2257111 | 36 | 1.933195 | 0.0611 |
| GenreCrime | 0.532489 | 0.3730757 | 36 | 1.427294 | 0.1621 |
| GenreDrama | 0.243692 | 0.2259823 | 36 | 1.078366 | 0.2880 |
| GenreHorror | -0.511771 | 0.3086486 | 36 | -1.658103 | 0.1060 |

Table 4.13.: Fixed effects results of Equation (4.4) substituting Budget for Star using $\mathrm{MPI}_{\alpha=0.8}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + Week + Week2 + Critic + Star + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 9.130248 | 0.5497303 | 295 | 16.608595 | 0.0000 |
| log(MPI) | -0.009399 | 0.0484279 | 295 | -0.194075 | 0.8463 |
| log(Vol) | 0.232494 | 0.0291077 | 295 | 7.987354 | 0.0000 |
| log(Screens) | 0.753427 | 0.0245619 | 295 | 30.674603 | 0.0000 |
| Week | -0.567951 | 0.0381556 | 295 | -14.885143 | 0.0000 |
| Week2 | 0.034636 | 0.0041788 | 295 | 8.288354 | 0.0000 |
| Critic | 0.017114 | 0.0052026 | 36 | 3.289444 | 0.0023 |
| Star | 0.38044 | 0.2266743 | 36 | 1.678357 | 0.1019 |
| MPAA NR | -1.170229 | 0.5442685 | 36 | -2.150096 | 0.0383 |
| MPAA PG-13 | 0.052212 | 0.4109105 | 36 | 0.127064 | 0.8996 |
| MPAA PG | 0.381536 | 0.3908728 | 36 | 0.976112 | 0.3355 |
| MPAA R | -0.071382 | 0.3927905 | 36 | -0.181729 | 0.8568 |
| GenreAdventure | 0.379258 | 0.3676869 | 36 | 1.031469 | 0.3092 |
| GenreAnimation | 0.202267 | 0.3676874 | 36 | 0.550105 | 0.5856 |
| GenreComedy | 0.430181 | 0.2269050 | 36 | 1.895866 | 0.0660 |
| GenreCrime | 0.503079 | 0.3742732 | 36 | 1.344148 | 0.1873 |
| GenreDrama | 0.223885 | 0.2265362 | 36 | 0.988295 | 0.3296 |
| GenreHorror | -0.500599 | 0.3102093 | 36 | -1.613745 | 0.1153 |

Table 4.14.: Fixed effects results of Equation (4.4) substituting Budget for Star using $\mathrm{MPI}_{\alpha=0.2}$

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.red4.ml | 1 | 25.00 | 207.03 | 303.41 | -78.52 | | | |
| m8.red4.nbud.ml | 2 | 25.00 | 210.37 | 306.75 | -80.19 | | | |
| m2.red4.ml | 1 | 25.00 | 207.71 | 304.09 | -78.86 | | | |
| m2.red4.nbud.ml | 2 | 25.00 | 211.05 | 307.43 | -80.53 | | | |

Table 4.15.: ML-based likelihood comparison between models substituting Budget for Star

for every film over time, it does not distinguish the effect of qualitative word of mouth for different 'kinds' of films.

Industry wisdom states that motion pictures produced on a large budget are generally heavily marketed and open on a large number of screens to draw in crowds early in a film's life cycle. In contrast, motion pictures produced on a small budget adopt a narrow release strategy. The film is released on a smaller number of (targeted) screens, where distributors hope it will build up positive word of mouth, which in turn allows them to release it on additional cinema screens and make it more successful. Some studies have found support for this (e.g. Yang et al., 2012).

In order to test for this phenomenon, the data set is divided into wide releases, defined as films that open on at least 1,000 screens, and narrow releases, defined as films that open on less than 1,000 screens. The model is subsequently run separately for both data sets in order to compare the results. Tables 4.16 and 4.17 present the fixed-effects results for wide releases and Tables 4.18 and 4.19 present the fixed-effects results for narrow releases.

Overall, the results are not very different from the results for the whole sample. Volume of word of mouth, number of screens, week, week-squared and critical reviews are all significant variables and point in the expected direction. The size of these coefficients is similar to the coefficients in the final model as well. However, the budget is insignificant both for wide and narrow releases. This is due to the fact that the budgets of films in each separate data set are more equally distributed and do not explain the differences in outcomes anymore.

In the case of wide releases, the comedy genre fares significantly better than any other genre, whereas the MPAA ratings do not have a significantly different effect on consumer decisions. In contrast, in the case of narrow releases, the genres do not differ in their impact on consumers, whereas a PG-rating positively influences consumer decisions.

Yet, contrary to industry wisdom, the valence of word of mouth is still insignificant, even for narrow releases that contain comparatively successful 'sleepers', such as *Winter's Bone*. A possible explanation for this is that both distributors and exhibitors are very quick to identify such phenomena and immediately increase the number of screens for a narrow-release film once it shows grows potential. For *Winter's Bone*, the number of screens is continuously increased after the opening week (four screens) through to week nine of its theatrical run (141 screens) and only slowly decreased thereafter.

A comparison of the likelihood values and information criteria, shown in Table 4.20, further supports this idea. The model fit for models using only wide releases (red4_wide) and models using only narrow releases (red4_narrow) is better than for the final model; the AIC, the BIC and the likelihood value are smaller. This suggests that the number of opening screens has a major influence on consumer decisions.

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 5.84 | 2.22 | 178 | 2.64 | 0.01 |
| log(MPI) | 0.07 | 0.10 | 178 | 0.76 | 0.45 |
| log(Vol) | 0.25 | 0.05 | 178 | 5.21 | 0.00 |
| log(Screens) | 0.75 | 0.04 | 178 | 20.37 | 0.00 |
| log(Budget) | 0.22 | 0.13 | 16 | 1.68 | 0.11 |
| Week | -0.56 | 0.04 | 178 | -14.97 | 0.00 |
| Week2 | 0.03 | 0.00 | 178 | 8.96 | 0.00 |
| Critic | 0.01 | 0.01 | 16 | 2.32 | 0.03 |
| MPAA PG-13 | 0.02 | 0.37 | 16 | 0.05 | 0.96 |
| MPAA PG | 0.14 | 0.31 | 16 | 0.45 | 0.66 |
| MPAA R | -0.42 | 0.33 | 16 | -1.27 | 0.22 |
| GenreAdventure | -0.09 | 0.39 | 16 | -0.23 | 0.82 |
| GenreAnimation | 0.27 | 0.29 | 16 | 0.92 | 0.37 |
| GenreComedy | 0.65 | 0.21 | 16 | 3.13 | 0.01 |
| GenreDrama | 0.42 | 0.26 | 16 | 1.63 | 0.12 |
| GenreHorror | -0.01 | 0.31 | 16 | -0.04 | 0.96 |

Table 4.16.: Fixed-effects results of Equation (4.2) for wide releases using $MPI_{\alpha=0.8}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 5.28 | 2.20 | 178 | 2.40 | 0.02 |
| log(MPI) | -0.17 | 0.09 | 178 | -1.81 | 0.07 |
| log(Vol) | 0.26 | 0.05 | 178 | 5.74 | 0.00 |
| log(Screens) | 0.75 | 0.04 | 178 | 20.34 | 0.00 |
| log(Budget) | 0.22 | 0.13 | 16 | 1.69 | 0.11 |
| Week | -0.55 | 0.04 | 178 | -14.97 | 0.00 |
| Week2 | 0.03 | 0.00 | 178 | 8.98 | 0.00 |
| Critic | 0.01 | 0.01 | 16 | 1.86 | 0.08 |
| MPAA PG-13 | -0.02 | 0.37 | 16 | -0.04 | 0.97 |
| MPAA PG | 0.10 | 0.31 | 16 | 0.33 | 0.75 |
| MPAA R | -0.45 | 0.33 | 16 | -1.37 | 0.19 |
| GenreAdventure | 0.00 | 0.40 | 16 | 0.00 | 1.00 |
| GenreAnimation | 0.30 | 0.30 | 16 | 1.01 | 0.33 |
| GenreComedy | 0.68 | 0.21 | 16 | 3.27 | 0.00 |
| GenreDrama | 0.46 | 0.26 | 16 | 1.74 | 0.10 |
| GenreHorror | 0.04 | 0.31 | 16 | 0.12 | 0.90 |

Table 4.17.: Fixed-effects results of Equation (4.2) for wide releases using $MPI_{\alpha=0.2}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 5.28 | 2.99 | 112 | 1.77 | 0.08 |
| log(MPI) | -0.05 | 0.06 | 112 | -0.81 | 0.42 |
| log(Vol) | 0.18 | 0.04 | 112 | 4.76 | 0.00 |
| log(Screens) | 0.66 | 0.04 | 112 | 15.26 | 0.00 |
| log(Budget) | 0.04 | 0.17 | 11 | 0.26 | 0.80 |
| Week | -0.51 | 0.09 | 112 | -5.42 | 0.00 |
| Week2 | 0.03 | 0.01 | 112 | 2.99 | 0.00 |
| Critic | 0.05 | 0.01 | 11 | 3.84 | 0.00 |
| MPAA PG-13 | 0.66 | 0.81 | 11 | 0.81 | 0.44 |
| MPAA PG | 2.14 | 1.01 | 11 | 2.12 | 0.06 |
| MPAA R | 1.25 | 0.73 | 11 | 1.73 | 0.11 |
| GenreAdventure | 0.48 | 1.01 | 11 | 0.48 | 0.64 |
| GenreComedy | 0.03 | 0.90 | 11 | 0.03 | 0.98 |
| GenreCrime | 0.65 | 0.90 | 11 | 0.72 | 0.49 |
| GenreDrama | 0.12 | 0.85 | 11 | 0.14 | 0.89 |
| GenreHorror | 0.02 | 0.93 | 11 | 0.02 | 0.98 |

