Reaching for the Stars

by applying a theoretically substantiated approach to star power in the context of box office success prediction.

Daniel Lang (40294)

STOCKHOLM SCHOOL OF ECONOMICS
Master Thesis in Business & Economics, Marketing & Media Management
Fall 2013

Abstract:
This thesis explores the effect of stars on the box office revenue of motion pictures. This is done by deriving a two-dimensional measure of stardom from the theoretical foundations provided by Sherwin Rosen and Moshe Adler. The two dimensions used and operationalized are talent on the one hand and consumption capital, a measure of previous knowledge of the movie cast's actors, on the other hand. Alongside previously established predictors of movie success these are entered in a regression to determine their influence on movie revenue. Both dimensions are shown to exert influence on the box office with consumption capital wielding the greater power.

Keywords: Motion picture, movie, star power, box office, talent, consumption capital

Supervisor: Anna Nyberg & Per Andersson
Opponent: Leah Steffensen
Examiner: Daniel Tolstoy
Presentation: Stockholm, December 18th, 2013

A special thanks to

Anna Nyberg & Per Andersson
for their caring advice and tutelage

Patric Andersson & Per Hedberg
for discussion and council

Christian Opitz & Kay Hofmann
for introducing me to this fascinating topic

Charlotta Knigge & Lars Velten
for reflection and advice on methodology

My family & friends
for enduring me during the last months

Table of contents

List of tables

List of figures

List of abbreviations

3D Three-dimensional.
bn Billion, as in $100bn.
CGI Computer generated imagery.
DVD Digital versatile disc.
e.g. Latin *exemplis gratia*: for the sake of an example.
et al. Latin *et alii*: and others.
etc. Latin *et cetera*: and so forth.
et seq. Latin *et sequentia*: on the following pages.
i.e. Latin *id est*: that is.
m Million, as in $100m.
n.a. Not available.
p. Page.
ROI Return on investment.
U.S. United States of America.

Glossary

Academy Award | An award of artistic achievement in the film industry also known as Oscar given for movies and performances of the previous year. Winners are voted on by approximately 6000 voting members of the Academy who are film industry professionals themselves (AMPAS, 2013a).

AMPAS | The Academy of Motion Picture Arts and Sciences is a professional organization dedicated to the advancement of filmmaking. It is most famous for its yearly → ***Academy Awards*** (AMPAS, 2013a).

Backend | An actor's participation in a movies profit aside from their fixed salary. Sometimes part of a more elaborate compensation contract it is commonly expressed as percent points of the producer's share of the → ***box office*** revenue (Pomerantz, 2010).

Blockbuster | An extraordinarily successful movie; derived from aerial bombs used in World War II which were capable of destroying, thus busting, an entire city block.

Box Office | Originally the location of sale for cinema tickets; by extension used as synonym for the revenue generated by the sale of cinema tickets.

Budget | The sum of funds available for realizing a movie project. Commonly separated in the production budget, which entails the cost associated with making the movie, and the marketing budget, which consists of the cost associated with selling the movie, like advertising. For an example of a budget sheet consult appendix p. 135.

Distribution company | A company responsible for making a movie available to consumers, through theatrical release, digital distribution or distribution of DVDs or Blu-rays. Large studios have their own distribution companies. Distribution companies have to be secerned from the theatrical distribution network which consists of the actual venues, the movie theatres, which the U.S. studios have been forbidden from owning since the Hollywood Antitrust Case of 1948 (Jacobs, 1983).

Director | Holds the main responsibility for the visualization of the script and guides the crew and cast in realising a vision of this visualization. Central role in the production phase of a movie production.

Franchise | Collective noun subsuming the entirety of movies of related story from one source, thus a wider definition as sequel and prequel (Hoffmann & Rose, 2005). "The Avengers" - technically not a → ***sequel*** to any other movie - would be classified as a franchise as it is part of the universe of Marvel-novel based movies produced by Disney.

Golden Globes | Award with similar ambition as the → ***Academy Award***, administered by the Hollywood Foreign Press Association. It is commonly considered of to be less prestigious than the Oscar (Pomerantz, 2013).

IMDb.com | Abbreviation of "Internet Movie Database", an online database for movie information owned and operated by Amazon.com. It contains detailed information on movies and allows users to rank movies on a scale of 1 to 10, the latter being the best.

Metacritic.com | Internet service gathering and accumulating reviews of movies. It differs from IMDb in considering only the verdict of professional critics, defined as staff writers for large print publications or members of selected critics' societies. Same concept as RottenTomatoes.com.

MPAA | The Motion Picture Association of America is a trade association representing the major Hollywood studios. It reports on industry development, promotes its members' interest in the political theatre and administers the → ***MPAA rating*** of movies' age suitability.

MPAA rating | Rating of age appropriateness administered by the → ***Motion Picture Association of America***. Currently there are five ratings (MPAA, 2013a): G (General audiences; all ages admitted), PG (Parental guidance advised), PG-13 (Parents strongly cautioned; inappropriate for children under 13), R (Restricted; children under 17 only admitted under parental supervision); NC-17 (No one under 17 admitted).

Oscars → ***Academy Awards***

Prequel | A movie that precedes another movie, either in terms of production chronology – "Shrek" is a prequel to "Shrek 2" – or in story chronology – "Star Wars Episode III" from 2005 is a story prequel to "Star Wars" from 1977.

Producer | Supervises the entire process of film production, encompassing the matching of story, → ***writer***(s), → ***director***(s), actor(s) and financer(s). Producers are often employed by a → ***studio*** in which case movies they produce are usually financed at least in part by that studio.

RottenTomatoes.com → Metacritic.com.

Sequel | A movie that succeeds another movie in production chronology, rarely applied to successors in story chronology that are → ***prequels*** in production chronology.

Studio | A company that owns and operates facilities to make movies which are used by film production companies. In common use as an integrated conglomerate that owns and operates facilities to produce films, as well as film distribution and production companies that employ → ***producers*** (Litman, 1998).

Style A poster | The most frequently used poster motive in the advertising of a specific movie; it consists of one sheet traditionally 27x40 inches.

Writer | Also called screenwriter or scriptwriter. Writes the script, a scene-by-scene sequence of events and dialogue that represents the foundational content of movies. Writers can work on the basis of their own imagination, producing an original screenplay, or transform an existing story or idea into a format that is suitable for cinematic use, called a screen-adaptation.

1 Introduction

" I want to be a ***big star*** *more than anything.*

It's something ***precious****. "*

Marilyn Monroe, quoted in Spoto (2001, p. 232)

Had Marilyn Monroe asked a hundred people on where she needs to go to fulfil that dream, the answer would have led her to the very same place she went to anyway: Hollywood. Over time this small district of Los Angeles, California has gained a colourful reputation that spans from outright fame to notoriety. Often referred to as the "Dream Factory" it is the perceived and actual epicentre of the Western world's motion picture industry. But while those "dreams" dominate the wildly stylised image the industry purposefully cultivates, there is no denying it is exactly that: an industry, or even more, a business.

1.1 The "Dream Factory"

The distinction seems meaningless at first, but at second glance it reveals a different focus of activity: "Film used to be an industry: its aim was to make films first, money second. Today, film is clearly a business." (Schumacher, 2000, p. 234). And what a business it is; revenue from movie theatre tickets in North America alone grossed the industry in excess of $10 billion in 2012 (MPAA, 2013b) more than twice the reported revenue of the music industry's sales from digital and physical recorded music in the same period and market (RIAJ, 2013). On a global scale the five years from 2008 to 2012 have witnessed 11 movies breaking the $1 billion dollar box office revenue barrier on their own, with the current leader of the scoreboard "Avatar" boasting $2.78 billion dollars in global ticket sales alone (MPAA, 2013b; The Numbers, 2013a). And while the number of sold tickets has been slowly decreasing since its peak at 1.58 billion units sold in North America in 2002, increased willingness to pay – partly incited by new formats like 3D movies (Faughnder, 2013) – has overcompensated those losses making 2012 at $34.7 billion global ticket revenue the most successful year on record in terms of revenue at the North American and global box office (The Numbers, 2013b; Zeidler, 2013). And while many try, no one gets even close to capturing the same share of this as the Hollywood studios who are responsible for two thirds of this global revenue (Hoad, 2013).

1.2 Hollywood nightmares

Revenue without cost however is a poor measure of economic viability if one calls to mind the idea of Hollywood as a business. And while those billion-dollar blockbusters have all more than earned their fair share of profit and provided their producers with magnificent rates of return, there have been more than a few million-dollar graves, or box office bombs. At this point it needs to be pointed out that only a part of the revenue generated by a movie at the box office is available to cover the cost of producing it. In practice the break-even point of a movie production lies at around twice its production budget as the theatrical distribution side of the business, i.e. the cinemas, retains around 50% of the revenue generated (Cinemark, 2011; Natale, 1999; Pomerantz, 2012; Weintraub, 1995). With this in mind the $282 million in box office revenue earned by Disney’s 2012 science-fiction adventure “John Carter” pale in comparison with the movie’s $250 million budget. Using the mentioned broad rule for profit estimates the movie incurred a loss in excess of $100 million. Surveying the top 10 financial failures in the 10 years between 2003 and 2012 shows that “John Carter” is not alone, in fact the year 2013 is set to enter 3 more movies into this hall of – questionable – fame:

Rank	Title	Year	Budget	Box office	Loss estimate
1	*Mars Needs Moms*	2011	$150m	$38m	-$130m
2	*John Carter*	2012	$250m	$282m	-$108m
3	*Sahara*	2005	$160m	$119m	-$100m
4	*Stealth*	2005	$135m	$76m	-$96m
5	*The Alamo*	2004	$107m	$25m	-$94m
6	*Green Lantern*	2011	$200m	$219m	-$90m
7	*Evan Almighty*	2007	$175m	$173m	-$88m
8	*The Nutcracker in 3D*	2010	$90m	$16m	-$81m
9	*The Wolfman*	2010	$150m	$139m	-$80m
10	*XXX: State of the Union*	2005	$113m	$71m	-$77m
Expected 2013 additions					
	R.I.P.D.	2013	$154m	$70m	-$118m
	The Lone Ranger	2013	$225m	$244m	-$102m
	Jack the Giant Slayer	2013	$185m	$197m	-$86m

Table 1: Largest absolute losses from Hollywood productions between 01/2003 and 12/2012; data: (Box Office Mojo, 2013a)

While the artistic merit of these movies will not be judged here, their financial failure is indisputable, and at the same time evidence that even the most seasoned executives in studios that have produced movies for more than a century are far from clairvoyant when it comes to predicting box office results of a movie project they are considering.

1.3 Dream catchers

As the industry itself is well aware of the fact, that failure of a movie cannot simply be ascribed to personal mistakes of those involved in its production, it has – since its beginnings – tried to devise methods of protecting itself. But when it comes to the question of how to protect oneself from such catastrophic cases of failure to launch “Hollywood is the land of hunch and the wild guess” (Dizard, 1994, p. 144). Thus, in their attempts to risk-proof their multi-million dollar investments studios will go to considerable lengths, in some cases not playing entirely by the book; from the invention of fictitious film critics like “David Manning” (Elsworth, 2005) – in fact only a nom de plume for a studio executive – to having studio employees pose as moviegoers and give favourable verdicts on their own company’s movies (Morris, 2005). Similarly awards like the coveted Oscars or Golden Globes, by now worth so much more than the recognition and accolade they once were intended to represent, have in their history seen their fair share of anything from the fixed 1929 award to Mary Pickford to a multitude of more or less louche attempts at nudging the prize toward a specific movie or actor (Bona, 2003; Litman & Ahn, 1998).

Aside from these approaches, other more innocuous techniques of risk mitigation have been observed, a premiere one of which is what could be referred to as the “Marilyn-Monroe-Strategy”. If one is to consider the connotations and ideas associated with her name more than 50 years after her premature demise, there will be few to argue against the notion that she succeeded in her endeavour to become something precious, to become a star. Thus, after her breakthrough to superstardom with “Niagara” in 1953, any movie she starred in was advertised with her name as though it was – or maybe because it was – a seal of guaranteed quality. This strategy has been ascribed to Adolph Zukor, founder of Paramount Pictures and film producer in the early 20th century:

> *"His product was the film star who became simultaneously a focal point for the construction of narratives within the film and for management coordination throughout the industry. The new star formed a synergistic link between film as an aesthetic form and as a product of corporate industry."*
>
> Kerr (1990, p. 387)

Likening this amalgamation of the star's identity with the movie to a celebrity endorsement for a product, a topic widely discussed in marketing research (Erdogan, 1999), however fails to recognize the fact that the celebrity, i.e. the star, derives his stardom from previous editions of the product he is embracing and is part of the product itself. The effect thus goes beyond the common conditioned response caused by the association with the endorser, as the star has, based on past performances, imputed competencies that affect the quality of the product itself.

1.4 Effectiveness of stars

To measure this effect is, unfortunately, a different issue. And while one could say that there is no harm in trying, the salaries those stars demand most certainly warrant a thorough consideration. Regularly demanding between $20 and $30 million, a figure exacerbated by the previously discussed break-through ratio between budget and box office, the piece of mind studio executives might be buying comes at a price. In addition to their salaries actors also increasingly explore so-called backend deals in which they participate in the revenue of the movie itself. While this may appeal as a type of risk sharing, it too, costs the studio dearly as the case of Johnny Depp in the "Pirates of the Caribbean" movies can illustrate: in addition to his total of $120 million fixed salaries (IMDb, 2013a), Depp received an impressive $350 million in a box office participation deal (Thompson, 2011). At a worldwide combined gross for the four movies of $3.7 billion (Box Office Mojo, 2013a) this results in one actor reaping roughly 25% of the cash flow going to those who, unlike him, stand the risk of actually losing money with the production of the movie. Thus the unknown benefits need to be weighed against the in fact very well-known costs of employing one or several stars. As is often the case in situations like these, the discussion then resorts to the reference of single cases:

> *"To me 'A guy stranded on an island' without Tom Hanks is not a movie. With another actor, it would gross $40 million. With Tom Hanks it grossed $200 million. There's no way to replace that kind of star power."*
>
> Bill Mechanic, former CEO of Fox, quoted in Bing (2002, p.1)

> *"But a star – or even two – is no guarantee. A studio can hire Pierce Brosnan and Geoffrey Rush, and buy a book by John Le Carré, and still bomb, as Sony's Columbia Pictures did with The Tailor of Panama."*
>
> Ackman (2003, p. 1)

Further even, Ackman (2004) imputes a confirmation bias and a selective memory to the industry when it comes to judging the usefulness of employing big Hollywood stars: picking up the poor opening weekend performance of "Troy" – starring Brad Pitt and at $175 million the most expensive production of 2004 – he speculates that "had Troy opened impressively, one can be sure Pitt would have gotten the credit" (Ackman, 2004, p. 1). Likening the attribution of success and blame-shifting for failure to behavioural patterns of corporate executives he comes to the conclusion that "even if Troy continues to flounder, Pitt certainly will not be blamed and will move on to the next massive payday" (Ackman, 2004, p. 1).

While "Troy" turned out to be what Hollywood refers to as a "sleeper hit" (Berger, 2011) – a movie that opens disappointingly but becomes highly profitable in the long run – both the studio executive and the journalist may be right within the confines of their own examples; a true answer however needs to rest upon a foundation of theory validated by quantitative analysis.

1.5 Purpose Statement

This then is the purpose of this thesis: to explore and illuminate the effect of star casts in motion pictures on the box office success of those motion pictures. This shall be done in a manner that not only contributes to and extends the general body of knowledge on the issue but also renders the results utilizable and useful for the people making movies by relying only on parameters that are available to them in the early stages of their decision making.

1.6 Delimitations

Due to the approach and data used all implications from this thesis have their full applicability only towards U.S. major studio productions with budgets above $5 million. The analysed market however will not be limited to the U.S. but instead be global, and the limitations linked only to the findings of the quantitative part of this thesis. Thus, it goes not to say that the concepts explored are not applicable outside of this realm. In fact the concept of stars is transcending geographical and cultural separations, and the theoretical foundations are also not specific in regards to their area of application, both in terms of geography but also in terms of the content they can be applied to.

1.7 Disposition

After this introduction a, first cursory, then in-depth look at the available literature will follow, identifying a gap in the existing body of research. Then the sample of movies used will be presented along with the set of control variables available for each of them. Particular focus will lie on the quantitative operationalization of the proposed determinant factors within the stardom concept. Based on this the methods employed for a subsequent regression analysis will be described, followed by a description of its outcome. Those results will then be discussed in detail, looping back to the literature review and the theoretical foundations. At this point also the limitations of the thesis will be reviewed. Based on this a conclusion will be drawn. Finally, an outlook towards further potential research areas applying the findings of the thesis will mark the end.

2 Literature review

This review will first give a broad overview of past research in the motion picture industry. It will then take a closer look at research models focussing on specific proposed determinants of box office success. Based on this, and picking up on research suggestions from the reviewed literature a research gap will be pointed out. Then, theoretical models suitable to fill this gap will be presented and analysed.

2.1 Broad box office prediction models

Success prediction models for motion pictures are plentiful – they scrutinize the performance of U.S. movies in the U.S. domestic and global market (Hennig-Thurau, Houston, & O'Neal, 2006; Hennig-Thurau, Walsh, & Wruck, 2001; Litman, 1983; Terry, Butler, & De'Armond, 2003; Walls, 2005), as well as abroad in markets like the United Kingdom (Elliott & Simmons, 2008), Germany (Dewenter & Westermann, 2005) or Italy (Cucco & Candeloro, 2011; Waterman & Jayakar, 2000). There have also been several comparative efforts offsetting different success factors of U.S. movies in different domestic markets (Craig, Greene, & Douglas, 2005; Elberse & Eliashberg, 2003). Equally the mega-industry that is Bollywood – long ignored by the Western research community – is now turning into a more thoroughly researched subject (Fetscherin, 2010; Tussu, 2008). Beyond that even smaller domestic movie markets have been examined such as success predictors within the domestic film industries of Germany (Hennig-Thurau & Wruck, 2000; Jansen, 2005) or Italy (Bagella & Becchetti, 1999).

While there are obviously extensive differences between these studies in terms of method, focus, and results, there are somewhat communal insights. These include the fairly straight-forward association between budget and box office revenue (Basuroy, Chatterjee, & Abraham, 2003; Litman & Ahn, 1998). Similarly the impact of the movie's genre has been scrutinized - from Anast (1967) showing increased revenue for violent and erotic movies, to Litman (1983) finding science-fiction movies to be more profitable, on to Neelamegham & Chintagunta (1999) finding the thriller genre to be the most popular across countries and continents. Others find evidence for the action genre being the biggest box office driver (Terry et al., 2003). Thus there is general support for genre being a determinant, but dissent as to which genres in specific improve box office performance. Similarly the effect of the age rating issued by the Motion Picture Association of America (MPAA) is unclear, as Ravid (1999) and Austin, Mark & Simonet (1981) find strong evidence of a positive impact of youth-friendly ratings on the rate of return of movie productions, while Sharda & Delen (2006) and Litman (1983) find no support for such a connection.

In short - no general shortage of research is to be observed when it comes to the success prediction of movies. However, it behooves one to take a second look at distinctions beyond the somewhat crude regional differentiations: one such distinction is whether the research is focussed on one issue or an attempt towards a general model. While many of the earlier approaches aim to explain movie success as a whole, the more recent publications usually focus on one particular area as the supposed driver of box office success.

2.2 Focussed prediction models

The general models simply aim at the highest possible percentage of explained variance. In this they propose no single construct with a supposedly paramount or special role, but instead aim at depicting and analysing the issue with a preferably very broad set of variables. The more specific approaches hypothesize on the impact of one singular aspect on the success of the examined movies and thus focus their investigation. An example of such a specific aspect is the impact of critical reviews on box office success.

2.2.1 Critical reviews

As expert judgements have been found to strongly influence market performance in various industries (Cameron, 1995), several investigations into the existence and extent of such an effect have been made regarding the motion picture industry (Eliashberg & Shugan, 1997; Gemser, Oostrum, & Leenders, 2006; Holbrook, 1999). Frequently these approaches utilize the advent of internet sites collecting and accumulating professional reviewers’ verdicts on movies such as Metacritic or Rotten Tomatoes. Applying the ratio of positive and negative reviews together with control variables such as the budget of the movie, the age rating it has been assigned by the MPAA, and awards it has received, Basuroy, Chatterjee & Abraham (2003) find proof of a negativity bias, meaning that the magnitude of the negative impact of negative reviews exceeds that of the positive impact of positive reviews. Extending on this Boatwright, Basuroy & Kamakura (2007) find critics to be influencers of success rather than predictors. This notion is supported by a finding of Brown, Camerer & Lovallo (2012) who have showed that “cold openings” of movies – cold opening meaning the movie has not been shown to professional critics prior to release – are correlated to a 10% to 30% increase in box office revenue.

2.2.2 Awards

A specific type of critical reception are awards which are endowed upon motion pictures. Smith & Smith (1986) survey the awards given out to the film and the actors in it and find different impacts on distributor rental agreements over different decades; however they do not include box office revenue in their model. Verifying earlier research of Dodds & Holbrook (1988), Nelson, Donihue, Waldman & Wheaton (2001) find that wins in the Academy Awards positively impact box office revenue for those movies still running in theatres when the winners are announced: a movie which has been released in the fourth quarter of the previous year[1] winning the best picture category will on average experience an increase in box office revenue of $16 million. Outside their monetary effects awards in general and the Academy Awards in particular have experienced a "gradual acceptance as an institutionalized measure of quality" (Levy, 1987, p. 330), an aspect which opens them up to further utilization in the prediction of film success.

2.2.3 Word of mouth and consumer ratings

Apart from the judgement of professional critics also consumers have been enabled to publicize their rating of a movie, mostly by using Internet platforms. One premiere locus of such consumer ratings is the Internet Movie Database (IMDb), a web service owned and run by Amazon.com. The site allows registered users to rate a movie on a scale from 1 to 10, the Top 250 movies alone received a total number of over 80 million ratings (IMDb, 2013b). These scores in turn are found to be significantly correlated to the long-term box office revenue by Hennig-Thurau, Houston & Sridhar (2006); a connection to the box office on the opening weekend was not found, which seems reasonable if the rating itself is hypothesized as an influencer as a certain latency is to be expected.

Similar to those focussing on the impact of judgements by expert and consumer critics other studies have concentrated on the effects of word-of-mouth from the consumer in the periods shortly before and after the release of movies (De'Vany & Lee, 2001; McKenzie, 2009). At this point the separation between consumer criticism and word of mouth is very fluent, a separation nigh impossible, a view supported by the findings of Oghina, Breuss, Tsagkias & Rijke (2012) who manage to adequately predict the user rating score on the Internet Movie Database (IMDb) using activity levels and valence on Twitter and YouTube, thus showing a

[1] Their results differ depending on the time in the year when the movie in question was released. As the Oscars are given out in the Academy Awards ceremony which takes place between the last week of February and the last week of March, effects are superior for movies which have been released late in the previous year.

linkage between the mere rating of a movie by a consumer and potential word-of-mouth utterances of that consumer.

Some of these approaches differentiate more precisely between the valence and amount of word-of-mouth. In this context Yong (2006) finds that the majority of the explanatory value stemming from word-of-mouth is in fact coming from its sheer amount and not its benevolence. Correspondingly Mishne & Glance (2006) find that the pre-release sentiment about a movie in the sphere of Internet blogs is insufficient to predict the box office result. In a more short-term oriented approach Asur & Huberman (2010) measure the determination of box office revenue from Twitter messages the day before the release. While the resulting coefficient of determination reaches 97% their model is somewhat limited by its sample size of only 24 movies. Consequently Wong, Sen & Chiang (2012) in a larger sample find a smaller predictive quality for the number of tweets and instead propose a so-called “hype-approval-factor”, which contains the ratio of “positive tweets before watching the movie” and “positive tweets after watching the movie” thus measuring in how far the movie lived up to the audience’s expectations.

2.2.4 Prequels

Looking at the top 10 highest grossing movies of 2012 quickly reveals that 7 of them have one thing in common: they are building upon the foundation of another movie, a prequel (Box Office Mojo, 2013b). The concept of producing not only one movie, but a whole series of them is only seldom the artistic foresight George Lucas showed when in 1978, one year after the release of his first Star Wars movie “Star Wars Episode IV: A New Hope”[2] he already publicly stated that Star Wars was going to be a series of movies with potentially as many as 10 parts (Donnelly, 1978). Instead these series or franchises are often only continued up to the point where they fail to be profitable or key actors drop the project. “Franchise” in a movie context denotes a somewhat wider definition containing sequels but also movies utilizing a pre-existing universe of characters or events. Based on the often selective continuation of financially successful projects, it is apparent why effects of this phenomenon on the success of the following movies have been repeatedly studied (Dhar, Sun, & Weinberg, 2012; Hennig-Thurau, Houston, & Heitjans, 2009; Prag & Casavant, 1994; Terry et al., 2003). While differing in operationalization the consensus of these is a positive effect of at least the existence of a prequel,

[2] While the release title in 1977 was in fact “Star Wars“, its sequel in 1980 was already called “Star Wars Episode V” implying a minimum of 3 additional movies. “Star Wars” was subsequently re-released in 1981 under the title “Star Wars Episode IV: A New Hope”.

which while often stronger for the short term box office on the release weekend (Hennig-Thurau, Houston, & O'Neal, 2006), is still significant for the lifetime box office.

2.2.5 User activity and search behaviour

Another recent approach going in yet another similar direction as word of mouth and critical reviews is that of predicting success based on search behaviour and activity in open knowledge platforms – more precisely Wikipedia. Measuring the editing frequency of the movie's Wikipedia article Mestyán, Yasseri & Kertész (2013) achieve a noteworthy 94% R^2 of explained variance in box office results within the week before the respective movies' release. Compared to the previously mentioned studies focussing on consumer sentiments this approach is intriguing as it provides high predictive value without requiring an assessment of the valence of the user activity.

A further study with the same advantage is a recently published Google investigation which achieves an equally remarkable 92% R^2 of explained variance on opening weekend U.S. domestic box office results based on only the amount of searches on Google, the amount of clicks on paid ads on Google, the number of theatres the movie will be released in, and the franchise status of the movie (Panaligan & Chen, 2013). In their regression Panaligan & Chen (2013) find for instance that an additional 20,000 clicks on paid advertising on Google is likely to add another $7.5 million in box office revenue.

2.2.6 Star power

The multitude of hypothesized and then statistically supported antecedents of a movie's success at the box office lends credibility to the thesis that there must be some underlying construct which influences those previously outlined determinants. The fact is that attempting to deduce the direct impact of aspects like the previously described critical reviews or word-of-mouth is in a way flawed as "such an approach is implicitly based on the assumption that each success factor influences movie success independently, but does not take into account the existence of inter-factor relationships, where one success factor exerts an influence on other success factors" (Hennig-Thurau, Houston, & O'Neal, 2006, p. 4). Moreover it disregards the possible existence of an indirect factor which has high explanatory value on the supposedly independent explanans factors while having an obscured relationship to the explanandum itself.

A construct that has been proposed for - and tested in - such a role is the impact of well-known actors on box office revenue (De'Vany & Walls, 1999; Elberse, 2007; Holbrook, 1999;

Neelamegham & Chintagunta, 1999; Prag & Casavant, 1994; Sochay, 1994; Wallace, Seigerman, & Holbrook, 1993), often referred to as star power. The idea that the presence of a previously successful and well known movie star induces consumers to talk or tweet about a movie, search for it online, or even for critics to issue a more benevolent verdict is at the very least a reasonable proposal for researchers to investigate. Earlier publications focus mostly on video rental revenue as the dependent variable and do not come to a unanimous conclusion concerning the impact of star power. While Wallace, Seigerman & Holbrook (1993) reveal a bankability of some actors, Prag & Casavant (1994) find star impact disappearing when advertising expenditures are entered into their model.

In so far the only study on the success of motion pictures aimed not at establishing direct antecedents but at illuminating their interrelationships, Hennig-Thurau, Houston & O'Neal (2006) employ a path analysis methodology, thus highlighting common variance between antecedents instead of obscuring it. Their model puts star power – in their study operationalized using a ranking of the industry magazine The Hollywood Reporter (Burman, 2006) – in a central yet ambivalent position: star power is found to influence the number of awards for the movie, the benevolence of professional critics, the amount of advertising expenses, and has - maybe surprisingly – a significant negative impact on the consumer-perceived quality of the movie. In their full success model star power's coefficients remain insignificant, while in a trimmed version the total effects turn significant, but with a negative impact. Based on these findings their most prominent future research implication is that

> *"…star power and a high production budget (…) are problematical, and a deeper understanding of stars' and budgets' relationships with box office and profitability has to be gained. As some stars obviously are successful at the box office, the factors that determine a star's influence on movie success must be identified."*
>
> Hennig-Thurau et al. (2006, p. 24)

There have been several attempts to address this research proposal, a prominent of which is an event study by Elberse (2007) observing cumulative abnormal returns in response to casting announcements, utilizing a market simulation from the multiplayer online game "Hollywood Stock Exchange". Despite technically being a game, the Hollywood Stock Exchange proves to have an impressive similarity to real stock exchanges in its ability to aggregate and evaluate information. Thus the final trading price for a movie the day before its release has a 94%

determination coefficient for its real world box office result. A key finding of the study is the fact that the engagement announcement of a star yielded roughly $3 million in game share price and thus close to that in real world box office revenue (Elberse, 2007), however the fact that the research design does not allow for any insights into what evokes the differences in effect between the actors, as well as the spread in magnitude of the abnormal returns prompt

> *"…further research (to) explore each of these aspects, thus advancing knowledge on the origins of stardom, e.g. Adler 1985 (and) Rosen 1981."*
>
> Elberse (2007, p. 119)

2.3 The research gap

2.3.1 Star power operationalization

The suggestions for further exploration by both Hennig-Thurau, Houston & O'Neal (2006) and Elberse (2007) point towards a void in existing research, that has yet to be filled. The first cause of this void is the shortcoming of the reviewed attempts to find a theoretical foundation for their operationalization of star power.

In their work investigating the mitigating effects of stars on the risk associated with the production of motion pictures De'Vany & Walls (1999) consult two star lists: one of them is the "100 Most Powerful People in Hollywood" which was regularly published by the by now discontinued industry magazine "Premiere". The other is the so-called "Ulmer Scale" a rating published by a journalist and industry consultant measuring a star's bankability "derived from polling dozens of behind-the-scenes international power brokers" (Ulmer, 2013). Anyone found on either of these lists was classified as a star for the purpose of the research (DeVany & Walls, 1999, p. 292). The flaw of this system is for one the unknown methodology behind the rating itself, but more importantly the assumption of stardom being a binary variable, an assumption which the underlying Ulmer Scale already discards as it scores the actors, separates them into tranches and differentiates bankability by production budget. The same binary classification of stardom, however on varying bases, is also found in Sharda & Delen (2006), Holbrook (1999), Neelamegham & Chintagunta (1999) and Sochay (1994).

In this light the operationalization employed by Hennig-Thurau, Houston & O'Neal (2006), using the bankability ranking of the Hollywood Reporter, at first seems like a more valid solution. It is based on a survey which "polled 114 executives at both major studios and independent companies, financiers and various industry players from around the world" (CBS

News, 2009), thus the sources are at least specified by occupation and the result is used not binary but scaled. However the distribution of movie returns with and without the employment of stars (De'Vany & Walls, 1999) as well as the fact that Elberse (2007) found widespread evidence of stars whose engagement in a movie significantly decreased the expected revenue of that movie[3], suggests that the industry insiders' understanding of the quality of movie stars is incomplete.

2.3.2 Temporal alignment

The second issue is one regarding the applicability of findings for the industry producing motion pictures. A vast majority of the entire body of research utilizes factors, events and data that is not available at the point in time when the decisions on the properties and dimensions of the product, the motion picture, are made by those who carry the economic risk of producing a movie. To better illustrate this the manifestations of different variables used in the literature described in the previous sections have been ascribed to the phases of the process of filmmaking. The basis for this is a five-step process of filmmaking with activities condensed from Litman (1998), Bordwell & Thompson (2008) and Byers, Cranor, Cronin, Korman & McDaniel (2004). To this the concepts used in the reviewed body of research were added in a timeline assigning their time of manifestation to the different phases. Whether these assignments flawlessly replicate the sequencing in the real world is discussible, they are however based on the respective research that utilizes the underlying factors to predict box office revenue or success. The critical point in this continuum, visualized in figure 1 on the next page, is the demarcation line where the principal influence of producer and studio executives mostly abates. This is arguably the case between pre-production and production (Yudelson, 2010), also referred to as principal photography. Thus during development while the producer and - possibly - involved writers or in some cases already stars try to secure financing in order to reach the first base of Hollywood, the green light for pre-production, they can only observe a fraction of the variables. As the story outline is commonly the first thing to have in a movie production, the genre is known.

[3] While the majority of negative impact events found by Elberse (for a list of prominent examples consult Elberse, 2007, p. 115) were cases of actors dropping out of movie productions – such as Tom Cruise deciding not to star in "Cold Mountain" which took $10 million off the expected revenue – there were also significant negative impacts caused by actors joining the cast of a movie – such as Jessica Biel joining "The Texas Chainsaw Massacre" which decreased expected revenue by $5.7 million.

Stages of movie project development				
Development	Pre-Production	Production	Post-Production	Distribution & Release
Typical tasks in stage				
- Idea generation - Securing rights - Securing financing - Finding writers - Vetting directors	- Assembling crew - Casting actors - Screenwriting - Storyboarding - Set design	- Recording video - Recording audio - Directing - Acting - Cinematography	- Synchronizing - Creating soundtrack - Editing - Visual effects - Processing	- Advertising - Stars in media - Channel choices - Physical distribution
Attribute manifestations				
Story & Genre	Immanent quality of movie			Critical reviews (e.g. Metacritic)
MPAA target	Market revenue anticipation (e.g. Hollywood Stock)			Awards (e.g. Oscars)
Budget		Collaborative and search activity (e.g. Wikipedia, Google)		
Writer(s)		Word of mouth (e.g. Twitter, blogs)		
Director(s)			Marketing expenditure	
Key Actor(s)			Actual MPAA rating	Consumer ratings (e.g. IMDb)
Cultural familiarity	Release target			Ticket sales

Figure 1: Movie production process and manifestations

The same goes for franchise status, or in a wider sense cultural familiarity[4]. While the actual MPAA rating is dependent on the MPAA seeing the film and thus only available in the last stage of post-production, the producers already during development have decided which rating they will aim for. Should the MPAA final rating not meet their target, they commonly have the film re-edited and resubmitted to achieve the desired, usually lower age rating (Mosk, 1997; Sperling, 2011). Similarly a preliminary budget is known already before pre-production starts, the final production budget usually before actual production (Litman, 1998). Finally the first two phases yield a set of writer, director, and headline actors (Hoppenstand, 1998; Lee & Holt, 2005).

