DOCTORAL DISSERTATION

Gábor Murai

Pécs, 2021

University of Pécs
Faculty of Business and Economics

Doctoral School of Regional Policy and Economics

Gábor Murai

Modeling the demand and supply of product related information; using evidence from YouTube

DOCTORAL DISSERTATION

Supervisor: Dr. habil. János Barancsuk

Pécs, 2021

Contents

List of Tables

Abstract

Despite the role of product related information in new product launches, our knowledge about its demand and supply on the product reviewer market is very limited. The dissertation aims to fill this gap in the literature by modeling the economy of product information using data from YouTube. Based on this gap, our main objective prior to the hypothesis formulation was to explore the product reviewer market on YouTube and identify the role, the demand, and the supply of product related information

We found, that based on the products the videos are reviewing, we can identify different information markets on the platform, and this segmentation significantly differentiates the performance of the videos posted on them as well. However, the topics' effect on the videos are diminishing over time.

Then, we were able to formulate hypothesizes regarding the characteristics of the demand and supply on the market and build model extensions aimed to answer them. First, related to the demand, we used the literature on information search to endogenize the overall interest towards the topic into a current state of topic awareness, which is mostly driven by the satiation of the audience. Our results indicate that both the topic awareness effect and its counterpart, the satiation effect are significant, having positive and negative relationship with the performance of the videos, respectively. Second, we also aimed to unfold the supply on the market and move away from the homogenous channels' assumption. We considered two factors that can differentiate these channels, their sizes, and their unobserved brand images, based on the personal branding literature. We found that the sizes of the channels have significant positive impact on the performance of the videos, while having significant negative effect on the above-defined demand effects. Our results suggest that the unobserved factors related to the image of the brand also significantly differentiates both the response variable, and the topic effects.

Finally, accounting for the long-term incentives of the channels, we aimed to derive a set of models examining their growth. Our main question to these models were whether and how the performance of the videos translates into subscriber counts. We accepted this hypothesis and found that the outstanding videos provide extra effects for the growth. In addition, we tested whether the video level reactions from the audiences can be related to the growth and found that the average ratio of likes to views and dislikes to views proved to be significant.

1. Introduction

1.1. Motivation

Product related information is one of the main drivers of new product launch successes. Therefore, its content, how it is presented, and how it is perceived are key information for managers at firms that launched, or plan to launch a product on the market. The literature differentiates three type of medium in which this information is distributed. The owned media, such as the website of the firm, the paid media, such as billboard advertisements and the earned media, such as amazon reviews or twitter posts coming from users or experts. While the firms essentially have a very high-level control over the owned and paid media, understanding earned media poses a great challenge for them. Nevertheless, marketers cannot simply avoid earned media and focus on the other two if they aim for success, as this medium could have immense effects on the market performances of their products (Erdem and Keane, 1996; Reinstein and Snyder, 2005; Wu et al. 2015; Li and Du, 2017). As Newman (2014) describes in his article:

> *"Earned media […] hardly ever works alone. You have to make it a part of your marketing ecosystem along with paid and owned media. The truth is: in today's digital landscape, they either work together or they don't work at all."*

Thus, if firms aim to understand how information about their product is going be reached by consumers, besides controlling paid and earned media, they also need to understand the drivers of earned media.

From the perspective of the firms, this challenge has steadily become even more difficult in recent decades. Along the widespread of the usage of the internet and social media, new platforms and possibilities emerged for those who aim to post product related information, making the earned media ecosystem increasingly more complex. In case of user reviews, the traditional word-of-mouth (WOM) of consumers who already bought the item now can be reached by almost anyone in the world in various forms, such as online ratings, or text-based feedback. Nowadays, most of the online ecommerce platforms have a segment for user feedbacks, but there are also websites dedicated only for such reviews.

The other main type of the product reviews is the feedback that does not come from the companies or users, but from some third-party information intermediary. Due to its function in the consumer learning process, we simplify the notation of this category as expert or professional reviews. In case of this type of reviews, we can observe that the domain has changed just as much as that of user reviews.

First, the traditional magazine or newspaper segments of product reviews has moved to blogs and websites. Then, with the emergence of organized online attention platforms, such as YouTube, the profession or expert reviews evolved into the complex ecosystem that we can observe nowadays. In this system, while the role of blogs and websites remained meaningful, the websites, where all the reviewers and consumers share the same platform has grown to be an integral source of product related information for consumers.

The different structure of these platforms has multiple consequences compared to that of previous model with separate websites, that resembled more the traditional newspaper or magazine model. The centralized supply provides easier access of information from more sources for consumers. Meanwhile, the properties of the platform make the entry to the market accessible for anyone who aim to pursue a carrier in this expertise. One can also argue that the centralized demand creates a completely different route to success than previous models.

Therefore, if firms and marketers want to understand how consumers access, gather, and ultimately learn about their products, they are facing an increasingly difficult challenge. They need to get a grasp on how product related information flows in the modern reviewer market and understand that reviewers nowadays may have different motives and incentives due to this complex ecosystem.

The literature on the evolution of user reviews and its effect on the consumers and the firms are well-documented in the marketing domain. However, our understanding about the expert review ecosystem is very limited in general and our knowledge is especially scarce regarding the modern shared platform reviewer market. Therefore, in this dissertation we aim to fill this gap in the literature and shed light of the main drivers of this complex market. In addition, we choose YouTube, one of the most popular organized online attention platforms to examine and model the expert review system.

Based on these arguments, our broad objectives prior to the research were the following:

1. Explore the role of product related information in the reviewer market.
2. Identify the key characteristics of the demand and supply in the market.
3. Examine the relationship between these characteristics and the information *"product",* which is the video containing the information

1.2. Related Domains

Examining product review channels on YouTube is a special field in the domain of marketing, as it lies in the intersection of multiple different literature streams. Hence, in this section we outline the most important connection points of the dissertation with the marketing and economic discipline. Then, in the next chapter, we will discuss the studies from these streams of literature that we are building on more thoroughly.

Product Reviews

Based on the content of these videos, the dissertation connects to the product review literature which can be divided into two parts in terms of the source of the reviews. The reviews that are coming from the users of the product who already used it and the ones that originates from professionals, defining the product review expertise.

The literature stream on user reviews is mostly focusing on the consumers' learning process when they are facing these feedbacks from other consumers. The results of this category had many implications for the dissertation. For instance, it describes the evolution of the individual and aggregate level of uncertainty and demand for information regarding a new product. Therefore, we will largely rely on this field during our model development process. However, we argue, that product reviewers on YouTube fit better to the field of expert reviews. The relatively narrow literature that is available for this domain are mainly examining the economic impact of the reviews on the firms (such as sales or market value). Nevertheless, the findings of these domains (both user and expert reviews) have shown that reviews in general play a crucial role in the consumers' quality perception and expectation about goods with uncertain quality. They also highlight the need for firms to understand these processes and acknowledge the role the earned media in this regard as well.

Behavior of the media

In contrast to the studies in the field of product reviews, the dissertation aims to model the expert review market itself, and not only examine its impact on the consumers or the firms. This includes the participants with their incentives and the dynamics in the market caused by the product related information. While we identified a gap in the marketing domain regarding to modeling the market of product relation information itself, we can find theoretical models from other disciplines where the agents have similar goals and incentives as in our approach. This stream models the behavior of various types of media. However, the framework of these studies is different compare to ours.

First, only a few studies examine similar decision variables of the actors in the market. Most of these studies investigate the decision regarding the objectivity, accuracy, political orientation, price, or programming variety of their content, which is not appliable to our model. However, perhaps the most important difference comes from the researchers' methodological choice, as this literature stream is building models on a theoretical level, while the dissertation uses quantitative models tested on data downloaded from YouTube. Nevertheless, this domain also points out important details for the dissertation as it unfolds the theories behind the different revenue models of the media. Based on this aspect, we can conclude that our approach builds on the model derived by Falkinger (2007) and Xiang and Sarvary's (2013). In their framework, they assume that news providers try to maximize ex ante expected audience size to achieve the optimum. This also means, that they have a fixed rate per viewer advertising and content revenue.