Table 4.18.: Fixed-effects results of Equation (4.2) for narrow releases using $MPI_{\alpha=0.8}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 5.41 | 2.99 | 112 | 1.81 | 0.07 |
| log(MPI) | 0.04 | 0.06 | 112 | 0.66 | 0.51 |
| log(Vol) | 0.17 | 0.04 | 112 | 4.54 | 0.00 |
| log(Screens) | 0.66 | 0.04 | 112 | 15.22 | 0.00 |
| log(Budget) | 0.05 | 0.16 | 11 | 0.30 | 0.77 |
| Week | -0.51 | 0.10 | 112 | -5.35 | 0.00 |
| Week2 | 0.03 | 0.01 | 112 | 2.95 | 0.00 |
| Critic | 0.05 | 0.01 | 11 | 3.86 | 0.00 |
| MPAA PG-13 | 0.70 | 0.81 | 11 | 0.86 | 0.41 |
| MPAA PG | 2.18 | 1.00 | 11 | 2.17 | 0.05 |
| MPAA R | 1.29 | 0.73 | 11 | 1.77 | 0.10 |
| GenreAdventure | 0.44 | 1.00 | 11 | 0.44 | 0.67 |
| GenreComedy | 0.01 | 0.90 | 11 | 0.01 | 0.99 |
| GenreCrime | 0.64 | 0.90 | 11 | 0.70 | 0.50 |
| GenreDrama | 0.09 | 0.85 | 11 | 0.11 | 0.91 |
| GenreHorror | 0.03 | 0.93 | 11 | 0.03 | 0.97 |

Table 4.19.: Fixed-effects results of Equation (4.2) for narrow releases using $MPI_{\alpha=0.2}$

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.red4 | 1 | 25.00 | 265.83 | 360.88 | -107.91 | | | |
| m2.red4 | 2 | 25.00 | 266.65 | 361.70 | -108.33 | | | |
| m8.red4_wide | 1 | 23.00 | 122.33 | 197.49 | -38.16 | | | |
| m2.red4_wide | 2 | 23.00 | 119.80 | 194.96 | -36.90 | | | |
| m8.red4_narrow | 1 | 23.00 | 168.79 | 233.48 | -61.40 | | | |
| m2.red4_narrow | 2 | 23.00 | 169.26 | 233.94 | -61.63 | | | |

Table 4.20.: Comparison of likelihood values between the full model and the models separating wide and narrow releases

# 5. Discussion

The results of the previous chapter have shown that it is important to distinguish between different dimensions of word of mouth, as they have different effects on consumer decisions. Past research has often claimed that word of mouth influences consumer decisions (e.g. De Vany and Lee, 2001; Moretti, 2011; Moul, 2007). Yet, it is not clear what exactly is meant by word of mouth in these studies. For example, De Vany and Walls (1996) state that "information transmission" among consumers causes the extreme dynamics of motion picture revenue distribution, but they do not elaborate on what this information entails.

In contrast to this strand of research and in line with research using online word of mouth data (e.g. Chintagunta et al., 2010; Duan et al., 2008b; Liu, 2006), this thesis clusters word of mouth into two dimensions: volume representing the amount of online posts and valence representing the aggregate opinion of consumers on a particular film. This distinction is important as it reflects differences in the manner by which consumers learn about motion pictures and form their expectations.

The results show that valence is insignificant signifying that individuals do not take other consumers' experiences into account when making their decision. Figures 4.2 and 4.3 (on page 86 and 87, respectively) show that consumers rate motion pictures very consistently over time suggesting that they largely agree on their quality. Yet, there is no discernible pattern between the qualitative evaluation of a film and its popularity.

The volume of word of mouth, on the other hand, does positively impact on consumer decisions. The volume is likely to raise consumer awareness about a film thus increasing the likelihood of people going to see it. It is intuitive that the more a product is being talked about, the more familiar people become with the brand name and the more likely they are to gather additional information about it, which potentially leads to subsequent purchase. Thus, the volume of word of mouth plays an important role in a multi-stage decision-making process similar to the one modelled by Zufryden (1996) who states that awareness leads to purchase intention which, in turn, leads to actual consumption.

In addition, the volume of word of mouth is strongly related to purchases or, in the case of motion pictures, to film revenues. This is also very intuitive, since people who go to see a film are the ones most likely to post their opinion about it. Thus, for the consumer today, volume is a representation of individuals who have previously gone to see a particular film.

This interpretation of the volume hints at a decision-making behaviour similar to the herding models described by Banerjee (1992) and Bikhchandani et al. (1992), where people choose the option that most people before them have chosen. Consumers may be directly influenced by the number of people who went to see the film in a "success drives success" manner (De Vany and Walls, 2004a). Thus, the role that word of mouth plays in shaping consumer decisions may have been overstated. Actually, they tend to simply follow the crowd.

It could be argued, though, that the volume of word of mouth implicitly contains a qualitative signal to consumers. Explicitly, it contains the information that the product is popular and has been consumed by many people. Implicitly, though, this is the case only because the film is of high quality. Consumers comparing different options are therefore likely to choose the most popular option. Such learning behaviour fits the definition of observational learning as decribed by Chen et al. (2011), where consumers partly infer a product's quality from its popularity.

## 5.1. The Role of Qualitative Word of Mouth Information

In the contemporary world with its possibilities of constant and continuous exchange of messages through various media, one would expect that consumers pass on quality signals about products they experience in order to inform other individuals about their experienced pleasure or disappointment, and that these communicated quality signals are in turn used by individuals in order to shape their decision. While the former is happening on a large scale with some motion pictures in the data set receiving a huge number of reviews, this study cannot find proof for the latter. To the contrary, it has shown that the evaluative aspect of word of mouth does not significantly impact on the decision-making of consumers.

There are a number of possible explanations as to why this may be the case. First, motion pictures' life cycles are generally very short; most films exhaust their revenue potential in the early weeks of their release and earn only small amounts in later weeks. Their revenue distribution over time describes a 'waterfall', something that is replicated both in this sample as well as in other samples of recent years (see Figure 1.1 on page 2).

With regards to consumers, this means that they decide to see a film early in its run when evaluative word of mouth is not available. Therefore, the influence of word of mouth is minimal when contrasted to a film's 'marketing power', expressed through its production budget and the number of opening screens, the two variables with the largest coefficients in the final model of this study. This finding is in sharp contrast to the findings of De Vany and Walls (1996, 2004b) who state that word of mouth is an important driver of film success and that it takes about four weeks for consumers to discover what they like.

Instead, this study suggests that consumers may rather rely more on publicly available information or use the evaluation of film critics to guide their decision. The results have shown that consumer opinions do not markedly differ from professional critics and that critical reviews positively influence consumer decsions, which is in congruence with previous research (e.g. Basuroy et al., 2003; Eliashberg and Shugan, 1997). Since critical reviews are often available prior to the release of a motion picture – in fact, many studios use short quotes from film critics in their advertising campaign, if the reviews are in the studio's favour – consumers may not 'need' further evaluative information. Thus, film consumers feel that, using their previous experience and assessing the publicly available characteristics of the film, they are able to reliably estimate whether they will enjoy a film or not.

Although they know that the pleasure (or disappointment) they will experience cannot be reliably assessed prior to consumption, and although they are likely to have experienced negative disconfirmation with motion pictures in the past, these disappointments are overridden by the expectations generated prior to film consumption. This also means that consumers are willing to take risks and, it could be argued, that this risk is an integral part of the consumption experience (cf. Sedgwick and Pokorny, 2010).

An alternative explanation is that word of mouth in fact builds up so fast and disseminates through the community so quickly that enough information is available on the first *day* of release for consumers to make a decision. This assumption is not unrealistic considering the modern possibilities that consumers have to spread their opinion; they can post their evaluation of a film to a networked community using their mobile device immediately after visiting the cinema.

This would explain the negative results of this study – if all the significant 'action' happens on the first day of release, the effect of word of mouth cannot be statistically captured anymore after the first week. This idea is supported by Hennig-Thurau et al. (2012) who use opening-*day* messages from Twitter and find a small but measurable influence on post-opening day film revenues. Unfortunately, the word of mouth data of this study is not rich enough to conduct such an analysis.

This also leads back to the findings of Moretti (2011) who argues that the fate of a motion picture is more or less decided after its opening. Films he characterises as "positive surprises" to exhibitors – films that more people went to see in the opening week than was expected – cause positive word of mouth to emerge, because the "underlying quality" of these films is superior, which subsequently leads to a slower concave decline of revenues. In contrast, "negative surprises" are films that generate negative word of mouth, leading to a faster convex decline in revenues.

From this, the question arises how consumers 'know' prior to the release of a film about its underlying quality. According to the conceptualisation of consumer decision-making in Figure 3.1 on page 60, 'knowledge' is created from the interaction between individual experience and product charactistics, since evaluative word of mouth is not available at this stage – this study has neglected word of mouth that arises prior to a film's release as largely anticipatory and thus containing no quality information (Asur and Huberman, 2010; Liu, 2006). Yet, the research on these characteristics – such as film stars, genres, advertising or critical reviews – has been shown to be largely inconclusive (see Section 2.3).