[4] Cultural familiarity includes the previously described concept of franchises but also content that is commonly known to the consuming public, e.g. the 2011 movie "Red Riding Hood" would not be classified as a franchise, because it is not part of a series like the Godfather trilogy, but as it is based on a commonly known fairy tale it would be classified as culturally familiar.

2.3.3 Requirements to fill the void

Based on this, any proposal to fill this void must fulfil two requirements:

1) In order for the results of the research to be useful to the industry that creates the analysed body of work, the research itself must be limited to the variables that have already manifested themselves, thus only knowing what is known at the time the decisions and choices are made. At this point it must be made clear, that this observation does by no means express or suppose any curtailment of the scientific validity and quality of the mentioned research, it merely focusses on the usability for the observed industry.

2) To understand the influence that stardom and thus stars exert on box office revenues, or in fact anything, first a thorough understanding of what a star is has to be found. Thus, a theory-grounded measure of stardom, which goes beyond binary scaling is required.

2.4 The stars

2.4.1 Defining the star ability

While the first of the two previously mentioned requirements is of a methodological nature, the second demands for a theoretical approach to stardom, which reveals that today's association of stardom with extensive media-coverage has forged our image of the star label to be that of someone who commands the media's attention (Redmond & Holmes, 2007). However, the earlier discussions of these phenomena focus on their ability to capture the – even by today's standards – most obvious reward of stardom: money. In this spirit famed economist Alfred Marshall ponders the ability of a star of his time, the opera singer Elizabeth Billington; while in his sibylline way he foresees the importance of technology for the future income of these exceptional artists, he fails to give specifics on the nature of the "exceptional ability to get very high prices" (Marshall, 1890, book VI, chapter XII, §11) that some artists possess and instead proposes rising income of the general population to be the source. Similarly, Max Weber's leadership typology has been applied to stardom (Redmond & Holmes, 2007), likening the abilities of Weber's charismatic leader type (Weber, 1922) to those required for stardom. This also however fails to provide an approach for operationalization of stardom. Fortunately Elberse (2007) in her research suggestion already points to a highly suitable source of theoretical foundations for further inquiries into stardom upon which a model examining its impact on box office profitability can be built: the works of Sherwin Rosen (1981) and Moshe Adler (1985).

2.4.2 Rosen's proposal

At first glance Rosen's idea in "Economics of Superstars" about what makes a star seems too simple to point out: talent. The gist of it however lies in the interrelation of talent and economic success: he observes that the utility of talent is inherently non-linear as "lesser talent often is a poor substitute for greater talent" thus "hearing a succession of mediocre singers does not add up to a single outstanding performance" (both Rosen, 2007, p. 846). This means that the function of revenue R in dependence on the quality q is strictly convex, thus $R''(q) > 0$. Further he elaborates on the technological progress which had been foreseen by Marshall (1890) which allows for virtually infinite reproduction with - if not zero - at least decreasing marginal cost. While talent itself remains a latent variable, the implication of Rosen's model is that it represents the ability to produce a product superior to the producer's peers without requiring additional resources. He concludes that

> *"...when the joint consumption technology and imperfect substitution features of preferences are combined, the possibility for talented persons to command both very large markets and very large incomes is apparent."*
>
> Rosen (1981, p. 847)

2.4.3 Adler's response

In his response to Rosen's much acclaimed theory, Adler goes one step further. At this point it must be emphasized that Adler does not intend to straight-out contradict Rosen, but instead to point out that it does not require the differences in talent for a market to be concentrated on stars. His argument is based on the theory of consumption capital developed by Stigler & Becker (1977) which stipulates that with the consumption of anything that is judged by taste, a body of knowledge is collected within the consumers which impacts their appreciation of future consumption. To this concept Adler adds a value of exchanging oneself with others on the topic of one's previous consumption:

> *" As an example, consider listening to music. Appreciation increases with knowledge. But how does one know about music? By listening to it, and by*

discussing it with other persons who know about it. In this learning process lies the key to the phenomenon of stars."

Adler (1985, p. 208)

Modelling a periodical game he shows conclusively that even in a set of artists with identical talent the consumers will, with time, converge to focus on a small number of stars. The drivers of this consumer utility function are twofold[5]: on the one hand the consumption capital, previously acquired knowledge which increases enjoyment of the consumption of an artist's performance and on the other hand the ability to exchange oneself with others on the topic of one's preferred artist which evokes something like a snowball effect (Adler, 1985; Stigler & Becker, 1977).

2.5 Application to movies

A key issue in applying Rosen's and Adler's theories to the motion picture industry, is the fact that the variable their models aim to explain is the income of the artists themselves. The purpose of this thesis however is to view the income they generate for the movies they star in. The actors however, unlike the comedians used by Rosen as an example category of artists, are a third-party service provider to an end product, the movie. The input they give adds revenue for the product – supposedly – but does come at a cost in the shape of their individual incomes, an information for which broad data availability is – despite regular media coverage of new salary records – poor. This means that a prediction of movies' net income would forcibly assume that the cost remains constant when swapping one actor for another, which due to differences in salaries and the potentially large share of total cost that these salaries make up, would spoil the point of the entire effort. In order to allow for changing properties related to stars which are not the measures of stardom themselves, such as variable cost of hiring the actors, the analytical goal must be the revenue they generate; otherwise the model would have an artificially fixed independent variable and any predictor coefficients would be worthless (appendix p. 134 contains a visualization of the problem). Any knowledge about revenue however can in turn – from the position of a researcher with more far-reaching access to data, or film producer who has the necessary information at hand – be filled in to allow for instance for an individual

[5] This twofold determination base is also the reason Adler's model cannot easily be reduced to absurdity by proposing that it would - after a long enough period of time - end up with only one worldwide "superstar"; as consumers incur switching cost in form of lost applicability of their "star knowledge" they can end up in a captive situation where their preferred star is not the big "superstar" (detailed in Adler, 1985, p. 212).

assessment of the financial feasibility of a particular project. A further difference between the assumptions of Adler and Rosen and the products to which they will be applied is the fact that movies may have more than one star, as stars often appear alongside other stars in the same movie. As such they form a combined body of resources, the cast. While the analytic target has to remain on the level of the movie, as the movie is what links explanantia and explanandum – the cast's attributes with the movie's revenue – it must incorporate the concept of both Adler and Rosen that more units of a lower quality are a poor substitute for fewer units of a higher quality. If the analysis is to purely focus on the cast as a single entity, it will fail to detect this phenomenon, thus also the impact of a potential fragmentation of the star qualities in a movie should according to the theories have an impact on the product performance.

The alternative to this setup would be to approach not a combination of cast and movie, but instead singular engagements of actors in movies. In this scenario one movie would, depending on the number of actors in its cast, appear multiple times in the analysis – this alone could be adjusted with a weighting mechanism. But with different independent variables determining the same dependent variable it would also be forcefully inducing diffusion into the coefficient. A remedy for this would be to use a mean or other type of average of attributes from the cast. This however would introduce punitive effects for movies with more than one leading actor. If a star's quality is someway measured and scaled, and for instance Julia Roberts is found to possess a quality of 8 and Johnny Depp a quality of 10, then a movie with Johnny Depp would, judged on an average of the attributes be viewed as superior to one with both Johnny Depp and Julia Roberts.

Instead of this in order to limit method-induced complexity the accumulation approach with a fragmentation measure was chosen. This way it is ensured that the combination of attributes from Johnny Depp and Julia Roberts cannot be lesser than each on their own, and that with the help of a fragmentation variable their effect can be compared to a third actor with the combined attributes. Thus if Tom Hanks has a quality of 18 it could be determined whether this, all else being equal, has a greater or smaller effect on a measure of movie success than the combination of 8 and 10.

2.6 Hypotheses

Subsequently this forms one of the suppositions this thesis will set out to test. Altogether five hypotheses were deducted from the theoretical concepts presented. The first and third relate to Rosen's talent conjecture, the second and fourth relate to Adler's proposed effects of consumption capital and the utility of exchanging with others on the topic of one's star[6]. The last hypothesis, as discussed, goes to test whether what is surmised on the level of the cast as a whole applies to its individual actors. The hypotheses thus are:

H1: Increasing talent in its cast increases the box office revenue of a movie.

H2: Increasing aggregate consumption capital in its cast increases the box office revenue of a movie.

H3: The marginal impact of talent in the cast on box office revenue is increasing.

H4: The marginal impact of consumption capital in the cast on box office revenue is increasing.

H5: The fewer stars the star qualities in the cast are concentrated on, the greater the box office revenue.

2.7 Proposed effect architecture

The general setup to test these hypotheses will follow the scheme that is depicted below. Throughout the next sections this simple version will continually be updated incorporating the elaborations of the employed and analysed data foundation.

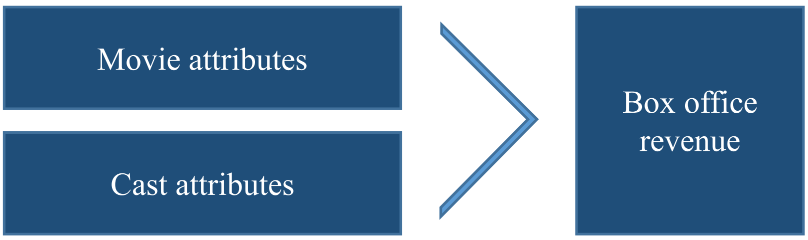


Figure 2: Basic model structure

[6] As they are deduced from one proposal, Stigler & Becker's (1977) idea of consumption capital and Adler's (1985) extended version of it, which is including the exchange utility, will henceforth be treated as one construct, referred to as consumption capital.

3 Data compilation

This section will start with a short description of the basic set of movies used as cases in this thesis. Then an overview of the variables that were linked to the movies will ensue. Further on, the method used to find the actors associated with each movie will follow, based on which the variables gathered regarding the individual actors will be reviewed.

3.1 Research strategy

In order to single out the hypothesized effect of star actors on box office results in movies, a broad spectrum of control variables is required. To this end, all movie-attributed variables available at the specified early stage of the project development and found to be predictors in the reviewed literature will be gathered, and then complemented by measures of stardom derived from the theories of Adler (1985) and Rosen (1981).

3.2 Research design

To test the previously formed hypotheses a linear regression model will be used. The fact that a linear model was chosen does not mean that a generally linear connection between determinants and dependent variables is assumed, as this would already foil any attempt to test the hypotheses of increasing marginal utility, which implies the existence of a connection best described by a power function. Instead altered terms will be used to test for these within the linear model where applicable. To strengthen the operationalization of the star measures, multiple variables will be used to operationalize each. High inter-correlations of these variable sets are to be expected, and are in fact desirable as they are a premiere signal of internal consistency (Cronbach, 1951; Tavakol & Dennick, 2011). As the resulting multicollinearity will however impact the two concepts' significance in the regression model and in fact the entire regression itself (Farrar & Glauber, 1967), the star measure variables will be submitted to a factor analysis, thus reducing the number of dimensions and ridding the analysis of multicollinearity caused by correlation among star-measurements (Jolliffe, 2002).

The remaining variables alongside the resulting factors will be entered into a regression with global box office revenue as dependent variable. As the total number of variables is generally unsuitable given the number of cases (Harrell, 2001), this model will be trimmed down to arrive at a stable and valid model.

3.3 Initial data set

As the delimitations section already point out, the scope of the quantitative part of this thesis is limited to large productions of U.S. major production and distribution studios. The output of this part of the industry is around 100 movies per year. As the operationalization of consumption capital requires data not only on the motion pictures themselves but also on the actors starring in them, using a timeframe reaching up to the present turns out to be problematic: the data providers used to collect the star-related data tend to have a considerable latency in introducing new actors into their databases. Thus the movie output of the last 5 years featured a significant number of stars on which data was not yet available. Therefore, in order to attain full data availability, the observation window was shifted 5 years back, thus incorporating the movies released in the 5 years between January 1st of 2003 and December 31st 2007. In this timeframe 512 movies were released by the major studios' distribution companies. Of these 18 fell below the $5 million minimum budget criterion and were thus not included in the regression analysis.

3.4 Movie-related variables

In the first phase of data gathering the variables associated with the motion pictures themselves were gathered. These variables aim to represent the entirety of the data foundation for the various focussed prediction models discussed in the literature review section in order to serve as control variables. Thus they include on the one hand the dependent variable, the box office revenue, the production budget, the MPAA-rating, genre and time of release, as well as information on the studio, potential prequels and the source of the story.

3.4.1 Box office revenue

The box office revenue was collected from two sources, the primary source being Box Office Mojo (2013), a service owned by Amazon.com. In cases of missing or noticeably corrupted data the secondary source used was The Numbers (2013c), a database maintained by another industry intelligence and consulting company. In the geographical dimension the data was collected for both the worldwide market, and the North-American market. In addition the revenue for the opening weekend in the North-American market was collected. Unless otherwise stated "box office" always refers to the global revenue over the entire lifespan of the movie. The timeframe of the data collection goes from release date to the date the movie was completely dropped from cinemas, thus, in order not to distort the data, the timeframe often extends beyond the end of 2007. Revenues are scaled in million $ and adjusted to 2003 price level to neutralize inflation effects.

3.4.2 Budget

The total budget spent on a movie is generally split into two sections: the production budget and the marketing and distribution budget. The latter consists of expenditures like advertising campaigns and promotion activities, but also of the production and distribution of the physical print copies of the movie if an analogous projector is used. While marketing and advertising of movies are gaining importance (Friedman, 2008), data on individual marketing budgets is scarce, but luckily the expenditure is consistently around one third of the budget (MPAA, 2008). Furthermore the marketing budget is not determined at the time the production is greenlighted, thus using it for success prediction would not be permissible.

The more important, more readily available, and also at the time available part of the total budget however is the production budget. This part subsumes costs for acquisition of rights to the story, payments for writers, directors, cast, expenses on set design, visual effects, transportation, music, post production and a plethora of other elements required in the production of a motion picture (for an example of a detailed production budget see appendix p. 135). A weakness of this variable is the fact that while the budget is set in the very early stages of a movie's development, cases of significant budget overruns are not unheard of; they are however not common (Munoz, 2006). Production budgets were gathered from Box Office Mojo (2013) and The Numbers (2013c), figures are in million $, and – as the box office revenue – are adjusted to 2003 price levels.

3.4.3 MPAA-rating

The MPAA-rating for the movies in the sample was obtained from the website of the consumer-oriented website of the MPAA (2013b). The ratings awarded by the MPAA and their respective definitions are (MPAA, 2013c):

- *G* - General Audiences. All Ages Admitted.
- *PG* - Parental Guidance Suggested. Some Material May Not Be Suitable For Children.
- *PG-13* - Parents Cautioned. Some Material May Be Inappropriate For Children Under 13.
- *R* - Restricted. Children Under 17 Require Accompanying Parent or Adult Guardian.
- *NC-17* - No One 17 and Under Admitted.

In addition two movies were not rated, as the submission of movies to the MPAA for rating purposes is technically voluntary. These movies were marked “NR” for “not rated”. While the MPAA-rating itself is an ordinal scaled variable as the suggested restrictions on admission or access increase with the amount of violence, sexual content, offensive language, drug abuse, or other aberrational behaviour, the multitude of determining dimensions already shows that the effect of the rating cannot be assumed to be ordinal, as a U-shaped or inverted U-shaped impact can easily be hypothesized. Therefore the rating variable was converted into a set of dummy variables representing each potential status.

3.4.4 Genre

For the genre assignment the IMDb database was used (IMDb, 2013c). There, every movie is assigned a genre or a list of genre attributions in order of descriptiveness. Thus every movie is assigned at least one, but up to four genres which are sorted by how well they describe it as judged by IMDb. This presented a problem, because using a total of ten genre attributions would result in 40 dummy variables associated with genre which in turn is impractical and in a later regression would elicit an R^2 adjustment. This issue at first sight leaves two options: The first is to drop the additional genres and lose the associated information. This seems drastic, as it would confound any chance to differentiate between different manifestations of one main genre. An example for the differentiation that would be lost is the distinction between “Bad Boys II” and “Shooter”: both are primarily action-movies, but while “Bad Boys II” features extensive use of comical elements and thus is classified “comedy” as secondary genre, “Shooter” due to less humorous storytelling is labelled “crime” as secondary and “drama” as tertiary genre. An alternative would be to use a shared dummy variable which would allow for multiple genre associations without distinguishing order, but this would also incur a loss of information as suddenly “Shooter” would be as much of a drama as “Schindler’s List”. To avoid these issues instead of binary dummy variables, ordinary variables were used for the genre attribution. As there is a maximum of four genres for each movie, a score of “4” was coded to all movies into for their primary genre into the reflecting variable. Where applicable scores of “3”, “2” and “1” were added in the ordinal dummies for secondary, tertiary, and quarternary genres. This way information loss was minimized while avoiding inadmissible assumptions on transitivity of genres. The disposable genres were: action, adventure, animation, comedy, documentary, drama, horror, music, romance and thriller.

3.4.5 Time of release

As the discussion in the awards section of the literature review revealed the time of release of motion pictures is not arbitrarily based on when the production of the movie is finished, but is instead timed by the producers. Therefore data on the time of the initial wide release of the movie was collected. The date was converted into dummy variables for each month and year, as no reasonable assumption of ordinality can be made in this case, as well as dummies for releases in the weeks before Thanksgiving and Boxing Day. Information on the time of the wide release was collected from Box Office Mojo (2013).

3.4.6 Studio

To control for differences between the production companies, the main associated studio was gathered and coded into dummy variables. A differentiation was made between major and minor studio companies, which are specialized subdivisions of major studios like for instance Pixar which is one of the animation-specialized companies owned by The Walt Disney Company. This results in 16 dummy variables representing the main studios and subdivisions of 20th Century Fox, The Walt Disney Company, DreamWorks, Paramount Pictures, Sony Pictures, Metro-Goldwyn-Mayer, Time Warner, and Universal Studios.

3.4.7 Prequels

As shortly discussed in the literature review the franchise status of a movie has also been used to predict its financial performance. To control for this, a dummy variable was created denoting whether the observed movie had a prequel within the last 10 years. The purpose of the timeframe requirement is to increase the likeliness that the association of the observed movie with its prequel or prequels is not merely one based on similarity of name, but instead on actual knowledge of the prequel. As success and renownedness cannot be assumed to be uniform among the prequels the accumulated box office results of the franchise, was introduced as a scaled measure of prequel renownedness. This measure fails to encompass the concept of cultural familiarity in its entirety as it is described by Hennig-Thurau, Houston & O'Neal (2006), but it does include all of the instances where cultural familiarity stems from recent use of familiar story elements or characters in previous movies. Prequel accumulated box office revenue is scaled in million $, data was retrieved from Box Office Mojo (2013). An overview tabulation comparing prequel and sequel success of movies in the sample can be found in the appendix on p. 117.

3.4.8 Material source

As prequels are only one type of source for a movie script, subsequently a variable encompassing alternative story sources was created and converted into dummies. All movies are categorized into one of 12 source categories, ranging from screen adaptations of books or short stories, on to graphic novels, historic real life events, musicals and theatre plays, fairytales, remakes, television shows, and finally original screenplays written exclusively for the movie. Data was gathered from Box Office Mojo (2013), IMDb (2013b) and The Numbers (2013c).

3.4.9 Overview of movie-related variables

After the incorporation of those elements the constructs relating to the movie itself that were discussed in the literature review are part of the setup. The status quo can be viewed below in figure 3. The following section will now elaborate the cast-related attributes used in the model.

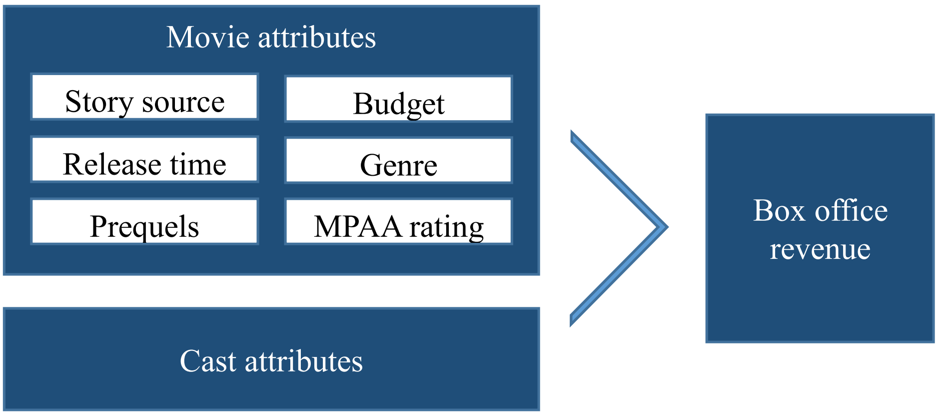


Figure 3: Model including movie attribute dimensions

3.5 Attribution of stars to movies

The challenge was to assign the star actors to these movies: while it is simple to establish the names of all actors playing a role of any sort in a movie, collecting data on all of them is not feasible; the number of actors and the availability of data on them are far too large and poor, respectively. The task then is to figure out which part of the cast entails the stars of the movie. This could be done by choosing the first “X” people from the cast list; this approach however is foiled by the fact that cast is frequently shown in order of appearance in the movie, and not in order of importance. Furthermore determining a generally applicable “X” would prove difficult due to the diverse nature of movie casts: applying the same “X” to “Ocean’s Eleven” as to “Sleuth” would either leave out a large number of stars or include an equally large number of not-at-all-known extras, as “Ocean’s Eleven” features a cast with 8 actors who have each grossed more than $500 million in their previous movies, while there are only 2 actors with speaking roles in “Sleuth”. The apparent option of choosing the actors at will would be arbitrary, thus an alternative materialization of an actor’s criticality to a movie was required and chosen: the placement of the name on the movie poster. To be more exact, the placement of the name outside the credit line (for clarification see figure 4 above), on the standard, one-sheet, A-style poster which usually represents the largest share of exhibited posters. This method of assignment was borrowed from Hennig-Thurau, Houston, & O’Neal (2006). The reasoning behind the overall choice is the combination of several desirable traits: for one the assignment of actors to movies is not arbitrary and can therefore be replicated. Secondly, the number of actors included is adapted to the movie, and thirdly it provides a weighting between the actor’s importance in the movie and his or her level of stardom which has been adjudicated by the people who have created the movie. The downside of the method is, that it leaves some films starless, if their posters do not

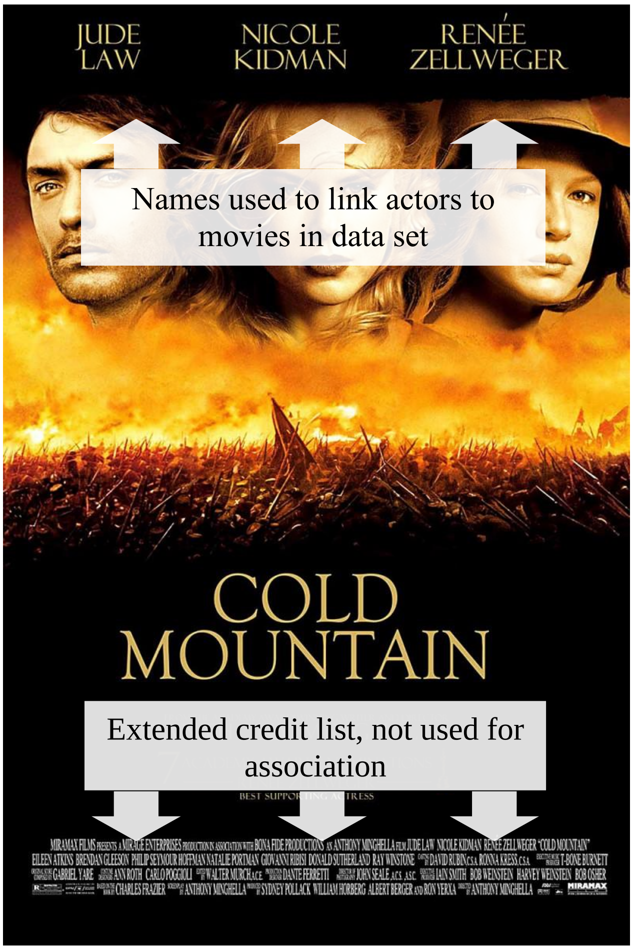


Figure 4: Attribution of stars to movies; poster from IMDb (2013)

advertise a star. This can be due to two reasons: either the absence of notable actors in the movie, or the absence of potentially notable actors' names from the poster.

The question whether the first of the two alternatives would pose an undesirable distortion to the data, is dependent on which of the proposed star-quality concepts is used: if the consumption capital and exchange value concept proposed by Adler (1985) is used, it does not pose a problem: having a star with no previous exposure in the population and thus no amassed consumption capital and no exchange value is effectively the same as having no star at all. Reflecting on Adler's definition that person would not even be considered a star in the first place. If however the talent concept put forward by Rosen (1981) is used, it causes a distortion as the lack of public recognition does not necessarily equate a lack of talent. An example to illustrate this is the 2007 movie "The Kite Runner": while previous exposure to the artists starring in the movie is minimal, critical acclaim and nominations for both Academy Awards and Golden Globes suggest they had the ability to contribute to an overall appealing product, a quality Rosen (1981) would call talent.

The second alternative, in contrast to the first, severely distorts the data regardless of the stardom concept used, as the actors' perceived lack of noteworthiness was not the motive to leave their name off the movie poster. Thus, whether the critical star quality be talent or consumption capital, the information on it would be lost. An example is the 2006 movie "World Trade Center" featuring Nicholas Cage in a leading role. Seeing that Nicholas Cage is an Academy Award winning actor whose movies have grossed close to $2 billion dollars in the U.S. alone, the motives for leaving him off the main poster (appendix p. 133) are likely to be found in the sensitivity of making a commercial movie on the topic of the September 11 attacks at all (Halbfinger, 2006) and not in his lack of manifestations of star quality by either type of definition.

As distinguishing between the two variants of the first alternative, and then yet again offsetting those against the second alternative is a task bound to end in what would be euphemistically referred to as a series of "judgement calls" but is in fact based on nothing but personal opinion, the movies affected by this were excluded from the analysis leaving the total of movies fit for analysis at a final 410 cases of the initial 512. After this, no further exclusions on methodological grounds, for data availability, or in fact any other reason were necessary.

3.6 Star-related variables

After the collection of the movie related variables and the assignment of the key actors to each movie, data on these actors was gathered. This subset of the variables will represent the theories on stardom applied to the stars in movies.

3.6.1 Consumption capital

Calling to mind the concept proposed by Adler (1985), appreciation of a performance increases with knowledge about the performer and the consumers ability to talk with others about her or him and recognize information perceived about her or him and thus feel like an expert. The operationalization then needs to feature past consumption, to represent the consumption capital, as well as a measure of public awareness of the actor, to encompass the exchange and recognition utility. For this purpose four variables were gathered: the first is the accumulated box office revenue from all movies prior to the one that is being investigated, the second is the number of movies the actor has starred in before, the third the accumulated number of cinemas the actor’s films have been shown in over his or her career, used as a proxy for the width and associated advertising power of the actor’s previous films. These three variables were collected using the databases of Box Office Mojo (2013) and The Numbers (2013c). The fourth measure is the number of news articles found by Google News (2013) featuring the actors’ name which were published more than 12 months prior to the release date of the movie in question. The purpose of this window of time is to attain a measure of media presence as close as possible to that at the actual time of release of the movie. However if the date of release was chosen to gather the data, it would be including the media coverage the actor has gained only through their part in the movie itself, thus using the information on media coverage from the movie to explain the success of the movie, an anticipation of information which would violate the requirement of temporal alignment set earlier. At this point it is obvious that these measures are all closely related to the two previously defined explanantia and are also in some ways overlapping each other: an actor whose previous movies sold a large number of tickets is likely to be well-covered in the media. Similarly the amount of previously generated box office, and thus sold tickets, can be expected to correlate to the number of cinemas the previous movies were shown in, and in the same way the number of those previous movies itself. It is therefore fortunate that this blurring of the lines of separation is acceptable as the breadth of phenomena found to be amalgamated here is encompassed in Adler's (1985) theory. While it encompasses not only consumption capital, but also the discussed exchange and recognition aspects proposed by Adler (1985), this variable set and the later resulting factor will henceforth be referred to as consumption capitals for ease of reading.

3.6.2 Talent

Being definitely the more challenging operationalization the concept of talent presents two questions: what is talent, and who is suitable to judge it? Regarding the first question Rosen (1981) fails to provide an explicit answer, and instead simply equates it with the ability to achieve an output of superior quality compared to outputs produced by lesser talent. Thus the problem could be transformed and simplified into judging the quality of outputs. This however still leaves the second question unanswered. One candidate for such a position are the consumers, a suitable manifestation of their judgement could be the rating on consumer-oriented rating sites like IMDb. This however would evoke two problems: the first on a more philosophical level is the age-old dispute over whether people are in fact consistent between their judgement of pleasures and goods on the one side, and their consumption and actions on the other[7]. More importantly however ratings like that of IMDb are notoriously unreliable as they suffer from social influence of previous votes, an effect aptly presented by Salganik, Dodds & Watts (2006) in a simulated market experiment. As an alternative measure which instead of socially influenced consumers features a panel of professional jurors who vote independently, by secret ballot and under the supervision of a neutral advisory company was chosen: the judgement of the Academy of Motion Picture Arts and Sciences. In its efforts to "reward the previous year's greatest cinema achievements as determined by some of the world's most accomplished motion picture artists and professionals" (AMPAS, 2013a) it bestows upon the winning artists an award sponsored by the Academy, most commonly known as the Oscar. While the media spectacle surrounding the ceremony may incite doubt whether the Oscars are in fact rewarding artistic merit or rather commercial success, examples like the defeat of the \$2.7 billion blockbuster "Avatar" to the less than \$50 million grossing "Hurt Locker" in the 82nd Academy Awards suggest that box office performance alone cannot buy an Oscar (Block, 2010). At this point it is important to bear in mind that this measure differs significantly from the research presented in the "Awards" section of the literature review in that it focusses not on

[7] This question originates in the disagreement between utilitarian theories in the works of Jeremy Bentham (1823) and John Stuart Mill (1861) who debate whether it can be said that anything that evokes higher amounts pleasure is inherently of higher quality or whether there are pleasures of higher and lower intrinsic value. Applied to the movie ratings and consumption it could be asked whether "Terminator 2: Judgement Day" is in fact superior in quality to "The Pianist" as the IMDb ratings suggest, or whether consumers are merely rating the pleasure the movie evokes and would answer differently if they were reflecting upon the intrinsic quality of the two movies.

the awards bestowed upon the investigated movie itself, but instead on the previous awards of the actors it features.

There remain however several other approaches to criticize the validity of this measure, which will need to be addressed. For instance, concerns regarding correlation between this talent measure and the consumption capital measures are justified at this point, but will be addressed by the rotation method employed in the factor analysis discussed in a later stage. In this context also the functional continuity of talent and its variable representation needs to be discussed. If talent is static – meaning an actor was born with a given amount of talent and cannot improve or deteriorate – it must either be recorded in a binary variable, as more awards would only be separate manifestations of the constant amount of talent, or it must be corrected for the number of movies the actor has starred in. If talent is completely dynamic – meaning actors can gain and lose talent over time – the explanatory value of previous exhibitions of talent is variable and dependent on latency or retention rate, as the actor may have lost their talent in the period between the last and the current movie. This would render the talent measure useless. For the purpose of this thesis then, in order to have a scaled measure of talent that has the theoretical potential for explanatory value, talent is assumed to be non-strictly monotonously increasing – meaning an actor’s talent remains at least constant, but has the potential to increase.

To alleviate the problem of potentially talented actors having too little chance to achieve manifestations of that talent due to the limited number of nominations and winners, the range was extended to encompass the same data that was collected for the Academy Awards also for the Golden Globe Awards. While obviously judged by a different decision panel than the Academy Awards, it has the same ambition of honouring artistic achievement, and is also awarded in the year after the movies theatrical run. Data on the awards was collected from the AMPAS database (2013) and the database of the Hollywood Foreign Press Association (2013). Data is captured in four variables, representing nominations and wins as leading or supporting actress or actor for each of the awards.

3.6.3 Age and gender

To control for potential moderating effects of actors’ personal attributes, gender and age at the time of release for each actor of the movie was incorporated. As gender and age discrimination are a widespread issue, it is not unthinkable to find them in the motion picture industry. While a pay disadvantage to female or older actors alone would – normative aspects aside – not be relevant for the model as it remains unobservably obscured within the budget variable, it would become so if it was based on the fact that they contribute less value to the movies and thus have a negative impact on revenue. As a pay difference is documented (Rose, 2008) and even a double-jeopardy effect, a combined discrimination effect of age and gender, has been suggested for Hollywood actors (Lincoln & Allen, 2004), age and gender constitute necessary control variables. Data was retrieved from IMDb (2013b).

3.7 Aggregation to final dataset

After collecting the data on the movies on the one hand, and creating the database containing the actors and their related information on the other hand, this data needed to be merged to arrive at a final database which could then be submitted to analysis. For this purpose, those variables relating to the stars of each movies where aggregated in all cases where more than one actor was associated with a movie. A variable counting the number of stars in each movie was added. This will be utilized in ascertaining the validity of hypothesis 5 which covers the fragmentation of the star qualities within a movie's cast. The gender of the individual actors was consolidated into the percentage of actors in the movie who are male, the ages of the actors were condensed into an arithmetic mean. With this task finalized the final setup can be seen below containing the two major classes of independent variables, those related directly to the movie itself and those related to the cast and the actors in it.