Personal Branding

Online personal branding is one of the trending topics in the marketing literature in the recent years. The main connection point here is the argument that YouTube review channels are creating, building, and managing their own brands as it is defined by this domain. The self-brands' unique property that their faces are the brand itself, it is built around the individual. For instance, we can mention the brands around popular figures such as Gordon Ramsey, LeBron James, or Calvin Klein, but the domain also considers the management of the brand of influencers as well.

From the perspective of the dissertation, the direction of persona-fied brands has the most relevant consequences. Essentially, we can conclude from these studies, that the image of the brand, the persona is a performed role by the individual who is the face of the brand. She/He is doing this to meet the expectation that she/he or her/his advisors considers to be connected to the profession of the brand. To successfully manage these types of brands, the channels need to appropriately merge different persona facets, features into a brand image and narrative.

Therefore, thorough the dissertation we consider these channels similarly to the brands in other industries. We develop our model to account for the differences among the channels' brand images, while we also consider the implications of our models to address the above described argument regarding the persona-fied self-brands.

YouTube and the video format

Finally, based on the chosen platform and format of the reviews, the dissertation also connects to the literature on YouTube and video content. However, important to note that the relation here is only methodological in its nature.

The literature on YouTube helps us to understand the unique features that only applies to this platform and can significantly alter our model if we do not account for them. The best example could be the control for the lifetime of the video when we estimate our model. Here, we can build on the studies that already examined the evolution of the views of the videos on YouTube.

The other important aspect of our chosen segment is the video format. As the related studies pointed out, that consumers react differently, they can be more influenced if they can actually see the product in someone's hand when they are using it, which strengthens our arguments regarding the product related elements of the video. In addition, we argue that this format enables more room for personal brand building than traditional text based expert reviews.

1.3. Modelling the product reviewer economy

Our approach to model the product reviewer economy is built around the product related information. Thus, our model development process starts by the definition and identification of the information markets on the platform that can be connected to new products on the market. Based on the volume of both new products and product review videos, we choose the smartphone industry to estimate our models. We define an information market in YouTube as the collection of the videos providing information about a given product on the market, and the demand for information as the audience's interest towards these videos, which we refer to as the information *products*. Therefore, we can measure the overall demand for information by the number of views the information *products* received in the past.

Making the ground for most of the models in the dissertation, we first hypothesize that the segmentation of the platform to different information markets is indeed significant. More specifically, the topic of the video, which is the product it is reviewing, has meaningful effect on its performance, denoted by the view count changes from one period to another. However, from the information market and product review literature we also know that as the uncertainty of the consumers decrease, the demand for information is decrease as well. Hence, we not only test the presence of the effect on the topic, but we also expect it to decrease over time.

H1:

A: The product reviewed in the video has significant effect on the performance of the video.

B: The effect of the product on the video is decreasing over time.

This approach was aimed to test the exogenous effect of the topic on the videos. We argue that while this is a very important aspect of the model, endogenous effects should be also represented. Thus, we aim to derive the endogenous measure(s) of topic interest from the aggregate behavior of the market participants. First, still relying on information economics, we assume that the individuals' interest towards a topic decreases over time due their satiation of the information. Therefore, after the point when they joined the market, they are gradually losing their interest over time. However, we do not assume that every viewer would become more and more satiated at the same rate.

Besides the individuals' satiation and topic interest effect, the channels are also an integral part of the market and their activity may also affect the performance of all the videos on the market. First, we assume that the channels are competing each other for views. We argue that based on the definition behind competition, it is only possible if there is scarcity regarding the focal resource. Hence, the competition among channels is connected to the satiation effect of the market, since that property shows that the viewers interest is finite. On the other hand, we also argue that as channels are posting videos on the market, they can also raise the overall interest towards the topic. This may work by directing some their unique fanbase to videos on competitor videos with the same topic. Among others, another possibility could be that they are making content such that it is interesting enough for the audience to incentivize them to remain aware and follow up on the topic. Either way, this effect essentially raises the pool of aware views on the market. Therefore, it is connected to our previously described topic awareness.

Based on the probabilistic properties of finding already satiated or still interested viewers, we derive a function that separates the viewership of the topic to recent views, representing the share of audience that is still interested, and to views that happened earlier, showing us the share that are already satiated.

With this function, we are able to introduce the current state of satiation and topic awareness into the model. Nevertheless, the properties of this function and the estimation of the model poses us a challenge to overcome. Hence, we formulate the following research questions regarding the endogenized topic interest:

RQ1: What is the resultant of the potential positive and negative endogenous topic effects on the YouTube product reviewer market?

RQ2: How can we separate the aggregate effect to represent the satiation and topic awareness of the consumers and the competition among channels?

Based on the answers to these questions, we can now hypothesize the main statement regarding the goals of these models:

H2: Recent topic views have a positive, while the ones that happened earlier have a negative impact on the performance of the videos.

In our model we differentiate three levels. The levels of the videos, the level of the product, which is the collection of the videos on the same topic, and finally, the level of the channel, which is the collection of the channels' videos on multiple topics. So far, we modelled the relation between the video and topic level, but we did not account for the channel level. As we outlined in the previous chapter, the personal branding literature shows us that we should not handle the supply on the market as a set of homogenous actors. Instead, we assume that we can observe heterogeneity among them from multiple aspect. First, we can understand the brands of the channels as a buffer in terms of the performances of their videos. In other words, the channels with more attractive brand images have a competitive advantage compare to other channels. Second, we also test the possibility that the brand image may not be independent from other factors of the model. This includes the topic effects we derived before. Meaning, that we test whether a channel with more attractive brand image have different relation to the topic information market than channels with worse image. Therefore, we hypothesize the following:

H3:

A: The unique channel characteristics have a significant effect on the performance of the videos.

B: The unique channel characteristics significantly differentiates the topic effects for the channels

The other differentiating factor among channels that we account for in this dissertation is the aspect that they have different sizes. Here, we build on the consideration that relies on the arguments of size dependent market power and possibilities of being a *"niche"* topic creator. Similarly to the previous differentiation, we test the effect of the size of channel from two perspective. It may be a buffer to the performance of the videos, but it can also alter the relations in the whole model. As an example, we may expect that bigger channels can facilitate the topic awareness effect better and grab more share from the common effect of the trending topic. Since we denote the size of the channels with the number of subscribers they have at the moment, we outlined the following hypothesizes:

H4:

A: The number of subscribers of the channels has a significant impact on the performance of their videos.

B: The number of subscribers of the channel has a significant interaction effect with the topic effects in the model.

The result of the Hypothesis 4A has another important implication for the channels. This is due to the aspect that the size of the channel could have multiplicative benefits for the channel if it is proven to be significant. The process in which this can work relies on the argument that there could be a relationship between the channels' size and the performance of their videos for both directions. If we find evidence that not only the channel size affects the views of the videos, but the views of the channel's videos also translates into subscribers at later periods, the channel size have multiplicative effect for the revenue of the channels. In this process the channel size affects the number of views its videos get, then the views translate into subscribers that causes even higher number of views in long term. Due to this potential connection and long-term incentives of the channel, our second set of models are built to explain the growth of the channels.

We derive the base model representing the discussed relationship where the performance of the videos can translate into subscribers. Then, we extend this approach into two directions. First, we argue that if the channels make videos such that it reaches outside of the usual viewership of the channel, it can generate a boost for the subscriber gaining process. Second, we try to explain this process by using the reactions from the audience towards the videos of the channels to understand some of the motives of the subscribers. Here, we use two different methodology, representing different consideration process behind the subscribing decision. We derive a model with an underlying assumption that the videos' properties are essentially the manifestations of the channels' overall properties. Hence, with the aggregation of the number of audience reactions across videos we can derive an average view for the channel. We can use three reactions for these models: the number of likes, dislikes, and comments. Our second approach takes the number of new view counts of the videos of the channel from period to period and assumes a process on a video contribution level. Thus, for the overall set of models, we formulated four hypothesizes, highlighting different aspects of the growth of the channels:

H5: The view count changes of the channels' videos has a significant positive effect on its subscriber number changes.

H6: Outlier videos of the channel in terms of their view counts have significant positive extra effect on the subscriber number changes of the channel.