Therefore, it is worthwhile revisiting the role that pre-release word of mouth might play in shaping consumer decisions. Berger et al. (2010) state that this kind of 'buzz' creates awareness amongst the population, a first step towards a purchase decision. The larger the amount of buzz the more likely individuals are to become aware of a film and subsequently decide to see it. It could be argued that motion pictures become news events or 'must-see's'. The fact that a lot of people see the film in its opening week may partly be strategic behaviour – people like to share experiences and to be able to talk about them with their peers (Barker and Brooks, 1999).

Rui et al. (2011) assume that pre-release buzz may actually be given high credibility – there has to be something special about a film if that many people talk about it. Thus, a large amount of such information raises consumer expectations prior to a film's release and leads to a large early cinema attendance.

Another unanswered question that arises is what actually causes pre-release word of mouth. It is likely to be related to marketing efforts undertaken by the production studio – early so-called 'teasers' are often produced to create interest in a motion picture –, yet no research has been conducted on this so far.[11] In this case, a large volume of pre-release buzz would be a reflection of expectations raised by the efforts put into a film's marketing. Contrary to Rui et al. (2011), it could be argued that buzz is merely a good predictor of film success, but not a direct cause for it. Whatever effect buzz has on consumer decisions, those studies that included it in their analysis found that it is positively related to opening revenues (e.g. Hennig-Thurau et al., 2012; Liu, 2006).

The discussion above has shown that the distinction of different dimensions of word of mouth is important to draw conclusions about consumer learning. While this study has clustered word of mouth into volume and valence, a further distinction may be needed, namely a distinction between pre-release word of mouth or buzz, containing merely anticipatory information and excluding any quality information, and post-release word of mouth, containing information about the experienced quality of a motion picture.

## 5.2. Different Effects for Different 'Kinds' of Motion Pictures

It has been stated before, but it needs to be reiterated that the motion picture industry is an industry of extreme or exceptional events. While this especially and intuitively counts

[11] Dellarocas and Narayan (2006) provide some proof for the propensity of post-purchase word of mouth to be positively related to marketing efforts, while both Berger and Schwartz (2011) and Sun et al. (2006) investigate personality traits that increase the likelihood of posting online reviews.

for the typical blockbusters that are widely talked about (*Titanic* and *Avatar* may serve as vivid examples), there are also exceptional or rare events at the other end of the long tail distribution. Some films manage to grow from small releases into 'sleepers' achieving recognisable success over their lifetime.

It has been argued that word of mouth is responsible for driving such 'small' films to success (Dellarocas et al., 2007). Yet, section 4.4 has analysed the different effect of word of mouth with regards to the film's opening number of screens, but found no significant difference. This suggests that the release strategy does not make a difference.

It is conceivable that a wide-release film featuring a large marketing budget and a targeted advertising strategy is able to override other signals available in the public domain. It has been suggested in section 5.1 that a waterfall behaviour of film revenues indicates that consumers do not wait for evaluative word of mouth to emerge. This is especially likely in the case of widely released blockbusters.

However, this is unlikely to happen in the case of narrow releases, which are often produced on a comparatively small budget and thus a small marketing budget. The question arises how sleepers are able to grow into successful films over time. One would expect the influence of qualitative word of mouth to be significantly larger for this 'kind' of film; however, this research could not find any proof for this hypothesis.

A possible explanation is that sleepers are somewhat driven by word of mouth, but not less so by an intelligent distribution strategy. Released in only few targeted cinemas – a distribution strategy that is more feasible on a small budget –, a film is aimed at creating fully booked screens and thus generating large revenues per screen. This, in turn, may lead to positive word of mouth from both consumers and exhibitors who can report fully booked screens. Obviously, this strategy does not work for every film released in this way, but once such a film is detected, distributors are very quick to increase the number of screens over the following weeks. The positive word of mouth from the (limited) number of early viewers helps carrying the film on to further success.

The reason why the model has not captured this phenomenon is twofold: first, as has been mentioned at the outset of this section, sleepers are exceptional events in contrast to many narrow-release films that taper out over time and do not grow into successful films. Therefore, these events are not easy to capture in statistical models explaining common patterns. Second, the increase in the number of screens overshadows the effect of positive word of mouth, as film distributors respond quickly in these cases.

Moretti (2011) argues that a positive-surprise film characterised by above-average revenues per screen in the opening week subsequently leads to (the emergence of positive word of mouth and) a slower concave decline in revenues. Following this argument, Figure 5.1 illustrates the development of weekly revenues, number of screens, and revenues per screen over the first eight weeks of release for four different films. The top half features two wide-release films, whereas the bottom half features two narrow releases.

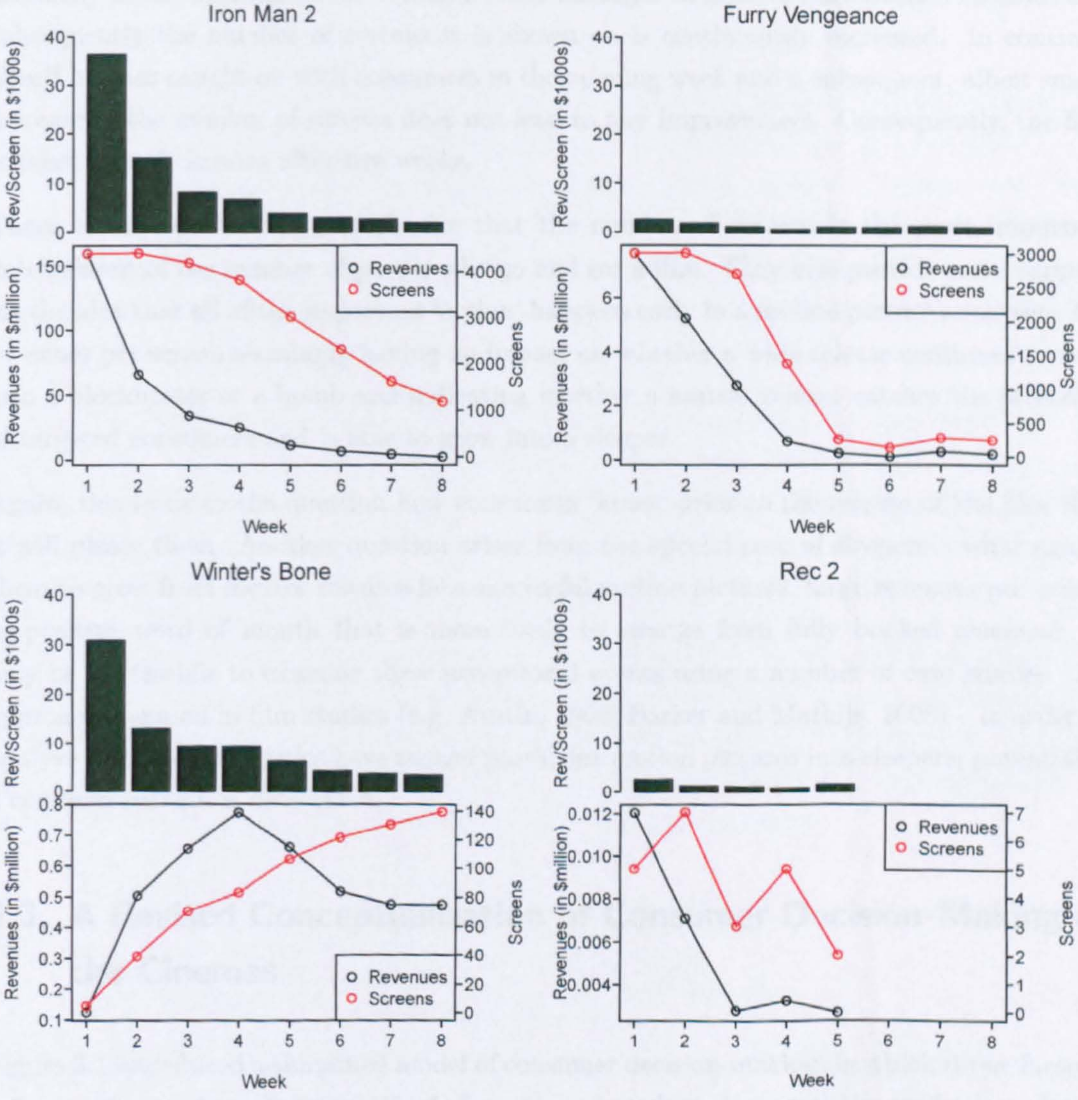



Figure 5.1.: Development of revenues, number of screens, and revenues per screen

The mean film in the data set generated $12,247 per screen in its opening week, the median lies at $9,201. As can be seen, *Iron-Man 2* not only generates large revenues but also above-average revenues per screen in the opening week, whereas *Furry Vengeance*, although released on many screens as well, generates far lower revenues per screen. Subsequently, the number of screens the film is shown on develops very differently. While it declines slowly for *Iron-Man 2*, it declines very quickly for *Furry Vengeance*. The latter film is still shown on the same number of screens in the second week – probably due to contractual agreements – but removed from the majority of cinemas by week five. At this point of its life cycle, *Iron-Man 2* is still shown on 3,007 screens.