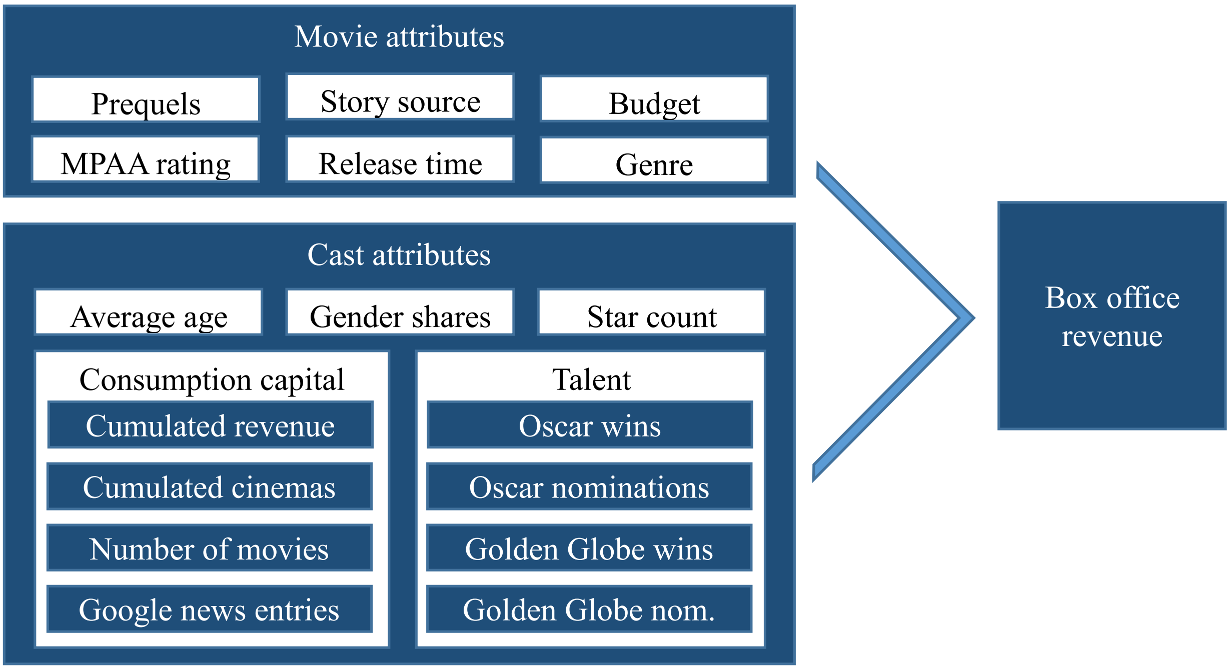


Figure 5: Model with all potential determinants

4 Data analysis

This section will set off with a description of the data in the aggregated dataset. After this the conducted factor analysis will be presented. The resulting factors will be carried on to the core of the quantitative analysis – the regression analysis. The regression model will be narrowed down from a broad initial model to a trimmed final version. Based on this a short description of the results and their implication for the previously stated hypotheses will follow.

4.1 Description of data

Despite the applied criteria the variety of productions in the sample was formidable and evident across all of the scale variables. In terms of box office revenue this variety ranges from just above $200.000 for the 2004 production “Levity” – which on a $7.5 million budget seems to have underperformed – to almost $1 billion for “Dead Man's Chest”, the second instalment of the Pirates of the Caribbean franchise, which however – to put things in perspective – had access to a budget of $225 million dollars to achieve this result. The overall mean for the global box office was $106 million, compared to an average $47 million of production budget, a ratio which using the previously mentioned rule of thumb for movie profits implies an average profit of around 12% on the initial investment.

In terms of star power in those movies, the differences were equally remarkable. They stretch from roughly 28% of movies which featured only one star on their poster, to star-heavy productions like “Love Actually” which featured 10 headline actors. In terms of previous exposure the range started at literally zero for movies like “House of Wax” despite featuring Paris Hilton, because while she scores high in Google News entries, it was her first appearance in a major movie production. On the other end of the scale lies the convention of commercially successful stars gathered for “Ocean’s Twelve” and “Ocean’s Thirteen”. Both movies contain actors whose accumulated previous box office revenues sum up to more than $8 billion in the U.S. alone. But while the two sequels to “Ocean’s Eleven” also score high in Google News entries they have to give way to the combined news power of Tom Cruise, Robert Redford and Meryl Streep – the headlined actors in “Lions for Lambs” who were mentioned in 128100 articles dating back more than 12 months from the movie’s release.

In the talent-oriented variables the big-budget productions however have to make way for the $10 million production “A Prairie Home Companion” starring among others Tommy Lee Jones, Kevin Kline, Lindsay Lohan, Meryl Streep, and John C. Reilly who in recognition of their previous displays of talent have together received 4 Oscars, 7 Golden Globes, and over 50 nominations for the two combined.

A tabulation of gender (see figure 6) reveals a remarkable imbalance between the number of female and male stars, as women account for only 36% of the total number of star actors. In line with this 60% of movies feature more male than female stars. More striking is the comparison between the number of productions featuring only females on the poster to those featuring only males: there are three times as many movies with an all-male cast than there are with an exclusively female cast (appendix p. 120). Plotted against the frequency of age brackets these gender differences become even more pronounced as the figure above shows. Despite being outnumbered by men almost by a 2:1 margin, women actually make up 60% of actors below the age of 30. In contrast to this only 20% of those above the age of 50 are female.

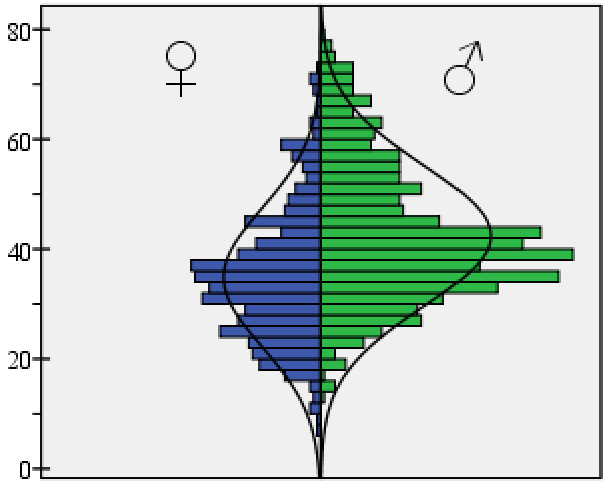


Figure 6: Population pyramid of star actors

Overall then the data show a sufficient variance across the variables to hope for both useful and interesting revelations. A full overview of the variables' descriptives and frequency statistics can be found on appendix p. 85 et seq.

4.2 Factor analysis

Based on the problem of multicollinearity and the discussed suitability of the method (Jolliffe, 2002), a factor analysis was executed[8]. The variables entered were all 8 variables gathered regarding the 2 star measures based on Rosen's and Adler's theories. The only parameter requirement was that the eigenvalue of the matrix was greater than 1, thus allowing the number of factors to be extracted and the assignment of the items to these factors to be determined freely. The result was rotated using Kaiser's Varimax method (Kaiser, 1958). While the assumed, and thus resulting orthogonality of this method is inferior in applicability to oblique

[8] While the assumption of the existence of a casual model suggests an exploratory factor analysis, the factual method used was a principal component analysis. The recent differentiation between the two formerly synonymous methods however does not apply to this case, because no cases of low inter-item correlations, a prerequisite for differing results between the two methods (Fabrigar, Wegener, MacCallum, & Strahan, 1999), can be observed.

alternatives like Promax or Oblimin (Browne, 2001) it promises an easier and more reliable interpretation (Abdi, 2003).

To ensure that the data is suitable for the method, tests for homoscedasticity and sampling adequacy were conducted. The null hypothesis of Bartlett's test of sphericity which alleges homoscedasticity is rejected at $P_{H0\ Bartlett} = 0.000$ significance level, thus indicating that the data is potentially suitable for the factor analysis (Dziuban & Shirkey, 1974) and should be further checked for sampling adequacy. As standard measure of sampling adequacy the Kaiser-Meyer-Olkin criterion was calculated, which at $MSA_{KMO} = 0{,}872$ is signalling more than sufficient sampling adequacy, scoring between "meritorious" and "marvellous" on Kaiser & Rice's (1974) scale. Finally a manual inspection of the anti-image covariance matrix confirms the suitability for this method (Dziuban & Shirkey, 1974) by producing only 2 out of 28 items which are not zero to the first decimal place. The cumulative explained variance is 86%, the distribution of variance explanation between the factors almost equal in the rotated solution (matrix in appendix on p. 96). While the resulting unrotated component matrix shows one general factor, as would be expected at this level of inter-item correlation, the rotated component matrix produces two separated factors. At this point the high inter-item correlations however act to the disadvantage of the analysis, because they dampen the discriminatory power in the component matrix as two of the items, Google news entries and Golden Globe nominations, have loadings' ratios between the two factors of below 2:1. A closer inspection of the correlation matrix shows however that the intra-factor correlations are distinctly higher than the inter-item correlations across the factor borders. The fact that despite the diversity of variables in the consumption capital section both factors are formed just as the theoretical foundation would predict it, speaks for theoretical consistency in the variable choices.

The statistical internal consistency of the resulting factors was checked using Cronbach's Alpha as a measure of reliability. While the absolute variance across variables in the talent factor is homogenous, this is not the case for the consumption capital factor due to the variety of phenomena they measure in different units. The ratio between mean and variance however is homogenous, thus prompting the use of Cronbach's standardized Alpha statistic. Using this statistic the scores for the factors are $\alpha = 0{,}925$ for talent and $\alpha = 0{,}949$ for consumption capital (full statistics in appendix on p. 96 et seq.), denoting "excellent" internal consistency for both factors (George & Mallery, 2005).

The resulting factor scores were computed and added to the database using Bartlett scores, as they are based on the maximum likelihood estimates which minimizes procedural bias (Distefano & Mîndrila, 2009). Per definition they are also uncorrelated[9] and standardized to z-values, thus have an arithmetic mean of 0 and a standard deviation of 1.

This creates an issue when creating the quadratic terms used for controlling hypotheses H3 and H4 which stipulate increasing marginal effects. To test for this a quadratic term will be added as a predictor which – provided its coefficient is significant – will provide information on the nature and orientation of a potential non-linear relationship. The simple issue in connection with the factor scores is that a squared factor would for instance lose discrimination between positive and negative scores. This was addressed by defining $z^2 = z * |z|$ as an operation that does not affect the algebraic sign of the z-scores. The complex issue is that it must be noted that the results of this transformation suffer from a parabolic distortion[10]. However the alternative of instead replicating the factors using squared versions of the original variables was tried, only to realise that it does not create factor scores compatible to the unsquared versions of the factors, and thus does not allow for interpretation of their coefficients.

At this point it should also be noted that the use of the factor analysis is the reason for the absence of an otherwise probably advisable weighting between for instance wins and nominations for awards, as the transformation would have eliminated any weighting effect. The yield of this analysis consists of the factors it produces, the understanding of which is elementary to the results presented later. For a full image the component matrices should be consulted, but the essence is this: the analyses searches for manifestations of underlying common constructs in the data. In this case it finds two such concepts which are associated with four variables each, the one with the four awards-related variables, the other with the four consumption capital variables. By rotating them a maximal discriminatory power is achieved without altering the data. Then the scores are computed using the component matrix as commandment. The scores are transformed to mean = 0 and standard deviation = 1 by subtracting the observed mean and dividing by the observed standard deviation, one unit in the factor scores represents one standard deviation of the construct variable. To get a better picture of which value combinations result in a particular score, examples can be examined in the following table:

---

9 The infinitesimal correlations that can be observed are due to the fact that the factor analysis was conducted including the movies under the $5 million budget cut-off criterion to fully utilize all available data.

10 Because using the transformation function "Z" (Kreyszig, 2006) on the empirical value "a" it is clear that even for positive values $[Z(a)]^2 \neq Z(a^2)$ and therefore $Z[Z(a) * |Z(a)|] \neq Z(a^2)$ unless $a = \mu$ and thus $Z(a) = 0$. Using a different factor score computation method would not have addressed the problem as both eligible alternatives, the Anderson-Rubin method and the Regression method are transformed (Field, 2000).

Movie	Google news hits	Prior box office $	Cumulated cinemas	Previous movies #	Consumption capital score
Mona Lisa Smile	24567	3342508269	87739	72	1,01009
The Lake House	16010	2850565071	100320	61	1,00282
Bringing Down the House	30520	1634295573	51893	44	1,00033
…					
The Rundown	8585	2476315201	98844	78	0,03026
Pirates of the Caribbean I	9796	2550524968	52642	44	0,01459
Tim Burton's Corpse Bride	14750	1918188562	46801	46	-0,00225
…					
25th Hour	3190	388393323	18865	11	-1,00812
A Man Apart	1050	594270398	14755	7	-1,01218
Underworld: Evolution	2735	572051505	7868	12	-1,01729

Movie	Golden Globes		Academy Awards		Talent score
	Wins	Nominations	Wins	Nominations	
Ocean's Twelve	7	16	3	6	1,11712
Open Range	5	10	1	9	1,06091
American Gangster	3	7	3	5	0,92348
…					
Solaris	5	2	1	3	0,02681
A Good Year	5	1	1	4	0,00637
Failure To Launch	8	4	0	4	0,00310
…					
Smokin' Aces	1	7	0	1	-0,97490
Twisted	0	8	0	2	-0,98129
Click	0	2	1	1	-1,02798

Table 2: Examples for variable combinations and resulting factor scores

4.3 Regression analysis

4.3.1 Assumptions of linear regression

In order for the following regression model to be interpreted, it has to be ascertained whether, and if not in what way, the model fulfils the methodological requirements for linear regression models. For this purpose the absence of interrelations between the independent variables was checked, along with their independence from the residuals of the regression. The residuals were further checked for autocorrelation, accuracy of their expected value, homogeneity of variance, and normal distribution. While partly dependent on the outputs of the following regression itself, this section provides more value situated before the regression.

The absence of connections between the independent variables was controlled by checking the correlation matrix. Within the scaled variables the only significant and critical correlation was that between the consumption capital factor and the number of stars in the movie (appendix, p. 100), a finding which is not entirely unexpected. While the correlation is slightly above .40 the impact was judged to not force an exclusion of this conceptually valuable information based on an analysis of the variance inflation factors (VIF) and derived tolerances which remain at $VIF_{Consumption\ Capital\ Factor}$ = 1,960 and $VIF_{Number\ of\ Stars}$ = 1,798 and thus well below even the stricter proposed limits of four and above for these indicators (Craney & Surles, 2002; O'Brien, 2007). As these reported variance inflation factors are calculated for the regression model featuring the whole battery of variables – which inflates the VIF-scores – and are even lower for the trimmed model no further action was taken. At this point it must be noted that the relative absence of multicollinearity-related issues would not be possible without the use of the factor analysis, at least not without the omission of a large share of the variables contained in its factors.

The absence of interdependence between the residuals from the regression and its independent variables is given for this dataset as a correlation matrix including the standardized regression residuals shows (appendix, p. 105), thus provides no indication of distortion of estimators. Equally the expected value of the residuals E(Res) = 0, thus showing no evidence of a systemic error or unobserved effect in the residuals. The uniformity of variance throughout the residuals proved to be at an acceptable level showing no pronounced heteroscedasticity. Due to the nature of the variables chosen, the timeframe, and the intense consumer interest in the analysed industry, there were also no cases of missing data.

The main problematic area instead was found to be the distribution of the residuals. As the stem-and-leaf plot visually suggests and both Kolmogorov-Smirnov and Shapiro-Wilk statistics confirm (appendix p. 106), the distribution is leptokurtic meaning the absolute values of the residuals of the regression are on average smaller than expected. The Q-Q plot of the standardized residuals (appendix p. 107) shows a slight S-shaped distortion with 10 pronounced outlier cases, 2 of which are box office failures while the rest are movies that exceeded their expected revenue by a visually noticeable margin. While this means the model performs better at spotting box office bombs than blockbusters, and may appear more desirable than platykurtosis, it still demands for a thorough look at those outliers to assess their criticality to the models significance test; this will follow later.

A correlation of the residuals with themselves was not observed as the Durbin-Watson statistic was computed at d_R = 1,828 which is close to d_R = 2 which is the value denoting no autocorrelation on the continuum between perfect positive and perfect negative autocorrelation at d_R = 0 and d_R = 4, respectively (Bhargava, Franzini, & Narendranathan, 1982). The existing deviation from the d_R = 2 does not pose evidence to suggest the presence of autocorrelation as it lies above the lower bound of $E(d_R)$ discussed by Durbin & Watson themselves (1950, p. 427). Also, while a non-linear function was only hypothesized for talent and consumption capital the other scaled variables were checked as well and did not show any significant coefficients or cases where a significant R^2 was greater with a non-linear interpretation than with a linear one.

4.3.2 Trimming towards final model

In order to fully utilize the breadth of available variables, yet in the end arrive at a model that only has a necessary level of complexity and is not burdened by an overload of variables, the number of variables was reduced from an initial model encompassing all available data to a final, trimmed version containing variables identified as core predictors. The first model contained a total of 70 variables. The relatively large number can be traced back to the highly prevalent necessity of dummy variables to include potential non-ordinal and non-scale connections, as there are 12 variables representing the material source, 16 variables representing the studio, 9 variables denoting the genre[11], 5 variables describing the MPAA-rating, and 19 variables deduced from the release date. The outputs of an initial regression model can be found in the appendix on p. 103.

A first look at this model's statistics shows that it possesses an adjusted determination coefficient of $R^2 = 0{,}639$ denoting that it manages to explain close to two thirds of the variance in the dependent variable. The entire model is highly significant, of the entered 70 variables 14 possess coefficients significant at 0,05-level or better. It is however, due to its large number of predictors in no way parsimonious and must be expected to carry a significant amount of feckless baggage. In order to on the one hand preserve this explanatory value of the model but on the other hand reduce its complexity and allow for easy interpretation, the forward regression method was chosen based on its ability to provide an idea of explanatory contribution (Field, 2000) and its suitability to identify relevant predictors in regression models with large ratios between predictors and observations (Wang, 2009). The downside of this method is that compared to the backward-stepwise method, or the simple stepwise method, it runs a higher risk of type II errors, thus leaving out actually influential predictors due to suppression effects (Field, 2009). This is in part owed to the fact that the method is essentially a greedy algorithm (Hastie, Tibshirani, & Friedman, 2009), meaning it is susceptible to premature satisfaction with a local optimum despite the existence of a superior global alternative (Cormen, Leiserson, Rivest, & Stein, 2009). As these are justified concerns threatening the validity of the model, a

[11] Due to excessive VIF-scores for the ordinal scaled genre variables, their use was discontinued and they were replaced by nominal scaled variables for the main genre, based on the conviction that a less discerning but reliable coefficient is superior to a nuanced, but unreliable one.

control mechanism of the method was introduced by utilizing the least absolute shrinkage and selection operator, or “lasso”, as an alternative selection method.

In this course, the predictors selected by the forward method using a $P(F) \leq 0{,}05$ inclusion requirement (results in the appendix on p. 108 et seq.) were compared to those selected by the lasso (appendix, p. 111). The model based on the forward algorithm was found in almost identical composition in the sequence of model compositions generated by the least absolute shrinkage operator. The lasso-generated model in question is, in terms of predictor inclusion, situated in between the optimal and the most parsimonious model proposed by lasso. In their identification of the three best predictors the two methods agree entirely, from there onwards a considerable consonance can be observed:

Forward method[a,b]	**Lasso**[a]	**Final model**[c]
Production budget	Production budget	Production budget
Consumption capital	Consumption capital	Consumption capital
Prequels' box office	Prequels' box office	Prequels' box office
Animation Genre	Talent	Talent
Based on historic events	September release	Average age of stars
September release	Talent squared	Number of stars
Number of stars	Percentage of stars male	Animation Genre
Talent	Animation Genre	Based on historic events
Average age of stars	Consumption capital squared	September release
Consumption capital squared	Number of stars	Consumption capital squared
December release	Based on historic events	Percentage of stars male
Drama genre	Average age of stars	Talent squared
July release	Drama Genre	
MPAA R-rated	Based on graphic novel	
Percentage of stars male	April release	
…	…	

[a] In order of inclusion as determined by the respective method
[b] Predictors in italics are not included by forward method; their order was determined by relaxing F-value requirements
[c] In order of absolute value of standardized beta coefficients

Figure 7: Predictor compositions for forward, lasso, and the final model

As figure 7 shows, the ten predictors included by the forward method are completely included in the first twelve predictors to be included by the lasso. The two additional predictors which are – according to the lasso method – missing in the forward model are the percentage of stars who are male, and the squared version of the talent factor. Based on this extensive congruence the two predictors suggested by the lasso were added to arrive at the final model, which was then executed without further selection algorithm.

4.3.3 Final regression model

After this arduous process – first ensuring the applicability of the used models, then applying the factor analysis, and finally reducing the number of predictors to a reasonable, yet significantly meaningful subgroup – at last a final model is ready for interpretation. The regression equation along with all outputs can be found in appendix from p. 112 onwards. The model is significant at 0,001 level. The determination coefficient indicating goodness of fit is at R_{adj}^2 = 0,64 which means the model has not lost explanatory power compared to the earlier version containing all available variables, but has in fact an increased R_{adj}^2 due to a smaller penalty adjustment for the number of predictors. The coefficient overview of the model is displayed below, visually separated into two main groups – those relating to attributes of the movie itself, and those containing attributes of the movie's cast.

Dependent Variable: film_box_global_inflation_adjusted	Unstandardized Coefficients		Standardized Coefficients	t	Sig
	B	Std. Error	Beta		
(Constant)	110,216	21,642		5,093	,000
film_budget_inflation_adjusted	2,043	,111	,613	18,404	,000
film_prequels_box_office	,100	,017	,188	5,912	,000
film_genre_animation	70,965	21,965	,097	3,231	,001
film_based_historic_events	-75,078	25,242	-,089	-2,974	,003
film_month_september	-36,110	12,533	-,086	-2,881	,004
star_consumption_cap_factor	38,957	4,769	,342	8,170	,000
star_consumption_cap_factor_squared	-1,001	,476	-,075	-2,100	,036
star_talent_factor	20,299	6,647	,168	3,054	,002
star_talent_factor_squared	-,122	,244	-,026	-,501	,617
star_average_age_at_release	-2,197	,487	-,162	-4,511	,000
star_number_of_stars	-9,055	2,506	-,135	-3,614	,000
star_percentage_of_stars_male	,217	,127	,057	1,709	,088

Table 3: Coefficients and significance levels of the final model

Of the coefficients all but two – the gender distribution within the cast and the squared version of the talent factor – produce t-values significant at the 0,05 level, the majority is even significant at 0,005 level or better. The final model coefficients were submitted to resampling, more particularly through bootstrapping which uses Monte Carlo generated samples to provide reliable standard errors and confidence intervals on a coefficient basis (Efron & Gong, 1983). The resulting bootstrap coefficients differ from those in the above depicted model only in suggesting that the positive effect of a higher percentage of males in the cast is in fact significant at 0,05 level.

Collating the β-coefficients which measure the impact in the comparable unit of standard deviations and also recalling to mind the R^2 contributions from the forward method regression, the most influential predictor overall is the budget. Within the subsection of movie-related variables the box office revenue earned by prequels turns out to improve a movie’s performance, the predicted benefit being an additional $1 million in box office for every $10 million in prequel box office. Furthermore animation movies are found to generate higher revenues, while movies based on historic events, and movies released in September have negative coefficients signalling a lower predicted revenue than movies which do not have those features.

In the group of star-related variables the largest influence is computed for the consumption capital of the cast. However in this context the linear and the quadratic term need to be interpreted jointly. Blinding out the other variables and the constant this combined function determining the box office caused by the consumption capital factor is:

$$f_{Box\ Office}\left(b_{Consumption\ Capital}\right) = 38{,}957 * b - 1{,}001 * |b| * b$$

Which means the vertex of the function at which F(b) is maximal and F(b)’= 0 is at

$$f_{Box\ Office}\left(b_{Consumption\ Capital}\right)' = 0 = 38{,}957 - 2 * 1{,}001 * |b|$$

This puts the vertex at a value of b = 19,459 for the consumption capital factor. As this value is many standard deviations above the maximum of the range for the variable it does in fact only mathematically denote an inversely U-shaped connection; in reality the predicate is that the marginal impact of the consumption capital factor is positive throughout the entire range of values which extends from -2 to few outliers close to +4, but is decreasing in size (for an illustration see appendix, p. 116). An example for the degree of that decrease can be made by comparing the effect of +1 standard deviation[12] added to the mean of 0 for the factor score, and to a score of already +4 standard deviations above that mean. The comparison shows that:

$$\frac{f_{Box\ Office}(1) - f_{Box\ Office}(0)}{f_{Box\ Office}(5) - f_{Box\ Office}(4)} = 1{,}267$$

This is indicating that the effect of 1 additional standard deviation of consumption capital in the cast is 26,7% greater when it is added onto a movie whose cast has the average amount of consumption capital, than when it is added to a movie that already has a score +4 standard deviations above the mean. The decreasing marginal effect then is not yet particularly prominent in the part of the curve.

For the talent factor in contrast only the linear component is significant meaning no change in marginal effect is observed, the linear effect itself is positive but smaller in terms of standardized coefficients than that of its consumption capital counterpart.

Further on in the star section of predictors, the average age of the cast has a negative impact on the predicted box office, a single year more in average cast age “costing” the movie as much as $2.2 million according to the model. Similarly the number of stars within the cast decreases the predicted revenue. If the coefficient of the percentage of male stars is in fact significant, as the bootstrap results suggest, it would mean that movies box office revenues benefit from an increasing share of male actors within the cast.

In terms of completeness and parsimony the final model incorporates a set of variables that, except for the quadratic term of talent and the percentage of male stars, were found to have significant R^2 contribution by the forward model and whose overall contribution was confirmed by the lasso. Together these variables – the control variables derived from the literature and the variables shaped to mirror the different aspects of stardom as derived from theory – provide sufficient basis to assess the veracity of the initial hypotheses.

4.3.4 Implication for hypotheses

Using this final model for its premiere purpose, the test of the hypotheses stipulated in the beginning of this thesis, the following verdicts can be rendered:

[12] The coefficients used are unstandardized, but as the variable itself is standardized, 1 unit is in fact 1 standard deviation.

H1: Hypothesis is accepted.

Increasing talent in the cast does increase the box office revenue of a movie.

The coefficient of cast talent is significant at the 0,005 level, and shows a positive connection between the amount of talent in the cast and the box office revenue generated by the movie the cast starred in. The variable also provides significant R^2 contribution.

H2: Hypothesis is accepted.

Increasing aggregate consumption capital and exchange utility in the cast do increase the box office revenue of a movie.

The coefficient of consumption capital is significant at the 0,005 level, and shows a positive connection between the amount of consumption capital contributed by the cast and the box office revenue generated by the movie the cast starred in. Like the coefficient of talent, also this predictor also provides significant R^2 contribution.

H3: Hypothesis is rejected.

No evidence of increasing marginal impact of talent in the cast on box office revenue is found.

The coefficient of the squared talent term is not significant by any standard and thus does not give any indication of an increasing marginal effect.

H4: Hypothesis is rejected.

No evidence of increasing marginal impact of consumption capital in the cast on box office revenue is found.

In fact a decreasing marginal effect of the consumption capital factor is found as the squared term is significant at 0,05 level and has a negative coefficient.

H5: Hypothesis is accepted.

The fewer stars the star qualities in the cast are concentrated on, the greater the box office revenue.

The number of stars the cast consists of has a negative coefficient which is significant at 0,001 level, thus a greater number of stars decreases the predicted box office revenue.

4.3.5 Outlier analysis

As stated earlier, the irregularity within the distribution of residuals prompts an investigation of outlier cases producing extraordinary residuals. In light of the frequently stated unpredictability of the box office (De'Vany & Walls, 1999) the purpose of this is to check for communalities among the extreme cases which have not been covered by the set of variables applied to the data in the analysis and check for extreme influence of single cases on the entire model. An exclusion of extreme cases seems inappropriate as no indication of flawed data was found, but instead only a failure of the model to explain the data.

For an overview of outlier cases all cases with large standardized residuals were extracted in table 4. Despite an expected value of 0 an asymmetry of large residuals was observed as suggested by the Q-Q plot discussed earlier. While there are 8 cases of standardized residuals greater than +3,00 the extreme value for negative residuals was -2,93 with only 3 cases having a residual of -2,00 or more.

Movie title	Predicted box office	Actual box office
Pirates of the Caribbean: Dead Man's Chest	$493m	$972m
Pirates of the Caribbean: The Curse of the Pearl	$305m	$654m
Movie 300	$80m	$404m
Ice Age: The Meltdown	$280m	$598m
Bruce Almighty	$183m	$484m
Transformers	$342m	$629m
War of the Worlds	$292m	$557m
I Am Legend	$296m	$519m
A Scanner Darkly	$165m	$7m
Stealth	$285m	$72m
Evan Almighty	$370m	$153m

Table 4: List of cases with large residuals

The largest residuals of all are commanded by the first two instalments of the "Pirates of the Caribbean" franchise which both earned roughly twice their predicted revenue. Being the second highest grossing movie franchise on a per-movie basis with an average of $925 million box office per movie (Box Office Mojo, 2013c), there is little use in post-rationalisation and instead these cases should be taken as the phenomena they are.

At the other end of the scale lies "Evan Almighty", a particularly interesting case seeing it is the sequel to "Bruce Almighty" which is in 5th place of the movies with the largest positive residuals. Even its predicted box office revenue would not have made it highly profitable on a budget of $175 million, which at the time made it the most expensive comedic movie ever made

(O'Neill, 2007), suggesting it was already before its release in a way untypical. Then, the pre- and postproduction of the film itself were rushed by the studio which had planned the movie a December release (Munoz, 2006). Furthermore, while it is a case of a movie based on a highly successful prequel, it features a different leading cast, a characteristic which is uncommon among franchise movies. Altogether then – and obviously in hindsight – the movie contains a group of unfavourable attributes which turned out to its financial disadvantage.

Another notable case seems to be "A Scanner Darkly" which despite its budget of just over $8 million was predicted to generate revenues of $165 million due mostly to its cast which featured Keanu Reeves, Robert Downey Jr., Woody Harrelson and Winona Rider, and the fact that it is an animated movie. But while stars are common for animated movies, they are usually not tracing paper versions of themselves (for an illustration see appendix p. 133). In fact within the sample the rotoscoping technique used is unique to this movie and is generally considered experimental (Roberts, 2009). On the other side of the verge lies the 2007 production "300" which, despite depicting undistorted images from traditional digital filming, was shot almost entirely with chroma-key technology which means the actors are real and shown without extensive distortion but the entire settings are computer generated and added through green screen overlaying. It represents the type of model failure opposite to "A Scanner Darkly" which was predicted to be highly successful and was in fact a losing investment; "300" was predicted to be a financially unviable production and turned out to be highly profitable. In this it differs from other high residual productions such as "Ice Age: The Meltdown" or "Transformers" which were both predicted to earn more than twice their budget, but far exceeded that expectation. Overall the observed prevalence of animated or highly CGI-intensive productions in this list as well as the different realizations of the technology point to an insufficiently detailed variable setup. Here a more differentiated coverage of the various attributes these movies have the capacity of possessing seems to be required in order to adequately decompose reality's complexity.

On a different note "A Scanner Darkly" also breaks with the pattern above average budgets common to the rest of the movies on this list. This connection seems reasonable given the influence of budget and the fact that the residuals are measured in standard deviations of the dependent variable. It may however lead to believe that the "true" residuals lie with cases not picked up by this measurement. While there certainly are cases where this is true, the model is by no means blind to smaller underdog productions such as the highly acclaimed "Little Miss Sunshine": while its $100 million box office revenue on a budget of $8 million was probably a positive surprise to its producers, its predicted revenue from the regression was $64 million, the residual thus only $36 million or roughly one third of a standard deviation of the predicted box office, meaning the model accurately predicted a highly successful movie, it only failed to concisely determine the size of the success.

Looking for extreme cases from a perspective of model independence from singular cases there are only four noteworthy cases in the entire sample when using Cook's distance as an indicator. Three of them are "Ice Age: The Meltdown", "Transformers" and "Pirates of the Caribbean: Dead Man's Chest" which produce distance statistics above a conservative control value (Fox, 1991) of $D = 4/(n - k - 1)$ but below the critical absolute value (Smith, 2005) of $D = 1$. The only movie to surpass this value at $D = 1{,}82$ is "Casino Royale" which has a relatively small negative residual of $117 million at a predicted value of $664 million. The reason for its Cook's distance lies not in its residual but in its uniqueness in regards to prequel history. The fact that the movie produces a negative residual in combination with having the strongest prequel history indicates that this case has a strong negative effect on the coefficient of the prequel box office. This hypothesis was checked by rerunning the final regression excluding "Casino Royale" from the analysis; the resulting coefficients and their significances are barely changed, except for that of prequel box office which increases by 80% from B=0,100 to B=0,180 thus confirming the previously stated hypothesis. As consequence of this observation the coefficient of prequel box office revenue should be viewed with reservation.

With this and based on the absence of further cases of large residuals or undue model influence, the outlier analysis – and in fact the entire analysis section – is concluded and a discussion of the results can follow.

5 Discussion

This section will feature an in-depth discussion of the results of the analysis issue by issue, integrating or where applicable contrasting them with the literature. Furthermore a note on modes of application for the findings will be supplied, along with a critical review of their limitations.

5.1 Findings

As the previous reservations caused by the irregularity of residuals subside, the essence of the analysis is that using an underlying set of only 16 variables[13] from an early stage of a movie project's development – 5 relating to movie attributes and 11 relating to the cast – around two thirds of the variance in that movie's box office revenue can be explained well in advance. Some frequently confirmed attributes like the MPAA rating and most of the genres as all of the studio variables do not show explanatory value. On the other hand regarding stardom, the gauges of both talent and consumption capital within the cast positively influence the box office. The results of the analysis including those measurements show that the concept of stardom which underlies the effects of stars is a lot more delicate, and apparently requires a consideration of this delicacy in order to have its effects revealed. This consideration is what the two factors are aimed at providing. The impact of those factors on the box office can be viewed in the dimensions magnitude and shape. In magnitude the consumption capital factor's impact is greater as denoted by its standardized coefficient. It, followed by all other concepts found to contribute to the explanation of the dependent variable, will now be discussed in detail.

5.1.1 Consumption capital

In the final model its coefficient denotes an additional $39 million in box office revenue for an additional standard deviation of consumption capital, thus indicating that the consumption capital associated with Nicolas Cage at the time of the release of "National Treasure: Book of Secrets" adds $39 million to the box office revenue of the movie when comparing it to a movie with the same properties that features Keanu Reeves who is associated with one standard deviation less than Cage. Accordingly an actor 2 standard deviations below Nicolas Cage like Viggo Mortensen will take in $78 million less. Again, for examples which combinations of

[13] There are 12 coefficients in the regression, 4 of which rely on the 2 factors, talent and consumption capital, which in turn are based on 4 variables each.

values evoke such a score, consult appendix p. 119. At this point it is important to recall to mind that the consumption capital scores are not static as they are dependent on the movies the cast stars in and on each cast member's media coverage and thus change over time. Thus the examples presented here are within a small margin of error as they are in this case from a range of one year. Furthermore the example chosen is limited to movies with only one star actor in them while the majority of movies in the sample has several.