H7: We can explain the channel growth better if we use the channels' average audience reaction metrics.

H8: We can explain the channel growth better if we use the video contribution audience reaction metrics.

We summarized the system of hypothesizes and research questions in Table 1, which provides a hierarchical ordering of the model effects relating to our statements and questions.

1.4. Outline

The dissertation builds up as follows. After the introduction, the second chapter describes the most important theories from the related disciplines. This includes the literature on product reviews and earned media, the domain of modeling news firms and agents, and finally, the literature stream of personal branding.

Our third chapter presents the prior technical approach based on our goals and purposes of the models. First, this consist the sources of data and its collection process. Second, it also includes the definitions of the time dimensions of the obtained dataset, and finally, the description of the methodology of hierarchical modeling, which will serve as a baseline model approach in this dissertation.

The fourth chapter serves as an initial exploration of our dataset with the goal of develop the baseline model for the upcoming chapters. In addition, at the end of the chapter we also aim to investigate our first hypothesis, since it is related to the overall assumption whether we can observe information markets on YouTube. The role of this question is crucial for the dissertation since it makes a ground for all other questions regarding the performance of the videos.

The fifth chapter emerges from the question whether we can endogenize the benefit of product review channels choosing the right product to review. We are doing so by introducing the behavior of the participants in the information market to the model. This relates to both the demand and the supply of information. From the demand side, we model the satiation and topic awareness dynamics of the audience. From the supply perspective, we model the competition and topic awareness buff effect by posting a video on a topic.

The sixth chapter is the last chapter dealing with view counts of the videos. The main motivation of this chapter mainly comes from the idea, that the supply of information is not homogenous, it can be differentiated. The motivation for the first differentiating factor comes from the personal branding literature. Based on this domain, we essentially assume that these product review channels can be different brands. This consideration could be a moderation factor for the effects we unfolded previously. Second, these channels are different in their sizes, which could result similar moderation effect to that of brand images. In addition, this chapter also serves as a base of the next chapter as it describes how the connection between the channel size and the performance of the channels' videos may be observed in both directions.

Finally, the seventh chapter extends our current set of models about the performance of the videos with another set of models exploring the subscription changes of YouTube product reviewers. The goal of these models is to investigate the order of the relation between subscription and views and examine the possibility of multiplicative growth through the subscription gathering process of the channels. The main question of the chapter is whether we can explain the growth of the channels by the performance of those of videos. Notwithstanding, we also aim to extend this approach and try to explain this effect with the audience reactions towards these videos and with the videos that reach outside of the channel's regular viewership

Table 1: Structure of the dissertation

Category	Chapter	Response variable	Code	Represented Effect	Variable	Direction
TOPIC EFFECTS	Ch. 4	VIEWS	H1: A-B	Popularity	$Random\ Intercept$	Direct
				Age	$age\ of\ topic_i$	Direct
	Ch. 5		H2; RQ1-2	Total Past Views	$\sum_l^N \sum_t^{\bar{T}} \Delta Views_{l,i}$	Direct
				Satiation	$\sum_l^N \sum_t^{\bar{T}} w_{\bar{T}}(t)\ \Delta Views_{l,i}$	Direct
				Topic Awareness	$\sum_l^N \sum_t^{\bar{T}} (1 - w_{\bar{T}}(t))\ \Delta Views_{l,i}$	Direct
CHANNEL EFFECTS	Ch.6		H3: A-B	Persona	$Random\ Intercept$	Direct
					$S_{j,t}$	Direct
					$TA_{j,t}$	Direct
			H4: A-B	Size	$Subscription_{k,t}$	Direct
					$(S_{j,t} * Subscription_{k,t})$	Interaction - Topic
					$(TA_{j,t} * Subscription_{k,t})$	Interaction - Topic
CHANNEL GROWTH	Ch. 7	SUBSCRIPTION	H5	Performance	$\sum_i^{N_{kt}} \Delta Views_{it}$	Direct
			H6	Reach	$\sum_i^{N_{kt}} \Delta Views_{it}^{\overline{Views}_{it}}$	Direct
			H7-8	Audience Reactions	$\frac{\sum_i^{N_{kt}} Audience\ Reaction\ Metric}{\sum_i^{N_{kt}} Views_{it}}$	Indirect
					$\sum_i^{N_{kt}} \frac{\sum_i^{N_{kt}} Audience\ R.M.}{Views_{it}} \Delta Views_{it}$	Direct

Source: own elaboration

2. Related Literature

2.1. The role and types of product related information

The role of product related information is especially important (for both firms and consumers) in case of consumer uncertainty that is present on the market due to the consumers' lack of sufficient knowledge about the quality of a given product or service (Oren and Schwartz, 1988; Roberts and Urban, 1988; Erdem and Keane, 1996; Iyengar et al., 2007; Narayanan and Manchanda, 2009; Zhao et al., 2013).

The reason behind this phenomenon relies on the theory of consumers' decision-making process. Although the marketing domain discovered many factors (e.g. Barone et al., 2000; Hall et al., 2001; Berger et al., 2007; Melewar et al., 2010) that can potentially influence the choice of consumers between two alternatives, the roots of the theory of choice can be found in the microeconomic literature (e.g. Friedman and Savage, 1948; Arrow, 1959; Debreu, 1959). According to these studies, consumers have a stable preference order over all the alternatives, which can be derived from their utility functions. However, there are many cases when consumers could be uncertain about this preference order, which implicates, that they cannot be sure about the optimality of their decision prior to the decision-making. For instance, such cases could arise from situations when the product is new on the market (e.g. Oren and Schwartz, 1988; Narayanan and Manchanda, 2009; Zhao et al., 2013) and consumers does not have enough and/or trusted information about its quality. Other cases could be if the consumer makes a menu choice with state-dependent utility function (Kreps, 1979; Dekel et al., 2001; Ahn and sarver, 2013) or in the presence of inherit product variability (Roberts and Urban, 1988). However, in this dissertation we focus on the first examples, the uncertainty due to new product launches.

In this case, we assume that consumers do not have a first-hand experience with the product, so they need to rely on other information sources to form an expectation about the properties of the unknown product, including its quality and essentially its marginal utility for the consumer. Then, based on these expectations, the consumer can compare the products and make decision. Consumers may have prior expectations about the unknown products before being exposed of any kind of information regarding the given product from a variety of sources, such as their peers through traditional word-of-mouth or advertisement. This prior expectation could come from prior experiences with the company's other products through brand related learning, or with different brands through cross-brand learning, but it can also come from information regarding products in the same category through category learning (Narayanan et al., 2005; Szymanowski and Gijsbrechts, 2012; Zhao et al., 2013). However, our main focus in this dissertation is the demand and supply of information about given unknown products, so from this stream of literature, we are mostly building on the studies examining the learning from information regarding the focal product.

Nevertheless, all the above-mentioned types of information sources are highly valuable for consumers since they can reduce the uncertainty of the decision-making processes. This means, that the decisions made on the expectation about the quality of an alternative, will be less risky, the probability of making wrong decisions becomes smaller. Important to note, that one usually assumes that the uncertainty of consumers cannot be reduced to zero if the product is unknown for the consumers until they gain first-hand experience. In this approach, the additional information has decreasing benefits for the consumers, which property of the process is closely related to the information search literature (Nelson, 1970; Stigler, 1961; Roos, 2013).

Another aspect of this area of the literature that is worth addressing is that consumers tend to differentiate among the information pieces collected over time about the same product in terms of their informativeness. In other words, they incorporate these information pieces into their expectations with different weights. These weights are essentially the manifestations of the consumers' trust and opinion about the credibility of the information. (Hu et al., 2011a, 2011b, 2012; Zhao et al., 2013;)

However, one of the most important aspect of the studies examining the role of product related information is its categorization by the relation and level of dependency between the product or firm and the source of the information. According to this differentiation, we distinguish three types of information sources (Stephen and Galak, 2012; Lovett and Staelin, 2016; Colicev et al., 2018). The company owned media, for instance the website of the channels containing information about the product in the format of specification comparison. The paid media, such as the advertisements about the product. Finally, the earned media created by independent, or quasi-independent sources of information, such as reviews, mentions, or ratings from the users of the products or experts.