With regards to narrow releases, a somewhat similar observation can be made. Both *Winter's Bone* and *[Rec]²* are released on a small number of screens, but the revenues per screen differ markedly in the opening week. *Winter's Bone* manages to achieve fully booked cinemas and subsequently the number of screens it is shown on is continuously increased. In contrast, *[Rec]²* has not caught on with consumers in the opening week and a subsequent, albeit small, increase in the number of screens does not lead to any improvement. Consequently, the film is taken out of cinemas after five weeks.

These examples once more emphasize that the number of screens is the most important determinant of the number of people who go and see a film. They also provide some support for the idea that all of the important 'action' happens early in a motion picture's release – the revenues per screen seemingly having an impact on whether a wide release continues to turn into a blockbuster or a bomb and indicating whether a narrow release catches the attention of targeted consumers and is able to grow into a sleeper.

Again, this leads to the question how consumers 'know' prior to the release of the film that it will please them. Another question arises from the special case of sleepers – what causes them to grow from narrow releases into successful motion pictures, large revenues per screen or positive word of mouth that is more likely to emerge from fully booked cinemas? It may be worthwhile to examine these exceptional events using a number of case studies – an approach common in film studies (e.g. Austin, 1999; Barker and Mathijs, 2008) – in order to analyse what characteristics have turned particular motion pictures into sleepers; potentially, a common pattern is detectable.

## 5.3. A Revised Conceptualisation of Consumer Decision-Making at the Cinemas

Figure 3.1 introduced a simplified model of consumer decision-making, in which three 'factors' influence the purchase decision. Word of mouth and product characteristics are both mediated by the third factor, previous experience. In the light of the findings of this study and the discussion so far, it may be worth revising this conceptualisation.

This study has defined word of mouth as an instrument through which qualitative information about past experiences can be communicated to other consumers. This means that word of mouth can only be disseminated once a product has been consumed. It is only after experiencing a film that a consumer is able to reliably evaluate its quality. Yet, the MPI – the aggregate measure of consumer opinion used in this study – does not significantly influence individuals' decisions. In contrast, the volume of word of mouth is positively related to film revenues and thus consumer decisions, but these two variables are likely to be interrelated.

A recurrent theme throughout this paper is that motion pictures generally make most of their revenues early in their life cycle, mostly during its opening. This indicates that a large proportion of decisions are made prior to the release of a motion picture and thus prior to the existence of evaluative word of mouth.

It is conceivable that pre-release buzz surrounding a film has an impact on consumer decisions. A few studies have included it in their model, and found it to be positively related to film success (e.g. Asur and Huberman, 2010; Liu, 2006). Rui et al. (2011) assume that this is partly due to the direct effect of such pre-consumption word of mouth; the people submitting anticipatory posts prior to the release of a film are likely to go and see it themselves. Yet, they do not rule out the possibility that such information is attributed high credibility and thus raises consumer expectations.

Figure 5.2 illustrates these ideas. Prior to a motion picture's release, the efforts put into its marketing create anticipatory word of mouth to arise, here labelled 'buzz', which in turn may influence consumer decisions. Consumers also gather additional information about the film and subsequently compare its features against their previous experience to form their expectations. If these are raised high enough, they go and watch the film.

Once a film is released, consumption by early viewers leads to the emergence of word of mouth. This word of mouth, in turn, influences late consumers (although a blockbuster may have exhausted a major share of its revenue potential already) in their decision, especially its volume. The valence may have an impact on consumer decisions, yet not necessarily for all 'kinds' of motion pictures, as the discussion in section 5.2 shows.

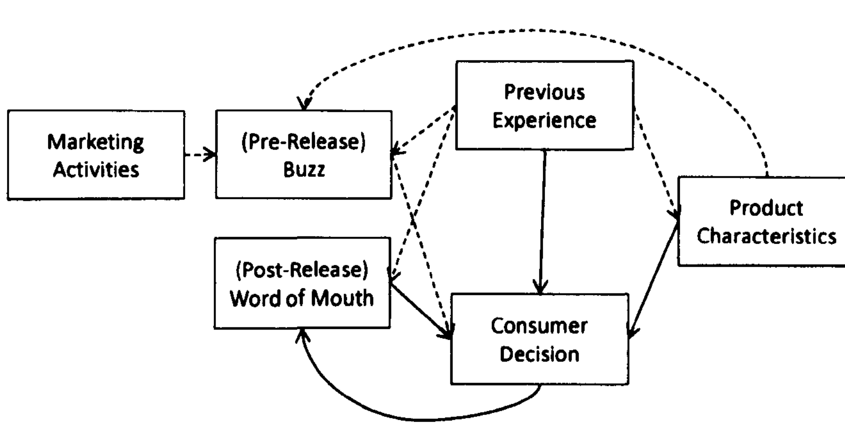



Figure 5.2.: A revised conceptualisation of the influences on consumer decision-making

Figure 5.2 also suggests some directions for further research highlighted by the dashed arrows. Especially the antecedents of pre-release buzz are worth investigating, as the volume of word of mouth seems to play a large role in determining film success. The marketing effort undertaken by film studios is a likely driver of buzz, but film characteristics may also play a role.

A valuable contribution to answer the question how previous experience mediates between product characteristics and word of mouth could be provided by a qualitative research analysing how consumers use their past experience to assess the likely quality of a particular film. This research could also shed more light on the role that film stars, genres and MPAA ratings play in decision-making, three variables for which the results of this study are not fully conclusive.

# 6. Conclusion

> "Box-office results are becoming increasingly front loaded, while the average gross in subsequent weekends is declining dramatically. Put simply, the people who want to see a movie are doing it during its opening weekend, while others are waiting until it comes onto DVD or Blu-ray." (Lang, 2010)

This study analyses the impact of online word of mouth on consumer decisions in the case of motion pictures. It investigates whether film consumers tend to engage in observational learning or word of mouth learning and whether it differs for different kinds of films. Previous research claims that word of mouth is an important driver of consumer decisions and thus the success of a product, yet how consumers learn from it is a question this previous research has not resolved.

In order to address this issue, a novel data set is created collecting online word of mouth data from a social network as a direct and empirical measure of word of mouth. Additionally, 132 motion pictures released between April and September 2010 are sampled. For each of these films, for the first eight weeks of their release, weekly data on both film-specific variables and the word of mouth they generate is collected.

Word of mouth data is clustered into volume and valence measures so as to account for the different manners by which consumers learn from word of mouth. In contrast to previous research on online word of mouth, a new approach to measure word of mouth valence is developed. The Movie Preference Index (MPI) accounts for the 'fuzziness' of rating data and counters the shortcomings of the approach to separate positive from negative sentiments and the approach of using the arithmetic mean, both of which have been used in previous research.

Explicitly, the MPI allows the analysis of how consumers weight different ratings to form their expectations of a motion picture. Through adjustment of a so-called orness degree positive weighting schemes, used by decision-makers who focus solely on whether a film received good ratings, can be distinguished from comparatively neutral and negative weighting schemes, employed by decision-makers who are more likely to shun films that received many bad ratings. The MPI thus allows to investigate the existence of a negativity effect of product ratings on consumer decisions.

The research uses mixed-effects models to analyse the data. This method introduces additional error terms (random effects) to capture the systematic variation of films or clusters of films around the population mean. A systematic model building and selection strategy is used to create a parsimonious model that explains the differing effects of film characteristics and word of mouth on consumer decisions.

The results show that it is important to cluster word of mouth into its constituent dimensions, volume and valence, as they have different effects on consumer decisions. Specifically, the volume of word of mouth has a significant and positive impact on consumer decisions. This indicates that consumers engage in observational learning meaning that once a motion picture has generated a large enough following – and thus a large volume of word of mouth – subsequent consumers will follow the crowd, since they infer that a popular film is likely to have high quality.

In contrast, the valence or the evaluative aspect of word of mouth does not directly affect consumer decisions. Most of the different weighting schemes calculated by the MPI prove to be insignificant. This means that consumers do not use quality information to shape their expectations – a conclusion that is somewhat intuitive looking at the number of people seeing a film at its opening –, yet the valence of word of mouth may have an impact on the adjustment of the number of screens certain films are shown on, as section 5.2 argues.

The implications of these results are twofold. First, for film marketers, they confirm the industry wisdom that all of the 'significant action' happens during a motion picture's opening, potentially even before a film is released onto the market. If a film manages to draw in a large enough crowd during its first weekend, its opening success will continue to drive demand over the subsequent weeks, because consumers are attracted to popular products.

There is some indication that the volume of pre-release word of mouth reflects consumer expectations held prior to the release of a film. Although further research is necessary, this type of word of mouth can be used in two ways. On the one hand, it can be utilised to forecast opening sales and to adjust the distribution strategy accordingly. On the other hand, a production studio can aim to generate a large volume of pre-release 'buzz' through various marketing efforts, which, in turn, raises consumer expectations and, subsequently, ticket sales.

Second, this study contributes to the literature on film consumers and online word of mouth research. It has shown that the distinctive features of online word of mouth affect consumer decisions differently and therefore point to different kinds of learning. In the case of motion pictures, further differentiating between pre-release buzz and post-release word of mouth is likely to lead to further improvement for model-building. Yet, a better understanding of the antecedents of pre-release buzz is necessary.

The MPI introduces a new tool to aggregate the valence of product ratings and create a consensus measure of consumer opinions. It is able to account for different weighting schemes that consumers may use when interpreting ratings and making purchase decisions. Although it has proven not to be significant as a measure of valence in the circumstance of motion pictures, it may nevertheless be a useful tool for research on products that are more likely to be affected by qualitative word of mouth. Books, for example, may prove to be such a good, as their life cycle is generally longer than that of motion pictures.