More importantly however, it also disregards the decreasing marginal impact of consumption capital which leads the discussion to the second dimension of impact, the shape. As described in the analysis the significance of the quadratic term suggests a decreasing marginal effect of consumption capital. As an example the cast's consumption capital contribution to revenue in one of the sample's movies will be compared to a fictional alternative version with one additional actor in its cast: the movie is "Ocean's Thirteen", the fictional cast addition Eddie Murphy. The combined cast of "Ocean's Thirteen" provides the movie with a consumption capital factor score of 3,70 suggesting that the consumption capital contribution to the box office is $3{,}70 * 39 - 3{,}70^2 * 1 = \$130$ million. Adding Eddie Murphy's individual attributes in the consumption capital dimension from his engagement in "Norbit" into the original cast's attributes propels the movie by half a standard deviation to a total score of 4,20 and thus to a consumption capital contribution to box office revenue of $4{,}20 * 39 - 4{,}20^2 * 1 = \$146$ million. Thus the original \$19,5 million of revenue differential have been reduced to \$16 million by the decreasing marginal impact. If all other parameters were to remain constant this would be the expected gain from the cast addition. This is arguably unlikely as the cast addition will expect to be paid. But leaving the perspective of revenue and turning once to profit, the additional cost according to the model is neutral as the coefficient of budget at B_{budget} = 2,043 combined with the earlier discussed 50% approximation shows that any profit will not come from the additional spending itself, but from the attributes that the spending buys. While this appears plausible, it also implies that it is impossible to overspend as the impact will always remain positive. While this is not actually accurate as for the movie to become profitable overall it would also need to cover its marketing budget, it is demonstrating the inadequacies that arise from the insufficient differentiation of the budget variable: a split into cast and non-cast expenses would be needed to allow for a true analysis of sense and sensibleness of single actor engagements. However, it also shows how despite revenue being the dependent variable the coefficients still can say something about profit. Overall, this seeming paradox should convey that the use of comparison between real and fictional cast compositions is for the purpose of coefficient illustration and is not a suitable mode of application for any of the findings as they predict consequences of independent decisions, an issue discussed later.

Putting this finding into perspective with results of the research presented in the literature review is challenging as there is little in terms of a common tenor. When comparing the impact to Hennig-Thurau et al. (2006) who find insignificant direct impact and significantly negative total impact of star power on box office revenue, especially the ratio between some of the standardized coefficients attracts attention: while they find the impact of star power to be not only negative, but also below 1:10 in magnitude comparing it to the standardized coefficient of budget, the model presented here suggests an impact that is for one positive and at above 1:2 ratio of standardized coefficients also more influential. A source of this divergence of results in this particular area is – considering the similarity of results in other dimensions – likely to stem from a different star quality operationalization, in this case newspaper star ratings. Moving to Elberse's (2007), the nature of analysis does not provide any comparable coefficient. However, a comparison of the top 10 largest cumulative abnormal returns shows that of those actors whose engagement announcement increased the movie's expected revenue the most, 6 are to be found among the top 20 in the list of all 597 actors ranked by an individual consumption capital factor (appendix p. 126). This suggests that at least some of the aspects of stardom Elberse (2007) marked out for further investigation using the theory of Adler (1985) are embedded in the consumption capital factor. Despite an entirely different method from both Elberse (2007) and this thesis, De'Vany & Walls (1999) also arrive at a list of 17 actors who significantly impact the hit probability of a movie. Among these, 14 of 17 have appeared in movies in the sample of this thesis, 11 of which are among the top 40 actors in the consumption capital ranking. Thus, while both De'Vany & Walls (1999) and Elberse (2007) have arrived at these lists by employing an approach on an actor-by-actor basis, their results show substantial overlap with a variable that is not firmly fixed to only one actor but shows general and measurable attributes of actors which evoke the effect or at least consistently coincide with it.

5.1.2 Talent

In comparison to consumption capital the talent factor impact is less complex as it does not follow a function with changing marginal utility. Instead it only has a linear coefficient indicating a $20 million positive effect on box office revenue for one standard deviation of previous talent manifestations in its cast. Although this based on the range of the scores indicates a theoretical spectrum of talent-evoked contribution to revenue of around $150 million, the majority of effects is by definition in the ± $20 million scope. Considering that one unit of the talent score consists of more than one award, the gain per award is decidedly smaller than those observed by Dodds & Holbrook (1988) and Nelson et al. (2001) who analysed gains from awards bestowed upon the analysed movie itself and not awards earned by the movie's actors in previous roles. This comes as no surprise as the monetary value determined this way is evoked by a combination of signalling effects and publicity of the award, but also arguably has a more direct force of expression regarding the expectable quality. But because at the time the nominations and later winners are announced the movie has already been in cinemas for months, from a movie lifecycle perspective the revenue generated by the awards announcement is an abnormal return linked to the event recognizing its quality. While a signalling effect of awards previously earned by actors can by no means be excluded, as distributors often even advertise academy award winning actors in an upcoming movie, at least the talent score does not eliminate the effects of the underlying talent's ability to produce performances of quality allowing the actors to receive those awards.

On a different note, in order to gain a better understanding of the internal score contribution the elements of the talent score provide, a look at the descriptives of the four award-related variables shows prevalence ratios that order the contribution to the factor score per additional unit from high to low starting with an Academy Award win, an Academy Award nomination, on to a Golden Globe win, and finally a Golden Globe nomination. This exclusively reflects the scarcity of those attributes in the sample – there are for instance only 40% of movies with an Academy Award winning actor but 80% which feature an actor who was at least nominated for a Golden Globe – and does not judge their impact on the dependent variable. However it contains the valuable information that one Academy Award win is increasing the talent score, which was found to be a determinant of the box office, by a far larger amount than a nomination for a Golden Globe.

A further aspect to the talent factor is the fact that it shows its effect only over the entire lifetime of a movie, a finding that was made by substituting the dependent variable from the lifetime box office to that of only the opening weekend (appendix p. 115). Here most of the other coefficients' significance remains unchanged, while the talent factor ceases to provide significant results. Overall this suggests a different timeframe for the effects of the two star qualities, where talent is only exerting its effect in the long run.

Another interesting observation is that the elite of movies in terms of the talent score of its cast not only profits in revenue from that talent, but also in profitability: comparing the means of the ratio between box office and budget of the top 5% of cases with the rest of the sample, the top 5% movies earn significantly more for every dollar spent on budgets than their less talent-laden counterparts (appendix, p. 128). The opposite, a significant negative effect for the bottom 5% was not found. A limitation to this finding is that it may well be caused by single individuals as for instance the frequency of movies starring Oscar record holders Meryl Streep and Jack Nicholson (O'Neil, 2010) among those movies with the highest talent scores is naturally high.

5.1.3 Budget

Even though consumption capital and talent have been given precedence in the order of the discussion, the single coefficient with the highest contribution to explaining revenue remains that of the budget. Already Litman (1983) considered it to be a proxy for the technical and artistic quality of a movie. While this quality remains unjudged here, the effect on gross revenue is undeniable at a rate of almost exactly $1 million in revenue for every $2 million in budget spending. But despite being a fixture in the literature, its attested proxy character has made budget susceptible to criticism for being an excuse for an inability to measure the true, latent reasons for movie success. Despite this De'Vany & Walls (1999) find budget to retain its impact on the probability of producing a hit movie also in the presence of star actors, defining a hit as a movie grossing more than $50 million. A finding confirmed in this thesis as even in the presence of more diversified measures of stardom, the budget variable maintains its position as strongest predictor. While usually the exact size of the unstandardized coefficient is somewhat immaterial in this case it deserves particular notice. This is because – as already broached in the section on consumption capital – the coefficient almost exactly corresponds to the revenue share available to the producer of the movie as discussed in the introduction. So while the dependent variable of the regression model is revenue, its force of expression by virtue of the coefficient of budget extends to profit as well. Although there are limitations to this, like the existence of the marketing expenses and the variation of the revenue distribution contracts between producers and the distribution channel, at its core it suggests that producers cannot "buy" profit by indiscriminately spending larger and larger amounts on production budgets in hopes of seeing positive marginal effects on revenue, a practice Hollywood has repeatedly been characterized as unsuccessfully attempting to employ (Kenigsberg, 2013; Olive, 2012). Using a different perspective on budget by including the manifestation of the film quality Basuroy et al. (2003) find its effect to be contingent upon the ratio of positive and negative reviews, indicating that a well-reviewed movie does not have to be expensive to excel at the box office, while one receiving mostly poor reviews is dependent on its budget. At this point a link between the scaled measures of star power and the critics' verdict would provide valuable insight but would also rely on variables the producers cannot have at an early stage of development. In essence then the budget provides something like a baseline against which the effects of the other variables determine if the movie generates not only revenue but also profit.

5.1.4 Size of the star cast

The negative coefficient of this variable may without context come as a surprise to anyone assuming that stars have part in determining how many people go to the cinema and by doing so generate box office revenue. In the presence of the other measures of stardom however this variable is, as was intended, reduced to a measure of fragmentation of stardom attributes. As per the cast definition there are no non-stars but instead only stars with a smaller or greater star quality, thus every movie is affected by this variable deducing from the get-go $9 million dollars from the constant of any movie for its first actor in the cast. For movies with a cast greater than one actor the additional negative effect of the star power fragmentation can take a significant share of the additional revenue brought in by that star. Returning to the earlier example of Eddie Murphy the coefficient of “number of stars” suggests that the distribution of the accumulated consumption capital on one additional star is expected to take $9 million off the $16 million in revenue from addition consumption capital leaving the additional revenue at $7 million, thus taking off more than half of the positive effect. Redirecting the focus towards the hypotheses the coefficient is a caveat to a categorical rejection of the increasing marginal utility of star qualities. Because while these hypotheses have to be rejected when applied on a cast level, the negative coefficient of number of stars basically states on the level of single actors that a higher degree of fragmentation of the star qualities goes along with a decrease in expected box office revenue, suggesting that:

$$f_{Box\ Office\ Effect}(Star\ Quality\ X) > n * f_{Box\ Office\ Effect}\left(\frac{Star\ Quality\ X}{n}\right)$$

This is all the more intriguing as no suppression effect of significance or coefficient of neither the normal nor the quadratic terms of the two star quality measures by the number of stars variable was observed. In essence then the negative coefficient does not affect the validity of the hypotheses rejection, but rejects the transfer of the findings from the effect of star qualities in a movie cast onto the single members of that cast. This limitation will be further discussed later and concludes the discussion of this coefficient, because unlike the other aspects, this element does not allow for comparison with the literature as no other study using the number of stars in the presence of a measure of their quality was found. Thus contrasting its negative impact with literature proposing a positive effect on movies' success (Kerr, 1990; McDonald, 2000) will not allow for a sensible judgement of either side of the argument, as in this context the expressive force of the coefficient extends to star quality fragmentation and is not used as a proxy for the amount of star qualities itself.

5.1.5 Prequel box office

The positive effect of the accumulated box office of prequels, or in the case of most movies in the sample the effect of its absence, coincide with findings of the literature presented earlier. While the coefficient denoting $1 million in additional box office for every $10 million in prequel box office may not seem overwhelmingly large, the implication is that for "Spiderman 3" which generated $890 million in ticket sales, around 18% or $160 million are owed exclusively to the success of its two prequels – even more if the earlier discussed impact of "Casino Royale" on the coefficient is eliminated. But contrasted with the path coefficient for sequel box office determined by Hennig-Thurau, Houston, & O'Neal (2006) even the higher bootstrap estimate is low when the standardized coefficient of budget is used as a reference. As Hennig-Thurau, Houston & O'Neal (2006) use the box office of only the most recent prequel movie, this suggests that the effect of prequels is subject to either a decay over time or a time-independent, general decreasing marginal utility. Sticking to Spiderman movies, the more directly comparable model of Hennig-Thurau et al. (2009) suggests that "Spiderman 2" benefited from $53 million in revenue in the U.S. market that was exclusively evoked by its continuation on "Spiderman" which – in that market – grossed $403 million, a value close to the $40 million contribution the coefficient at b = 0,10 would predict.

Interestingly when the earlier mentioned overview of prequel and sequel revenue is transformed to instead show the return on investment rate based on the discussed profit estimate (as done on p. 118 in the appendix), a strong pattern can be observed. Of the 36 movies in the sample with prequel history 24 have earned more than twice their budget, thus suggesting they are financially at least roughly neutral. However of the 12 movies that did not accomplish this, 7 mark the bottom of the list when sorted by prequels' return on investment estimate. In fact, not a single movie whose prequel has generated less than 50% return on investment has produced enough revenue to cover its budget. Whereas producers seem to have a somewhat intuitive rule about not producing sequels to movies that lost money, the market threshold below which a movie has insufficient momentum to support a sequel seems to be higher. And while the subsample is too small for definitive statements, the observation that using a 50% return on investment minimum requirement reduces the odds of producing a financially unviable sequel from 33% to 17% is certainly worth consideration as well as further exploration[14].

Reflecting upon explanation models for the effect of prequel box office, the propinquity to consumption capital sticks out. As only a small part of the sequels in the sample are stand-alone sequels which do not further develop or depend on their prequel story, the majority presuppose a certain knowledge in the consumer to unfold the full enjoyment. As this special knowledge increasing the utility of further consumption is an archetypical example of consumption capital, this coefficient would surely have to be included in the consumption capital factor were it not for the lack of suitability due to a large share of movies without prequels in the sample which would negatively impact the internal consistency of the factor.

5.1.6 Animated movies

Although the standardized coefficient of animated movies is smaller than that of most other variables, the highly significant $70 million revenue premium found for those movies still deserves noticing. Animation movies in general are a somewhat odd breed which, one could have argued, need to be excluded from the sample because they are fundamentally different from the rest of the movies. This may be an intuitive notion, but one it is hard to find support for as an independent samples t-test shows no significant differences between animated movies and the supposedly "normal" movies in terms of the number of stars in the cast, the amount of consumption capital or talent in the cast, the age of those stars, or the amount of box office revenue generated by the animated movies' prequels (appendix, p. 128). The only significant difference lies in the dependent variable, the box office they generate, providing no data driven reason for an exclusion. They are by now, and have been for most of the sample timeframe, an established compartment of the industry that is responsible for around 10% of global ticket revenue as a small cross-compilation of data from the sample, the MPAA, and revenue tracking sites (Box Office Mojo, 2013d; MPAA, 2008, 2013b; The Numbers, 2013d) can show:

[14] While this was only observed after finishing the analysis its effect in the final model was tested by substituting it for the actually used measure of prequel success; the standardized coefficient of the return on investment estimate was surprisingly slightly smaller than that of the cumulated prequel box office.

Year	Animated movies Cumulated budgets[a]	Cumulated box offices[a]	Box offices / budgets	All movies Cumulated box offices[b]	Box office / budgets[c]	Animated box office share
2012	$1067m	$4094m	3,84	$34700m	n.a.	11,80%
2011	$1286m	$3465m	2,70	$32600m	n.a.	10,63%
2010	$1106m	$3962m	3,58	$31800m	n.a.	12,46%
2009	$1144m	$3401m	2,97	$29400m	n.a.	11,57%
2008	$867m	$2680m	3,09	$27700m	n.a.	9,68%
2007	$874m	$2744m	3,14	$28100m	2,50	9,77%
2006	$1034m	$2839m	2,75	$26700m	2,44	10,63%
2005	$457m	$1877m	4,11	$25400m	1,98	7,39%
2004	$652m	$2489m	3,82	$24900m	2,06	10,00%
2003	$287m	$896m	3,12	$20100m	2,35	4,46%

[a] Data from Box Office Mojo (2013d) and The Numbers (2013d)
[b] Data from MPAA (2008) and MPAA (2013b)
[c] Based on movies in the sample

Table 5: Animated movies as established industry segment

More captivating than that market share however is the ratio of budgets and revenues for animated movies compared to that for all movies: it confirms the core statement of the coefficient as animated movies consistently generate a surplus of revenue that is not absorbed by a greater budget. While this notion is also contained in the model as budget effects are controlled for by the budget coefficient, the illustration in table 5 conveys how far the box office superiority of animated movie extends to the odds of being profitable movies. In fact for some studios such as the Disney subsidiary “Pixar Animation Studios” the question so far has not been if a new movie is going to be financially successful but only how successful it will be, as no Pixar-produced movie has ever lost money (Zeitschik, 2013).

Surprisingly then this is not a universal finding, as for instance Elliott & Simmons (2008) find no significant effect on total revenue for animated movies. In contrast to this Sharda & Delen (2006) using a sensitivity analysis find a high amount of technical effects, a category which includes animated movies but also live-action movies with extensive use of computer generated imagery, to be one of three major predictors for box office revenue. In fact, even before the 2000s’ wave of animated movies, Simonoff & Sparrow (2000) found animated movies to be better performing in a comparison of animated and live-action films targeted towards children. The context of technical effects and the reminder that “animated movies” is not really a genre but rather a technical aspect that contains a whole subset of genres only one of which is children-oriented films, recalls to mind the problematic nature of animation movies explored in the outlier analysis. But while effects of MPAA rating within animated movies have been researched – Kaimann (2012) finds “G” and “PG” rated movies to achieve higher revenue in the U.S. domestic market – no differentiation tailored to animated movies is known. This is all the more adverse as aspects unique to animated movies were proposed earlier as explanations for the large residuals some of them produce in the regression. Upon this realization and perusing the list of animated movies in the sample a small test (appendix p. 130) was conducted checking whether there are significant differences between animated movies with predominantly human characters such as “The Adventures of Tintin”, and animated movies with predominantly non-human characters such as “Madagascar” or “Cars”. Using data on 104 animated movies (Box Office Mojo, 2013d) from 2003-2012 a significant difference of means of global box office revenue was found; averages are $210 million for the human-character movies and $326 million for non-human character movies. Budgets did not differ significantly, being on average $77 million for human character movies, and $91 million for non-human character movies. While this finding is in a way superficial, it should lend sufficient credibility to the earlier stated hypothesis that there really are differentiation attributes of animated movies that remain to be explored and thus, that this claim is not merely an excuse for the large residuals some of the animated movies produce in the regression.

5.1.7 Historic events

With the approximately opposite effect of being an animated movie, the negative coefficient of $75 million for movies based on historic events accounts for the fact that while their budgets are $21 million higher than average, their observed box offices are slightly below the global average of $106 million. An interesting observation is that a around half of those movies are biopics, detailing the entirety or selected sections of the life of – exclusively male – figures such as aerospace pioneer Howard Hughes, narcotics trafficker Franc Lucas, journalist Edward Murrow, or emperor Alexander the Great. In this context consumption capital may actually prove to be a treacherous friend, if producers rely on the fact that many consumers have heard of these figures, but disregard that mere recognition of a name is unlikely to increase the expected or actual enjoyment of consumption. Also a historical paragon has been discussed as limiting movies’ entertainment value as “filmmakers must juggle their artistic sensibilities and desire for historical accuracy with the requirements of the marketplace, the expectations and values of the audience” (Weinstein, 2001, p. 28). While this is a common theory, Sharda & Delen (2006) in a quantitative assessment find a “Historic Epic Drama” dummy variable to have no significant contribution to predicting motion picture success. As no further exploration of the issue was found, an analysis of a greater timeframe and thus a greater number of history-based movies should contribute to confirming and understanding this finding.

5.1.8 Release month

Although frequently discussed in the past, the investigation of release dates has mostly been focussed on assessing the effects of holiday release dates (Chiou, 2008; Radas & Shugan, 1998) – which here were not found to provide significant explanatory value – or on adversity effects of movies similar in MPAA rating or genre being released simultaneously (Ainslie, Dreze, & Zufryden, 2005) – which were not included into the model as this would require beforehand knowledge of all other studios' release plans. Rather the only significant release time related factor, the negative impact of a release in September, is without reference to other releases. This was found sufficient to warrant an inspection of monthly means for box office which shows evidence of a seasonality of box office revenues, only one aspect of which is a high average box office revenue for holiday releases as figure 8 shows.

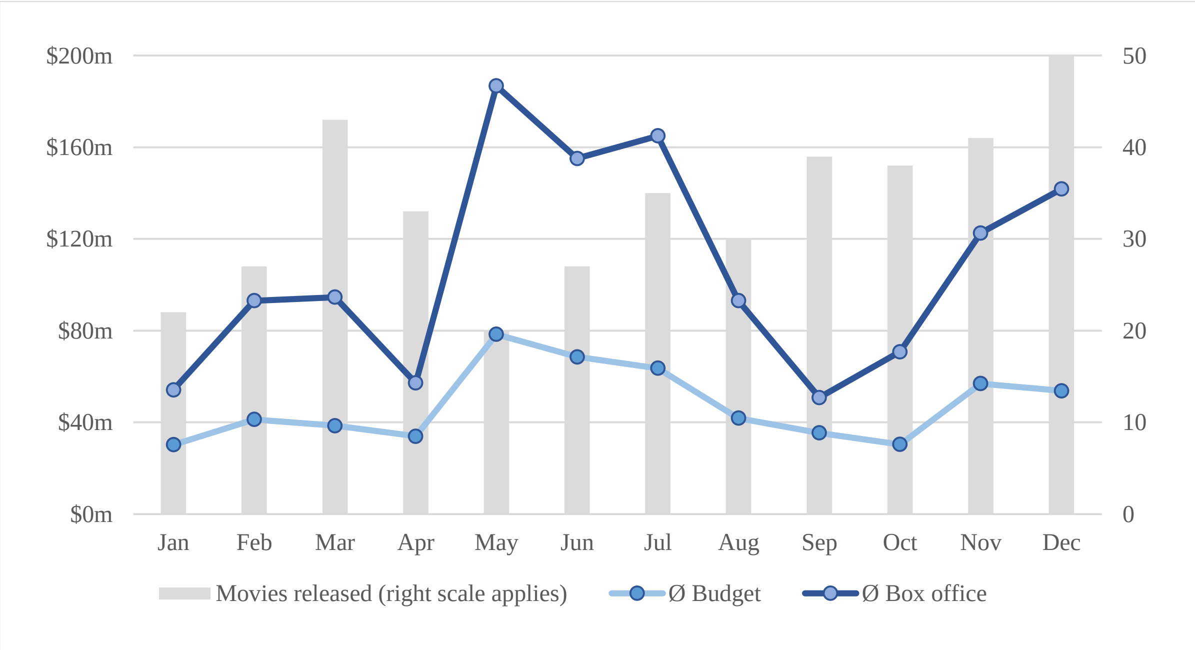


Figure 8: Seasonality of releases, box office and budgets

The number of movies released as detailed on the secondary vertical axis is distributed around two peaks, one in March and one in December, with low points in January and May. While the average box office revenue per film follows this distribution in the winter months as market demand seems to spike, it diverges between March and September during which time the months with high release density yield low average revenues as would be suggested in market with uniform demand. More interesting than this observation which has already been made by Radas & Shugan (1998) is the ratio between average box offices and budgets which three quarters of the year is pegged between 2,2 and 2,6. However for January (1,8) and April (1,7) and then finally September (1,4) it drops to almost half the ratio found during December (2,6). As budget is the variable with the highest contribution to the determination coefficient this fluctuation explains the significant disadvantage for September releases. But as the more general seasonality of revenue was found not to be fully dictated by the market but instead purposively induced by the studios' decisions on quantity and quality of releases (Einav, 2007) suggesting a market strategy based on an adaptation to these trends is difficult as the studios' actions obscure the nature of the unaltered underlying market demand. Thus the take-away from this coefficient is for one an awareness of the phenomenon, but also a proposal to revisit it from a different perspective like for instance game theory, as it appears the protagonists do not actually know the field they are competing on in its full complexity.

5.1.9 Age and gender

The findings regarding age and – if relying on the bootstrapping significance level or working with a laxer 0,10 significance threshold – gender show a pattern of discrimination by the audience towards on the one hand older casts, and on the other hand casts with a stronger female presence. But while age discrimination is usually accompanied by negative connotations, this must not be entirely the case for this occurrence. If the motivation to see a movie is analysed using Ryan & Deci's (2000) prominent approach to self-determination theory, the motivation would be viewed as an attempt to serve the innate need for relatedness. The value of entertainment in this context is derived from "the exploration of relationships through simulations that permit individuals to identify with substitute agents and thus create the subjective experience of relationships" (Voderer, Steen, & Chan, 2006, p. 14). This identification is aided by similarities between both oneself and the perceived character, and also between one's ideal of oneself and the perceived character (Igartua, 2010; Voderer et al., 2006). Thus the demographics of cinema goers would already suggest a negative impact of age, seeing that more than 50% of movie tickets are sold to consumers under 30 years of age (MPAA, 2013b). In fact Addis & Holbrook (2010) in an analysis of demographic determinants of movie ratings by consumers determine that at least the age of lead actors of the opposite sex affects a consumer's rating of a movie: independent of the consumer's age: same-age or younger stars of the opposite gender have a positive effect compared to older opposite-sex protagonists.

However, this approach does not provide an explanation for any gender differences as sexes are almost exactly level in terms of ticket sales (MPAA, 2013b). Furthermore the question arises about what was there first, a predominantly young theatre audience or the movies that cater to it. And while a simple interaction term between percentage of males and average age of the cast was tested and showed no significant results, a gender-moderated effect of age and a potentially non-linear connection could still be hypothesized. Furthermore in regards to gender a schism between the two stardom dimensions can be observed as female actors are found to have significantly lower scores in all variables relating to consumption capital, while at the same time having received significantly more wins and nominations for the used awards (full results in appendix, p. 132). If this finding is used to explain the coefficient of percentage of male actors it suggests that the superiority of women in providing talent to a movie as shown by their received awards is insufficient to compensate for the negative effect of their significantly lower consumption capital.

5.2 Application of findings

Another aspect demanding consideration is the concept of performativity (Callon, 1998) – the fact that a model like this does not exist in a separate dimension from its observed phenomenon, and thus can impact the phenomenon it sets out to explore. Just like the field of economics "in the broad sense of the term, performs, shapes and formats the economy, rather than observing how it functions" (Callon, 1998, p. 2) a model predicting movies' box office performances can only provide truly unbiased prediction so long as the model's attributes remain unknown to all parties having any direct or indirect impact on any of its variables. But while the notion that submitting this thesis negatively impacts the accuracy of its findings is more of a philosophical topic, more importantly it must be noted that the model cannot serve as a "recipe for the successful movie". To illustrate one must only consider that if the model were used as a recipe, it would suggest that cinemas should be shut in September, that "Schindler's List" would have generated a greater box office had it been an animated movie, and that movie posters should show only one actor who must be male. This simple reductio ad absurdum shows that this use of the model is dangerous. In light of the fact that the products of the movie industry have aside from the financial one also an artistic dimension, it seems reasonable that some issues remain outside the scope of the variables such as for instance a fit between an actor and a role. Thus the model is most suitable to be applied not during the process of deciding on a movie's attributes but in the process of deciding whether the movie will be actually realized. This point of the process, the decision on the ever-anticipated green light where the model's variables have been manifested, is where the model is most aptly applied and can unfold its predictive capabilities. Here also lies the reason why the discussion of the results is mostly void of often called-for management implications from individual aspects of the model; the management implication is to use these findings to consider whether to do something, not how to do it. If a movie project provides a revenue expectation that after deduction of the revenue share absorbed by the channel exceeds the cost of its production by a sufficient margin to justify the amount of money it takes to generate that revenue, then the project is viable. To a contemplating producer the two unknowns that remain in this statement, the exact retained revenue share and the desired return on investment threshold, are well known, thus rendering the model applicable in the depicted way. Like most rules naturally also this one has its caveats; if for instance a situation arises where several potential actors have the same fit for a role in a movie that a producer is planning on making, comparing those actors from a perspective of talent and consumption capital is certainly an advisable approach to making a choice. The somewhat timid prescription thus is intended to hint at the dangers of blindly following the results without applying the common "due diligence" of the motion picture industry. While this consideration is a requisite theoretically inherent in many models based on observation of past occurrences with uncaptured latent aspects, there are further limitations which are specific to only this model.

5.3 Limitations

Some of those lie within the sample; while the delimitation section has already limited the data foundation to major studio productions with a budget over $5 million the attribution of actors to movies further narrows this down to movies advertising their cast on the poster. In retrospect a factor analysis on the level of individual actors' engagements instead of on a movie level could, with a manual calculation of individual scores possibly based on the factor components but without transformation around a zero mean, have provided a more comprehensible way of presenting the star attributes' impact. While it also comes with undesired side effects, this would spare anyone using the coefficients the part of the tedious calculation of factor scores on the basis of the component matrix and the variable means. Those means in fact present another limitation as not all of them are static and thus may differ from the means in the sample timeframe: the number of Google News entries for instance is definitely dynamic as the internet generates more new data than it loses, thus the average number of news entries for the cast of the average movie in 2013 is bound to be greater than that in the sample timeframe. Another issue is the dependent variable: the discussed problem of using the budget which is partly determined by the unknown cast salaries as part of a composite dependent variable, such as net profit or return on investment, limits the use for practitioners as from a business perspective the stake required to generate a profit is an essential consideration. This shortcoming is however somewhat alleviated by the discussed expressive force of the coefficients towards profits due to the budget coefficient's similarity to the revenue share of the producers. Further, while they were identified as available in the stipulated timeframe, information on directors and writers due to data availability was not included in the analysis despite findings indicating that directors can have significant impact on movie success (De'Vany & Walls, 1999). Regardless of this, the most important and grave limitations definitely lie in the operationalization of the measurements of star qualities in the cast.

5.3.1 Operationalization of consumption capital

On the side of the consumption capital, it could be argued that cinema revenues are a poor measure of actual consumption capital as visiting the cinema is only one way of gathering consumption capital on an artist. Alternatively consumers might have seen previous performances on TV, on Blu-ray or DVD, via online streaming services, or they might even have downloaded them illegally. While this distribution across different channels in itself does not invalidate the use of box office revenue as a sole measure of consumption capital, a divergence of the relative view shares across these different options will at the very least undermine it. Following this notion a look at for instance a list of the most rented movies of all time on Netflix (Day & Godley, 2011) – by revenue the largest DVD rental-subscription and online-streaming service in North America – reveals a disparity that can be seen in table 6:

	Top 10 movies by box office	Top 10 Netflix rentals [a]	Top 10 Blu-ray discs [a]
1	Avatar *($2782m)*	The Blind Side *($309m)*	Avatar *($2782m)*
2	Titanic *($2187m)*	Crash *($98m)*	Star Trek *($386m)*
3	Harry Potter VIII *($1342m)*	The Bucket List *($175m)*	Inception *($826m)*
4	Transformers III *($1124m)*	Curious Case of Benjamin B. *($334m)*	The Hangover *($467m)*
5	The Lord of the Ring III *($1120m)*	The Hurt Locker *($49m)*	Beauty and the Beast *($425m)*
6	Pirates of the Caribbean II *($1066m)*	The Departed *($290m)*	Harry Potter VII *($960m)*
7	Toy Story III *($1063m)*	Sherlock Holmes *($524m)*	The Lion King *($987m)*
8	Pirates of the Caribbean IV *($1046m)*	Inception *($826m)*	Harry Potter VIII *($1342m)*
9	Jurassic Park *($1029m)*	Iron man *($585m)*	Despicable Me *($543m)*
10	Star Wars: Episode I *($1027m)*	No Country For Old Men *($172m)*	Harry Potter VI *($934m)*

[a] Numbers in brackets represent box office revenue in million $

Table 6: Top 10 movies by box office, Netflix rentals, and Blu-ray discs sold; for comparability only until 12/2011

This comparison of data from Day & Godley (2011) and The Numbers (2013a, 2013e) shows that he average box office revenue of the top 10 movies on Netflix is at $336 million around a quarter of the average $1,38 billion of the top 10 blockbusters of all time. Even more, the Netflix top list does not share a single movie with the box office top list. Instead it features movies like "Crash" or "The Hurt Locker" which have box office results even below the sample average. And while the communalities between the most successful movies at the box office and the best-selling Blu-ray discs of all time (Nash, 2013) are at least existent as they share "Avatar" as leader and both feature the last Harry Potter movie, the discrepancies are still substantial. In essence then this comparison shows that consumption patterns across different channels do vary considerably. The only real mitigation to this limitation is the fact that the targeted dependent variable was the box office and thus a consistency between the source of the consumption capital – having seen a movie in cinemas – and acting upon it by watching another movie in cinema was achieved. This in turn however is constrained by the fact that in this type of analysis it is impossible to tell if those consumers who contributed to the revenue of a movie have in fact seen previous performances of the cast which is starring in it.

5.3.2 Operationalization of talent

Like the operationalization of consumption capital, also the variables chosen to represent talent are susceptible to criticism. For one the awards measure is weakened by the fact that it does not account for the number of opportunities an actor has had the chance to display their potential talent. Thus one could argue that while both Frances Conroy in "Broken Flowers" and Claire Danes in "The Family Stone" both had earned one Golden Globe award prior to the release of those movies, Frances Conroy had more chances to display her talent over the 40 years of acting career that lay behind her than Claire Danes who was merely 26 years old when "The Family Stone" was released.

Furthermore awards are not based on a transitive relation of performance quality. This means that even if they manage to detect the best performance and honour it, they do so for performances rendered within a one year timeframe. Thus one year's winner could be objectively worse than a movie that was not even nominated in another year. As an example one could argue that Nicholas Cage's chances to win an Academy Award might have slimmed if his movie "Leaving Las Vegas" had been released a few months earlier and he would therefore have been competing with Tom Hanks' performance in "Forrest Gump" as well as with John Travolta in "Pulp Fiction" and Morgan Freeman in "The Shawshank Redemption".

Also the fact that an institution like the Academy which is to its core American is used as a general measure of talent seems inappropriate – not because it is American, but because cultural differences in taste might suggest that the ability to cater to them is not universal, and thus no single culturally influenced judgement can have universal applicability. That the industry's different institutions have greatly differing interpretations of what kind of performance is worth honouring was shown by Popik (2011) and can in fact be demonstrated for the sample data using correlations between accumulated wins across different awards[15]:

	Golden Globes	Oscars	Cannes	Berlinale	BAFTA
Golden Globes	1	,762**	,326**	,263**	,553**
Oscars	,762**	1	,483**	,278**	,659**
Cannes	,326**	,483**	1	,234**	,283**
Berlinale	,263**	,278**	,234**	1	,085
BAFTA	,553**	,659**	,283**	,085	1

**. Correlation is significant at the 0.01 level (2-tailed).

Table 7: Correlations of number of wins across different awards

This most certainly raises to question the use of only Oscars and Golden Globes and calls for an integration of more diverse measures of talent. Nonetheless as an upside of this operationalization remains, that even if one was to categorically reject the validity of using awards to measure talent, a meaningful and strong quintessence would remain intact: the amount of awards and nominations accumulated by a movies cast in their previous performances positively influences the observed movie's box office performance.