As the dissertation is aimed to enrich the literature on product related information by earned media, we mostly rely on the studies in this category. Hence, in the following chapters, we discuss the studies and directions from the earned media literature on which the dissertation is built.

From the different possible types of earned media, we first describe the literature on user reviews. Then, we extend our scope with the considerably narrow literature on expert reviews. However, the studies in these two sub-categories mostly focus on the impact of the information on some metric regarding the performance of the product or the firm that sells the product. In contrast, the main objective of the dissertation is to examine the demand and supply of the product related information itself, and not its effect on the demand or supply of the given product. Finally, we also explore the possibility of earned media coming from news agents, media firms or other professional information mediators. Since these studies also model the behavior of the information mediators and information mediation process, they are related to our models in this respect. However, these models are examining theoretical concepts, while this thesis have different objectives. We aim to use empirical data to look for evidence for the questions and hypothesizes arising from the identified gap in the literature regarding the demand and supply of product related information.

2.1.1. Directions in the literature on earned media

In this section we outline two types of product reviews differentiated along the perspective that the source of the reviews are users that are already familiar with the product, posting reviews, or professional, third-party reviewers, posting with the intention to achieve profit. However, the incentives behind the review is not the only aspect these two types of content are different. For instance, a differentiation can be made from the consumers attitude towards the review in terms of informativeness or credibility, but we may also observe that these two types are different in terms of the entertainment factor of the review or the moral and ethical responsibility of the reviewer.

We describe the literature and arguments related to these domains from three point of view, answering three different, but interrelated and perhaps equally important questions for firms. “Why examining earned media is important for the firms?” “How consumers use product related information in regard to their perception about the focal product?" "What are the incentives behind posting the content?"

2.1.1.1. User Generated Content

One of the main types of earned media a given company can receive for their products is the content that is generated by the consumers who have already used the focal product. The possible forms of this content has a wide range, from simple mentions (e.g. Stephen and Galak, 2012), through ratings (e.g. Chevalier and Mayzlin, 2006; Zhao et al., 2013, Wu et al., 2015) to the detailed text-based reviews (e.g. Tirunillai and Tellis, 2012, Hu et al., 2012).

Multiple studies have shown that these reviews could have an immense effect on the perception of the consumers those are still uncertain about the product(s) on the market. Hence, it is crucial for firms to understand the nature of this information market as it is shown by the literature (Dellarocas et al. 2007; Chevalier and Mayzlin 2006; Zhao et al. 2013; Wu et al., 2015). We describe the most important findings more thoroughly below.

Zhao et al. (2013) modelled consumer learning from both own experience with the same genre (books) and learning from online reviews. Their results shown that 1. consumers learn more from online reviews than from their experiences, 2. fake reviews increase the consumers' uncertainty regarding the underlying product, 3. online reviews has an impact on the firms' profit, 4. this impact is diminishing as the number of reviews are increasing.

Wu et al.'s (2015) proposed a model of online reviews and derived the economic value of reviews from this model. They estimated the model on restaurant dining data and reviews from Dianping.com, a popular Chinese user review website. They found that the economic reviews are beneficial for both the consumers and the restaurants. Consumers on average gain around 6.7 CNY (Chinese yuan) value from the reviews. Moreover, they also found that contextual reviews, comments are more valuable for consumers than ratings. For restaurants, consumer reviews increase the probability of consumer visits, thus, increasing their profit by 8.6 CNY on average.

Reflecting to these results He and Chen (2017) derived an optimal pricing strategy for the firms that assume that consumers learn about their products quality from consumer reviews.

The methodology in which this stream of literature models the consumers information incorporation process is the Bayesian update mechanism (Erdem and Keane, 1996; Miller, 1984; Roberts and Urban, 1988; Szymanowski and Gijsbrechts, 2012, 2013; Wu et al., 2015; Zhao et al., 2013). This method allows the researches to examine how uncertainty regarding a product is evolving over time on an individual level by learning additional information pieces from various sources. Moreover, it also enables to investigate the credibility corresponding to the reviews (Zhao et al., 2013), for instance with a question of whether consumers indeed acknowledge the fact that there could be fake reviews.

The examination of the incentives behind the posting decisions of these reviews can be considered as a relatively small domain. Only a handful of studies examined the users' motives behind their expression to share their opinion in the form of product information. (Nardi et al. 2004a, 2004b; Mackiewicz, 2008, 2010)

Mackiewitz (2008, 2010) discusses three potential drivers behind consumers decision to take an effort an express their opinion regarding the product. First, consumers may see the reviews beneficial for them because they seek a sense of efficacy. According to these reasoning, consumers may write these reviews to have a feeling that they had some impact on the world. Second, consumers may share their information based on pure altruism. This means, that they simply want to help others making better decisions and for this goal, they even take the time and effort to write these reviews. Finally, a possible explanatory driver could be the tendency of humans to crave attention, and their need to be heard.

In conclusion, we can assume that users are usually buy the products for their own usage and then they share their opinion about it offline by word-of mouth (WOM) or online by electronic WOM (eWOM) in the form of mentions, recommendations and/or text or rating product reviews. Then we can infer from the literature discussed above, that the willingness of the consumers to post reviews and express their opinions is based on utilities derived from various psychological "rewards", such as altruism, need for attention and/or for the feeling that they have affected the world somehow.

Overall, this is an important point that differentiates these reviews from those of made by posters whose profession is to review these products, in which case the posting decision of the reviews is corresponding to monetary incentives. Other factors separating these two types of reviewers could be considerations that consumers may also perceive expert reviews to be different in terms of its informativeness or credibility.

The most important aspect in which the dissertation relies on the studies in this stream of literature is that they show how the uncertainty evolves over time on an individual level. This happens by consumers updating their expectation about the product and uncertainty regarding this expectation after incorporating more and more information about the products from others.

2.1.1.2. Third-party or expert reviews

The literature on professional or expert consumer reviews is relatively small in the marketing domain compare to that of on other sources of product information and it examines the reviews empirically with data from only a handful of industries.

The most researched area in this domain examines the reviews' effect on the sales performance in the motion picture industry (Reinstein and Snyder, 2005; Eliashberg and Shugan, 1997; Basuroy et al., 2003, 2008; Boatwright et al., 2007; Prag and Casavant, 1994; Gemser et al. 2007; Henning-Thurau et al., 2012; Terry et al., 2011), while Cox (2015) and Hilger et al. (2011) showed similar effects in case of the video game and the wine industry, respectively.

Other approaches showed the effects of the reviews on the firm strategy in case of printers and running shoes (Chen and Xie, 2005) or the effect on firm value in the movie (Chen et al., 2012) and consumer electronics (Tellis and Johnson, 2007) industry. One exception from this is Kim et al's (2019) paper, focusing on the reviewer's psychological trade-off between being objective or helping the brands.

These studies highlighted how important expert reviews are for firms regarding their performance in general. In recent decades, with the widespread of the internet, the professional reviewer system, and hence this expertise has started to evolve to a more complex system. In the offline era, professional reviews were first either a part of, or they were separate printed media. Then, TV and radio stations had their possibilities to have segments dedicated to these professionals. Examples to this kind of professional reviews could be book, movie or museum review sections in the magazine of *"The World Today"* (https://www.chathamhouse.org/publications/the-world-today), *"It Its Innovation (i3)"* magazine (https://cta.tech/Resources/i3-Magazine) by the Consumer Technology Association (CTA) or the popular TV show *"Top Gear"* (https://www.topgear.com/), focusing on reviewing primarily motor vehicles.

The expert reviews published or broadcasted in an offline medium meant, that becoming a professional reviewer has entry costs, and it is not something that anyone can immediately start to pursue. This barrier has changed with the internet. While some of the offline media, containing expert reviews, has launched an online extension or fully moved to an online format, the biggest difference was that now everyone could become a professional reviewer by creating websites or blogs dedicated to reviewing typically one or just a couple of product categories. We can mention websites that were born from previously printed media such as *"goodhousekeeping.com"*, reviewing housekeeping appliances or *"expertreviews.co.uk"*, which is a collection of product reviews on a few different categories. An example for a website that did not have a prior offline media behind it is *"GSMArena.com"*, which will be also one of our information sources in the data collection process.