The results of this thesis are surprising in so far, as (online) word of mouth and its influence on consumer decisions have received a lot of attention over the last years, from hotel and travel websites to Amazon's recommendation algorithm. In many of these industries significant positive effects of consumer ratings have been found, e.g. for hotel bookings (Ye et al., 2011) or restaurant visits (Zhang et al., 2010).

In contrast, the motion picture industry is exceptional in that quality signals appear to be less important in influencing consumer behaviour. This may have something to do with the nature of film as a product – on the supply side, its infinite and nearly costless reproducibility encourages the major studios to invest heavily in both production values and marketing; on the demand side, the requirement of novelty is coupled with the need for accessibility, as it is comparatively rare for film consumers to repeat-consume films (at the cinema). Since the fraction of revenues from theatrical release is not as large as it used to be and consumers continuously search for novel experiences, the revenues of films have become increasingly front-loaded, driven by marketing and buzz.

# Bibliography


Ainslie, A., Drèze, X., and Zufryden, F. (2005). Modeling movie life cycles and market share. *Marketing Science*, 24(3): pp. 508–517.

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In: *International Symposium on Information Theory*, pp. 267–281. Tsahkadsor.

Akerlof, G.A. (1970). The market for "Lemons": quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, 84(3): pp. 488–500.

Albert, S. (1998). Movie stars and the distribution of financially successful films in the motion picture industry. *Journal of Cultural Economics*, 22(4): pp. 249–270.

Alexa Internet (2010). Rottentomatoes.com site info. URL: http://www.alexa.com/siteinfo/rottentomatoes.com#.

Anderson, L.R. and Holt, C.A. (1997). Information cascades in the laboratory. *The American Economic Review*, 87(5): pp. 847–862.

Arndt, J. and May, F. (1981). The hypothesis of a dominance hierarchy of information sources. *Journal of the Academy of Marketing Science*, 9(4): pp. 337–351.

Asur, S. and Huberman, B.A. (2010). Predicting the future with social media. *Arxiv preprint arXiv:1003.5699*. URL: http://arxiv.org/PS_cache/arxiv/pdf/1003/1003.5699v1.pdf.

Austin, T. (1999). 'Desperate to see it': Straight men watching Basic Instinct. In: *Identifying Hollywood's Audiences: Cultural Identity and the Movies*, pp. 147–161. British Film Institute, London.

Bagella, M. and Becchetti, L. (1999). The determinants of motion picture box office performance: Evidence from movies produced in italy. *Journal of Cultural Economics*, 23(4): pp. 237–256.

Bakker, G. (2003). Building knowledge about the consumer: The emergence of market research in the motion picture industry. *Business History*, 45(1): pp. 101–127.

Bandura, A. (1977). *Social learning theory*. Prentice Hall, Englewood Cliffs, NJ.

Banerjee, A. (1992). A simple model of herd behavior. *Quarterly Journal of Economics*, 107(3): pp. 797–817.

Barker, M. and Brooks, K. (1998). *Knowing Audiences: Judge Dredd, Its Friends, Fans and Foes.* University of Luton Press, Luton.

Barker, M. and Brooks, K. (1999). Bleak futures by proxy. In: M. Stokes and R. Maltby (eds.), *Identifying Hollywood's Audiences: Cultural Identity and the Movies*, pp. 162–174. British Film Institute, London.

Barker, M. and Mathijs, E. (eds.) (2008). *Watching the Lord of the Rings. Tolkien's World Audiences.* Peter Lang, New York.

Basuroy, S., Chatterjee, S., and Ravid, S.A. (2003). How critical are critical reviews? the box office effects of film critics, star power, and budgets. *Journal of Marketing*, 67: pp. 103–117.

Basuroy, S., Desai, K.K., and Talukdar, D. (2006). An empirical investigation of signaling in the motion picture industry. *Journal of Marketing Research*, 43(2): pp. 287–295.

Berger, J. and Schwartz, E.M. (2011). What drives immediate and ongoing word of mouth? *Journal of Marketing Research*, 48(5): pp. 869–880.

Berger, J., Sorensen, A.T., and Rasmussen, S.J. (2010). Positive effects of negative publicity: When negative reviews increase sales. *Marketing Science*, 29(5): pp. 815–827.

Bettman, J.R., Johnson, E.J., and Payne, J.W. (1991). Consumer decision making. In: T.S. Robertson and H.H. Kassarjian (eds.), *Handbook of Consumer Behavior*, pp. 50–84. Prentice-Hall, Englewood Cliffs, NJ.

Bikhchandani, S., Hirshleifer, D., and Welch, I. (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of Political Economy*, 100(5): pp. 992–1026.

Bourdieu, P. (1984). *Distinction: A Social Critique of the Judgement of Taste.* Harvard University Press, Cambridge, MA.

Bourdieu, P. (1993). *The Field of Cultural Production: Essays on Art and Literature.* Columbia University Press, New York.

Box Office Mojo (2012). Yearly box office. URL: http://boxofficemojo.com/yearly/.

Burzynski, M.H. and Bayer, D.J. (1977). The effect of positive and negative prior information on motion picture appreciation. *Journal of Social Psychology*, 101(2): pp. 215–218.

Chakravarty, A., Liu, Y., and 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): pp. 185–197.

Chang, B.H. and Ki, E.J. (2005). Devising a practical model for predicting theatrical movie success: Focusing on the experience good property. *Journal of Media Economics*, 18(4): pp. 247–269.

Chen, Y., Wang, Q., and Xie, J. (2011). Online social interactions: A natural experiment on word of mouth versus observational learning. *Journal of Marketing Research*, 48(2): pp. 238–254.

Cherry, B. (1999). Refusing to refuse to look: Female viewers of the horror film. In: M. Stokes and R. Maltby (eds.), *Identifying Hollywood's Audiences: Cultural Identity and the Movies*, pp. 187–203. British Film Institute, London.

Chintagunta, P.K., Gopinath, S., and 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): pp. 944–957.

Clement, M., Fabel, S., and Schmidt-Stolting, C. (2006). Diffusion of hedonic goods: A literature review. *JMM: The International Journal on Media Management*, 8(4): pp. 155–163.

De Vany, A.S. and Lee, C. (2001). Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues. *Journal of Economic Dynamics and Control*, 25(3-4): pp. 593–614.

De Vany, A.S. and Walls, W.D. (1996). Bose-einstein dynamics and adaptive contracting in the motion picture industry. *The Economic Journal*, 106(439): pp. 1493–1514.

De Vany, A.S. and Walls, W.D. (1997). The market for motion pictures: Rank, revenue, and survival. *Economic Inquiry*, 35(4): pp. 783–797.

De Vany, A.S. and Walls, W.D. (1999). Uncertainty in the movie industry: Does star power reduce the terror of the box office? *Journal of Cultural Economics*, 23(4): pp. 285–318.

De Vany, A.S. and Walls, W.D. (2002). Does Hollywood make too many R-rated movies? Risk, stochastic dominance, and the illusion of expectation. *The Journal of Business*, 75(3): pp. 425–451.

De Vany, A.S. and Walls, W.D. (2004a). Motion picture profit, the stable paretian hypothesis, and the curse of the superstar. *Journal of Economic Dynamics and Control*, 28(6): pp. 1035–1057.

De Vany, A.S. and Walls, W.D. (2004b). Quality evaluations and the breakdown of statistical herding in the dynamics of box-office revenue. In: A.S. De Vany (ed.), *Hollywood Economics*, pp. 48–63. Routledge, London.

Dellarocas, C. and Narayan, R. (2006). A statistical measure of a populations propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, 21(2): pp. 277–285.

Dellarocas, C., Zhang, X., and Awad, N.F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing*, 21(4): pp. 23–45.

Duan, W., Gu, B., and Whinston, A.B. (2008a). Do online reviews matter? - An empirical investigation of panel data. *Decision Support Systems*, 45(4): pp. 1007–1016.

Duan, W., Gu, B., and Whinston, A.B. (2008b). The dynamics of online word-of-mouth and product sales - an empirical investigation of the movie industry. *Journal of Retailing*, 84(2): pp. 233–242.

Elberse, A. and Anand, B. (2007). The effectiveness of pre-release advertising for motion pictures: An empirical investigation using a simulated market. *Information Economics and Policy*, 19(3-4): pp. 319–343.

Elberse, A. and Eliashberg, J. (2003). Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures. *Marketing Science*, 22(3): pp. 329–354.

Eliashberg, J., Jonker, J.j., Sawhney, M.S., and Wierenga, B. (2000). MOVIEMOD: an implementable decision-support system for prerelease market evaluation of motion pictures. *Marketing Science*, 19(3): pp. 226–243.

Eliashberg, J. and Shugan, S.M. (1997). Film critics: Influencers or predictors? *The Journal of Marketing*, 61(2): pp. 68–78.

Ellison, G. and Fudenberg, D. (1995). Word of mouth communication and social learning. *The Quarterly Journal of Economics*, February: pp. 93–125.

Engle, R. (2001). GARCH 101: The use of ARCH/GARCH models in applied econometrics. *The Journal of Economic Perspectives*, 15(4): pp. 157–168.

Faber, R.J. and O'Guinn, T.C. (1984). Effect of media advertising and other sources on movie selection. *Journalism Quarterly*, 61(2): pp. 371–377.