[15] Data for Berlinale, Cannes, and the British Academy of Film and Television Arts (BAFTA) was obtained from a privately maintained database (Shin, 2013), which proved to have minor flaws and was thus not included in the analysis; however the accuracy was deemed sufficient to give an overview on the awards' interrelations.

6 Conclusion

Remaining as concise as possible, this section will first synthesize and distil the outcome of the thesis to its core. Upon this, an outlook towards further research opportunities derived from the findings will be provided.

6.1 Synthesis of findings

The key finding of this thesis is the presented impact of talent and consumption capital in movie casts on the box office revenue of those movies. The journey that lead to this result commenced with a characterization of star utilization in motion pictures throughout the industry's history. Upon reviewing the multitude of aspects proposed as determinants of box office success, particular focus was placed on the existent body of research regarding the impact of stars. Based on the identified gap within that field, theoretical foundations suitable for a more differentiated approach were reviewed and then operationalized alongside a battery of control variables. The findings of the subsequent analysis were discussed in the previous section and now an attempt at merging and contemplating them will follow.

Considering the combined effect of consumption and talent, the observed impact puts the thesis at odds with some of those studies investigating effects of star power referenced in the literature review and the discussion. In light of the width of the potential predictors and control variables tested, this discrepancy is likely to emanate from not only the choice of representations of talent and consumption capital but also from their scaled nature. Pitched against a simple binary variable denoting whether an actor was on a star list or had played parts in more than 5 films – both being star power representations used in the discussed literature – the scaled metering has proven sufficiently sensitive to detect aspects of stardom which allow for a partial prediction of movie revenue. Incidentally the two measures also entail different attribute locations: while talent – as it is viewed here – is located within the cast itself, the consumption capital is actually not owned by or located in the cast, but rather as the term "capital" already implies the outcome of the individual cast members' previous efforts in movies. Therefore instead of with the actors, the capital lies with those consumers that have heard of or seen their previous movies. Thus in its full state the reason for the divergence of results from previous research, is presumed to lie in the combination of internal and externally associated attributes of the star cast, covering consumption capital and talent as well as demographic aspects, while using measurements that are not only multidimensional but also where applicable scaled.

With this outcome the thesis has served its purpose as it was stated in the introduction: it has shed new light on the effect of star casts on box office results by employing a hitherto unutilized approach to the qualities that lend actors the power to attract audiences. By relying only on data that is already available in an early stage of a movie's genesis the presented model also allows for use by the industry it observes. While the used analytic tools complicate this application slightly, they compensate for that by presenting the reader with two clear and clearly separated concepts of box-office-affecting star power.

Taking the separate elements first found to be predictors in the analysis part and then scrutinized in the discussion section, and evaluating them jointly, the relative importance of cast-related effects compared to their movie-related counterparts is sticking out. Were it not for the budget, which is in fact also related to the cast as it comprises the actors' pay, the latter would have to be considered of secondary importance. In fact with the similarity between the approximate revenue proportion going to the producers and the size of the budget coefficient the relative importance of the cast-related effects actually reaches that level of importance. This becomes all the more apparent when the number of movie-related predictors is called to mind that have previously been found to significantly impact movie success and failed to exhibit such an effect here, in the presence of the two dimensions of star qualities in the movies' cast. Thus after the individual contextualisation of the different predictors in the discussion, the outcome of the thesis when regarded in unison is contrary to the results of endeavours like that of De'Vany & Walls (1999) who despite finding that some actors increase the success probability conclude that "the real star is the movie" (De'Vany & Walls, 1999, p. 285). Rather the mentioned absence of findings actually marks an important finding in itself that has not been comprehensively covered in the discussion. To arrive at this conclusion one only needs to observe that the MPAA rating which is a completely solid variable, in the sense that it is not open for interpretation or alternative assessment like for instance a genre variable, does not show a significant effect in even one of its categories despite an absence of strong correlations with the scaled variables used in the final regression. Given this observation and the earlier referenced findings of box office impact of the rating, the results connote that the measures of star quality have explanatory value and impact that supersedes those of several of the attributes more closely linked to the movie than its cast. Overall the implication is that aspects of the movie itself and the star actors in the cast are of at least comparable importance. In this way Peter Bart and Peter Guber, former chairmen of Paramount Pictures and Sony Entertainment respectively, have likely managed to strike a reasonable balance between blind star-reliance and categorical denial of star effects when suggesting that “the movies need stars and the stars need movies” (Bart, 2003, p. 2). This statement not only strikes a balance, but also accounts for the different locations of the star qualities: the stars need the movies as canvas and opportunity to display their talent, the movies need the stars to activate and capitalize on the consumption capital that is associated with the stars, but is actually located inside the consumer base.

Turning to the theoretical foundations, the observation that while the marginal effect of talent is linear, that of consumption capital is decreasing yet remaining positive is a realisation on its own. But combined with the negative impact of the number of stars on the box office it suggests that what Adler (1985) and Rosen (1981) proposed for the individual star’s ability to generate income for themselves, is true for the individual movie star’s ability to generate revenue for the movie they are a cast-member in. Moreover the observed absence of a significant effect of talent in the very short timeframe of the opening weekend combined with the continuation of the effect of consumption capital asserts Adler's (1985) proposal that stars can concentrate consumption on themselves even in the presence of equality of talent or the absence of additional consumer utility from talent. Reconnecting this to the postulated and reviewed hypotheses it reinforces the validity of the statement that the rejection of hypotheses 3 & 4, which are relating to the increasing marginal utility of talent and consumption capital, reflects exclusively on a setting where they are applied to already aggregated qualities, as was the case in the thesis where the qualities were assessed on the level of the cast which mostly consisted of more than one individual.

With this recourse to the implications of the hypotheses’ verdicts on their theoretical foundations, the assessment of the stated purpose and the achieved result, and a short synthesis of the findings, it is time to turn the perspective towards the future. Because based on the findings discussed and gauged in the last two sections, the establishment of the effect of the two-dimensional stardom measure – which is most likely the core contribution of this thesis – opens up multiple approaches to existing problems as well as kindling new questions and ideas.

6.2 Outlook

One such instance is the use and extension of the findings on consumption capital by capturing a wider variety of channels through which consumption capital is generated, like online streaming, has the potential to illuminate differences in transfer-values of consumption capital across consumption channel boundaries. Also the range of manifestations could be extended, instead of scanning the Twittersphere for tweets on a new movie, the actors' number of followers could provide a clue on the social aspect of consumption capital long before the first scene is shot. Sources of data are plentiful and more importantly, accessible. This could also help comprehend the stark contrast between the most successful movies across channels seen earlier.

Taking yet another step back, the sources of consumption capital are also less clear than they may seem. Finding that Jennifer Aniston's cumulated box office is actually below sample average may be a surprise – but only until realizing that over 10 seasons she has starred in more than 200 episodes of one of the most successful television shows in history; "Friends" with over 50 million viewers for its final episode in the U.S. alone (Kinon, 2009) is most certainly a considerable source of consumption capital that a box office perspective will remain blind to. Seeing how the lines of separation between cinema and television are blurring rapidly as television actors like Jennifer Aniston "graduate" to cinema, while established actors like two-time Oscar recipients Kevin Spacey and Dustin Hoffman turn to television formats in "House of Cards" and "Luck", the idea that consumption capital only works within the confines of its own content format seems highly unlikely. The fact that "House of Cards" is technically not even a television series, as it was produced and aired online by Netflix, reinforces the need for a comprehensive, cross-channel, cross-source review of consumption capital.

Introducing time as a further dimension to consumption capital, also its erosion is an issue that will allow for a better understanding of hype phenomena on a longer term basis. The current research on hype effects in the movie industry is mostly focussed on short term analyses of consumers' behaviour (Reddy, Kasat, & Jain, 2012; Uri & Dholakia, 2012). Applying a depreciation rate that exceeds mere inflation adjustment on consumption capital in the consumer base could provide an appreciation of how a momentum in actors' careers can be utilized in the industry. As an example, a look at Ryan Gosling's career shows only 3 movies with a box office of over $20 million in his 8-years of acting up to 2011; in contrast his last 5 movies since then have consecutively scored well above that mark totalling a combined revenue of above $435 million. While this puts him far below the sample average in terms of total previous box office, it puts him ahead in terms of box office in a 3-year period – ahead of revenue heavyweights like Tom Hanks who has had several slow years in terms of movie revenue. With a sensitive rate of decay this differentiation can provide an enhanced view at consumption capital that would otherwise be lost, and provide a fact-based counterpoint to the media's perpetual discussion which actress or actor are "hottest" right now – "hot" in this context in the sense of career momentum not in terms of physical appeal. Also an understanding of this decay is potentially applicable outside the world of motion pictures or even outside the entertainment world. In fact, if the assumption of monotonous increase for talent is reviewed, the restriction on the applicability of a decay function is dissolved.

Calling to mind the partially low correlations among the different awards also a more differentiated understanding of talent from the institutional side could provide dimensions of talent that are so far underrepresented in this measure. Because while he definitely is high in all dimensions of consumption capital, judging Arnold Schwarzenegger's ability to deliver what Rosen (1981) would consider a superior performance only based on his two Golden Globe nominations may not do his talent justice. This does by no means go unnoticed by institutionalized surveyors of quality performances, as he has been nominated numerous times for the Saturn Award presented by the Academy of Science Fiction, Fantasy & Horror Films, and won awards such as the MTV Movie Award or the Nickelodeon Kids' Choice Award. To develop this train of thought further a more versatile definition of talent as the earlier formulated "ability to achieve an output of superior quality" is needed, as this is rested upon an assumption of universality of quality. In reviewing this assumption and refining the definition of talent a matching mechanism between the past manifestations of talent and the movie in which it will be applied could prove useful. In this approach, that could be equally useful for consumption capital, a discount factor between the source genre and the target genre of star power comes to mind. In practice this could mean reviewing whether, and if how, an actor who has been honoured in the Berlin or Cannes film festivals is better suited for leading a drama to financial success at the box office than a recipient of the previously mentioned Saturn Award.

The observations on gender and age effects, especially set in context with entertainment through identification or immersion, prompt hypotheses for other types of immutable attributes of the cast such as their physical appearance or race. This would require extensive collection of first hand data, but has the potential to shine a light on what is more important to the viewer when evaluating a screen character: a high degree of fit with that character to aid the immersion, or desirable relationships that character has with other characters. Applied to movies this would mean knowing whether a male watching a romantic movie derives greater utility from having a male character in the movie that strongly resembles himself or his ideal of himself and thus makes it easy to immerse, or from having a female character he perceives as having traits that make a relationship with her desirable. While this may initially not be linked to the core findings, the fact that consumption capital can be viewed as a degree of familiarity with an actor's screen persona predestines the question to an application of this star quality indicator.

Further, seeing that a considerable number of the efforts investigating box office success have used variations of survival games (Ainslie et al., 2005; De'Vany & Walls, 1997; De'Vany & Walls, 1999; McKenzie, 2009) which aim to find characteristics that allow a movie to survive as long as possible on the cinemas' programme rosters, the disappearance of the significant effect of talent during the opening weekend implies a usefulness in such a survival analysis. Applying the two established star-quality factors in a survival game similar to De'Vany & Walls' (1999) method might thus help to better understand differences in revenue development paths between movies; this could shed light on why – despite both movies grossing around $360 million – "The Fast and the Furious" generated more than 50% of its domestic gross within the first week of its runtime, while "My Big Fat Greek Wedding" did not even gross 2% of its domestic gross within the first month after its release (Box Office Mojo, 2013a).

Overall then, based on the core contribution of this thesis the opportunities to continue augmenting the understanding of this fascinating field are manifold and ought to provide sufficient foundation for a sequel, or two.

References

Abdi, H. (2003). Factor Rotations in Factor Analyses. In M. Lewis-Beck, A. Bryman, & T. Futing (Eds.), *Encyclopedia of Social Sciences Research Methods* (1st ed., pp. 1–8). Thousand Oaks, CA: Sage Publications Ltd.

Ackman, D. (2003). Hollywood's Star Power Failure. *Forbes Magazine Online Edition*. Retrieved October 09, 2013, from www.forbes.com/2003/06/18/cx_da_0619stars.html

Ackman, D. (2004). The Myth of Brad Pitt. *Forbes Magazine Online Edition*. New York, NY. Retrieved September 10, 2013, from www.forbes.com/2004/05/17/cx_da_0517topnews.html

Addis, M., & Holbrook, M. B. (2010). Consumers' identification and beyond: Attraction, reverence, and escapism in the evaluation of films. *Psychology and Marketing*, *27*(9), 821–845. doi:10.1002/mar.20359

Adler, M. (1985). Stardom and talent. *The American Economic Review*, *75*(1), 208–212.

Ainslie, A., Dreze, X., & Zufryden, F. (2005). Modeling Movie Life Cycles and Market Share. *Marketing Science*, *24*(3), 508–517. doi:10.1287/mksc.1040.0106

AMPAS - Academy of Motion Pictures and Arts and Sciences. (2013a). About the Academy Awards. *The Academy of Motion Picture Arts and Sciences*. Retrieved October 20, 2013, from www.oscars.org/awards/academyawards/about

AMPAS - Academy of Motion Pictures and Arts and Sciences. (2013b). AMPAS Database. *The Official Academy Awards® Database*. Retrieved September 20, 2013, from http://awardsdatabase.oscars.org/ampas_awards

Anast, P. (1967). Differential Movie Appeals as Correlates of Attendance. *Journalism & Mass Communication Quarterly*, *44*(1), 86–90.

Asur, S., & Huberman, B. A. (2010). Predicting the Future with Social Media. In *Proceedings of the 2010 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology*. (Vol. 1, pp. 492–499). Computers and Society; Physics and Society.

Austin, B., Mark, N., & Simonet, T. (1981). M. P. A. A. Ratings and the Box Office: Some Tantalizing Statistics. *Film Quarterly*, *35*(2), 28–30.

Bagella, M., & Becchetti, C. (1999). The Determinants of Motion Picture Box Office Performance: Evidence from Movies Produced in Italy. *Journal of Cultural Economics*, *23*(4), 237–256.

Bart, P., & Guber, P. (2003). The Trouble with Stars. *The Guardian Online Edition*. Retrieved December 01, 2013, from http://www.theguardian.com/film/2003/may/02/artsfeatures

Basuroy, S., Chatterjee, S., & Abraham, D. (2003). How Critical Are Critical Reviews? The Box Office Effects of Film Critics, Star Power, and Budgets. *Journal of Marketing*, *67*(4), 103–117.

Bentham, J. (1823). *An Introduction to the Principles of Morals and Legislation*. (H. Frowde, Ed.). London, United Kingdom: Clarendon Press.

Berger, J. (2011). The enduring life of the slow-burner. *The Independent*. Retrieved October 23, 2013, from http://www.independent.co.uk/arts-entertainment/films/features/the-enduring-life-of-the-slowburner-2355331.html

Bhargava, A., Franzini, L., & Narendranathan, W. (1982). Serial Correlation and the Fixed Effects Model. *The Review of Economic Studies*, *49*(4), 533. doi:10.2307/2297285

Bing, J. (2002). Actors savor star bucks. *Variety Magazine*. Retrieved October 04, 2013, from www.variety.com/2002/film/news/actors-savor-star-bucks-1117864688/

Block, A. (2010). Why Oscar didn’t embrace “Avatar.” *The Hollywood Reporter*. Retrieved October 04, 2013, from www.hollywoodreporter.com/news/why-oscar-didnt-embrace-avatar-21390

Boatwright, P., Basuroy, S., & Kamakura, W. (2007). Reviewing the reviewers: The impact of individual film critics on box office performance. *Quantitative Marketing and Economics*, *5*(4), 401–425.

Bona, D. (2003). Oscar Films: 75 Years of Bribes, Lies and Overkill. *The New York Times*. Retrieved October 02, 2013, from www.nytimes.com/2003/03/09/movies/oscar-films-75-years-of-bribes-lies-and-overkill.html

Bordwell, D., & Thompson, K. (2008). *Film Art and Filmmaking* (6th ed., pp. 1–51). New York, NY: Mcgraw-Hill.

Box Office Mojo. (2013a). Box Office Statistics. Retrieved October 10, 2013, from www.boxofficemojo.com

Box Office Mojo. (2013b). 2012 Domestic Grosses. *Box Office Mojo*. Retrieved November 22, 2013, from http://boxofficemojo.com/yearly/chart/?yr=2012

Box Office Mojo. (2013c). Franchise Index. Retrieved November 27, 2013, from http://www.boxofficemojo.com/franchises/

Box Office Mojo. (2013d). Animation Movies 1980-present. *Box Office Mojo Database*. Retrieved November 23, 2013, from http://www.boxofficemojo.com/genres/chart/?id=animation.htm

Brown, A. L., Camerer, C. F., & Lovallo, D. (2012). To Review or Not to Review? Limited Strategic Thinking at the Movie Box Office. *American Economic Journal: Microeconomics*, *4*(2), 1–26.

Browne, M. W. (2001). An Overview of Analytic Rotation in Exploratory Factor Analysis. *Multivariate Behavioral Research*, *36*(1), 111–150. doi:10.1207/S15327906MBR3601_05

Burman, J. (2006). Star Power 2002. *The Hollywood Reporter*. Retrieved September 20, 2013, from www.hollywoodreporter.com/news/star-power-2002-credits-139392

Byers, S., Cranor, L. F., Cronin, E., Korman, D., & McDaniel, P. (2004). An analysis of security vulnerabilities in the movie production and distribution process. *Telecommunications Policy, 28*(7-8), 619–644.

Callon, M. (1998). Introduction: The Embeddedness of Economic Markets in Economics. In M. Callon (Ed.), *The Laws of the Markets* (2nd ed., pp. 1–57). Oxford: Blackwell.

Cameron, S. (1995). On the role of critics in the culture industry. *Journal of Cultural Economics*, *19*(4), 321–331. doi:10.1007/BF01073994

CBS News. (2009). Bankable Stars. *CBS News*. Retrieved September 30, 2013, from http://www.cbsnews.com/stories/2002/02/05/entertainment/main328301.shtml

Chiou, L. (2008). The Timing of Movie Releases: Evidence from the Home Video Industry. *International Journal of Industrial Organization*, *26*(5), 1059–1073. doi:10.1016/j.ijindorg.2007.11.005

Cinemark. (2011). *Form 10-K Annual Report SEC Filing Cinemark Holdings, Inc.* (Vol. 15). Plano, TX. Retrieved from www.sec.gov

Cormen, T. H., Leiserson, C. E., Rivest, R. L., & Stein, C. (2009). *Introduction to Algorithms* (3rd ed.). Cambridge, Massachusetts & London, United Kingdom: The MIT Press. doi:10.1002/9781119990413

Craig, C. S., Greene, W. H., & Douglas, S. P. (2005). Culture Matters : Consumer Acceptance of of U. S. in Foreign Markets. *Journal of Internation Marketing*, *13*(4), 80–103.

Craney, T. a., & Surles, J. G. (2002). Model-Dependent Variance Inflation Factor Cutoff Values. *Quality Engineering*, *14*(3), 391–403. doi:10.1081/QEN-120001878

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Cucco, M., & Candeloro, J.-P. (2011). Italian Feature Films on National Public and Commercial Broadcasters. *Observatorio Lugano Journal*, *5*(4), 31–48.

Day, P. K., & Godley, C. (2011). Netflix's 10 Most Rented Movies of All Time. *The Hollywood Reporter*. Retrieved November 14, 2013, from http://www.hollywoodreporter.com/gallery/netflixs-10-rented-movies-ever-239740

De'Vany, A., & Lee, C. (2001). Quality signals in information cascades and the dynamics of the distribution of motion picture box office revenues. *Journal of Economic Dynamics & Control*, *25*(1), 593–614.

De'Vany, A. S., & Walls, W. D. (1997). The Market for Motion Pictures: Rank, Revenue, and Survival. *Economic Inquiry*, *35*(4), 783–797. doi:10.1111/j.1465-7295.1997.tb01964.x

De'Vany, A., & Walls, D. (1999). Uncertainty in the Movie Industry : Does Star Power Reduce the Terror of the Box Office ? *Journal of Cultural Economics*, *23*(4), 285–318.

Dewenter, R., & Westermann, M. (2005). Cinema Demand In Germany. *Journal of Cultural Economics*, *29*(3), 213–231. doi:10.1007/s10824-005-6421-0

Dhar, T., Sun, G., & Weinberg, C. B. (2012). The Long-Term Box Office Performance of Sequel Movies. *Marketing Letters*, *23*(1), 13–29. doi:10.1007/s11002-011-9146-1

Distefano, C., & Mîndrila, D. (2009). Understanding and Using Factor Scores: Considerations for the Applied Researcher. *Practical Assessment, Research & Evalutation*, *14*(20), 1–11.

Dizard, W. (1994). *Old Media New Media: Mass Communications in the Information Age* (1st ed., pp. 129–144). New York, NY: Longman.

Dodds, J., & Holbrook, M. (1988). What's an Oscar worth? An empirical estimation of the effects of nominations and awards on movie distribution and revenues. *Current research in film: Audiences, economics and the law*, *4*, 72–88.

Donnelly, A. (1978). George Lucas' Galactic Empire. Get ready for Star Wars II, III, IV, V. *TIME Magazine*.

Durbin, B. Y. J., & Watson, G. S. (1950). Testing for Serial Correlation in Least Squares Regression: I. *Biometrika*, *37*(3/4), 409–428.

Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, *81*(6), 358–361. doi:10.1037/h0036316

Efron, B., & Gong, G. (1983). A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, *37*(1), 36–48. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00031305.1983.10483087

Einav, L. (2007). Seasonality in the U.S. motion picture industry. *The RAND Journal of Economics*, *38*(1), 127–145. doi:10.1111/j.1756-2171.2007.tb00048.x

Elberse, A. (2007). The Power of Stars : Do Star Actors Drive the Success of Movies ? *Journal of Marketing*, *71*(October), 102–120.

Elberse, A., & Eliashberg, J. (2003). Demand and Supply Dynamics for Sequentially Released Products in International Markets : The Case of Motion Pictures. *Marketing Science*, *22*(3), 329–354.

Eliashberg, J., & Shugan, S. M. (1997). Film Critics: Influencers or Predictors? *Journal of Marketing*, *61*(2), 68. doi:10.2307/1251831

Elliott, C., & Simmons, R. (2008). Determinants of UK Box Office Success: The Impact of Quality Signals. *Review of Industrial Organization*, *33*(2), 93–111. doi:10.1007/s11151-008-9181-0

Elsworth, C. (2005). Sony ordered to pay $1.5m for film-goer hoax. *The Telegraph*. Retrieved October 09, 2013, from http://www.telegraph.co.uk/news/worldnews/northamerica/usa/1495591/Sony-ordered-to-pay-1.5m-for-film-goer-hoax.html

Epstein, E. (2013). Budget Sheet Tomb Raider II. Retrieved December 03, 2013, from http://www.edwardjayepstein.com/documents/tombraider.swf

Erdogan, B. Z. (1999). Celebrity Endorsement: A Literature Review. *Journal of Marketing Management*, *15*(4), 291–314. doi:10.1362/026725799784870379

Fabrigar, L. R., Wegener, D. T., MacCallum, R. C., & Strahan, E. J. (1999). Evaluating the use of exploratory factor analysis in psychological research. *Psychological Methods*, *4*(3), 272–299. doi:10.1037//1082-989X.4.3.272

Farrar, D., & Glauber, R. (1967). Multicollinearity in regression analysis: The problem revisited. *The Review of Economics and Statistics*, *49*(1), 92–107.

Faughnder, R. (2013). Movie ticket prices rise, propelled by 3-D and Imax films. *L.A. Times*. Retrieved September 24, 2013, from http://articles.latimes.com/2013/jul/19/entertainment/la-et-ct-movie-ticket-prices-3d-20130719

Fetscherin, M. (2010). The Main Determinants of Bollywood Movie Box Office Sales. *Journal of Global Marketing*, *23*(5), 461–476. doi:10.1080/08911762.2010.521117

Field, A. (2000). *Discovering Statistics Using SPSS for Windows: Advanced Techniques*. (P. Beana, Ed.) (1st ed.). London, United Kingdom: Sage Publications Ltd.

Field, A. (2009). *Discovering Statistics using SPSS*. (P. Beana, Ed.) (3rd ed.). London, United Kingdom: Sage Publications Ltd.

Fox, J. (1991). *Regression Diagnostics - Quantitative Applications in the Social Sciences*. (D. Foster, Ed.) (2nd ed.). London, United Kingdom: Sage Publications Ltd.

Friedman, J. (2008). Movie ticket sales hit record. *L.A. Times*. Retrieved October 20, 2013, from http://articles.latimes.com/2008/mar/06/business/fi-boxoffice6

Gemser, G., Oostrum, M., & Leenders, M. a. a. M. (2006). The impact of film reviews on the box office performance of art house versus mainstream motion pictures. *Journal of Cultural Economics*, *31*(1), 43–63. doi:10.1007/s10824-006-9025-4

George, D., & Mallery, P. (2005). *SPSS for Windows Step-by-Step: A Simple Guide and Reference* (6th ed.). Boston, MA: Allyn & Bacon, Pearson Education.

Google. (2013). Google News. Retrieved from http://news.google.com

Halbfinger, D. (2006). Where angels fear to tread. *The Guardian Online Edition*. Retrieved October 19, 2013, from www.theguardian.com/film/2006/jul/28/1

Harrell, F. E. J. (2001). *Regression modeling strategies*. (P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg, I. Olkin, N. Wermuth, & S. Zeger, Eds.) (2nd ed., Vol. 64). New York, NY: Springer Statistics.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *Springer Series in Statistics The Elements of Statistical Learning: Data Mining, Inference, and Prediction* (2nd ed.). New York, NY: Springer Statistics.

Hennig-Thurau, T., Houston, M. B., & Heitjans, T. (2009). Conceptualizing and Measuring the Monetary Value of Brand Extensions: The Case of Motion Pictures. *Journal of Marketing*, *73*(6), 167–183. doi:10.1509/jmkg.73.6.167

Hennig-Thurau, T., Houston, M. B., & O'Neal, J. (2006). Determinants of Motion Picture Box Office and Profitability : An Interrelationship Approach Determinants of Motion Picture Box Office and Profitability : An Interrelationship Approach. *Review of Managerial Science*, *24*(December), 1–37.

Hennig-Thurau, T., Houston, M. B., & Sridhar, S. (2006). Can good marketing carry a bad product? Evidence from the motion picture industry. *Marketing Letters*, *17*(3), 205–219. doi:10.1007/s11002-006-7416-0

Hennig-Thurau, T., Walsh, G., & Wruck, O. (2001). An Investigation into the Factors Determining the Success of Service Innovations : The Case of Motion Pictures. *Academy of Marketing Science Review*, *6*(6), 1–23.

Hennig-Thurau, T., & Wruck, O. (2000). Warum Wir ins Kino gehen: Erfolgsfaktoren von Kinofilmen [Why We go to the Movies: Success Factors of Motion Pictures]. *Marketing ZFP Journal*, *22*(3), 241–258.

Hoad, P. (2013). Hollywood's hold over global box office – 63% and falling. *The Guardian Online Edition*. Retrieved November 20, 2013, from http://www.theguardian.com/film/filmblog/2013/apr/02/hollywood-hold-global-box-office

Hoffmann, L., & Rose, L. (2005). Most Lucrative Movie Franchises. *Forbes Magazine Online Edition*. Retrieved from http://www.forbes.com/2005/06/15/batman-movies-franchises-cx_lh_lr_0615batman.html

Holbrook, M. B. (1999). Popular Appeal versus Expert Judgments of Motion Pictures. *Journal of Consumer Research*, *26*(2), 144–155.

Hollywood Foreign Press Association. (2013). Golden Globes Awards Database. Retrieved September 20, 2013, from http://www.goldenglobes.org/awards-search

Hoppenstand, G. (1998). Hollywood and the Business of Making Movies: The Relationship between Film Content and Economic Factors. In B. R. Litman (Ed.), *The Motion Picture Mega-Industry* (1st ed., pp. 222–241). New York, NY: Allyn & Bacon, Pearson Education.

Igartua, J.-J. (2010). Identification with characters and narrative persuasion through fictional feature films. *Communications, 35*(4), 347–373. doi:10.1515/comm.2010.019

IMDb. (2013a). Johnny Depp Biography. Retrieved November 20, 2013, from http://www.imdb.com/name/nm0000136/bio

IMDb. (2013b). IMDb Top 250 Movies of all Time. Retrieved October 02, 2013, from www.imdb.com/chart/top

IMDb. (2013c). The Internet Movie Database. Retrieved September 01, 2013, from www.imdb.com

Jacobs, L. (1983). The Paramount Case and the Role of the Distributor. *Journal of the University Film and Video Association*, *35*, 44.

Jansen, C. (2005). The Performance of German Motion Pictures, Profits and Subsidies: Some Empirical Evidence. *Journal of Cultural Economics*, *29*(3), 191–212. doi:10.1007/s10824-005-1157-4

Jolliffe, I. (2002). *Principal Component Analysis*. (P. Bickel, P. Diggle, S. Fienberg, K. Krickeberg, I. Olkin, N. Wermuth, & S. Zeger, Eds.) (2nd ed.). New York, NY: Springer Statistics.

Kaimann, D. (2012). *To infinity and beyond! - A genre-specific film analysis of movie success mechanisms*. University of Paderborn.

Kaiser, H. F. (1958). The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, *23*(3), 187–200. doi:10.1007/BF02289233

Kaiser, H., & Rice, J. (1974). Little Jiffy, Mark IV. *Educational and Psychological Measurement*, *34*(1), 111–117. doi:10.1177/001316447403400115

Kenigsberg, B. (2013). Summer of the Mega-Flop. *Slate - Financing Update*. Retrieved November 25, 2013, from http://www.slate.com/articles/business/moneybox/2013/07/steven_spielberg_hollywood_imploding_how_he_predicted_a_disastrous_summer.html

Kerr, C. (1990). Incorporating the star: The intersection of business and aesthetic strategies in early American film. *Business History Review*, *64*(3), 383–410. Retrieved from http://journals.cambridge.org/production/action/cjoGetFulltext?fulltextid=8339413

Kinon, C. (2009). The most watched TV episode of the decade was the series finale of "Friends." *Daily News New York*. Retrieved November 30, 2013, from http://www.nydailynews.com/entertainment/tv-movies/watched-tv-episode-decade-series-finale-friends-article-1.431833

Kreyszig, E. (2006). *Advanced Engineering Mathematics*. (D. Grace, Ed.) (9th ed.). Columbus, OH: John Wiley & Sons, Inc.

Lee, J., & Holt, R. (2005). *The Producer's Business Handbook: The Roadmap for the Balanced Film Producer*. (A. Gillen, Ed.) (3rd ed.). Burlington, MA: Focal Press, Elsevier Group.

Levy, E. (1987). *And the Winner is.. The History and Politics of the Oscar Awards.* (1st ed., p. 330). New York, NY: Frederick Ungar Publishing Company.

Lincoln, A. E., & Allen, M. P. (2004). Double Jeopardy in Hollywood: Age and Gender in the Careers of Film Actors 1926-1999. *Sociological Forum*, *19*(4), 611–631. doi:10.1007/s11206-004-0698-1

Litman, B. R. (1983). Predicting Success of Theatrical Movies: An Empirical Study. *The Journal of Popular Culture*, *16*(4), 159–175. doi:10.1111/j.0022-3840.1983.1604_159.x

Litman, B. R. (1998). The Current Structure of the Industry. In B. R. Litman (Ed.), *The Motion Picture Mega-Industry* (1st ed., pp. 22–43). New York, NY: Allyn & Bacon, Pearson Education.

Litman, B. R., & Ahn, H. (1998). Predicting Financial Success of Motion Pictures: The Early 90s Experience. In B. R. Litman (Ed.), *The Motion Picture Mega-Industry* (1st ed., pp. 172–197). New York, NY: Allyn & Bacon, Pearson Education.

Marshall, A. (1890). *Principles of Economics* (8th ed.). London: Macmillan & Co., Limited.

McDonald, P. (2000). *The Star System: Hollywood's Production of Popular Identities* (2nd ed.). New York, NY: Wallflower Press.

McKenzie, J. (2009). Revealed word-of-mouth demand and adaptive supply: survival of motion pictures at the Australian box office. *Journal of Cultural Economics*, *33*(4), 279–299.

Mestyán, M., Yasseri, T., & Kertész, J. (2013). Early prediction of movie box office success based on wikipedia activity big data. *PLOS one*, *8*(8), 1 – 8. doi:10.1371/journal.pone.0071226

Mill, J. S. (1861). *Utilitarianism*. London, United Kingdom: Fraser's Magazine.

Mishne, G., & Glance, N. (2006). Predicting Movie Sales from Blogger Sentiment. In *Proceedings of AAAI 2006 Spring Symposium on Computational Approaches to Analysing Weblogs AAAI-CAAW* (pp. 155–158).

Morris, J. (2005). Sony Confesses Phony Patriot Plugs, Faces Lawsuit. *ABC News*. Retrieved October 09, 2013, from http://abcnews.go.com/Entertainment/story?id=104207&page=1

Mosk, R. (1997). The Jurisprudence of Ratings Symposium Part I: Motion Picture Ratings in the United States. *Cardozo Arts & Entertainment Law Journal*, *15*, 135–145.

MPAA. (2008). *MPAA Theatrical Market Statistics 2007* (p. 9).

MPAA. (2013a). MPAA Consumer Filmratings. *Motion Picture Association of America Inc.* Retrieved September 10, 2013, from www.filmratings.com

MPAA. (2013b). *MPAA Theatrical Statistics Summary 2012*.

MPAA. (2013c). MPAA - What Our Ratings Mean. *Motion Picture Association of America Inc.* Retrieved September 20, 2013, from www.mpaa.org/ratings/what-each-rating-means

Munoz, L. (2006). Budget Overruns of Biblical Proportions. *L.A. Times*. Retrieved October 12, 2013, from http://articles.latimes.com/2006/oct/09/business/fi-evan9

Nash. (2013). All Time Best-Selling Blu-ray Titles in the United States. *The Numbers Online Database*. Retrieved November 14, 2013, from http://www.the-numbers.com/alltime-bluray-sales-chart

Natale, R. (1999). Company Town : Company Town Film Profit Report. *L.A. Times*. Retrieved September 20, 2013, from www.articles.latimes.com/1999/sep/08/business/fi-7832

Neelamegham, R., & Chintagunta, P. (1999). A Bayesian Model to Forecast New Product Performance in Domestic and International Markets. *Marketing Science*, *18*(2), 115–136. doi:10.1287/mksc.18.2.115

Nelson, R., Donihue, M., Waldman, D., & Wheaton, C. (2001). What's an Oscar worth? *Economic Inquiry*, *39*(1), 1–16. Retrieved from http://onlinelibrary.wiley.com/doi/10.1111/j.1465-7295.2001.tb00046.x/full

O'Brien, R. M. (2007). A Caution Regarding Rules of Thumb for Variance Inflation Factors. *Quality & Quantity*, *41*(5), 673–690. doi:10.1007/s11135-006-9018-6

O'Neil, T. (2010). Meryl Streep sets new record with 16th Oscar nomination. *L.A. Times*. Retrieved November 27, 2013, from http://goldderby.latimes.com/awards_goldderby/2010/02/meryl-streep-adds-to-record-at-oscars-with-nod-no-16.html

O'Neill, P. (2007). Evan Almighty. *The Guardian Online Edition*. Retrieved November 27, 2013, from http://www.theguardian.com/film/2007/aug/03/morganfreeman.stevecarrell

Oghina, A., Breuss, M., Tsagkias, M., & Rijke, M. De. (2012). Predicting IMDB Movie Ratings Using Social Media. In *ECIR 2012: 34th European Conference on Information Retrieval*. (pp. 503–507).