The professional review market has developed even further in the recent decade with the widespread of the usage of social media and organized online attention platforms, such as YouTube (Smith, 2020). These websites essentially give platforms for the demand and supply of information to meet each other. This means, that it is easier to become a reviewer on the supply side, making the entry to the market even easier for anyone aiming to pursue a carrier in this expertise. However, it could be also beneficial for the consumers, as it is easier to get information from multiple sources from various reviewers.

Hence, we argue, that the expert review system has been evolving from a simple, more segmented market to a more complex ecosystem where all the reviewers and consumers share the same platform. In this platform it is easier to become a reviewer on the supply side, and easier to get information from more reviewers on the demand side, while the older, more traditional sources (e.g. user rating, advertisements, etc.) of information still play an important role in the consumer decisions. Therefore, if a firm aim to understand how their target consumers access, gather, and learn about their products from experts in these platforms, they are facing an increasingly difficult challenge. They need to understand how the product related information flows in the platform, how consumers seek for information and what are the incentives of the reviewers on the market in this market.

In contrast, each study described above corresponding to one of the two main types (user or expert) of reviews is focusing on some economic impact on the firms (such as sales or market value) or the product (purchase intention) and not the demand and the supply of the product information itself. Therefore, we aim to fill this gap in the literature by modeling the product reviewer economy, including both the motives of the consumers of the information and the incentives of the experts providing this information, and finding empirical evidence collected data from YouTube, one of the emerging platforms of product related information.

In addition, the marketing domain lacks literature that aims to model the incentives of the expert reviewers in general. The most closely related studies are in streams of literature that explores the behavior of media firms, news providers and other entities that aim to attract the attention of the audience. Another stream that guides our hypothesizes is the consideration that the incentives of these reviewers may have some of the incentives with that of influencers on different, but more importantly on the same platform. Hence, in the following chapters we outline the novelties and most important consequences of the findings in these two areas to the dissertation.

2.1.2. Theoretical models on information mediators

The literature stream that examines the behavior of the information mediators is a domain which consist of studies with multiple different assumptions regarding the goals and incentives of the entities modelled by them. Hence, we can also observe that the decision variables of the information mediators, derived from these assumptions, are also different in these papers.

There is a considerable number of studies focusing on the objectivity, accuracy, or political orientation of the presented content (e.g. Mullainathan and Shleifer. 2005; Xiang and Sarvary, 2007; Battagion and Vaglio, 2015; Gabszewicz et al., 2001; 2002; 2004), but there are also studies concerning the decision of the information mediators with respect to the price to access information (Godes et al., 2009), programming variety (Gal-Or and Dukes, 2003) and presented information signal (Falkinger, 2007; Xiang and Soberman, 2014).

However, these models are not only different in the perspective of the decision variables of the information mediators but also in terms of their source of revenue. While Gal-Or and Dukes (2003) assumes only advertising revenue, Godes et al. (2009) assumes content and advertising revenues as well. Our approach in this regard is most closely related to Falkinger (2007) and Xiang and Sarvary's (2013) study, assuming that news providers try to maximize ex ante expected audience size to maximize their revenue. This means, that agents have a fixed rate per viewer advertising and content revenue. Important to note, that thoroughly the dissertation we derive, that YouTube channels could essentially have two objective: maximizing the size of audience that watches their content and to maximize the size of audience that become subscriber for the channel. However, out of these two potential goals, only one results direct revenue for the channel, the audience that watches their content. The other objective only contributes to the revenue indirectly, through the first objective.

The last segment of this domain that we are building on during the development of our models is the studies concerning attention economies partly (Smith, 2020) or entirely (Falkinger, 2007). These studies highlighted how different these markets are from traditional markets with a clear demand and supply definition, based on the approach that YouTube channels, media firms or similar information mediation entities are trying to attract the attention of the audience. Assuming different attention capacities for every audience members and competing information signal sellers, with their decision to choose the strength of the signal, Falkinger (2007) was able to derive the equilibrium audience sizes. His findings rely on the theorems proved on a theoretical model that may be applied to platforms and fields where the supply side aims to attract attention from the audience members. Therefore, Falkinger's (2007) model can be easily translated to the case of YouTube. The "*family of information signal sender*" -Falkinger (2007) is essentially the supply of information, which equals to the set of YouTube channels in this platform. The set of information signal receivers is the set consisting individual audience members, in other words, the aggregate audience.

In addition, these studies also show how attention grabbed by a channel can create more attention later for their or others' posted content. These findings, along with the personal building literature (Chapter 2.2) will further motivate us to represent effects that relates to the information signal sellers' spill-over effect on the market, or their capability to build a follower base for long-term benefits in chapters 7. and 5., respectively.

Nonetheless, there is a key difference between this domain and the dissertation. Besides Smith's (2020) paper, the results of the studies discussed above were derived from theoretical models without the usage of empirical data. In contrast, as stated in our main objectives, we aim to explore the research questions and hypothesizes by developing empirical models using data downloaded from YouTube (Chapter 3). To our knowledge, this is a major novelty in this domain, as it is the first analysis approaching our objectives this way.

2.2. Personal Branding

Our last building block in this dissertation is the domain of personal brand building. The studies in this field examine brands that are built around an individual. (Thomson, 2006; Dion and Arnould, 2011, 2016; Kerrigan et al., 2011; Bendisch et al., 2013; Parmentier et al., 2013; Moulard et al., 2015; Duffy and Hund, 2015; Scolere et al., 2018; Fournier and Eckhardt, 2019; Smith, 2020). These personal brands may compete with firms in the same industry, or it can be an extension of a firm on the market, but there could be also cases, when the supply only consist individual brands. We can observe examples to these types in various industries such as it was shown by Hewer & Brownlie (2013) and Dion & Arnould (2016) with the case of chefs Joël Robuchon, Gordon Ramsay or Jamie Oliver in the restaurant and cuisine-related markets, but we can find examples in the homemaking industry, examined by Fournier and Eckhardt (2019) and Murphy (2010) on the Martha Stewart brand. In addition, we can also list plenty of personal brands centered around popular athletes (LeBron James, Tom Brady, Serena Williams, etc.) or famous fashion designers (Calvin Klein, Donna Karan, etc.).

Similarly, to the literature on expert reviews (Chapter 2.1.1.2) the emergence of the internet, social media, and organized online attention platforms opened new directions in the field of personal branding as well. In these platforms, individuals can build their follower- or fanbase. Whether intentionally or not, this fanbase building often leads to similar personal brands to that of traditional figures, discussed above. We can mention examples for such brand building processes in case of bloggers (Duffy and Hund, 2015; Delisle and Parmentier, 2016; McQuarrie et al., 2012) or like in our case, in case of YouTube channels (Chen, 2013).

We also considered that it is important to mention the literature on influencers and celebrity endorsement (e.g. Lee and Watkins, 2016; Sokolova and Kefi, 2020; Burke, 2017; Munnukka et al., 2019). This stream of literature is strongly connected to the domain of personal branding as well, and there are some aspects of influencers and celebrities which resemble the YouTube channels we put in focus in this dissertation. Such similarities could be for instance the follower and subscriber gathering incentives for both type of content creators. However, these papers are focusing on the effect of the activity of the influencers or celebrities on the performance of a product on the market or a firm, such as purchase intention or sales.

The dissertation is focusing on the YouTube product review channels, that intentionally chosen this expertise and built their brands in the previous decade. Therefore, we only highlight findings from studies in this domain, in which individuals intentionally invest in the process of creating and building their own brand image to achieve success on the long term and become popular self-brands on the market.