Fullér, R. and Majlender, P. (2003). On obtaining minimal variability OWA operator weights. *Fuzzy Sets and Systems*, 136(2): pp. 203–215.

Gemser, G., Oostrum, M., and Leenders, M.A.A.M. (2007). The impact of film reviews on the box office performance of art house versus mainstream motion pictures. *Journal of Cultural Economics*, 31(1): pp. 43–63.

Godes, D. and Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science*, 23(4): pp. 545–560.

Granovetter, M.S. (1973). The strength of weak ties. *The American Journal of Sociology*, 78(6): pp. 1360–1380.

Haavelmo, T. (1943). The statistical implications of a system of simultaneous equations. *Econometrica*, 11(1): pp. 1–12.

Hadida, A.L. (2009). Motion picture performance: A review and research agenda. *International Journal of Management Reviews*, 11(3): pp. 297–335.

Handel, L.A. (1953). Hollywood market research. *The Quarterly of Film Radio and Television*, 7(3): pp. 304–310.

Hanson, W.A. and Putler, D.S. (1996). Hits and misses: Herd behavior and online product popularity. *Marketing Letters*, 7: pp. 297–305.

Hennig-Thurau, T., Wiertz, C., and Feldhaus, F. (2012). Exploring the 'Twitter effect:' an investigation of the impact of microblogging word of mouth on consumers early adoption of new products. *SSRN eLibrary*. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2016548.

Hidalgo R, C.A., Castro, A., and Rodriguez-Sickert, C. (2006). The effect of social interactions in the primary consumption life cycle of motion pictures. *New Journal of Physics*, 8(4): pp. 1–11.

Hirschman, E.C. and Holbrook, M.B. (1982). Hedonic consumption: Emerging concepts, methods and propositions. *Journal of Marketing*, 46(3): pp. 92–101.

Holbrook, M.B. (1999). Popular appeal versus expert judgments of motion pictures. *Journal of Consumer Research*, 26(2): pp. 144–155.

Holbrook, M.B. and Hirschman, E.C. (1982). The experiential aspects of consumption: Consumer fantasies, feelings, and fun. *Journal of Consumer Research*, 9(2): pp. 132–140.

IMDb (2012). The internet movie database. URL: http://www.imdb.com/.

Jedidi, K., Krider, R., and Weinberg, C. (1998). Clustering at the movies. *Marketing Letters*, 9(4): pp. 393–405.

Joeckel, S. (2007). The impact of experience: The influences of user and online review ratings on the performance of video games in the US market. In: *DiGRA*, pp. 629–638.

Kahneman, D. (2011). *Thinking, Fast and Slow.* Penguin Books, London.

Kahneman, D. and Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, 47(2): pp. 263–291.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics*, 38(4): pp. 963–974.

Lang, B. (2010). Study: The 'Twitter effect' does not exist. URL: http://www.thewrap.com/media/column-post/thegrill-twitter-effect-myth-21035?page=0,1.

Lazarsfeld, P.F. (1947). Audience research in the movie field. *Annals of the American Academy of Political and Social Science*, 254: pp. 160–168.

Li, X. and Hitt, L.M. (2008). Self-selection and information role of online product reviews. *Information Systems Research*, 19(4): pp. 456–474.

Litman, B.R. (1983). Predicting success of theatrical movies: An empirical study. *Journal of Popular Culture*, 16(4): pp. 159–175.

Litman, B.R. and Kohl, L.S. (1989). Predicting financial success of motion pictures: The '80s experience. *Journal of Media Economics*, 2(2): pp. 35–50.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, 70(3): pp. 74–89.

Mahajan, V., Muller, E., and Kerin, R.A. (1984). Introduction strategy for new products with positive and negative word-of-mouth. *Management Science*, 30(12): pp. 1389–1404.

Maltby, R. (1999). Sticks, hicks and flaps: Classical hollywood's generic conception of its audiences. In: M. Stokes and R. Maltby (eds.), *Identifying Hollywood's Audiences: Cultural Identity and the Movies*, pp. 23–41. British Film Institute, London.

McFadden, D.L. and Train, K.E. (1996). Consumers' evaluation of new products: Learning from self and others. *Journal of Political Economy*, 104(4): pp. 683–703.

McPhee, W.N. (1963). *Formal Theories of Mass Behavior.* Free Press of Glencoe, London.

Metacritic (2012). How we create the metascore magic. URL: http://www.metacritic.com/about-metascores.

Mishra, D.P., Heide, J.B., and Cort, S.G. (1998). Information asymmetry and levels of agency relationships. *Journal of Marketing Research*, 35(3): pp. 277–295.

Moretti, E. (2011). Social learning and peer effects in consumption: Evidence from movie sales. *The Review of Economic Studies*, 78(1): pp. 356 -393.

Morrell, C.H. (1998). Likelihood ratio testing of variance components in the linear mixed-effects model using restricted maximum likelihood. *Biometrics*, 54(4): p. 1560.

Morrell, C.H., Pearson, J.D., and Brant, L.J. (1997). Linear transformations of linear mixed-effects models. *The American Statistician*, 51(4): pp. 338–343.

Moul, C.C. (2007). Measuring word of mouth's impact on theatrical movie admissions. *Journal of Economics and Management Strategy*, 16(4): pp. 859–892.

MPAA (2011). What each rating means. URL: http://mpaa.org/ratings/what-each-rating-means.

MPAA (2012). Theatrical market statistics 2011. URL: http://www.mpaa.org/Resources/5bec4ac9-a95e-443b-987b-bff6fb5455a9.pdf.

Narayan, V., Rao, V.R., and Saunders, C. (2011). How peer influence affects attribute preferences: A bayesian updating mechanism. *Marketing Science*, 30(2): pp. 368–384.

Neelamegham, R. and Chintagunta, P. (1999). A bayesian model to forecast new product performance in domestic and international markets. *Marketing Science*, 18(2): pp. 115–136.

Nelson, P. (1970). Information and consumer behavior. *Journal of Political Economy*, 78(2): pp. 311–329.

Nelson, R.A., Donihue, M.R., Waldman, D.M., and Wheaton, C. (2001). What's an Oscar worth? *Economic Inquiry*, 39(1): pp. 1–6.

O'Hagan, M. (1988). Aggregating template or rule antecedents in real-time expert systems with fuzzy set logic. In: *Proceedings of the Twenty-Second Annual IEEE Asilomar Conference on Signals, Systems and Computers*, vol. 2, pp. 681–689. Pacific Grov, CA.

Ohmer, S. (1999). The science of pleasure: George gallup and audience research in Hollywood. In: M. Stokes and R. Maltby (eds.), *Identifying Hollywood's Audiences: Cultural Identity and the Movies*, pp. 61–80. British Film Institute, London.

Pinheiro, J.C. and Bates, D.M. (2000). *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer Verlag, New York.

Pokorny, M. and Sedgwick, J. (2010). Profitability trends in Hollywood, 1929 to 1999: Somebody must know something. *The Economic History Review*, 63(1): pp. 56–84.

Prag, J. and Casavant, J. (1994). An empirical study of the determinants of revenues and marketing expenditures in the motion picture industry. *Journal of Cultural Economics*, 18(3): pp. 217–235.

Quigley Publishing (2011). Top ten money making stars. URL: http://www.quigleypublishing.com/MPalmanac/MP_listing_07.html.

Ravid, S.A. (1999). Information, blockbusters, and stars: A study of the film industry. *The Journal of Business*, 72(4): pp. 463–492.

Reinstein, D.A. and Snyder, C.M. (2005). The influence of expert reviews on consumer demand for experience goods: A case study of movie critics. *The Journal of Industrial Economics*, 53(1): pp. 27–51.

Rigby, R.A. and Stasinopoulos, D.M. (2005). Generalized additive models for location, scale and shape (with discussion). *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 54(3): pp. 507–554.

Rogers, E.M. (1983). *Diffusion of Innovations*. Free Press, New York, 3rd ed.

Rosen, S. (1981). The economics of superstars. *The American Economic Review*, 71(5): pp. 845–858.

Rui, H., Liu, Y., and Whinston, A.B. (2011). Whose and what chatter matters? the impact of tweets on movie sales. *NET Institute Working Paper No. 11-27.* URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1958068.

Sakamoto, Y., Ishiguro, M., and Kitagawa, G. (1986). *Akaike Information Criterion Statistics.* Reidel, Hingham, MA.

Sawhney, M.S. and Eliashberg, J. (1996). A parsimonious model for forecasting gross box-office revenues of motion pictures. *Marketing Science*, 15(2): pp. 113–131.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics*, 6(2): pp. 461–464.

Sedgwick, J. (2002). Product differentiation at the movies: Hollywood, 1946 to 1965. *The Journal of Economic History*, 62(03): pp. 676–705.

Sedgwick, J. (2007). A shacklean approach to the demand for movies. In: M. Bianchi (ed.), *The Evolution of Consumption: Theories and Practices*, no. 10 in Advances in Austrian Economics, pp. 77–91. Elsevier, Oxford.

Sedgwick, J. and Pokorny, M. (1998). The risk environment of film making: Warner bros in the inter-war years,. *Explorations in Economic History*, 35(2): pp. 196–220.

Sedgwick, J. and Pokorny, M. (2005). The characteristics of film as a commodity. In: *An Economic History of Film*, no. 26 in Routledge explorations in economic history, pp. 6–23. Routledge, Abingdon & New York.