Olive, D. (2012). John Carter syndrome: Why big budget Hollywood movies are financial busts. *The Star Online Edition*. Retrieved November 25, 2013, from http://www.thestar.com/business/2012/03/23/john_carter_syndrome_why_big_budget_hollywood_movies_are_financial_busts.html

Panaligan, R., & Chen, A. (2013). Quantifying Movie Magic with Google Search. *Google Whitepaper - Industry Perspectives and User Insights*, *1*(June), 1–11.

Pomerantz, D. (2010). Hollywood's New Deal. *Forbes Magazine Online Edition*. Retrieved December 05, 2013, from http://www.forbes.com/2010/03/23/actor-pay-hollywood-business-entertainment-movies.html

Pomerantz, D. (2012). The Biggest Box Office Flops Of 2012. *Forbes Magazine Online Edition*. Retrieved September 20, 2013, from http://www.forbes.com/sites/dorothypomerantz/2012/11/14/2012s-biggest-movie-turkeys/

Pomerantz, D. (2013). Will The Golden Globes Help "Argo" Win Best Picture? *Forbes Magazine Online Edition*. Retrieved December 05, 2013, from http://www.forbes.com/sites/dorothypomerantz/2013/01/14/will-the-golden-globes-help-argo-win-best-picture/

Popik, Z. (2011). *The Film Industry Award Series - Corroboration or Competition*. Erasmus University Rotterdam.

Prag, J., & 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), 217–235. doi:10.1007/BF01080227

Radas, S., & Shugan, S. M. (1998). Seasonal Marketing and Timing New Product Introductions. *Journal of Marketing Research*, *35*(3), 296–315.

Ravid, S. A. (1999). Information , Blockbusters , and Stars : A Study of the Film Industry. *The Journal of Business*, *72*(4), 463–492.

Reddy, S. S. A., Kasat, P., & Jain, A. (2012). Box-Office Opening Prediction of Movies based on Hype Analysis through Data Mining. *International Journal of Computer Applications, 56*(1), 1–5. Retrieved from http://adsabs.harvard.edu/abs/2012IJCA...56a...1S

Redmond, S., & Holmes, S. (2007). *Stardom and Celebrity: A Reader*. (S. Redmond & S. Holmes, Eds.) (1st ed., pp. 1–13). SAGE Publications Ltd.

RIAJ. (2013). *RIAJ Yearbook 2013 The Recording Industry in Japan*. Retrieved from www.riaj.or.jp/e/issue/pdf/RIAJ2013E.pdf

Roberts, R. (2009). A Scanner Darkly. *Film Journal International*. Retrieved November 10, 2013, from http://www.filmjournal.com/filmjournal/esearch/article_display.jsp?vnu_content_id=1002801068

Rose, L. (2008). Hollywood's Best-Paid Actors. *Forbes Magazine Online Edition*. Retrieved October 10, 2013, from http://www.forbes.com/2008/07/22/actors-hollywood-movies-biz-media-cx_lr_0722actors.html

Rosen, Sherwein. (2007). The Economics of Superstars Sherwin Rosen. *American, The Review, Economic Dec, No*, *71*(5), 845–858.

Rosen, Sherwin. (1981). The economics of superstars. *The American economic review*, *71*(5), 845–858. Retrieved from http://www.jstor.org/stable/10.2307/1803469

Ryan, R., & Deci, E. (2000). Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary educational psychology*, *25*(1), 54–67. doi:10.1006/ceps.1999.1020

Salganik, M. J., Dodds, P. S., & Watts, D. J. (2006). Experimental study of inequality and unpredictability in an artificial cultural market. *Science (New York, N.Y.)*, *311*(5762), 854–6.

Schumacher, M. (2000). *Francis Ford Coppola: A Filmmaker's Life* (2nd ed.). London: Bloomsbury Publishing.

Sharda, R., & Delen, D. (2006). Predicting box-office success of motion pictures with neural networks. *Expert Systems with Applications*, *30*(2), 243–254. doi:10.1016/j.eswa.2005.07.018

Shin, Y. (2013). Ya Shin's Award List. *Ya Shin's List*. Retrieved October 20, 2013, from http://www.ya-shin.com/awards/awards.html

Simonoff, J., & Sparrow, I. (2000). Predicting movie grosses: Winners and losers, blockbusters and sleepers. *Chance*, *13*(3), 15–24.

Smith, S., & Smith, V. (1986). Successful movies: A preliminary empirical analysis. *Applied Economics*, *18*(5), 501–507. Retrieved from http://www.tandfonline.com/doi/abs/10.1080/00036848608537445

Smith, W. (2005). *Experimental Design for Formulation: Statistics and Applied Probability* (1st ed., p. 196). Alexandria, Va: SIAM: Society for Industrial and Applied Mathematics.

Sochay, S. (1994). Predicting the Performance of Motion Pictures. *Journal of Media Economics*, *7*(4), 1–20. doi:10.1207/s15327736me0704_1

Sperling, N. (2011). MPAA awards PG-13 rating to alternative version of "The King"s Speech'. *L.A. Times*. Retrieved October 02, 2013, from www.latimesblogs.latimes.com/awards/2011/02/mpaa-awards-pg-13-rating-to-alternate-version-of-the-kings-speech.html

Spoto, D. (2001). *Marilyn Monroe: The Biography* (1st ed.). New York, NY: Cooper Square Press.

Stigler, G., & Becker, G. S. (1977). De Gustibus Non Est Disputandum. *The American Economic Review*, *67*(2), 76–90.

Tavakol, M., & Dennick, R. (2011). Making sense of Cronbach's alpha. *International Journal of Medical Education*, *2*, 53–55.

Terry, N., Butler, M., & De'Armond, D. (2003). The Determinants of Domestic Box Office Performance in the Motion Pictury Industry. *Southwestern Economic Review*, *32*(1), 137–148.

The Numbers. (2013a). All Time Box Office Records. *The Numbers Online Database*. Retrieved September 20, 2013, from www.the-numbers.com/movies/records/#world

The Numbers. (2013b). Domestic Movie Theatrical Market Summary 1995 to 2013. *The Numbers Online Database*. Retrieved September 20, 2013, from http://www.the-numbers.com/market/

The Numbers. (2013c). The Numbers Movie Database. *Nash Information Services, LLC*. Retrieved September 10, 2013, from www.the-numbers.com/

The Numbers. (2013d). Box Office History for Digital Animation Movies. *The Numbers Online Database*. Retrieved November 23, 2013, from http://www.the-numbers.com/market/production-method/Digital-Animation

The Numbers. (2013e). All-Time Best-Selling Blu-ray Titles in the United States. *The Numbers Online Database*. Retrieved November 20, 2013, from http://www.the-numbers.com/alltime-bluray-sales-chart

Thompson, P. (2011). Nice little earner: Johnny Depp signs on to fifth Pirates movie as it’s revealed he's earned £200m from the franchise. *The Daily Mail Online Edition*. Retrieved October 11, 2013, from http://www.dailymail.co.uk/tvshowbiz/article-2012091/Johnny-Depp-signs-Pirates-5-revealed-hes-earned-200m-franchise.html

Tussu, D. K. (2008). The Globalization of Bollywood. In *Global Hollywood* (2nd ed., pp. 97–117). British Film Institute.

Ulmer, J. (2013). Ulmer Scale. Retrieved September 20, 2013, from www.ulmerscale.com

Uri, D., & Dholakia, N. (2012). Hype , Hope , and Hit in Movies : A Contribution to the Metatheory of Bubbles Proceedings of the 37th Macromarketing Conference. In N. Dholakia & R. Turcan (Eds.), *37th Macromarketing Conference* (pp. 202–206). Berlin.

Voderer, P., Steen, F., & Chan, E. (2006). View on Motivation. In J. Bryant & P. Voderer (Eds.), *Psychology of Entertainment* (1st ed., pp. 2–17). Mahwah, NJ: Lawrence Erlbaum Associates.

Wallace, W. T., Seigerman, A., & Holbrook, M. B. (1993). The role of actors and actresses in the success of films: how much is a movie star worth? *Journal of Cultural Economics*, *17*(1), 1–27. doi:10.1007/BF00820765

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

Wang, H. (2009). Forward Regression for Ultra-High Dimensional Variable Screening. *Journal of the American Statistical Association*, *104*(488), 1512–1524. doi:10.1198/jasa.2008.tm08516

Waterman, D., & Jayakar, K. P. (2000). The Competitive Balance of the Italian and American Film Industries. *European Journal of Communication*, *15*(4), 501–528. doi:10.1177/0267323100015004003

Weber, M. (1922). *Wirtschaft und Gesellschaft*. (P. Siebeck, Ed.). Tübingen: Mohr.

Weinstein, P. B. (2001). Movies as the Gateway to History: The History and Film Project. *The History Teacher*, *35*(1), 27–48.

Weintraub, B. (1995). “Waterworld” Disappointment As Box Office Receipts Lag. *The New York Times*. Retrieved September 20, 2013, from http://www.nytimes.com/1995/07/31/business/waterworld-disappointment-as-box-office-receipts-lag.html

Wong, F. M. F., Sen, S., & Chiang, M. (2012). Why watching movie tweets won’t tell the whole story? In *Proceedings of the 2012 ACM workshop on Workshop on online social networks*

- *WOSN '12* (pp. 61–66). New York, New York, USA: ACM Press. doi:10.1145/2342549.2342564

Yong, L. (2006). Word of Mouth for Movies : Its and Impact on Box Office Dynamics Revenue. *Journal of Marketing*, *70*(3), 74–89.

Yudelson, J. (2010). The impact of actors and producers in studio-financed movie deals. *Journal of Behavioral Studies in Business*. Retrieved from www.aabri.com/manuscripts/10698.pdf

Zeidler, S. (2013). Global box office 2012 sales up 6 percent; China sales spike. *Reuters Online Edition U.S.* Retrieved October 20, 2013, from http://www.reuters.com/article/2013/03/21/entertainment-us-movies-boxoffice-idUSBRE92K1D020130321

Zeitschik, S. (2013). Monsters University: Is Pixar's quality-success link over? *L.A. Times*. Retrieved from http://www.latimes.com/entertainment/movies/moviesnow/la-et-mn-monsters-university-box-office-pixar-20130624,0,4404887.story

Appendix

Descriptive statistics of scaled variables

Descriptive Statistics

	N	Range	Mean	Std. Deviation
film_budget_adj	410	223,6819	47,020954	36,9378325
star_avg_age_at_release	410	58,0000	39,361412	9,0953454
star_percentage_of_stars_male	410	100,00%	65,9623%	32,40277%
star_no_stars	410	10	2,75	1,841
star_cc_factor	410	7,07385	-,0173222	1,08051519
star_talent_factor	410	7,76270	-,0024668	1,01688745
film_prequels_number	410	20	,16	1,051
film_prequels_box_office	410	3901,669	37,90302	232,136672
film_box_global_adj	410	972,6713	106,446105	123,1290943
star_tal_academy_noms_ONLY	410	17	2,04	3,031
star_tal_academy_wins_ONLY	410	5	,64	,971
star_tal_gg_noms_ONLY	410	28	4,27	4,748
star_tal_gg_wins_ONLY	410	9	1,30	1,864
star_tal_BAFTA_wins	410	5	,44	,858
star_tal_berlinale_wins	410	2	,12	,391
star_tal_cannes_wins	410	2	,11	,361
star_tal_MTV_wins_ONLY	410	10	,80	1,425
star_cc_bo_prior_to_release	410	8629538949		
star_cc_google_news_entries	410	128015	18121,35	18086,727
star_cc_kum_cinemas_prior	410	283900	58546,86	43549,260
star_cc_prior_appearances	410	202	44,78	35,233
Valid N (listwise)	410			

film_date_year

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	2003	85	20,7	20,7	20,7
	2004	81	19,8	19,8	40,5
	2005	88	21,5	21,5	62,0
	2006	76	18,5	18,5	80,5
	2007	80	19,5	19,5	100,0
	Total	410	100,0	100,0	

film_date_month

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	22	5,4	5,4	5,4
	2	27	6,6	6,6	12,0
	3	43	10,5	10,5	22,4
	4	33	8,0	8,0	30,5
	5	20	4,9	4,9	35,4
	6	27	6,6	6,6	42,0
	7	35	8,5	8,5	50,5
	8	30	7,3	7,3	57,8
	9	39	9,5	9,5	67,3
	10	38	9,3	9,3	76,6
	11	41	10,0	10,0	86,6
	12	55	13,4	13,4	100,0
	Total	410	100,0	100,0	

film_date_thanks_giving

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	394	96,1	96,1	96,1
	1	16	3,9	3,9	100,0
	Total	410	100,0	100,0	

film_date_boxing_d

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	384	93,7	93,7	93,7
	1	26	6,3	6,3	100,0
	Total	410	100,0	100,0	

film_date_friday_release

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	60	14,6	14,6	14,6
	1	350	85,4	85,4	100,0
	Total	410	100,0	100,0	

film_mpaa

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	G	3	,7	,7	,7
	NR	2	,5	,5	1,2
	PG	73	17,8	17,8	19,0
	PG-13	195	47,6	47,6	66,6
	R	137	33,4	33,4	100,0
	Total	410	100,0	100,0	

film_genre

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	action	74	18,0	18,0	18,0
	adventure	31	7,6	7,6	25,6
	animation	12	2,9	2,9	28,5
	comedy	138	33,7	33,7	62,2
	drama	94	22,9	22,9	85,1
	horror	10	2,4	2,4	87,6
	music	1	,2	,2	87,8
	romance	13	3,2	3,2	91,0
	thriller	37	9,0	9,0	100,0
	Total	410	100,0	100,0	

film_genre_2

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		32	7,8	7,8	7,8
	Action	26	6,3	6,3	14,1
	Adventure	43	10,5	10,5	24,6
	Adventures	1	,2	,2	24,9
	Animation	2	,5	,5	25,4
	Biography	10	2,4	2,4	27,8
	Comedy	39	9,5	9,5	37,3
	Crime	34	8,3	8,3	45,6
	Drama	87	21,2	21,2	66,8
	Family	21	5,1	5,1	72,0
	Fantasy	14	3,4	3,4	75,4
	History	6	1,5	1,5	76,8
	Horror	5	1,2	1,2	78,0
	Music	9	2,2	2,2	80,2
	Musical	2	,5	,5	80,7
	Mystery	12	2,9	2,9	83,7
	Romance	40	9,8	9,8	93,4
	Sci-Fi	5	1,2	1,2	94,6
	Sport	6	1,5	1,5	96,1
	Suspense	5	1,2	1,2	97,3
	Thriller	10	2,4	2,4	99,8
	War	1	,2	,2	100,0
	Total	410	100,0	100,0	

film_genre_3

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		126	30,7	30,7	30,7
	Action	2	,5	,5	31,2
	Adventure	4	1,0	1,0	32,2
	Animation	1	,2	,2	32,4
	Biography	2	,5	,5	32,9
	Comedy	22	5,4	5,4	38,3
	Crime	28	6,8	6,8	45,1
	Drama	41	10,0	10,0	55,1
	Family	21	5,1	5,1	60,2
	Fantasy	15	3,7	3,7	63,9
	History	5	1,2	1,2	65,1
	Horror	8	2,0	2,0	67,1
	Music	2	,5	,5	67,6
	Musical	2	,5	,5	68,0
	Mystery	17	4,1	4,1	72,2
	Romance	52	12,7	12,7	84,9
	Sci-Fi	10	2,4	2,4	87,3
	Sport	11	2,7	2,7	90,0
	Thriller	36	8,8	8,8	98,8
	War	4	1,0	1,0	99,8
	Western	1	,2	,2	100,0
	Total	410	100,0	100,0	

film_genre_4

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid		269	65,6	65,6	65,6
	Adventure	2	,5	,5	66,1
	Comedy	1	,2	,2	66,3
	Crime	4	1,0	1,0	67,3
	Drama	9	2,2	2,2	69,5
	Family	15	3,7	3,7	73,2
	Fantasy	11	2,7	2,7	75,9
	History	3	,7	,7	76,6
	Horror	1	,2	,2	76,8
	Music	1	,2	,2	77,1
	Musical	1	,2	,2	77,3
	Mystery	12	2,9	2,9	80,2
	Romance	7	1,7	1,7	82,0
	Sci-Fi	8	2,0	2,0	83,9
	Science Fiction	2	,5	,5	84,4
	Sport	6	1,5	1,5	85,9
	Thriller	51	12,4	12,4	98,3
	War	5	1,2	1,2	99,5
	Western	2	,5	,5	100,0
	Total	410	100,0	100,0	

film_studio_para

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	366	89,3	89,3	89,3
	1	44	10,7	10,7	100,0
	Total	410	100,0	100,0	

film_studio_sony

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	367	89,5	89,5	89,5
	1	43	10,5	10,5	100,0
	Total	410	100,0	100,0	

film_studio_dream

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	376	91,7	91,7	91,7
	1	34	8,3	8,3	100,0
	Total	410	100,0	100,0	

film_studio_mgm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	385	93,9	93,9	93,9
	1	25	6,1	6,1	100,0
	Total	410	100,0	100,0	

film_studio_fox

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	381	92,9	92,9	92,9
	1	29	7,1	7,1	100,0
	Total	410	100,0	100,0	

film_studio_warner

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	351	85,6	85,6	85,6
	1	59	14,4	14,4	100,0
	Total	410	100,0	100,0	

film_studio_disney

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	361	88,0	88,0	88,0
	1	49	12,0	12,0	100,0
	Total	410	100,0	100,0	

film_studio_univers

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	363	88,5	88,5	88,5
	1	47	11,5	11,5	100,0
	Total	410	100,0	100,0	

film_studio_paramm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	360	87,8	87,8	87,8
	1	50	12,2	12,2	100,0
	Total	410	100,0	100,0	

film_studio_sonymm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	358	87,3	87,3	87,3
	1	52	12,7	12,7	100,0
	Total	410	100,0	100,0	

film_studio_dreammm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	376	91,7	91,7	91,7
	1	34	8,3	8,3	100,0
	Total	410	100,0	100,0	

film_studio_mgmmm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	380	92,7	92,7	92,7
	1	30	7,3	7,3	100,0
	Total	410	100,0	100,0	

film_studio_foxmm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	362	88,3	88,3	88,3
	1	48	11,7	11,7	100,0
	Total	410	100,0	100,0	

film_studio_warnermm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	308	75,1	75,1	75,1
	1	102	24,9	24,9	100,0
	Total	410	100,0	100,0	

film_studio_disneymm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	332	81,0	81,0	81,0
	1	78	19,0	19,0	100,0
	Total	410	100,0	100,0	

film_studio_universmm

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	348	84,9	84,9	84,9
	1	62	15,1	15,1	100,0
	Total	410	100,0	100,0	

film_based_book_short_story

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	322	78,5	78,5	78,5
	1	88	21,5	21,5	100,0
	Total	410	100,0	100,0	

film_based_comic_gn

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	399	97,3	97,3	97,3
	1	11	2,7	2,7	100,0
	Total	410	100,0	100,0	

film_based_magazine_article

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	409	99,8	99,8	99,8
	1	1	,2	,2	100,0
	Total	410	100,0	100,0	

film_based_musical_opera

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	408	99,5	99,5	99,5
	1	2	,5	,5	100,0
	Total	410	100,0	100,0	

film_based_play

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	406	99,0	99,0	99,0
	1	4	1,0	1,0	100,0
	Total	410	100,0	100,0	

film_based_tv

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	400	97,6	97,6	97,6
	1	10	2,4	2,4	100,0
	Total	410	100,0	100,0	

film_based_original_screenplay

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	218	53,2	53,2	53,2
	1	192	46,8	46,8	100,0
	Total	410	100,0	100,0	

film_based_sequel

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	371	90,5	90,5	90,5
	1	39	9,5	9,5	100,0
	Total	410	100,0	100,0	

film_based_traditional_legend_fairytale

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0	406	99,0	99,0	99,0
	1	4	1,0	1,0	100,0
	Total	410	100,0	100,0	

Film_based_historic_eventsx

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	,000	388	94,6	94,6	94,6
	1,000	22	5,4	5,4	100,0
	Total	410	100,0	100,0	

Factor analysis outputs

Communalities

	Initial	Extraction
star_cc_google_eng_media_presence_12m_prior	1,000	,773
star_cc_bo_prior_to_release	1,000	,937
star_cc_kum_cinemas_prior	1,000	,963
star_cc_prior_appearances_on_BO_mojo	1,000	,926
star_tal_gg_wins_ONLY	1,000	,846
star_tal_gg_noms_ONLY	1,000	,817
star_tal_academy_wins_ONLY	1,000	,788
star_tal_academy_noms_ONLY	1,000	,840

Extraction Method: Principal Component Analysis.

Total Variance Explained

Component	Initial Eigenvalues			Extraction Sums of Squared Loadings			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	5,790	72,380	72,380	5,790	72,380	72,380	3,528	44,098	44,098
2	1,099	13,741	86,121	1,099	13,741	86,121	3,362	42,023	86,121
3	,367	4,592	90,713						
4	,306	3,827	94,540						
5	,181	2,266	96,805						
6	,127	1,585	98,390						
7	,091	1,139	99,530						
8	,038	,470	100,000						

Extraction Method: Principal Component Analysis.

Component Matrix[a]

	Component	
	1	2
star_cc_google_eng_media_presence_12m_prior	,879	,003
star_cc_bo_prior_to_release	,871	-,421
star_cc_kum_cinemas_prior	,845	-,500
star_cc_prior_appearances_on_BO_mojo	,879	-,393
star_tal_gg_wins_ONLY	,818	,420
star_tal_gg_noms_ONLY	,885	,184
star_tal_academy_wins_ONLY	,770	,442
star_tal_academy_noms_ONLY	,853	,335

Extraction Method: Principal Component Analysis.

a. 2 components extracted.

Rotated Component Matrix[a]

	Component	
	1	2
star_cc_google_eng_media_presence_12m_prior	,437	,778
star_cc_bo_prior_to_release	,324	,908
star_cc_kum_cinemas_prior	,270	,946
star_cc_prior_appearances_on_BO_mojo	,329	,892
star_tal_gg_wins_ONLY	,880	,266
star_tal_gg_noms_ONLY	,764	,443
star_tal_academy_wins_ONLY	,857	,224
star_tal_academy_noms_ONLY	,844	,321

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.

a. Rotation converged in 3 iterations.

KMO and Bartlett's Test

Kaiser-Meyer-Olkin Measure of Sampling Adequacy.		,872
Bartlett's Test of Sphericity	Approx. Chi-Square	3966,933
	df	28
	Sig.	,000

Descriptive Statistics

	N	Mean	Std. Deviation
star_cc_google_eng_media_presence_12m_prior	410	18121,35	18086,727
star_cc_bo_prior_to_release	410	1862706807	1422139040
star_cc_kum_cinemas_prior	410	58546,86	43549,260
star_cc_prior_appearances_on_BO_mojo	410	44,78	35,233
star_tal_gg_wins_ONLY	410	1,30	1,864
star_tal_gg_noms_ONLY	410	4,27	4,748
star_tal_academy_wins_ONLY	410	,64	,971
star_tal_academy_noms_ONLY	410	2,04	3,031
Valid N (listwise)	410		

Reliability test of factor consumption capital

Case Processing Summary

		N	%
Cases	Valid	410	100,0
	Excluded[a]	0	,0
	Total	410	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,000	,949	4

Inter-Item Correlation Matrix

	star_cc_google_eng_media_presence_12m_prior	star_cc_bo_prior_to_release	star_cc_kum_cinemas_prior	star_cc_prior_appearances_on_BO_mojo
star_cc_google_eng_media_presence_12m_prior	1,000	,758	,701	,711
star_cc_bo_prior_to_release	,758	1,000	,936	,891
star_cc_kum_cinemas_prior	,701	,936	1,000	,939
star_cc_prior_appearances_on_BO_mojo	,711	,891	,939	1,000

Reliability test of factor talent

Case Processing Summary

		N	%
Cases	Valid	410	100,0
	Excluded[a]	0	,0
	Total	410	100,0

a. Listwise deletion based on all variables in the procedure.

Reliability Statistics

Cronbach's Alpha	Cronbach's Alpha Based on Standardized Items	N of Items
,830	,925	4

Inter-Item Correlation Matrix

	star_tal_gg_wins_ONLY	star_tal_gg_noms_ONLY	star_tal_academy_wins_ONLY	star_tal_academy_noms_ONLY
star_tal_gg_wins_ONLY	1,000	,733	,785	,762
star_tal_gg_noms_ONLY	,733	1,000	,664	,862
star_tal_academy_wins_ONLY	,785	,664	1,000	,725
star_tal_academy_noms_ONLY	,762	,862	,725	1,000

Correlations among award variables

Correlations

		1	2	3	4	5	6	7	8	9	10	11	12
1 star_tal_gg_wins_ONLY	Pearson Correlation	1	,73**	,86**	,79**	,76**	,81**	,40**	,24**	,39**	,33**	,26**	,55**
	Sig. (2-tailed)		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
2 star_tal_gg_noms_ONLY	Pearson Correlation	,73**	1	,98**	,66**	,86**	,86**	,33**	,23**	,33**	,42**	,26**	,59**
	Sig. (2-tailed)	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
3 star_tal_gg_wins_AND_noms	Pearson Correlation	,86**	,98**	1	,74**	,88**	,89**	,37**	,25**	,37**	,41**	,28**	,61**
	Sig. (2-tailed)	,000	,000		,000	,000	,000	,000	,000	,000	,000	,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
4 star_tal_academy_wins_ONLY	Pearson Correlation	,79**	,66**	,74**	1	,72**	,83**	,31**	,11*	,26**	,36**	,36**	,61**
	Sig. (2-tailed)	,000	,000	,000		,000	,000	,000	,021	,000	,000	,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
5 star_tal_academy_noms_ONLY	Pearson Correlation	,76**	,86**	,88**	,72**	1	,98**	,29**	,16**	,27**	,48**	,28**	,66**
	Sig. (2-tailed)	,000	,000	,000	,000		,000	,000	,001	,000	,000	,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
6 star_tal_academy_noms_AND_wins	Pearson Correlation	,81**	,86**	,89**	,83**	,98**	1	,31**	,16**	,28**	,48**	,31**	,68**
	Sig. (2-tailed)	,000	,000	,000	,000	,000		,000	,001	,000	,000	,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
7 star_tal_MTV_noms_ONLY	Pearson Correlation	,40**	,33**	,37**	,31**	,29**	,31**	1	,49**	,91**	,088	,21**	,20**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000		,000	,000	,077	,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
8 star_tal_MTV_wins_ONLY	Pearson Correlation	,24**	,23**	,25**	,11*	,16**	,16**	,49**	1	,81**	-,01	,054	,079
	Sig. (2-tailed)	,000	,000	,000	,021	,001	,001	,000		,000	,916	,275	,112
	N	410	410	410	410	410	410	410	410	410	410	410	410
9 star_tal_MTV_wins_AND_noms	Pearson Correlation	,39**	,33**	,37**	,26**	,27**	,28**	,91**	,81**	1	,056	,17**	,17**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,000		,255	,001	,001
	N	410	410	410	410	410	410	410	410	410	410	410	410
10 star_tal_cannes_wins	Pearson Correlation	,33**	,42**	,41**	,36**	,48**	,48**	,088	-,01	,056	1	,23**	,28**
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,077	,916	,255		,000	,000
	N	410	410	410	410	410	410	410	410	410	410	410	410
11 star_tal_berlinale_wins	Pearson Correlation	,26**	,26**	,28**	,36**	,28**	,31**	,21**	,054	,17**	,23**	1	,085
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,275	,001	,000		,084
	N	410	410	410	410	410	410	410	410	410	410	410	410
12 star_tal_BAFTA_wins	Pearson Correlation	,55**	,59**	,61**	,61**	,66**	,68**	,20**	,079	,17**	,28**	,085	1
	Sig. (2-tailed)	,000	,000	,000	,000	,000	,000	,000	,112	,001	,000	,084	
	N	410	410	410	410	410	410	410	410	410	410	410	410

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Correlations among consumption capital variables

Correlations

		star_cc_google_eng_media_presence_12m_prior	star_cc_bo_prior_to_release	star_cc_kum_cinemas_prior	star_cc_prior_appearances_on_BO_mojo	star_cc_av_cinemas_prior_films
star_cc_google_eng_media_presence_12m_prior	Pearson Correlation	1	,758**	,701**	,711**	,476**
	Sig. (2-tailed)		,000	,000	,000	,000
	N	410	410	410	410	410
star_cc_bo_prior_to_release	Pearson Correlation	,758**	1	,936**	,891**	,713**
	Sig. (2-tailed)	,000		,000	,000	,000
	N	410	410	410	410	410
star_cc_kum_cinemas_prior	Pearson Correlation	,701**	,936**	1	,939**	,797**
	Sig. (2-tailed)	,000	,000		,000	,000
	N	410	410	410	410	410
star_cc_prior_appearances_on_BO_mojo	Pearson Correlation	,711**	,891**	,939**	1	,697**
	Sig. (2-tailed)	,000	,000	,000		,000
	N	410	410	410	410	410
star_cc_average_cinemas_prior_films (NOT USED)	Pearson Correlation	,476**	,713**	,797**	,697**	1
	Sig. (2-tailed)	,000	,000	,000	,000	
	N	410	410	410	410	410

**. Correlation is significant at the 0.01 level (2-tailed).

Correlation matrix of scaled variables in final model

Correlations

		1	2	3	4	5	6	7
(1) film_budget_adj	Pearson Correlation	1	,076	,001	,350**	,034	-,081	,200**
	Sig. (2-tailed)		,124	,985	,000	,496	,101	,000
	N	410	410	410	410	410	410	410
(2) star_cc_factor	Pearson Correlation	,076	1	-,012	,044	,309**	,520**	,127*
	Sig. (2-tailed)	,124		,801	,370	,000	,000	,010
	N	410	410	410	410	410	410	410
(3) star_talent_factor	Pearson Correlation	,001	-,012	1	-,007	,255**	,222**	-,117*
	Sig. (2-tailed)	,985	,801		,894	,000	,000	,018
	N	410	410	410	410	410	410	410
(4) film_prequels_box_office	Pearson Correlation	,350**	,044	-,007	1	,016	-,038	,077
	Sig. (2-tailed)	,000	,370	,894		,744	,441	,120
	N	410	410	410	410	410	410	410
(5) star_avg_age_at_release	Pearson Correlation	,034	,309**	,255**	,016	1	,079	,345**
	Sig. (2-tailed)	,496	,000	,000	,744		,109	,000
	N	410	410	410	410	410	410	410
(6) star_no_stars	Pearson Correlation	-,081	,520**	,222**	-,038	,079	1	-,077
	Sig. (2-tailed)	,101	,000	,000	,441	,109		,120
	N	410	410	410	410	410	410	410
(7) star_percentage_of_stars_male	Pearson Correlation	,200**	,127*	-,117*	,077	,345**	-,077	1
	Sig. (2-tailed)	,000	,010	,018	,120	,000	,120	
	N	410	410	410	410	410	410	410

**. Correlation is significant at the 0.01 level (2-tailed).

*. Correlation is significant at the 0.05 level (2-tailed).