Dion & Arnould (2016) analyzes the persona-fied brands, which term is summarized the best by the definition: *"[...] they show and do what they are, simultaneously performing distinctively but with reference to a normative schema recognised by networks of stakeholders (Bode, 2010; D'Adderio, 2008; Durand, Rao, & Monin, 2007; Feldman & Pentland, 2003; Kjellberg & Helgesson, 2006)."* - Dion and Arnould's (2016). In other words, the image of the brand, the persona is performed, played by the actual individual to meet the social expectation towards the profession corresponding to the activity of the brand. The authors then argue that in order to successfully manage these types of brands, they need to appropriately integrate different persona facets, features into a brand narrative. Duffy and Hund (2015) shows how fashion bloggers thinks about the ideal persona, built by them; what is their *"having it all"* perception about this profession. Fournier and Eckhardt (2019) examines the human elements of the personal brand, highlighting the role and management of characteristics such as hubris, unpredictability, and social embeddedness. They argue that these human factors may compromise brand value, but with the right management, they can also benefit the brand through creating a perception in the consumers regarding intimacy and authenticity. Finally, Scolere et al. (2018) shows that platform dependency of the developed personal brands by the individuals and highlight key elements such as the platform features, audience in the platform, and the producer's own self-concept.

In conclusion, the described studies crucially pointed out how important role the image of the brand, the persona could play in the market we aim to model. Therefore, in the model development and hypothesis formulation parts of the dissertation, we allow for effects related to this aspect of the channels.

3. Data and Methodology

3.1. Data collection procedure

3.1.1. Identifying product reviewers on YouTube

The overall goals set up by this dissertation can be investigated on many different sets of observations, coming from reviews on different categories of products. The only condition which the chosen product category must fulfill, is the presence of enough product reviewer channels to obtain sufficient number of observations to derive reliable results.

Notwithstanding, there are multiple products that can serve as potentially suitable category for our research. For instance, beauty products, technology, board games, sneakers, headphones, or speakers. Motivated by our prior knowledge about the category we decided to test our hypothesizes on the technology, more specifically the smartphone subcategory of product reviews.

Driven by the goals of this dissertation, our first task was to collect potential YouTube Product reviewers to have a list of YouTube channels that will be the central focus of our empirical analysis. Hence, we used YouTube API channel search option with keywords that fits to the product review genre. We built up the phrases to contain at least two words, one that specifies the category we are looking for, which aimed to narrow the channels around the technology and smartphone genre. Hence, our category phrases were:

- Technology
- Tech
- Smartphone
- Phone

The other part of the search phrase contained the relevant channel type keywords, aimed to filter out the channels that are not oriented around the product review genre. Here, we also used multiple keywords, that we considered as related to product reviewers. These phrases were the following:

- Product Review (counted as one keyword)
- Unboxing

- Review

To have more reliable search results, we not only searched with pairing one product category and one channel type keyword, but we also used every combination of at least one phrase from each of the two categories of keywords, but with a maximum 3 word limit. We also sent the channel search request to the YouTube API with the channel language option restricted to English only. There was no other restrictions or options of the requests. These searches resulted 1642 channels as potential subjects for our research. However, the distribution of the subscriber count of these channels is highly skewed, as we observe exponentially more channels as the channel size decreases.

Hence, we use a cutoff value on the subscriber counts of the channels to decide which channels will be included in the dataset. In Table 2 we divided the channels into five groups according to their subscriber counts to have decide . Based on this table, we decided that the threshold value for channels to be represented in the dataset will be 10 000 subscribers.

Table 2: Number of channel search results per subscriber count groups

Subscriber Count	Number of Channels
0 - 999	985
1 000 – 9 999	334
10 000 – 99 999	189
100 000 – 999 999	101
1 000 000 -	33

Source: own elaboration based on data from YouTube API

However, after double checking the channels by taking a random sample of channels and screen the validity of the search result manually, we noticed the following.

1. Our results are indeed product review channels; we did not observe any type II error in the sample.
2. Some of the channels are incorrectly labelled as English language channels.

The reason behind the second observation could be that

a. the channels are incorrectly state that they are making English content or

b. Google’s API regarding the option to filter according to the language of the channel did not work correctly.

Therefore, we manually screened all the channels from the previous list. In this way we could filter out the channels creating non-English content to finally end up with 78 channels overall. Important to note, that essentially our goal is to bind the consumers uncertainty due to new product launches with reviews on YouTube. Presumably not every channel on this list makes content about the new products on the market. Thus, we expect more channels to drop out from the final list in the model.

3.1.2. Observing the reviewer market

The main objective of the dissertation is to examine the demand and supply of the product related information on the product reviewer market. One of the main features of this approach is the way these metrics evolve over time. Therefore, in contrast to the cross-sectional data, we collected our data on the daily basis.

However, to understand our data gathering process, we also need to understand the structure of our chosen platform, YouTube. First, start with the definition we already outlined in the previous section, the set of information suppliers on the market, which translates to the set of product reviewer channels in this platform. Second, we define the information *“products”* on the market, which contains the product related information on a smartphone. This information products are essentially the videos the above-described channels are posting regularly on the platform. Finally, the demand for information which comes from the audience can be identified by the information of how many members was interested about the given information *products.* Therefore, we can measure this by the number of views a given video received. Note, both the views gathered from the first day of the video on the market up until the observation and the number of views it received compare to the last observation could contain information for us.

Since we have the list of channels, the next step is to gather the information *products* they posted on the platform, which could be done by collecting all the video IDs the given channel posted from a given date. We have chosen to start collecting the video IDs from 01 May 2020, which meant a 47-day time window between the date when the first videos in the dataset were posted, and the day when the daily observation began.

Our motivation behind the chosen date relies on the goal that we aim to model videos about new products, which makes the collection of data about older videos irrelevant. However, more details about this process can be found in the next chapter, describing the product list collecting process.

Then, we have both the channel and video IDs, we can collect the observations regarding these sets. Regarding the information *products,* we observe the views the video received up until that point in time. This is our most important variable thorough the dissertation, since it shows us the demand for product related information. Besides this information, we also have the possibility to collect the aggregate number of reaction measure, such as the number of likes, dislikes, and comment, that a given video received up until that point. In addition, for the purpose of identifying the content of the video, which will be important in the next section of the chapter, we also downloaded the title and description corresponding to the videos.

Regarding the channels, we collect the information about their follower base at a given period, measure by the number of subscribers the focal channel has at this period.

As we mentioned before, in contrast to the one-time collection set of channel IDs itself (Chapter 3.1), we acquire data regarding the list of videos and data about each of the channel IDs and video IDs on the daily basis. Hence, every day, we checked whether new video(s) was/were posted on the market compare to previous observation day. If there was/were, we added it/them to the list of videos, then repeated the downloading process for every channels ID and for the updated list of video IDs. The download process took place from 16 June 2020 to 01 October 2020 and was held at the property of Rotterdam School of Management, Erasmus University.

Table 3: Descriptive statistics for the total video dataset

Total Video Dataset							
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
views	294,890	261,964	674,237	0	8,096	149,816	10,774,304
likes	294,890	10,126	29,475	0	283	4,975	643,652
dislikes	294,890	358	1,403	0	12	209	40,846
comments	294,890	1,150	3,352	0	58	816	71,007

Source: own elaboration based on data from YouTube API

Unfortunately, an issue with the collection of the data was occurred during the time window due to technical difficulties regarding the automatized handling of the continuously growing list of video IDs, resulting a gaps in the dataset from 7 August 2020 to 9 August 2020, when we could not observe the market.

In addition, one channel has been removed from the dataset, because his/her channel was no longer accessible on the platform due to unknown reasons.

Notwithstanding, 294 890 number of observations was collected for the video and 8320 number for the channel dataset over the course of the 108-day period. The full channel and video dataset will be used for modelling the channel's subscription number. Descriptive statistics of these dataset can be found in Table 3 and Table 4, respectively.