Sedgwick, J. and Pokorny, M. (2010). Consumers as risk takers: Evidence from the film industry during the 1930s. *Business History*, 52(1): pp. 74–99.

Sharda, R. and Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications*, 30(2): pp. 243–254.

Sheth, J.N. and Venkatesan, M. (1968). Risk-reduction processes in repetitive consumer behavior. *Journal of Marketing Research*, 5(3): pp. 307–310.

Smith, S.P. and Smith, V.K. (1986). Successful movies: A preliminary empirical analysis. *Applied Economics*, 18(5): pp. 501–507.

Sochay, S. (1994). Predicting the performance of motion pictures. *Journal of Media Economics*, 7(4): pp. 1–20.

Sun, T., Youn, S., Wu, G., and Kuntaraporn, M. (2006). Online word-of-mouth (or mouse): An exploration of its antecedents and consequences. *Journal of Computer-Mediated Communication*, 11(4): pp. 1104–1127.

Terry, N., Butler, M., and DeArmond, D. (2005). The determinants of domestic box office performance in the motion picture industry. *Southwestern Economic Review*, 32: pp. 137–145.

The Numbers (2012). Movie box office data, film stars, idle speculation. URL: http://www.the-numbers.com/.

Thode, H.C. (2002). *Testing For Normality*. Marcel Dekker, Inc., New York, NY.

Triantaphyllou, E. (2010). *Multi-Criteria Decision Making Methods: A Comparative Study*. Springer, Dordrecht.

Variety (2012). Weekly box office. URL: http://www.variety.com/charts/film/domestic-film-box-office/weekly/.

Vogel, H.L. (2001). *Entertainment Industry Economics: A Guide for Financial Analysis*. Cambridge University Press, Cambridge, UK, 5th ed.

Voudouris, V., Gilchrist, R., Rigby, R., Sedgwick, J., and Stasinopoulos, D. (2012). Modelling skewness and kurtosis with the BCPE density in GAMLSS. *Journal of Applied Statistics*, 39(6): pp. 1279–1293.

Wang, Y.M., Luo, Y., and Hua, Z. (2007a). Aggregating preference rankings using OWA operator weights. *Information Sciences*, 177(16): pp. 3356–3363.

Wang, Y.M., Luo, Y., and Liu, X. (2007b). Two new models for determining OWA operator weights. *Computers & Industrial Engineering*, 52: pp. 203–209.

Wang, Y.M. and Parkan, C. (2005). A minimax disparity approach for obtaining OWA operator weights. *Information Sciences*, 175(1-2): pp. 20–29.

West, B., Welch, K.B., and Galecki, A.T. (2007). *Linear Mixed Models: A Practical Guide Using Statistical Software*. Taylor & Francis, Boca Raton, FL.

Wu, P.F., Heijden, H.V.D., and Korfiatis, N. (2011). The influences of negativity and review quality on the helpfulness of online reviews. In: *International Conference on Information Systems*. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1937664.

Yager, R.R. (1988). On ordered weighted averaging aggregation operators in multicriteria decisionmaking. *IEEE Transactions on Systems, Man, and Cybernetics*, 18(1): pp. 183–190.

Yang, J., Kim, W., Amblee, N., and Jeong, J. (2012). The heterogeneous effect of WOM on product sales: why the effect of WOM valence is mixed? *European Journal of Marketing*, 46(11/12): pp. 1523–1538.

Ye, Q., Law, R., Gu, B., and Chen, W. (2011). The influence of user-generated content on traveler behavior: An empirical investigation on the effects of e-word-of-mouth to hotel online bookings. *Computers in Human Behavior*, 27(2): pp. 634–639.

Zhang, Z., Ye, Q., Law, R., and Li, Y. (2010). The impact of e-word-of-mouth on the online popularity of restaurants: A comparison of consumer reviews and editor reviews. *International Journal of Hospitality Management*, 29(4): pp. 694–700.

Zhou, S.M., Chiclana, F., John, R.I., and Garibaldi, J.M. (2008). Type-1 OWA operators for aggregating uncertain information with uncertain weights induced by type-2 linguistic quantifiers. *Fuzzy Sets and Systems*, 159(24): pp. 3281–3296.

Zufryden, F.S. (1996). Linking advertising to box office performance of new film releases - a marketing planning model. *Journal of Advertising Research*, 36(4): pp. 29–41.

# A. Using the Arithmetic Mean to Calculate the Valence of Ratings

In a preliminary analysis, the effect of word of mouth valence on consumer decisions was tested using the arithmetic mean film rating posted by consumers. Equation (A.1) presents the mixed-effects model, in which weekly film revenues depend on average rating, volume of ratings, number of screens, budget, week of release, week-squared, critical reviews, and dummy variables for sequel, star, MPAA rating, and genre.

Since the model uses the logarithm of the arithmetic mean rating, a value of one was added to the arithmetic mean rating to avoid a mis-definition; the logarithm of zero is not defined.

$$
\begin{aligned}
\log(Rev_{it}) &= \beta_0 + \beta_1 \log(Avg + 1_{it}) + \beta_2 \log(Vol_{it}) + \beta_3 \log(Screens_{it}) \\
&+ \beta_4 \log(Budget_i) + \beta_5 Week_{it} + \beta_6 Week_{it}^2 + \beta_7 Critic_i \\
&+ \beta_8 Sequel_i + \beta_9 Star_i + \beta_1 0 MPAA_i + \beta_1 1 Genre_i + b_{0i} \\
&+ b_{1i} Week_{it} + b_{2i} Week_{it}^2 + \epsilon_{it} \\
i &= 1 \ldots n_i, \quad b_i \sim \mathcal{N}(0, D), \quad \epsilon_i \sim \mathcal{N}(0, \sigma^2 I)
\end{aligned}
\tag{A.1}
$$

The results are presented in Table A.1. The arithmetic mean rating is insignificant indicating that the valence of word of mouth does not influence consumer decisions. Strikingly, the coefficient is negative, which would mean that a positive rating deters consumers from seeing a film, if it were significant. This finding is counter-intuitive, as it can be assumed that a positive rating influences purchases positively, whereas a negative rating influences subsequent purchases negatively.

| Fixed effects: log(Rev) ~ log(Avg + 1) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + Sequel + Star + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 5.457582 | 1.7737327 | 295 | 3.076891 | 0.0023 |
| log(Avg + 1) | -0.046068 | 0.0459137 | 295 | -1.003370 | 0.3165 |
| log(Vol) | 0.226561 | 0.0293638 | 295 | 7.715660 | 0.0000 |
| log(Screens) | 0.743692 | 0.0249871 | 295 | 29.763001 | 0.0000 |
| log(Budget) | 0.203104 | 0.0952982 | 34 | 2.131249 | 0.0404 |
| Week | -0.564349 | 0.0379394 | 295 | -14.875024 | 0.0000 |
| Week2 | 0.033478 | 0.0040814 | 295 | 8.202570 | 0.0000 |
| Critic | 0.023393 | 0.0054941 | 34 | 4.257813 | 0.0002 |
| Sequel | 0.040001 | 0.2589804 | 34 | 0.154454 | 0.8782 |
| Star | 0.213216 | 0.2461207 | 34 | 0.866308 | 0.3924 |
| MPAA NR | -0.931888 | 0.5384843 | 34 | -1.730576 | 0.0926 |
| MPAA PG-13 | 0.084264 | 0.4216240 | 34 | 0.199855 | 0.8428 |
| MPAA PG | 0.417466 | 0.4062782 | 34 | 1.027536 | 0.3114 |
| MPAA R | 0.039228 | 0.4009782 | 34 | 0.097830 | 0.9226 |
| GenreAdventure | 0.264264 | 0.3871766 | 34 | 0.682542 | 0.4995 |
| GenreAnimation | 0.050693 | 0.4239236 | 34 | 0.119580 | 0.9055 |
| GenreComedy | 0.595983 | 0.2417079 | 34 | 2.465714 | 0.0189 |
| GenreCrime | 0.808167 | 0.4399413 | 34 | 1.836988 | 0.0750 |
| GenreDrama | 0.433513 | 0.2624511 | 34 | 1.651785 | 0.1078 |
| GenreHorror | -0.269811 | 0.3314193 | 34 | -0.814107 | 0.4212 |

Table A.1.: Fixed effects for the full model using the arithmetic mean

# B. Removing only one dummy variable from the model

This section analyses whether either of the dummy variables *MPAA* and *Genre* can be removed from the reference model presented in Equation (4.2). Since many levels of the dummy variables are insignificant in this model (see Table 4.8 and Table 4.9), it may be possible to omit them without significantly affecting the goodness of the model.

First, it is tested whether *MPAA* can be removed from the model by comparing the reduced model omitting *MPAA*(red5) to the reference model (red4). Table B.1 presents the results of the ML-based likelihood comparison. The AIC and the likelihood value prefer the model with both dummy variables, while the BIC prefers the reduced model. Due to the significance of the LRT statistic, it is decided to keep *MPAA* as a dummy variable.

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.red4.ml | 1 | 25.00 | 207.03 | 303.41 | -78.52 | | | |
| m8.red5.ml | 2 | 21.00 | 211.47 | 292.43 | -84.74 | 1 vs 2 | 12.44 | 0.01 |
| m2.red4.ml | 1 | 25.00 | 207.71 | 304.09 | -78.86 | | | |
| m2.red5.ml | 2 | 21.00 | 212.28 | 293.24 | -85.14 | 1 vs 2 | 12.57 | 0.01 |

Table B.1.: ML-based likelihood comparison between models with both dummy variables and with Genre only

In a second step, it is tested whether *Genre* can be removed from the model. The coefficients of the model show that only *Comedy* and *Crime* affect consumer decisions significantly different from the reference level *Action* (see Table 4.8 and Table 4.9).