Absence of effects of MPAA rating

Correlations

	film_mpaa_G	film_mpaa_NR	film_mpaa_PG	film_mpaa_PG_13	film_mpaa_R
film_budget_adj	,065	-,002	,015	,147	-,180
	,191	,971	,758	,003	,000
	410	410	410	410	410
film_prequels_box_office	,006	-,011	,000	,097	-,101
	,898	,817	,993	,051	,040
	410	410	410	410	410
film_based_historic_events	-,013	-,010	,017	-,009	,000
	,795	,832	,727	,850	,996
	410	410	410	410	410
film_genre_animation	,325	-,012	,222	-,136	-,092
	,000	,806	,000	,006	,062
	410	410	410	410	410
film_month_sep	-,028	-,023	-,042	,024	,017
	,574	,647	,394	,626	,730
	410	410	410	410	410
star_avg_age_at_release	,027	,149	-,020	-,064	,057
	,592	,003	,691	,197	,251
	410	410	410	410	410
star_cc_factor	-,014	,036	-,052	-,057	,099
	,771	,466	,294	,253	,045
	410	410	410	410	410
star_cc_factor_SQ	-,019	-,015	,070	-,015	-,035
	,700	,767	,158	,765	,476
	410	410	410	410	410
star_talent_factor	,011	,001	-,038	-,011	,040
	,830	,989	,446	,829	,420
	410	410	410	410	410
star_talent_factor_SQ	-,014	-,011	-,047	,014	,027
	,784	,820	,346	,772	,589
	410	410	410	410	410
star_no_stars	-,050	,010	-,071	-,052	,120
	,308	,845	,151	,294	,015
	410	410	410	410	410
star_percentage_of_stars_male	,002	,002	-,067	-,057	,114
	,970	,975	,176	,252	,021
	410	410	410	410	410

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	114,382	22,109		5,174	,000
	film_budget_adj	2,011	,113	,603	17,819	,000
	film_prequels_box_office	,098	,017	,185	5,823	,000
	film_genre_animation	71,174	24,030	,098	2,962	,003
	film_based_historic_events	-75,633	25,245	-,090	-2,996	,003
	film_month_sep	-36,481	12,561	-,087	-2,904	,004
	star_cc_factor	39,944	4,798	,351	8,324	,000
	star_cc_factor_SQ	-1,121	,481	-,084	-2,332	,020
	star_talent_factor	21,375	6,667	,177	3,206	,001
	star_talent_factor_SQ	-,151	,244	-,032	-,619	,536
	star_avg_age_at_release	-2,218	,494	-,164	-4,487	,000
	star_no_stars	-8,725	2,516	-,130	-3,467	,001
	star_percentage_of_stars_male	,256	,129	,067	1,983	,048
	film_mpaa_G	-32,227	45,771	-,022	-,704	,482
	film_mpaa_NR	-16,782	53,066	-,010	-,316	,752
	film_mpaa_PG	3,396	10,520	,011	,323	,747
	film_mpaa_R	-14,996	8,520	-,058	-1,760	,079

a. Dependent Variable: film_box_global_adj

Untrimmed regression model output

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,834[a]	,696	,639	73,9331104	2,015

a. Predictors: (Constant), film_studio_warnermm, film_year_2005, film_month_mar, film_based_magazine_article, film_genre_animation, star_percentage_of_stars_male, film_mpaa_NR, film_based_musical_opera, film_based_real_life_events, film_based_play, film_genre_music, film_based_tv, film_date_thanksgiving, film_prequels_box_office, film_based_historic_events, film_genre_romance, film_month_apr, film_genre_adventure, star_cc_factor_SQ, film_month_may, film_genre_thriller, star_talent_factor_SQ, film_month_aug, film_based_comic_gn, film_based_remake, film_studio_mgmmm, film_date_boxing_day, film_mpaa_PG_13, film_studio_disney, film_genre_horror, film_month_jun, film_based_traditional_legend_fairytale, film_year_2005, film_month_feb, film_studio_fox, film_month_jan, film_studio_sony, film_based_book_short_story, film_studio_para, film_mpaa_G, film_month_jul, film_year_2004, film_studio_dreammm, film_month_okt, film_genre_action, star_avg_age_at_release, film_studio_univers, star_no_stars, film_based_sequel, film_month_sep, film_mpaa_PG, film_year_2006, film_genre_drama, star_cc_factor, film_budget_adj, film_studio_warner, film_studio_disneymm, film_month_nov, film_studio_foxmm, star_talent_factor, film_studio_universmm, film_studio_mgm, film_studio_sonymm, film_studio_paramm

b. Dependent Variable: film_box_global_adj

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4314950,343	64	67421,099	12,334	,000
	Residual	1885806,162	345	5466,105		
	Total	6200756,505	409			

a. Dependent Variable: film_box_global_adj

Excluded Variables[a]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics
						Tolerance
1	film_month_dez	.	.	.	.	,000
	film_year_2003	.	.	.	.	,000
	film_based_original_screenplay	.	.	.	.	,000
	film_genre_comedy	.	.	.	.	,000
	film_mpaa_R	.	.	.	.	,000
	film_studio_dream	.	.	.	.	,000

a. Dependent Variable: film_box_global_adj

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.	95,0% Confidence Interval for B Lower Bound	95,0% Confidence Interval for B Upper Bound
1	(Constant)	139,210	32,121		4,334	,000	76,031	202,388
	film_budget_adj	1,938	,154	,582	12,619	,000	1,636	2,241
	film_prequels_box_office	,094	,020	,178	4,778	,000	,056	,133
	star_avg_age_at_release	-2,618	,539	-,193	-4,857	,000	-3,678	-1,558
	star_cc_factor	43,076	5,273	,378	8,169	,000	32,704	53,447
	star_cc_factor_SQ	-1,479	,508	-,111	-2,912	,004	-2,478	-,480
	star_talent_factor	23,742	7,420	,196	3,200	,002	9,148	38,336
	star_talent_factor_SQ	-,244	,267	-,051	-,915	,361	-,768	,280
	star_no_stars	-8,067	2,789	-,121	-2,892	,004	-13,554	-2,581
	star_percentage_of_stars_male	,230	,140	,061	1,641	,102	-,046	,506
	film_month_jan	-40,863	23,010	-,075	-1,776	,077	-86,121	4,395
	film_month_feb	-35,162	21,911	-,071	-1,605	,109	-78,257	7,934
	film_month_mar	-26,320	19,608	-,066	-1,342	,180	-64,887	12,246
	film_month_apr	-45,903	20,868	-,102	-2,200	,028	-86,948	-4,859
	film_month_may	-25,917	23,300	-,045	-1,112	,267	-71,744	19,911
	film_month_jun	-46,985	21,436	-,095	-2,192	,029	-89,146	-4,823
	film_month_jul	-8,351	19,835	-,019	-,421	,674	-47,364	30,663
	film_month_aug	-31,810	20,960	-,067	-1,518	,130	-73,036	9,415
	film_month_sep	-68,038	19,875	-,162	-3,423	,001	-107,130	-28,947
	film_month_okt	-41,919	20,119	-,099	-2,084	,038	-81,490	-2,348
	film_month_nov	-33,731	22,547	-,082	-1,496	,136	-78,077	10,615
	film_year_2004	-17,857	12,263	-,058	-1,456	,146	-41,978	6,263
	film_year_2005	-8,488	11,788	-,028	-,720	,472	-31,674	14,697
	film_year_2005	-8,708	12,712	-,028	-,685	,494	-33,711	16,295
	film_year_2006	3,837	12,767	,012	,301	,764	-21,275	28,949
	film_based_book_short_story	-9,135	10,923	-,030	-,836	,404	-30,618	12,349
	film_based_comic_gn	21,270	26,247	,028	,810	,418	-30,354	72,894
	film_based_historic_events	-70,708	27,956	-,084	-2,529	,012	-125,695	-15,722
	film_based_magazine_article	-,112	79,634	,000	-,001	,999	-156,740	156,517
	film_based_musical_opera	-34,903	57,307	-,020	-,609	,543	-147,619	77,812
	film_based_play	18,790	40,502	,015	,464	,643	-60,871	98,451
	film_based_real_life_events	10,040	23,914	,014	,420	,675	-36,995	57,075
	film_based_remake	-3,528	14,596	-,008	-,242	,809	-32,236	25,181
	film_based_sequel	24,077	16,698	,057	1,442	,150	-8,767	56,920
	film_based_traditional_legend_fairytale	-38,069	41,216	-,030	-,924	,356	-119,136	42,998
	film_based_tv	13,692	25,501	,017	,537	,592	-36,464	63,849
	film_date_boxing_day	-17,944	22,280	-,036	-,805	,421	-61,766	25,878
	film_date_thanksgiving	-13,519	25,476	-,021	-,531	,596	-63,628	36,589
	film_genre_action	-10,049	13,249	-,031	-,758	,449	-36,108	16,010
	film_genre_adventure	-14,563	16,188	-,031	-,900	,369	-46,404	17,277
	film_genre_animation	71,432	27,816	,098	2,568	,011	16,722	126,142
	film_genre_drama	-13,066	12,472	-,045	-1,048	,296	-37,596	11,465
	film_genre_horror	-3,843	26,822	-,005	-,143	,886	-56,598	48,912
	film_genre_music	-95,099	79,683	-,038	-1,193	,234	-251,824	61,627
	film_genre_romance	24,870	23,278	,035	1,068	,286	-20,915	70,655
	film_genre_thriller	4,219	15,018	,010	,281	,779	-25,320	33,758
	film_mpaa_G	-59,515	51,037	-,041	-1,166	,244	-159,898	40,868
	film_mpaa_NR	-7,178	56,182	-,004	-,128	,898	-117,680	103,324
	film_mpaa_PG	15,213	13,421	,047	1,134	,258	-11,184	41,610
	film_mpaa_PG_13	13,899	9,799	,056	1,418	,157	-5,374	33,172
	film_studio_disney	44,854	19,291	,118	2,325	,021	6,911	82,797
	film_studio_disneymm	13,112	17,998	,042	,729	,467	-22,288	48,511
	film_studio_dreammm	15,608	17,529	,035	,890	,374	-18,869	50,086
	film_studio_fox	23,966	24,173	,050	,991	,322	-23,579	71,511
	film_studio_foxmm	21,584	23,378	,056	,923	,357	-24,397	67,565
	film_studio_mgm	-39,763	39,261	-,077	-1,013	,312	-116,985	37,459
	film_studio_mgmmm	34,800	38,157	,074	,912	,362	-40,250	109,850
	film_studio_para	-18,537	34,865	-,047	-,532	,595	-87,112	50,038
	film_studio_paramm	35,193	34,740	,094	1,013	,312	-33,137	103,522
	film_studio_sony	-30,497	29,621	-,076	-1,030	,304	-88,757	27,763
	film_studio_sonymm	27,490	30,130	,074	,912	,362	-31,772	86,751
	film_studio_univers	-2,818	23,880	-,007	-,118	,906	-49,786	44,150
	film_studio_universmm	21,968	23,476	,064	,936	,350	-24,207	68,143
	film_studio_warner	1,533	16,591	,004	,092	,926	-31,099	34,165
	film_studio_warnermm	13,459	18,852	,047	,714	,476	-23,621	50,539

a. Dependent Variable: film_box_global_adj

Residuals Statistics[a]

	Minimum	Maximum	Mean	Std. Deviation	N
Predicted Value	-63,755062	691,650513	106,446105	102,7131970	410
Residual	-218,9195862	411,7869568	0E-7	67,9026731	410
Std. Predicted Value	-1,657	5,697	,000	1,000	410
Std. Residual	-2,961	5,570	,000	,918	410

a. Dependent Variable: film_box_global_adj

Residual statistics

Correlations

		Unstandardized Residual
Unstandardized Residual	Pearson Correlation	1
	N	410
film_budget_adj	Pearson Correlation	,000
	Sig. (2-tailed)	1,000
	N	410
film_prequels_box_office	Pearson Correlation	,000
	Sig. (2-tailed)	1,000
	N	410
star_avg_age_at_release	Pearson Correlation	,000
	Sig. (2-tailed)	1,000
	N	410
star_cc_factor	Pearson Correlation	,000
	Sig. (2-tailed)	1,000
	N	410
star_talent_factor	Pearson Correlation	,000
	Sig. (2-tailed)	1,000
	N	410
star_no_stars	Pearson Correlation	,000
	Sig. (2-tailed)	1,000
	N	410
star_percentage_of_stars_male	Pearson Correlation	,000
	Sig. (2-tailed)	1,000
	N	410

Descriptive Statistics

	N	Sum	Mean	Std. Deviation
Unstandardized Residual	410	,00000	0E-7	67,90267309
Valid N (listwise)	410			

Descriptives

			Statistic	Std. Error
Unstandardized Residual	Mean		0E-7	3,35347400
	95% Confidence Interval for Mean	Lower Bound	-6,5921957	
		Upper Bound	6,5921957	
	5% Trimmed Mean		-4,1190612	
	Median		-5,3855176	
	Variance		4610,773	
	Std. Deviation		67,90267309	
	Minimum		-218,91958	
	Maximum		411,78696	
	Range		630,70655	
	Interquartile Range		62,32446	
	Skewness		1,609	,121
	Kurtosis		7,121	,240

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Unstandardized Residual	,125	410	,000	,877	410	,000

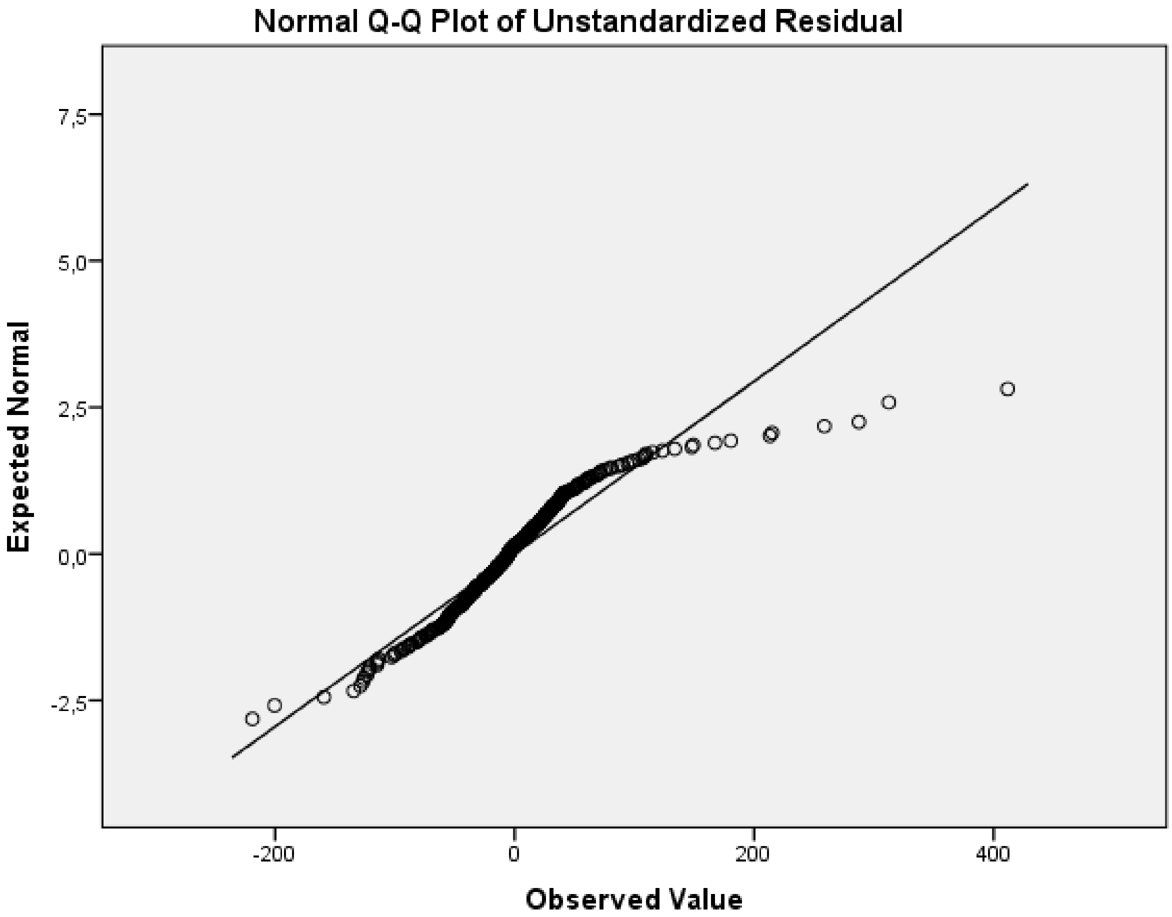
a. Lilliefors Significance Correction

```
Unstandardized Residual Stem-and-Leaf Plot

 Frequency    Stem &  Leaf

     4,00 Extremes    (=<-134)
     7,00       -1 .  2222222
     7,00       -1 .  0001111
    10,00       -0 .  8888889999
    18,00       -0 .  666666666777777777
    44,00       -0 .  44444444444444444444555555555555555555555555
    59,00       -0 .  22222222222222222222222222233333333333333333333333333333333
    77,00       -0 .  00000000000000000000000000000000000000000011111111111111111111111111111111111
    63,00        0 .  000000000000000000000000000000011111111111111111111111111111111
    57,00        0 .  222222222222222222222222222233333333333333333333333333333
    23,00        0 .  44444444444555555555555
    11,00        0 .  66666677777
     8,00        0 .  88889999
     6,00        1 .  000001
    16,00 Extremes    (>=123)

 Stem width:   100,0000
 Each leaf:        1 case(s)
```


Normal Q-Q Plot of Unstandardized Residual
7,5
5,0
2,5
0,0
-2,5
Expected Normal
-200
0
200
400
Observed Value

Forward regression output

Model Summary[k]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,726[a]	,528	,527	84,7220343	
2	,750[b]	,563	,561	81,5768613	
3	,771[c]	,595	,592	78,6539100	
4	,778[d]	,605	,601	77,7763244	
5	,783[e]	,613	,608	77,0756787	
6	,787[f]	,620	,614	76,5007599	
7	,791[g]	,626	,619	75,9623762	
8	,796[h]	,634	,626	75,2675079	
9	,804[i]	,647	,639	73,9786059	
10	,807[j]	,651	,642	73,6735587	2,107

a. Predictors: (Constant), film_budget_adj

b. Predictors: (Constant), film_budget_adj, star_cc_factor

c. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office

d. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office, film_genre_animation

e. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office, film_genre_animation, film_based_historic_events

f. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office, film_genre_animation, film_based_historic_events, film_month_sep

g. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office, film_genre_animation, film_based_historic_events, film_month_sep, star_no_stars

h. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office, film_genre_animation, film_based_historic_events, film_month_sep, star_no_stars, star_talent_factor

i. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office, film_genre_animation, film_based_historic_events, film_month_sep, star_no_stars, star_talent_factor, star_avg_age_at_release

j. Predictors: (Constant), film_budget_adj, star_cc_factor, film_prequels_box_office, film_genre_animation, film_based_historic_events, film_month_sep, star_no_stars, star_talent_factor, star_avg_age_at_release, star_cc_factor_SQ

k. Dependent Variable: film_box_global_adj

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	3272204,682	1	3272204,682	455,877	,000
	Residual	2928551,823	408	7177,823		
	Total	6200756,505	409			
2	Regression	3492259,296	2	1746129,648	262,387	,000
	Residual	2708497,209	407	6654,784		
	Total	6200756,505	409			
3	Regression	3689062,857	3	1229687,619	198,772	,000
	Residual	2511693,648	406	6186,438		
	Total	6200756,505	409			
4	Regression	3750848,067	4	937712,017	155,015	,000
	Residual	2449908,438	405	6049,157		
	Total	6200756,505	409			
5	Regression	3800729,764	5	760145,953	127,956	,000
	Residual	2400026,741	404	5940,660		
	Total	6200756,505	409			
6	Regression	3842252,902	6	640375,484	109,422	,000
	Residual	2358503,603	403	5852,366		
	Total	6200756,505	409			
7	Regression	3881102,899	7	554443,271	96,086	,000
	Residual	2319653,607	402	5770,283		
	Total	6200756,505	409			
8	Regression	3929012,207	8	491126,526	86,692	,000
	Residual	2271744,298	401	5665,198		
	Total	6200756,505	409			
9	Regression	4011622,854	9	445735,873	81,445	,000
	Residual	2189133,651	400	5472,834		
	Total	6200756,505	409			
10	Regression	4035066,998	10	403506,700	74,341	,000
	Residual	2165689,507	399	5427,793		
	Total	6200756,505	409			

a. Dependent Variable: film_box_global_adj

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-7,416	6,778		-1,094	,275
	film_budget_adj	2,422	,113	,726	21,351	,000
2	(Constant)	-4,792	6,543		-,732	,464
	film_budget_adj	2,374	,110	,712	21,673	,000
	star_cc_factor	21,529	3,744	,189	5,750	,000
3	(Constant)	1,756	6,414		,274	,784
	film_budget_adj	2,153	,113	,646	19,118	,000
	star_cc_factor	21,140	3,610	,186	5,855	,000
	film_prequels_box_office	,101	,018	,190	5,640	,000
4	(Constant)	1,798	6,343		,283	,777
	film_budget_adj	2,103	,112	,631	18,705	,000
	star_cc_factor	20,262	3,581	,178	5,659	,000
	film_prequels_box_office	,104	,018	,196	5,875	,000
	film_genre_animation	73,804	23,093	,101	3,196	,002
5	(Constant)	1,479	6,286		,235	,814
	film_budget_adj	2,151	,113	,645	19,096	,000
	star_cc_factor	21,050	3,559	,185	5,915	,000
	film_prequels_box_office	,100	,018	,189	5,680	,000
	film_genre_animation	70,225	22,919	,096	3,064	,002
	film_based_historic_events	-76,429	26,376	-,091	-2,898	,004
6	(Constant)	6,045	6,471		,934	,351
	film_budget_adj	2,124	,112	,637	18,917	,000
	star_cc_factor	21,204	3,533	,186	6,002	,000
	film_prequels_box_office	,099	,017	,187	5,675	,000
	film_genre_animation	73,419	22,779	,101	3,223	,001
	film_based_historic_events	-78,936	26,196	-,094	-3,013	,003
	film_month_sep	-34,567	12,977	-,082	-2,664	,008
7	(Constant)	25,204	9,788		2,575	,010
	film_budget_adj	2,086	,112	,626	18,555	,000
	star_cc_factor	26,844	4,127	,236	6,505	,000
	film_prequels_box_office	,098	,017	,185	5,657	,000
	film_genre_animation	74,265	22,621	,102	3,283	,001
	film_based_historic_events	-77,364	26,019	-,092	-2,973	,003
	film_month_sep	-35,471	12,891	-,085	-2,752	,006
	star_no_stars	-6,270	2,417	-,094	-2,595	,010
8	(Constant)	30,907	9,894		3,124	,002
	film_budget_adj	2,075	,111	,623	18,617	,000
	star_cc_factor	28,717	4,139	,252	6,937	,000
	film_prequels_box_office	,098	,017	,185	5,716	,000
	film_genre_animation	73,396	22,416	,101	3,274	,001
	film_based_historic_events	-75,942	25,785	-,090	-2,945	,003
	film_month_sep	-33,280	12,795	-,079	-2,601	,010
	star_no_stars	-8,217	2,486	-,123	-3,305	,001
	star_talent_factor	11,077	3,809	,091	2,908	,004
9	(Constant)	104,946	21,395		4,905	,000
	film_budget_adj	2,069	,110	,621	18,884	,000
	star_cc_factor	34,945	4,373	,307	7,991	,000
	film_prequels_box_office	,098	,017	,185	5,798	,000
	film_genre_animation	69,669	22,053	,095	3,159	,002
	film_based_historic_events	-73,319	25,353	-,087	-2,892	,004
	film_month_sep	-34,186	12,578	-,082	-2,718	,007
	star_no_stars	-10,029	2,488	-,150	-4,031	,000
	star_talent_factor	15,845	3,940	,131	4,022	,000
	star_avg_age_at_release	-1,741	,448	-,129	-3,885	,000
10	(Constant)	110,204	21,456		5,136	,000
	film_budget_adj	2,079	,109	,624	19,034	,000
	star_cc_factor	38,863	4,745	,341	8,189	,000
	film_prequels_box_office	,100	,017	,188	5,908	,000
	film_genre_animation	70,789	21,969	,097	3,222	,001
	film_based_historic_events	-73,273	25,248	-,087	-2,902	,004
	film_month_sep	-35,347	12,539	-,084	-2,819	,005
	star_no_stars	-9,255	2,505	-,138	-3,694	,000
	star_talent_factor	16,327	3,930	,135	4,154	,000
	star_avg_age_at_release	-1,878	,451	-,139	-4,164	,000
	star_cc_factor_SQ	-,987	,475	-,074	-2,078	,038

a. Dependent Variable: film_box_global_adj

Output lasso selection

Credit

Catreg Version 3.0 by Data Theory Scaling System Group (DTSS) Faculty of Social and Behavioral Sciences Leiden University, The Netherlands

Model Summary

Multiple R	R Square	Adjusted R Square	Regularization "R Square" (1-Error)	Apparent Prediction Error	Expected Prediction Error		
					Estimate[a]	Std. Error	N
,783	,614	,592	,602	,398	,464	,049	410

Penalty ,140
Dependent Variable: film_box_global_adj
a. Mean Squared Error (10 fold Cross Validation).

Lasso Paths

X-axis reference lines at optimal model and at most parsimonious model within 1 Std. Error.

Final regression outputs

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,808[a]	,653	,643	73,5827522	2,094

a. Predictors: (Constant), star_percentage_of_stars_male, film_genre_animation, star_talent_factor_SQ, film_month_sep, star_cc_factor_SQ, film_based_historic_events, film_prequels_box_office, star_avg_age_at_release, film_budget_adj, star_no_stars, star_cc_factor, star_talent_factor

b. Dependent Variable: film_box_global_adj

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	4051231,200	12	337602,600	62,352	,000[b]
	Residual	2149525,305	397	5414,421		
	Total	6200756,505	409			

a. Dependent Variable: film_box_global_adj

b. Predictors: (Constant), star_percentage_of_stars_male, film_genre_animation, star_talent_factor_SQ, film_month_sep, star_cc_factor_SQ, film_based_historic_events, film_prequels_box_office, star_avg_age_at_release, film_budget_adj, star_no_stars, star_cc_factor, star_talent_factor

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	110,216	21,642		5,093	,000		
	film_budget_adj	2,043	,111	,613	18,404	,000	,787	1,271
	film_prequels_box_office	,100	,017	,188	5,912	,000	,867	1,154
	film_genre_animation	70,965	21,965	,097	3,231	,001	,963	1,038
	film_based_historic_events	-75,078	25,242	-,089	-2,974	,003	,965	1,036
	film_month_sep	-36,110	12,533	-,086	-2,881	,004	,977	1,024
	star_cc_factor	38,957	4,769	,342	8,170	,000	,499	2,005
	star_cc_factor_SQ	-1,000	,476	-,075	-2,100	,036	,688	1,455
	star_talent_factor	20,299	6,647	,168	3,054	,002	,290	3,451
	star_talent_factor_SQ	-,122	,244	-,026	-,501	,617	,330	3,030
	star_avg_age_at_release	-2,197	,487	-,162	-4,511	,000	,675	1,482
	star_no_stars	-9,055	2,506	-,135	-3,614	,000	,622	1,607
	star_percentage_of_stars_male	,217	,127	,057	1,709	,088	,780	1,282

a. Dependent Variable: film_box_global_adj

Bootstrap for Coefficients

Model		B	Bootstrap[a] Bias	Std. Error	Sig. (2-tailed)	95% Confidence Interval Lower	Upper
1	(Constant)	110,216	2,558	17,604	,005	79,407	147,701
	film_budget_adj	2,043	-,017	,191	,005	1,644	2,433
	film_prequels_box_office	,100	,025	,056	,015	,067	,280
	film_genre_animation	70,965	2,506	44,449	,097	-7,862	166,703
	film_based_historic_events	-75,078	7,883	33,529	,025	-136,260	1,763
	film_month_sep	-36,110	-,041	7,837	,005	-49,749	-16,600
	star_cc_factor	38,957	,521	4,649	,005	29,074	48,653
	star_cc_factor_SQ	-1,000	-,138	,601	,050	-2,548	-,367
	star_talent_factor	20,299	2,625	7,830	,035	9,018	40,612
	star_talent_factor_SQ	-,122	-,067	,299	,662	-,795	,429
	star_avg_age_at_release	-2,197	-,030	,412	,005	-3,030	-1,476
	star_no_stars	-9,055	-,272	2,403	,005	-14,256	-4,071
	star_percentage_of_stars_male	,217	,004	,093	,010	,051	,405

a. Unless otherwise noted, bootstrap results are based on 200 bootstrap samples

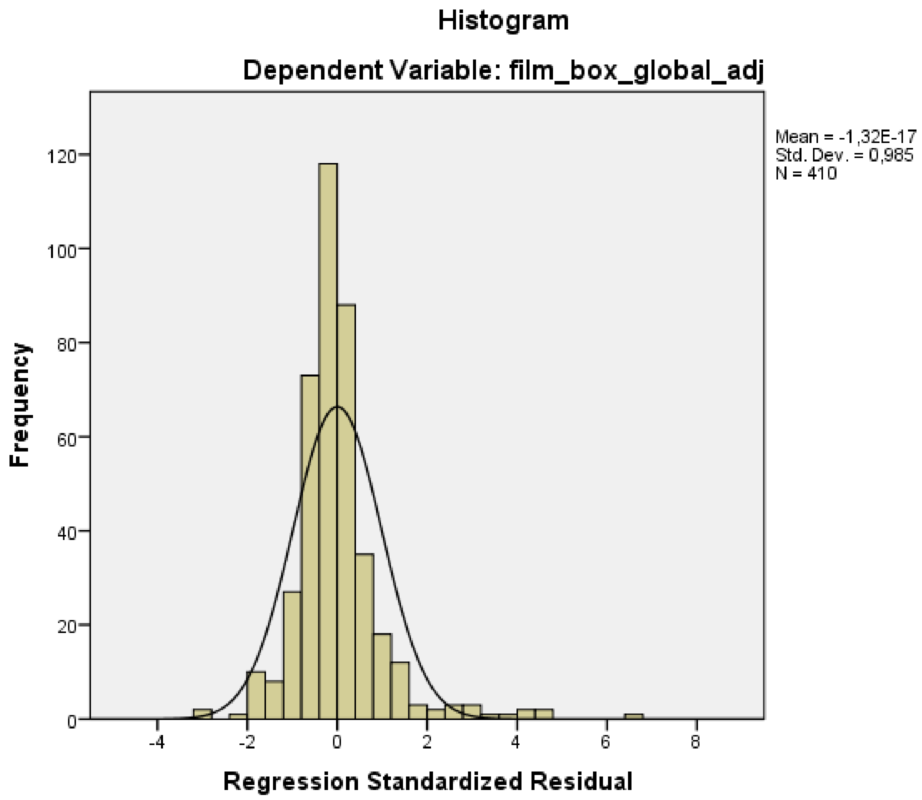
Residuals Statistics[a]

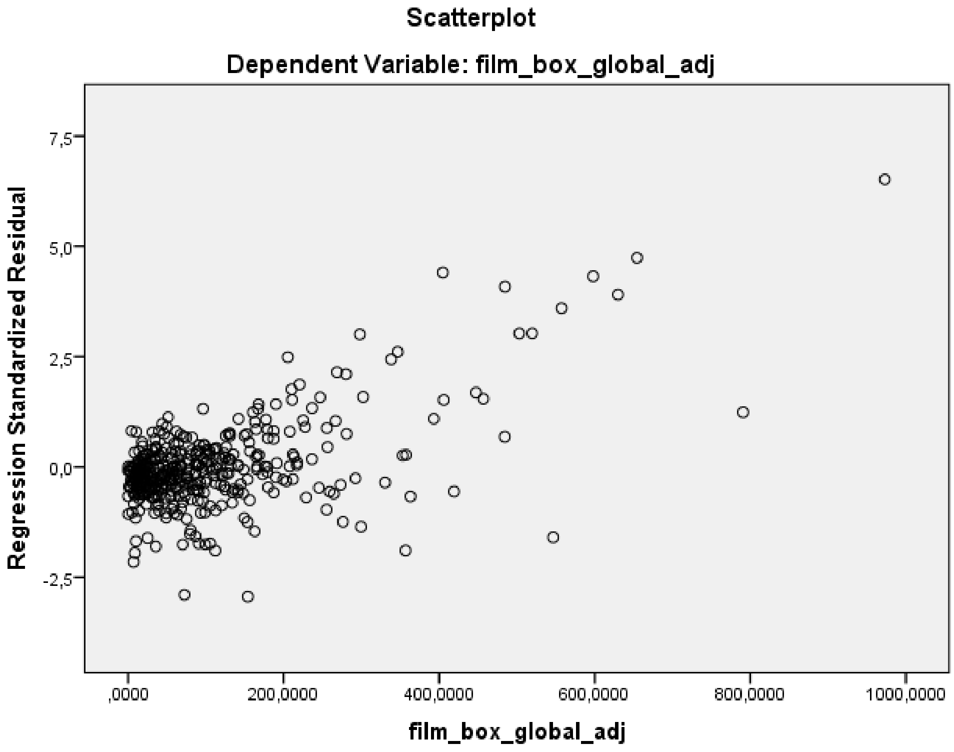
		Statistic	Bootstrap[b] Bias	Std. Error	95% Confidence Interval Lower	Upper
Predicted Value	Minimum	-55,879108				
	Maximum	699,577393				
	Mean	106,446105	,195770	6,118095	96,445591	120,411058
	Std. Deviation	99,5249253	1,4377497	10,6860542	83,3149515	122,5951941
	N	410	0	0	410	410
Residual	Minimum	-216,2650909				
	Maximum	479,4762573				
	Mean	0E-7	0E-7	0E-7	0E-7	0E-7
	Std. Deviation	72,4952626	-2,1368210	5,5861367	57,5944747	80,7637283
	N	410	0	0	410	410
Std. Predicted Value	Minimum	-1,631				
	Maximum	5,960				
	Mean	,000	,000	,000	,000	,000
	Std. Deviation	1,000	,000	,000	1,000	1,000
	N	410	0	0	410	410
Std. Residual	Minimum	-2,939				
	Maximum	6,516				
	Mean	,000	,000	,000	,000	,000
	Std. Deviation	,985	,000	,000	,985	,985
	N	410	0	0	410	410

a. Dependent Variable: film_box_global_adj

b. Unless otherwise noted, bootstrap results are based on 200 bootstrap samples

$$
\begin{aligned}
y = {} & \beta_{budget} * budget \\
& + \beta_{prequel\ box\ office} * prequel\ box\ office \\
& + \beta_{animation} * animation\ status \\
& + \beta_{historic\ events} * historic\ events\ status \\
& + \beta_{september} * september\ release\ status \\
& + \beta_{consumption\ capital} * consumption\ capital \\
& + \beta_{consumption\ capital\ squared} * |consumption\ capital| * consumption\ capital \\
& + \beta_{talent} * talent \\
& + \beta_{talent\ squared} * |talent| * talent \\
& + \beta_{average\ cast\ age} * average\ cast\ age \\
& + \beta_{number\ of\ stars} * number\ of\ stars \\
& + \beta_{percentage\ of\ stars\ male} * percentage\ of\ stars\ male
\end{aligned}
$$