Table 4: Descriptive statistics for the total channel dataset

Total Channel Dataset							
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
subscription_number	8,320	1,424,399	2,816,021	16,300	184,000	1,230,000	17,200,000
number_of_videos	8,320	1,289	1,334	0	392	1,624	7,283
channel_views	8,320	329,808,624	679,984,586	3,733,352	30,170,818.0	258,302,594	4,073,463,818

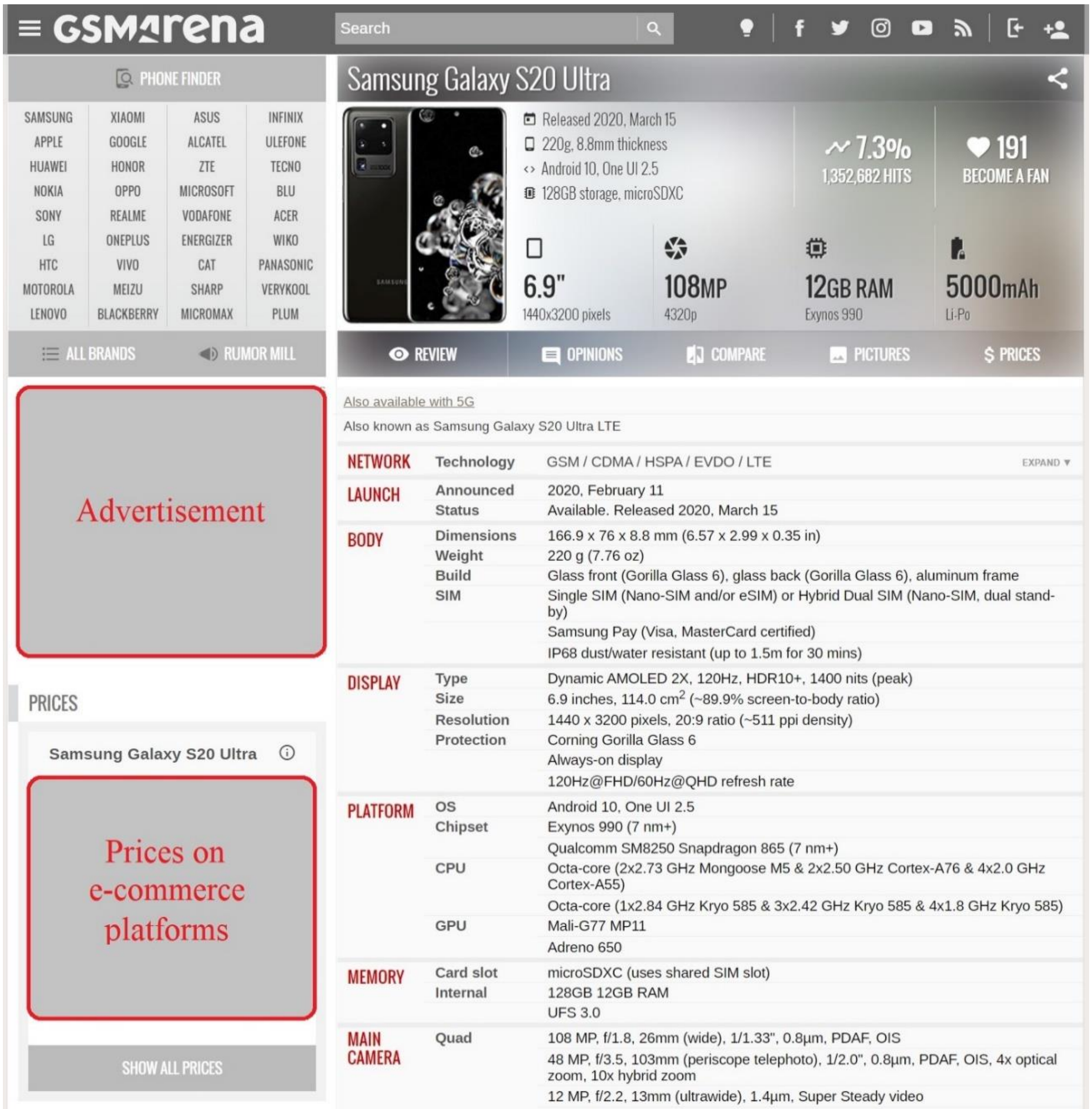
Source: own elaboration based on data from YouTube API

3.1.3. Collecting the list of new products

In the previous sections, we acquired two panel datasets, containing the channel and the video related metrics. In contrast to the channel dataset, the video dataset will be restricted. The reason behind this procedure comes from the goal of the dissertation to examine the demand and supply of information relating to new products on the market. Hence, in this chapter we aim to collect the list of new products in the smartphone industry in our specified data collection time window. Then, we can use this list of products with our dataset of videos and for every video decide, whether its content can be matched with a product on the list or not.

The collection of new products examined in this dissertation is obtained by using a popular technology specification webpage, GSMArena.com. The decision to choose this page relied on the wide variety and highly accurate information for a large collection of smartphones. Our main interest among these specifications was – trivially – the date when the phone was launched. GSMArena.com performs especially well in this aspect, as they publish not only the release, but also the date of the announcement of the given smartphones. Unfortunately, the webpage does not have an API. Hence, we built a web scraper to obtain the dates corresponding to the products from the product specification pages, shown in Figure 1. For the scraping procedure, we used rvest package[1] written in R language.

Figure 1: Example page of GSMArena.com, our source of list of new smartphones



Source: https://www.gsmarena.com/samsung_galaxy_s20_ultra_5g-10040.php

Note: The advertisements and the prices were hidden to avoid any unfair representation of the products and e-commerce platforms

[1] https://cran.r-project.org/web/packages/rvest/index.html

Following the successful collection of the release dates for each product, the next task is to match the ones with recent launch date (9 January 2020) to the videos in the dataset. Since the matching relies on the names of the products, a potential issue arises regarding the strings that contain special characters and/or notes that may be not used by the products reviewers. For instance while the official name being Apple iPhone SE (2020), the version that product reviewers are using could be simply Apple iPhone SE, as it is trivial for them that it is the 2020 version and not the one being released in 2019, based on the upload date of the video. We observed similar issue with other version or extra specification declaration words with the phrases: *"5G"*, *"4G"*, *"2019"*, *"T-Mobile"*, *"NFC"*, *"16+32"*, *"48+40"*, *"India"*, *"Verizon"*, *"3 cameras"*, *"China"*, *"Indonesia"*, *"UW"*, and *"Aluminum"*. Hence, we removed these words from the product names.

Our approach to match the videos to certain products is based on the title and description of the videos. Based on these, we used the following algorithm:

- 1. First, check whether the title contains one of the products from the list of all the new products on the market in the given time window.
- 2. If it contains one product, match that product to the video, if it contains more than one product, remove the video from the dataset.
- 3. If the title does not contain any of the product on the list, screen the description of the videos.
    - A. If the description contains one product from the list, match that product to the video, if it contains more than one product, remove the video from the list.
    - B. If the description does not contain any of the product on the list, remove the video from the dataset.

The reason behind the removal following the multiple product matches in case of both the titles and the subscriptions comes from the consideration that we want to match one and only one product for each video. In this way, we filter out for instance comparison videos, but also the videos, in which the channels are advertising other products in the description.

Table 5: Descriptive statistics for the dataset containing videos about new smartphones

Product-matched Video Dataset							
Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
views	44,015	150,980	493,003	21	4,743	94,704	7,768,909
likes	44,015	4,966	16,473	0	174	2,627.5	180,681
dislikes	44,015	172	525	0	8	104	6,780
comments	44,015	617	1,770	0	38	409	23,710

Source: own elaboration based on data from YouTube API

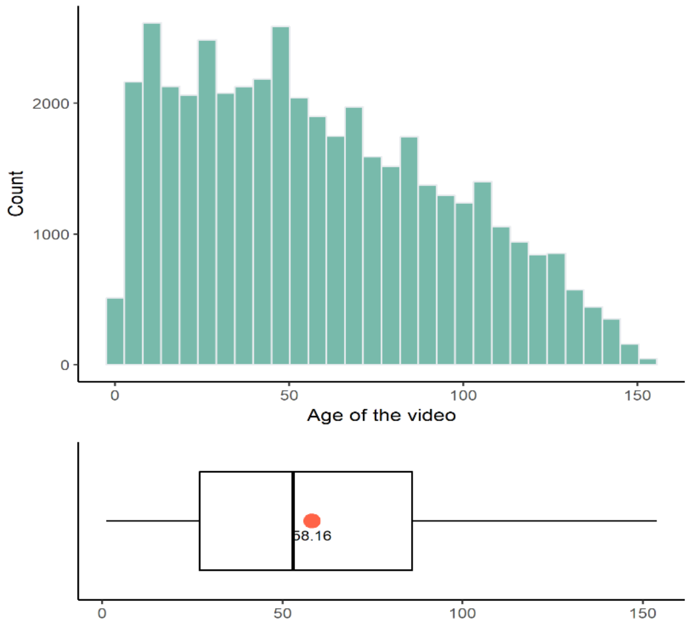
Nevertheless, our final product related video dataset contains 44 015 observation. Descriptive statistics of the dataset can be found in Table 5.