Table B.2 presents the results of the ML-based likelihood comparison between the reference model (red4) and a reduced model omitting *Genre* (red6). The AIC and the likelihood value indicate that the reference model should be preferred, while the BIC indicates that the reduced model is the more parsimonious one. Again, due to the significance of the LRT statistic, it is decided to keep *Genre* as a dummy variable.

| | Model | df | AIC | BIC | logLik | Test | L.Ratio | p-value |
|---|---|---|---|---|---|---|---|---|
| m8.red4.ml | 1 | 25.00 | 207.03 | 303.41 | -78.52 | | | |
| m8.red6.ml | 2 | 19.00 | 214.01 | 287.26 | -88.01 | 1 vs 2 | 18.98 | 0.00 |
| m2.red4.ml | 1 | 25.00 | 207.71 | 304.09 | -78.86 | | | |
| m2.red6.ml | 2 | 19.00 | 214.12 | 287.37 | -88.06 | 1 vs 2 | 18.40 | 0.01 |

Table B.2.: ML-based likelihood comparison between models with both dummy variables and with MPAA rating only

# C. Re-Running the Model Without the Identified Outliers

The model fit analysis in Chapter 4.2 identified two main outliers, namely *Survival of the Dead* and *Jonah Hex*. Subsequently, the outliers' data was inspected in more detail. At first glance, the two films are fairly different from each other. *Survival of the Dead* is a horror film produced with a relatively small budget ($4 million), opening on only few screens (20) and thus achieving small revenues. *Jonah Hex*, in contrast, is a big-budget ($47 million) action film with a large opening (2,825 screens) and comparatively large revenues.

However, they both share a characteristic that is only sporadically observed in the film industry. They both open on the largest number of screens during their run, and this number gradually declines over the next weeks. Their weekly revenues follow this pattern. Yet, in the fifth and sixth week, respectively, the number of screens increases again. The resulting – albeit small – increase in revenues may not have been appropriately captured by the model. Figure C.1 illustrates the weekly change in both revenues and number of screens the respective films are shown on.

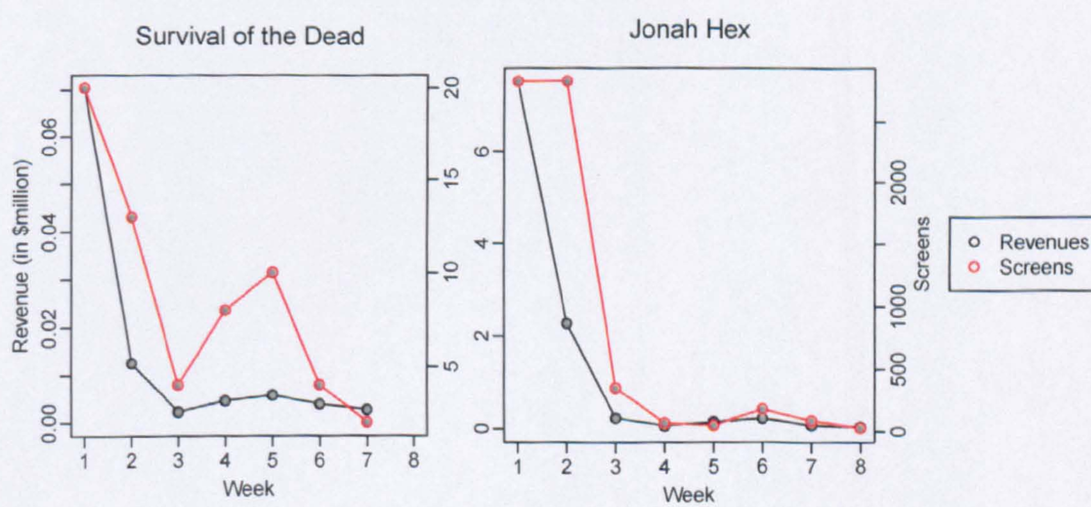



Figure C.1.: Weekly revenues and number of screens for the identified outliers

It was decided to re-run the model without these outliers and test whether these films have a significant impact on the model's results. Table C.1 and Table C.2 show the results of the final model (red4) omitting the two outliers.

It is noteworthy that the coefficient for the intercept is lower for the model omitting the outliers, whereas the coefficient for the budget is higher. The other coefficients do not change significantly. This can be explained by the fact that both of the films were rather unsuccessful at the box office considering their production budget. *Survival of the Dead* generated only $101,740 (2.5% of its production costs) and *Jonah Hex* generated only $10,545,758 (26.4% of its production costs). In contrast, the average rate of return of films in the data set for which the production budget is available is 101%.

Omitting two comparatively unsuccessful films therefore pushes the importance of the production budget upwards, whereas the intercept can decrease to accomodate better for low-budget productions with low revenues. However, since none of the other coefficients change significantly, the outliers are retained in the data set. The influences on consumer decision-making remain well-explained.

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 3.721102 | 1.8024969 | 282 | 2.064415 | 0.0399 |
| log(MPI) | -0.053297 | 0.0495767 | 282 | -1.075041 | 0.2833 |
| log(Vol) | 0.203928 | 0.0263792 | 282 | 7.730644 | 0.0000 |
| log(Screens) | 0.734168 | 0.0245736 | 282 | 29.876323 | 0.0000 |
| log(Budget) | 0.286331 | 0.0941134 | 34 | 3.042404 | 0.0045 |
| Week | -0.560322 | 0.0366813 | 282 | -15.275418 | 0.0000 |
| Week2 | 0.030805 | 0.0036574 | 282 | 8.422730 | 0.0000 |
| Critic | 0.0275 | 0.0057446 | 34 | 4.787053 | 0.0000 |
| MPAA NR | -0.964708 | 0.5599360 | 34 | -1.722889 | 0.0940 |
| MPAA PG-13 | 0.14127 | 0.4595557 | 34 | 0.307406 | 0.7604 |
| MPAA PG | 0.422365 | 0.4331889 | 34 | 0.975014 | 0.3364 |
| MPAA R | 0.002945 | 0.4283870 | 34 | 0.006875 | 0.9946 |
| GenreAdventure | 0.187922 | 0.4186123 | 34 | 0.448916 | 0.6563 |
| GenreAnimation | 0.014221 | 0.4206108 | 34 | 0.033811 | 0.9732 |
| GenreComedy | 0.630901 | 0.2589630 | 34 | 2.436261 | 0.0202 |
| GenreCrime | 0.882944 | 0.4422574 | 34 | 1.996449 | 0.0539 |
| GenreDrama | 0.38809 | 0.2697853 | 34 | 1.438513 | 0.1594 |
| GenreHorror | -0.230667 | 0.3703885 | 34 | -0.622771 | 0.5376 |

Table C.1.: Fixed-effects results of Equation (4.2) omitting the outliers using $MPI_{\alpha=0.8}$

| Fixed effects: log(Rev) ~ log(MPI) + log(Vol) + log(Screens) + log(Budget) + Week + Week2 + Critic + MPAA + Genre | | | | | |
|---|---|---|---|---|---|
| | Value | Std.Error | DF | t-value | p-value |
| (Intercept) | 3.832246 | 1.8207067 | 282 | 2.104812 | 0.0362 |
| log(MPI) | -0.013925 | 0.0437297 | 282 | -0.318432 | 0.7504 |
| log(Vol) | 0.201707 | 0.0265306 | 282 | 7.602812 | 0.0000 |
| log(Screens) | 0.73452 | 0.0245682 | 282 | 29.897176 | 0.0000 |
| log(Budget) | 0.288547 | 0.0946358 | 34 | 3.049028 | 0.0044 |
| Week | -0.559367 | 0.0366326 | 282 | -15.269676 | 0.0000 |
| Week2 | 0.030713 | 0.0036580 | 282 | 8.396066 | 0.0000 |
| Critic | 0.026378 | 0.0057955 | 34 | 4.551419 | 0.0001 |
| MPAA NR | -0.959309 | 0.5654627 | 34 | -1.696503 | 0.0989 |
| MPAA PG-13 | 0.119234 | 0.4622581 | 34 | 0.257938 | 0.7980 |
| MPAA PG | 0.44085 | 0.4348087 | 34 | 1.013895 | 0.3178 |
| MPAA R | 0.006099 | 0.4307044 | 34 | 0.014160 | 0.9888 |
| GenreAdventure | 0.168503 | 0.4203852 | 34 | 0.400830 | 0.6911 |
| GenreAnimation | -0.026146 | 0.4225041 | 34 | -0.061883 | 0.9510 |
| GenreComedy | 0.626718 | 0.2608085 | 34 | 2.402981 | 0.0219 |
| GenreCrime | 0.855882 | 0.4444758 | 34 | 1.925598 | 0.0626 |
| GenreDrama | 0.37201 | 0.2714967 | 34 | 1.370220 | 0.1796 |
| GenreHorror | -0.228653 | 0.3729577 | 34 | -0.613081 | 0.5439 |

Table C.2.: Fixed-effects results of Equation (4.2) omitting the outliers using $MPI_{\alpha=0.2}$