Final regression with opening weekend as dependent variable

Opening weekend is North-American market only

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,756[a]	,571	,558	10,4769345	2,030

a. Predictors: (Constant), star_percentage_of_stars_male, film_genre_animation, star_talent_factor_SQ, film_month_sep, star_cc_factor_SQ, film_based_historic_events, film_prequels_box_office, star_avg_age_at_release, film_budget_adj, star_no_stars, star_cc_factor, star_talent_factor

b. Dependent Variable: film_box_opening_weekend_adj

ANOVA[a]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	58024,779	12	4835,398	44,052	,000[b]
	Residual	43577,164	397	109,766		
	Total	101601,943	409			

a. Dependent Variable: film_box_opening_weekend_adj

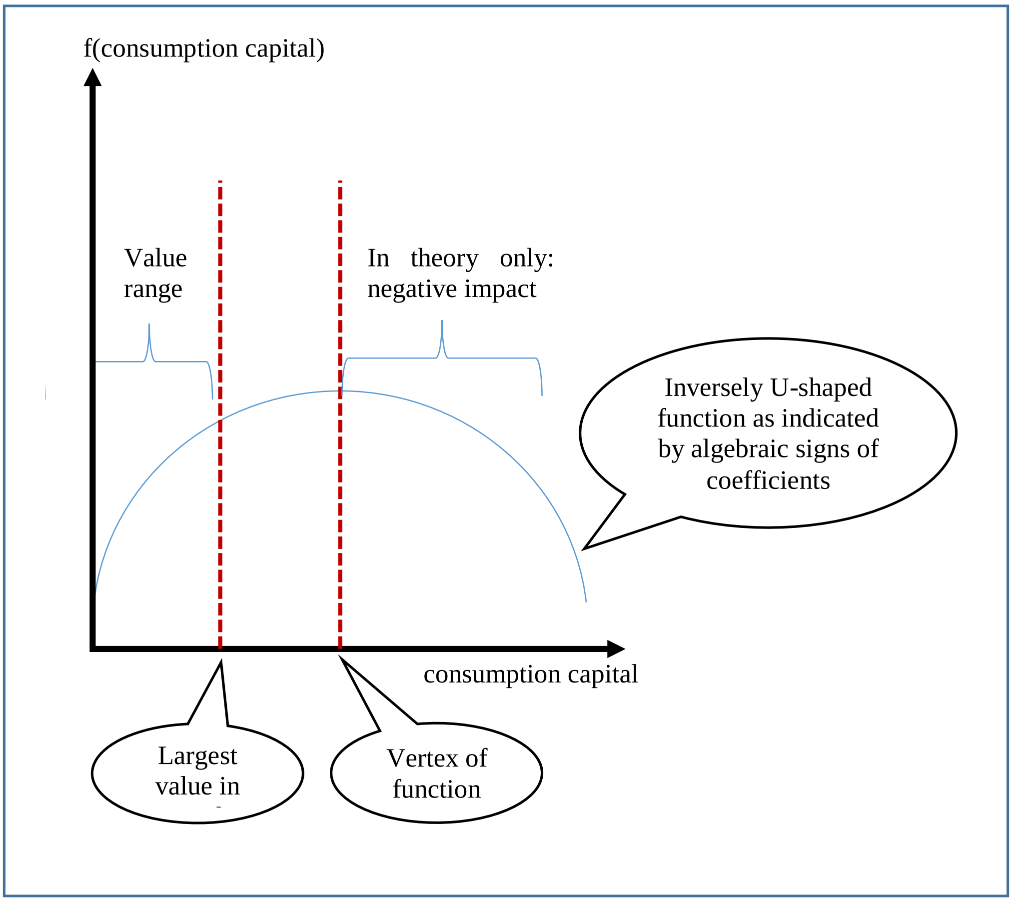
b. Predictors: (Constant), star_percentage_of_stars_male, film_genre_animation, star_talent_factor_SQ, film_month_sep, star_cc_factor_SQ, film_based_historic_events, film_prequels_box_office, star_avg_age_at_release, film_budget_adj, star_no_stars, star_cc_factor, star_talent_factor

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	14,486	3,082		4,701	,000
	film_budget_adj	,264	,016	,618	16,685	,000
	film_prequels_box_office	,010	,002	,149	4,230	,000
	film_genre_animation	2,460	3,127	,026	,786	,432
	film_based_historic_events	-11,896	3,594	-,111	-3,310	,001
	film_month_sep	-4,397	1,784	-,082	-2,464	,014
	star_cc_factor	3,429	,679	,235	5,050	,000
	star_cc_factor_SQ	-,059	,068	-,034	-,863	,388
	star_talent_factor	,221	,946	,014	,233	,816
	star_talent_factor_SQ	,022	,035	,035	,620	,536
	star_avg_age_at_release	-,259	,069	-,149	-3,731	,000
	star_no_stars	-1,180	,357	-,138	-3,308	,001
	star_percentage_of_stars_male	,031	,018	,063	1,690	,092

a. Dependent Variable: film_box_opening_weekend_adj

Illustration of inversely U-shaped function of consumption capital



Prequel and sequel global box offices

Title	Cumulated prequel box office	Box office of observed movie
2 Fast 2 Furious	$207.283.925	$236.350.700
Agent Cody Banks 2: Destination London	$58.795.814	$28.061.300
Bad Boys II	$141.407.024	$273.339.600
Barbershop 2: Back in Business	$77.063.924	$64.236.900
Be Cool	$115.101.622	$89.673.700
Before Sunset	$5.535.405	$15.572.200
Big Momma's House 2	$173.959.438	$126.160.400
Blade: Trinity	$286.193.562	$125.516.400
Casino Royale	$3.901.669.026	$546.625.000
Charlie's Angels: Full Throttle	$264.105.545	$259.175.800
Cheaper by the Dozen 2	$190.212.113	$121.649.500
Deuce Bigalow: European Gigolo	$92.938.755	$42.479.300
Evan Almighty	$484.592.874	$153.933.200
Fantastic Four:Rise Of The Silver Surfer	$330.579.719	$256.570.000
Garfield: A Tail Of Two Kitties	$200.804.534	$129.302.300
Ice Age: The Meltdown	$383.257.136	$598.036.800
Legally Blonde 2: Red, White & Blonde	$141.774.679	$124.914.800
Legend of Zorro	$250.288.523	$134.097.000
Meet the Fockers	$330.444.045	$503.060.300
Miss Congeniality 2: Armed and Fabulous	$212.742.720	$95.481.500
Mission: Impossible III	$1.004.084.464	$363.035.200
National Treasure: Book Of Secrets	$347.512.318	$405.974.500
Ocean's Thirteen	$813.461.430	$276.333.100
Ocean's Twelve	$450.717.150	$353.207.700
Once Upon a Time in Mexico	$27.446.365	$98.185.600
Pirates Of The Caribbean: Dead Man's Chest	$654.264.015	$972.881.000
Rush Hour 3	$591.712.666	$229.030.500
Santa Clause 3: The Escape Clause	$362.688.422	$101.075.100
Shanghai Knights	$99.274.467	$88.323.500
Spider-Man 3	$1.605.474.892	$790.772.100
The Bourne Supremacy	$214.034.224	$280.915.500
The Bourne Ultimatum	$502.534.441	$393.067.800
The Chronicles of Riddick	$53.187.659	$112.729.000
The Princess Diaries 2: Royal Engagement	$165.335.153	$131.192.300
Underworld: Evolution	$95.708.457	$101.597.600

Comparison of prequel and sequel return on investment

List is sorted by return on investment estimate based on the estimate

ROI = ((Global box office / 2) – Budget) / Budget

Title	Prequels' box office	Prequels' ROI	Film ROI
Legally Blonde 2: Red, White & Blonde	$141.774.679	294%	39%
Bad Boys II	$141.407.024	272%	5%
Ice Age: The Meltdown	$383.257.136	225%	310%
Barbershop 2: Back in Business	$77.063.924	221%	83%
The Princess Diaries 2: Royal Engagement	$165.335.153	218%	50%
Meet the Fockers	$330.444.045	200%	223%
Evan Almighty	$484.592.874	199%	-50%
Big Momma's House 2	$173.959.438	190%	73%
Deuce Bigalow: European Gigolo	$92.938.755	173%	3%
2 Fast 2 Furious	$207.283.925	173%	18%
Ocean's Twelve	$450.717.150	165%	65%
Casino Royale	$3.901.669.026	161%	100%
Mission: Impossible III	$1.004.084.464	145%	33%
Rush Hour 3	$591.712.666	141%	-8%
Cheaper by the Dozen 2	$190.212.113	138%	-8%
Spider-Man 3	$1.605.474.892	137%	73%
Miss Congeniality 2: Armed and Fabulous	$212.742.720	136%	13%
Pirates Of The Caribbean: Dead Man's Chest	$654.264.015	134%	167%
Underworld: Evolution	$95.708.457	118%	11%
Ocean's Thirteen	$813.461.430	109%	48%
Santa Clause 3: The Escape Clause	$362.688.422	108%	-8%
Garfield: A Tail Of Two Kitties	$200.804.534	101%	42%
Once Upon a Time in Mexico	$27.446.365	96%	69%
X-Men: The Last Stand	$704.051.076	90%	9%
Be Cool	$115.101.622	90%	-10%
The Bourne Ultimatum	$502.534.441	86%	101%
The Bourne Supremacy	$214.034.224	78%	92%
National Treasure: Book Of Secrets	$347.512.318	74%	108%
Fantastic Four:Rise Of The Silver Surfer	$330.579.719	65%	45%
Blade: Trinity	$286.193.562	45%	-1%
Charlie's Angels: Full Throttle	$264.105.545	42%	-4%
Legend of Zorro	$250.288.523	32%	-11%
The Chronicles of Riddick	$53.187.659	16%	-45%
Before Sunset	$5.535.405	11%	-20%
Agent Cody Banks 2: Destination London	$58.795.814	5%	-45%
Shanghai Knights	$99.274.467	-10%	-12%

Examples for variable-combinations and their scores

Movie	Google news hits	Prior box office $	Cumulated cinemas	Previous movies #	Consumption capital score
Mona Lisa Smile	24567	3342508269	87739	72	1,01009
The Lake House	16010	2850565071	100320	61	1,00282
Bringing Down the House	30520	1634295573	51893	44	1,00033
...					
The Rundown	8585	2476315201	98844	78	0,03026
Pirates of the Caribbean I	9796	2550524968	52642	44	0,01459
Tim Burton's Corpse Bride	14750	1918188562	46801	46	-0,00225
...					
25th Hour	3190	388393323	18865	11	-1,00812
A Man Apart	1050	594270398	14755	7	-1,01218
Underworld: Evolution	2735	572051505	7868	12	-1,01729

Movie	Golden Globes		Academy Awards		Talent score
	Wins	Nominations	Wins	Nominations	
Ocean's Twelve	7	16	3	6	1,11712
Open Range	5	10	1	9	1,06091
American Gangster	3	7	3	5	0,92348
...					
Solaris	5	2	1	3	0,02681
A Good Year	5	1	1	4	0,00637
Failure To Launch	8	4	0	4	0,00310
...					
Smokin' Aces	1	7	0	1	-0,97490
Twisted	0	8	0	2	-0,98129
Click	0	2	1	1	-1,02798

Frequencies of the variable percentage of stars male

star_percentage_of_stars_male

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	0,00%	47	11,5	11,5	11,5
	25,00%	1	,2	,2	11,7
	28,57%	1	,2	,2	12,0
	33,33%	17	4,1	4,1	16,1
	40,00%	8	2,0	2,0	18,0
	42,86%	2	,5	,5	18,5
	44,44%	1	,2	,2	18,8
	45,45%	1	,2	,2	19,0
	50,00%	85	20,7	20,7	39,8
	57,14%	2	,5	,5	40,2
	60,00%	7	1,7	1,7	42,0
	66,67%	52	12,7	12,7	54,6
	71,43%	2	,5	,5	55,1
	75,00%	26	6,3	6,3	61,5
	80,00%	6	1,5	1,5	62,9
	83,33%	2	,5	,5	63,4
	85,71%	1	,2	,2	63,7
	87,50%	3	,7	,7	64,4
	90,00%	1	,2	,2	64,6
	100,00%	145	35,4	35,4	100,0
	Total	410	100,0	100,0	

Top 50 movies by budget

Spider-Man 3
Superman Returns
X-Men: The Last Stand
Pirates Of The Caribbean: Dead Man's Chest
Troy
The Polar Express
The Golden Compass
Van Helsing
Evan Almighty
Alexander
Charlie and the Chocolate Factory
The Last Samurai
Pirates of the Caribbean: The Curse of the Black Pearl
Casino Royale
Mission: Impossible III
Master and Commander: The Far Side of the World
Charlie's Angels: Full Throttle
Stealth
I Am Legend
Transformers
Bee Movie
Bad Boys II
War of the Worlds
Rush Hour 3
Miami Vice
Sahara
Lemony Snicket's Unfortunate Events
The Island
The Cat in the Hat
Around the World in 80 Days
Ocean's Twelve
The Alamo
The Chronicles of Riddick
The Good Shepherd
2 Fast 2 Furious
Bourne Ultimatum
National Treasure: Book Of Secrets
The Aviator
Hidalgo
Catwoman
Kingdom of Heaven
Constantine
Fun With Dick & Jane
Ocean's Thirteen
Blood Diamond
Eragon
Fred Claus
American Gangster
Fantastic Four: Rise Of The Silver Surfer
The Stepford Wives

Top 50 movies by global box office

Pirates Of The Caribbean: Dead Man's Chest
Spider-Man 3
Pirates of the Caribbean: The Curse of the Black Pearl
Transformers
Ice Age: The Meltdown
War of the Worlds
Casino Royale
I Am Legend
Meet the Fockers
Bruce Almighty
Troy
The Last Samurai
Charlie and the Chocolate Factory
X-Men: The Last Stand
National Treasure: Book Of Secrets
Movie 300
Bourne Ultimatum
Mission: Impossible III
Shark Tale
Superman Returns
Ocean's Twelve
Hitch
National Treasure
The Golden Compass
Enchanted
The Polar Express
The Devil Wears Prada
Van Helsing
The Bourne Supremacy
The Pursuit Of Happyness
Ocean's Thirteen
Bad Boys II
Wedding Crashers
Something's Gotta Give
The Departed
Charlie's Angels: Full Throttle
Fantastic Four: Rise Of The Silver Surfer
Bridget Jones: Edge of Reason
Bee Movie
Love Actually
Robots
American Gangster
2 Fast 2 Furious
Rush Hour 3
Eragon
Wild Hogs
Elf
Constantine
Click
The Terminal

Top 50 movies by opening weekend

Spider-Man 3
Pirates Of The Caribbean: Dead Man's Chest
X-Men: The Last Stand
I Am Legend
Bruce Almighty
Movie 300
Transformers
Ice Age: The Meltdown
Bourne Ultimatum
War of the Worlds
Charlie and the Chocolate Factory
Fantastic Four: Rise Of The Silver Surfer
The Bourne Supremacy
2 Fast 2 Furious
Van Helsing
Superman Returns
Pirates of the Caribbean: The Curse of the Black Pearl
Bad Boys II
Shark Tale
Troy
Meet the Fockers
Longest Yard
Rush Hour 3
Mission: Impossible III
Talladega Nights: Ballad Of Ricky Bobby
Hitch
Daredevil
National Treasure: Book Of Secrets
50 First Dates
American Gangster
The Cat in the Hat
Ocean's Twelve
Charlie's Angels: Full Throttle
Casino Royale
S.W.A.T.
Click
The Break-Up
Wild Hogs
National Treasure
Robots
Bee Movie
Ocean's Thirteen
Wedding Crashers
Elf
Bringing Down the House
Enchanted
I Now Pronounce You Chuck And Larry
Norbit
Blades Of Glory
Dodgeball: A True Underdog Story

Top 50 movies by talent score

Something's Gotta Give
A Prairie Home Companion
Rendition
Manchurian Candidate
Lions For Lambs
Meet the Fockers
The Departed
The Bucket List
Finding Neverland
Prime
The Devil Wears Prada
Secondhand Lions
Shark Tale
Ocean's Thirteen
Charlie Wilson's War
All The King's Men
The Stepford Wives
Inside Man
Stranger Than Fiction
Broken Flowers
Don't Come Knocking
I Heart Huckabees
Georgia Rule
In Her Shoes
North Country
The White Countess
Cold Mountain
Rumor Has It
Children Of Men
Ocean's Twelve
Open Range
American Gangster
Welcome to Mooseport
Superman Returns
The Recruit
The Terminal
Cinderella Man
The Life of David Gale
Million Dollar Baby
The Interpreter
The Family Stone
Levity
The Human Stain
Calendar Girls
Robots
The Good Shepherd
Bee Movie
Birth
Love Actually
The Prestige

Top 50 movies by consumption capital score

Hairspray
Ocean's Twelve
Wild Hogs
Ocean's Thirteen
Robots
Good Night, and Good Luck
Syriana
Shark Tale
Mystic River
The Family Stone
Click
A Scanner Darkly
Eternal Sunshine of the Spotless Mind
All The King's Men
Love Actually
The Good Shepherd
Black Snake Moan
Smokin' Aces
Gone, Baby, Gone
Shall We Dance?
You, Me And Dupree
Alexander
Mr. Brooks
Home Of The Brave
Closer
The Holiday
Into The Wild
The Departed
The Hoax
Million Dollar Baby
Stardust
August Rush
Little Miss Sunshine
Enchanted
A History of Violence
Babel
The Stepford Wives
Anything Else
Gothika
S.W.A.T.
The Terminal
Charlie Wilson's War
Troy
Music And Lyrics
No Country For Old Men
Ice Age: The Meltdown
A Prairie Home Companion
Snakes On A Plane
Identity
Dreamgirls

Top 50 actors by individual consumption capital score

Samuel L. Jackson
Tom Cruise
Robin Williams
Eddie Murphy
Matt Damon
Johnny Depp
Tom Hanks
Bruce Willis
John Travolta
Ben Stiller
Will Ferrell
Morgan Freeman
Robert De Niro
Brad Pitt
Christopher Walken
Harrison Ford
Owen Wilson
Drew Barrymore
Julia Roberts
Jim Carrey
Nicolas Cage
Steve Martin
Will Smith
Keanu Reeves
Adam Sandler
Sylvester Stallone
Nicole Kidman
Jack Black
Ben Affleck
Halle Berry
Kevin Costner
Kirsten Dunst
Denzel Washington
Anthony Hopkins
Dennis Quaid
Danny DeVito
Vince Vaughn
Sandra Bullock
George Clooney
Clint Eastwood
Tommy Lee Jones
Queen Latifah
Chris Rock
Robert Redford
Cameron Diaz
John Goodman
Martin Lawrence
Viggo Mortensen
Antonio Banderas
Paul Giamatti

Top 50 actors by individual talent score

Meryl Streep
Jack Nicholson
Al Pacino
Shirley MacLaine
Dustin Hoffman
Michael Caine
Jane Fonda
Vanessa Redgrave
Jessica Lange
Gene Hackman
Tom Hanks
Marlon Brando
Robin Williams
Robert De Niro
Robert Duvall
Sissy Spacek
Julie Andrews
Denzel Washington
Diane Keaton
Helen Mirren
Jodie Foster
Susan Sarandon
Barbra Streisand
Renée Zellweger
Glenn Close
Anjelica Huston
Nicole Kidman
Holly Hunter
Emma Thompson
Ben Kingsley
Julia Roberts
Anthony Hopkins
Tom Cruise
Sean Penn
Morgan Freeman
Bette Midler
Hilary Swank
Geoffrey Rush
Cate Blanchett
Russell Crowe
James Woods
Sarah Jessica Parker
Kate Winslet
Alan Arkin
Ed Harris
Kevin Spacey
Sigourney Weaver
Annette Bening
Michelle Pfeiffer
Leonardo DiCaprio

T-test of box office to budget ratio between top 5% talent movies and the remaining 95%

Group Statistics

	talent_top_5_percent	N	Mean	Std. Deviation	Std. Error Mean
global_box_office_divided_by_budget	1,000	20	3,25597	2,227577	,498101
	,000	390	2,25539	2,123661	,107536

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
global_box_office_divided_by_budget	Equal variances assumed	,972	,325	2,05	408	,041	1,000580	,488024	,041224	1,959936
	Equal variances not assumed			1,96	20,8	,063	1,000580	,509577	-,059733	2,060892

T-test of independent variables for animated movies and non-animated movies

Group Statistics

	film_genre_animation	N	Mean	Std. Deviation	Std. Error Mean
star_no_stars	1	12	3,17	1,992	,575
	0	398	2,73	1,837	,092
film_prequels_box_office	1	12	31,93809	110,636805	31,938095
	0	398	38,08287	234,895739	11,774260
star_avg_age_at_release	1	12	38,775595	7,1627466	2,0677068
	0	398	39,379075	9,1538837	,4588427
star_cc_factor	1	12	,5093538	1,57140891	,45362668
	0	398	-,0332018	1,06101188	,05318372
star_talent_factor	1	12	,0751715	1,02388335	,29556966
	0	398	-,0048077	1,01788088	,05102176
film_box_global_adj	1	12	247,352400	199,4491610	57,5760134
	0	398	102,197674	117,8922606	5,9094051
film_month_sep	1	12	,17	,389	,112
	0	398	,09	,291	,015
film_based_historic_events	1	12	,00	,000	,000
	0	398	,02	,149	,007

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
star_no_stars	Equal variances assumed	,768	,381	,803	408	,423	,433	,540	-,628	1,494
	Equal variances not assumed			,743	11,6	,472	,433	,582	-,841	1,707
film_prequels_box_office	Equal variances assumed	,029	,864	-,090	408	,928	-6,14477	68,097436	-140,01040	127,72085
	Equal variances not assumed			-,181	14,2	,859	-6,14477	34,039317	-79,062176	66,772634
star_avg_age_at_release	Equal variances assumed	,895	,345	-,226	408	,821	-,603480	2,6679845	-5,8481915	4,6412318
	Equal variances not assumed			-,285	12,1	,781	-,603480	2,1180057	-5,2136058	4,0066460
star_cc_factor	Equal variances assumed	6,10	,014	1,72	408	,087	,542556	,31583282	-,07830708	1,1634183
	Equal variances not assumed			1,19	11,3	,259	,542556	,45673370	-,45941548	1,5445267
star_talent_factor	Equal variances assumed	,215	,643	,268	408	,789	,079979	,29828126	-,50638071	,66633912
	Equal variances not assumed			,267	11,7	,794	,079979	,29994107	-,57562359	,73558200
film_box_global_adj	Equal variances assumed	7,64	,006	4,10	408	,000	145,155	35,3982803	75,568950	214,74050
	Equal variances not assumed			2,51	11,2	,029	145,155	57,8784795	18,086616	272,22284
film_month_sep	Equal variances assumed	2,46	,117	,856	408	,392	,074	,086	-,096	,243
	Equal variances not assumed			,650	11,4	,528	,074	,113	-,175	,322
film_based_historic_events	Equal variances assumed	1,16	,282	-,526	408	,599	-,023	,043	-,107	,062
	Equal variances not assumed			-3,03	397	,003	-,023	,007	-,037	-,008

T-test of animated movies' box offices and budgets by character "humanness"

Movies with predominantly human lead characters compared to movies with mostly non-human characters. Classification was done by hand based on IMDb entries of the respective movies.

Group Statistics

	human_vs_non_human	N	Mean	Std. Deviation	Std. Error Mean
Global box office	human	47	209743178,7	198751529,7	28990890,19
	non-human	57	326222761,0	283330516,3	37528043,66
Budget	human	47	76,877660	65,2146271	9,5125310
	non-human	57	90,562632	51,6840265	6,8457165

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Global BO	Equal variances assumed	8,239	,005	-2,376	102	,019	-1,165E+8	49015472	-213701540	-19257625
	Equal variances not assumed			-2,456	99,600	,016	-1,165E+8	47421786	-210567683	-22391482
Budget	Equal variances assumed	3,027	,085	-1,194	102	,235	-13,68497	11,462552	-36,420889	9,0509453
	Equal variances not assumed			-1,168	86,850	,246	-13,68497	11,719730	-36,979773	9,6098292

T-test of age difference between male and female actors

The test was conducted not on movie basis, but on basis of the individual engagements of all 1186 actors mentioned on the movie posters of the sampled movies

Group Statistics

	con_disc_gender	N	Mean	Std. Deviation	Std. Error Mean
age_at_release	male	760	42,41	12,869	,467
	female	426	34,83	12,590	,610

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
age_at_release	Equal variances assumed	1,517	,218	9,805	1184	,000	7,578	,773	6,062	9,095
	Equal variances not assumed			9,866	896,4	,000	7,578	,768	6,071	9,086

T-test of differences in star attributes by gender

This test was also performed on an actor basis

Group Statistics

	con_disc_gender	N	Mean	Std. Deviation	Std. Error Mean
google_eng_media_presence	male	761	7187,97	8656,561	313,800
	female	426	4999,42	6648,913	322,141
box_office_prior_to_release	male	612	948157097,0	652478373,6	26374875,03
	female	326	607828810,5	421532874,6	23346546,65
cumulated_cinemas	male	611	29518,18	16533,224	668,863
	female	326	19775,89	11874,109	657,646
prior_appearances	male	611	22,19	12,820	,519
	female	326	16,15	9,152	,507
golden_globe_wins_ONLY	male	761	,40	,899	,033
	female	426	,58	1,138	,055
golden_globe_noms_ONLY	male	761	1,46	2,031	,074
	female	426	1,67	2,558	,124
academy_wins_ONLY	male	761	,20	,507	,018
	female	426	,27	,519	,025
academy_noms_ONLY	male	761	,70	1,286	,047
	female	426	,79	1,763	,085

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
google_eng_media_presence	Equal variances assumed	23,891	,000	4,524	1185	,000	2188,555	483,761	1239,430	3137,679
	Equal variances not assumed			4,867	1073,64	,000	2188,555	449,717	1306,131	3070,978
box_office_prior_to_release	Equal variances assumed	40,184	,000	8,517	936	,000	3,403E+8	39957942	261910757,4	418745815,6
	Equal variances not assumed			9,662	902,237	,000	3,403E+8	35223505	271198750,0	409457823,0
cumulated_cinemas_prior	Equal variances assumed	26,579	,000	9,421	935	,000	9742,290	1034,143	7712,780	11771,801
	Equal variances not assumed			10,386	856,715	,000	9742,290	938,017	7901,210	11583,371
prior_appearances	Equal variances assumed	31,916	,000	7,546	935	,000	6,043	,801	4,471	7,614
	Equal variances not assumed			8,333	859,702	,000	6,043	,725	4,619	7,466
golden_globe_wins_ONLY	Equal variances assumed	22,609	,000	-3,012	1185	,003	-,181	,060	-,298	-,063
	Equal variances not assumed			-2,821	724,572	,005	-,181	,064	-,306	-,055
golden_globe_noms_ONLY	Equal variances assumed	5,722	,017	-1,576	1185	,115	-,213	,135	-,478	,052
	Equal variances not assumed			-1,478	727,284	,140	-,213	,144	-,496	,070
academy_wins_ONLY	Equal variances assumed	11,849	,001	-2,218	1185	,027	-,069	,031	-,129	-,008
	Equal variances not assumed			-2,203	863,027	,028	-,069	,031	-,130	-,007
academy_noms_ONLY	Equal variances assumed	4,006	,046	-1,019	1185	,308	-,091	,089	-,266	,084
	Equal variances not assumed			-,935	681,961	,350	-,091	,097	-,282	,100

Movie poster A-style "World Trade Center"

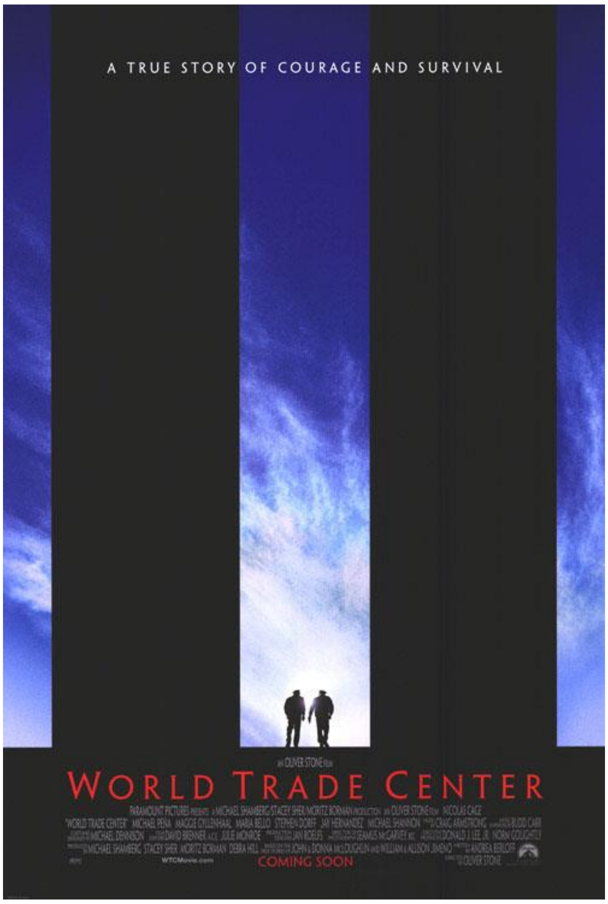
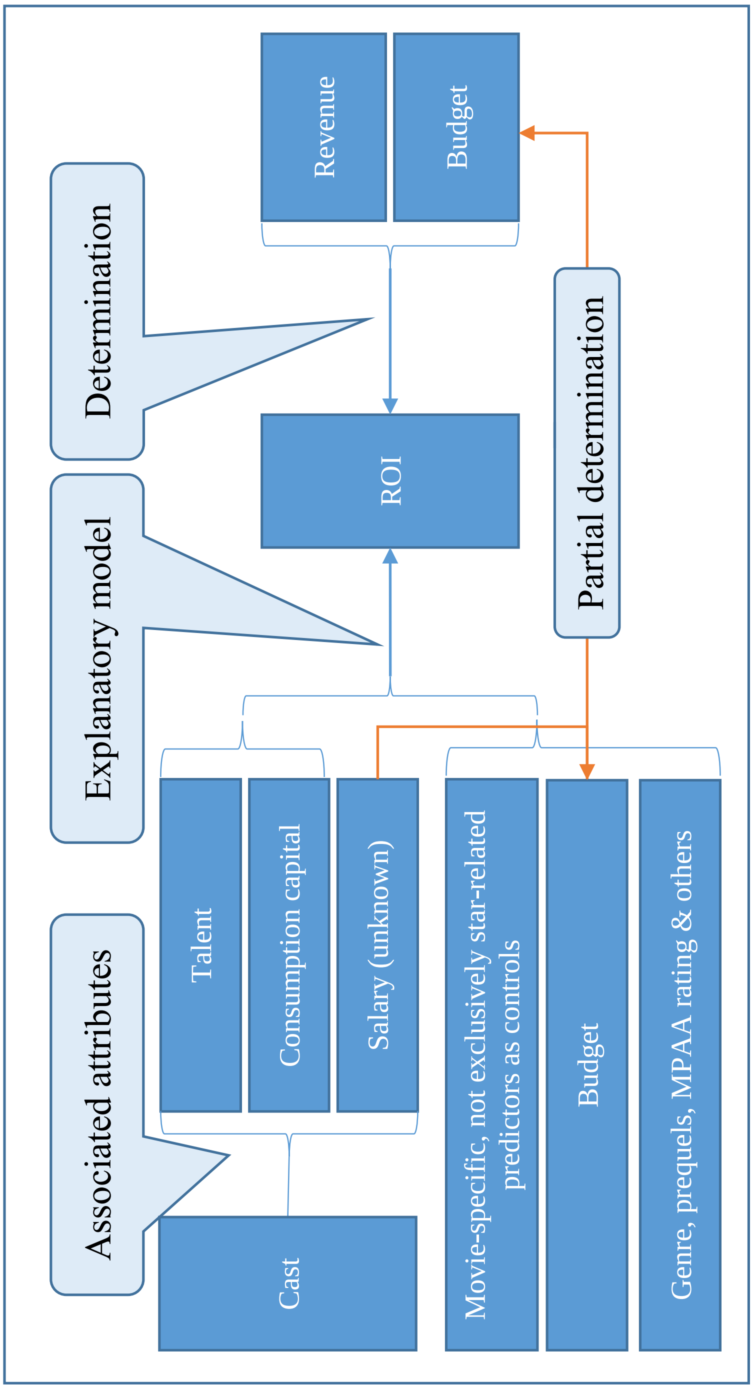


Image from (IMDb, 2013c)

Example for rotoscoping animation technique: still image from "A Scanner Darkly"

Inadmissible circular connection



Example of a budget sheet: Tomb Raider II

"TOMB RAIDER II"
WORKING PASS #5

Exchange Rates applied are as follows:
US$ 1.54= UK£1
US$ 1.00 = Euro .97 (For Greece)
US$ 1.00 = Kenya Shills 71.95 (For Africa)
US$ 1.00 = Hong Kong Dollar 7.80

budget dated: August 16, 2002

Main Unit Shoot 08/26/02 - 02/07/03 (Incl 3 wks Hiatus)
GREECE LOC - 1.2 WKS (6 DYS)
AFRICA LOC - 2.0 WKS (10 DYS)
UK STUDIO/LOCAL LOC 15.6 WKS (78 DYS)
U.K. DISTANT LOC - 1.2 WKS (6 DYS)
TOTAL SHOOT 21 WKS (100 DYS)
TVL/PREP/TURNAROUND - (14 DYS)

Acct#	Category Title	Page	Total
600-00	STORY	1	$4,052,130
610-00	PRODUCER	2	$4,302,758
620-00	DIRECTOR	4	$4,999,797
630-00	CAST	7	$17,063,709
650-00	TRAVEL & LIVING	14	$1,375,064
640-00	**Total Fringes**		$783,443
	Total Above-The-Line		$32,576,901
700-00	EXTRA TALENT	19	$261,646
705-00	PRODUCTION STAFF	20	$2,468,044
710-00	CAMERA	26	$2,052,914
715-00	SET DESIGN	29	$2,342,583
720-00	SET CONSTRUCTION	34	$10,333,239
721-00	SET STRIKE	37	$462,000
722-00	VISUAL EFFECTS	37	$13,001,942
725-00	SET OPERATIONS	39	$2,051,914
730-00	ELECTRICAL	44	$2,956,293
735-00	SPECIAL EFFECTS	48	$6,651,637
740-00	2ND/ACTION UNIT	57	$5,159,576
741-00	PRODUCTION GRAPHICS	68	$777,700
742-00	PLATE/P'UP/BEAUTY UNIT	69	$253,946
743-00	GREECE LOC- LOCAL SPEND	69	$1,148,532
745-00	SET DRESSING	76	$4,206,948
750-00	PROPERTIES	82	$1,179,100
751-00	UNDERWATER/AERIAL	85	$771,655
755-00	WARDROBE	87	$1,911,974
760-00	MAKEUP & HAIRSTYLISTS	90	$703,364
765-00	PRODUCTION SOUND	92	$540,916
770-00	TRANSPORTATION	94	$1,883,389
772-00	PICTURE VEHICLES	97	$950,902
775-00	LOCATION EXPENSE	100	$3,266,100
781-00	AFRICA LOC-LOCAL SPEND	105	$1,610,589
785-00	PRODUCTION DAILIES	112	$1,042,776
790-00	TRAVEL & LIVING EXPENSE	113	$1,211,981
797-00	TESTS	114	$238,190
798-00	FACILITIES FEES	114	$6,928,277
795-00	**Total Fringes**		$2,511,179
	Total Production		$78,879,506
800-00	EDITING	117	$2,902,692
810-00	MUSIC	120	$3,322,410
820-00	POST PRODUCTION SOUND	121	$662,316
830-00	STOCK SHOTS	122	$76,000
840-00	TITLES	122	$107,800
850-00	OPTICALS, MATTES, INSERTS	122	$77,000
860-00	LABORATORY PROCESSING	122	$423,120
870-00	**Total Fringes**		$240,418
	Total Post Production		$7,810,756
910-00	ADMINISTRATIVE EXPENSES	123	$2,167,323
912-00	PPC INTERNAL	124	$0
920-00	PUBLICITY	124	$137,030
940-00	**Total Fringes**		$0
	Total Other		$2,304,353

Retrieved from Epstein (2013), used with permission