3.2. The construction of the time related variables

One can argue that our dataset is especially peculiar in a sense that multiple dimensions of time can be observed in the dataset that may be correlated with the observed variables. However, before the distinction and definition of these dimensions, we should define the universal measure of time in the data. Even though we know exactly the posting time of the videos at a seconds level of precision, our measures are gathered on a daily basis. Hence, in our data, we define one period as one calendar day, regardless of weekends and weekdays.

The first time dimension we define is the absolute time, which will be calculated from the first day of the downloading process, 16 June 2020 to that of last day, 01 October 2020. Since we followed every video that the focal channels are posted after the predefined starting date of relevant videos, the number of videos we followed is increasing over the absolute time. Therefore, the slices of the datasets along the absolute time will be exponentially larger as we approach the last day of the data gathering process.

Figure 2: Histogram and Box plot for the age of the videos in the dataset



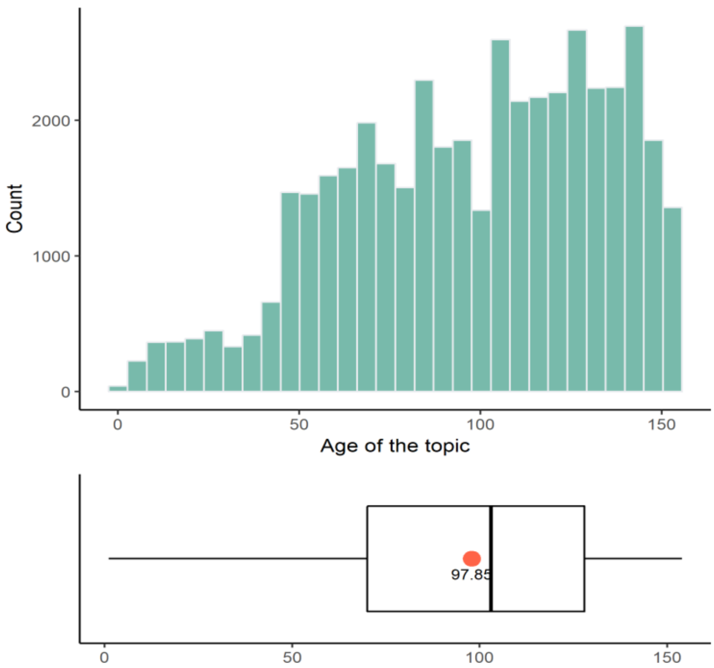
Source: own elaboration based on data from YouTube API

The second time dimension is starting for each video on its corresponding posting day. Trivially, this dimension will be only the same for videos that were posted on the same day. The goal of the representation of this dimension is to examine the evolution of the videos along their lifetimes. There are multiple benefits of calculating this variable for every video. For instance, we can mention the important controlling role at later stages of the dissertation, but we can also introduce a video specific unique shock to the viewership for its first day on the market, regardless of the absolute time.

The third dimension is motivated by the goals to identify product information markets on YouTube with the topics of the videos on the platform. As one information market could contain multiple videos, the lifetime of a topic will be different from the previous time dimensions. Hence, all information market could have its own unique lifetime, which creates our final time dimension. There are multiple possible ways of determining the appropriate starting date for each product. One could be the announcement date of the product, motivated by the idea, that consumers may start to seek for information on that day. It could be also the release date of the product, in which case one can argue that is the first time point when actual review videos can be done by product reviewers. However, important to note, that some firms are using these influencers as a strategy tool and send them the products before the launch date. Our method to determine the starting day is simply approaching this variable as the age of the topic on YouTube only, and define the first day of the topic as the posting date of the first video that was posted on this topic. This approach relies on the fact, that even if we define earlier date than the first product review video, that would only result empty slices from the dataset for the first periods until the first video appears, while defining the starting point later would leave out videos from the information market.

With the purpose of illustrating the difference between the second and third time dimension, we visualized the distribution of the ages of the videos and topics in the dataset in Figure 2 and Figure 3, respectively.

Figure 3: Histogram and Box plot for the age of the topics in the dataset



Source: own elaboration based on data from YouTube API

3.3. *Methodology*

3.3.1. Motivation

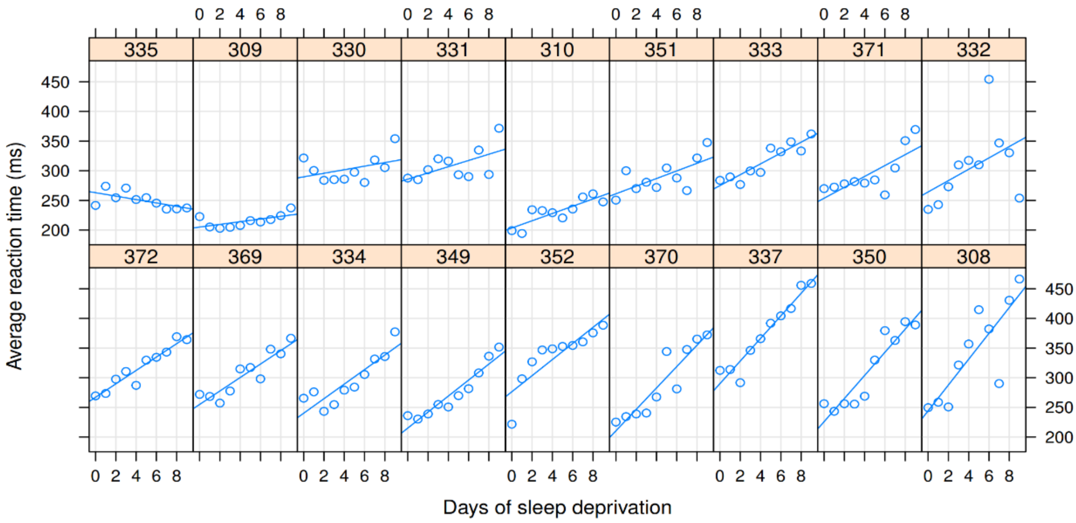
Based on the background theories, highlighted in Chapter 2, we can assume that there may be underlying hierarchical or nested structure(s) in the dataset. For instance, one nesting factor could be the topic of the videos, but we can also mention the channels as another grouping factor.

From the perspective of building and estimating models with regressions, the nesting structure in general is caused by unobserved characteristics which sorts our examined variables into separate groups with significantly different estimated regression equation(s).

Classic examples of such hierarchical structures could be the frog-pond theory (Hox, 1995) where unobserved environmental characteristics of the ponds provides significantly different sizes of frogs. Another example could be Belenky et al.'s (2003) sleep deprivation study where unobserved biological characteristics of the individuals creates different regression equation for each group (Figure 4). Notice, significantly different regression equations could be present for each group due to different intercepts, different slope parameter(s) for the independent variable(s), or both. In this dissertation we identify two potential nesting structures. First, the videos could be nested in a product related information market. In this case, the characteristics of the given topic could contain the products' and the brands' exogenous popularity or historical perception. Second, the videos could be nested by their corresponding content creators. The unobserved factors here could be the channels' presentation or title giving style, but we can list all the factors that is part of the channels' persona (Chapter 2.2.) and we do not measure it.

However, we also take advantage of this methodology in different grouping variables, such as the time horizon. With the methodology briefly outlined in the following sections, we can assume and test multiple different, even very complicated hierarchical systems in the data, each corresponding to different hypothesizes regarding the underlying structures in the product information market on YouTube. Moreover, the benefits of this method cannot be grabbed fully by being able to define complicated hierarchy into the regression. It also provides us the tool to control for unobserved factors that may result in spurious regressions if it would remain unhandled.

Figure 4: Illustration of different intercepts and slopes estimated for different groups



Source: Bates et al. (2004)

3.3.2. Random Effects

We identify two main type of random effects in this dissertation, depending on the assumption about the relation between the grouping variable and the observed independent variables. In this chapter we describe these two main types on a simple example with channel characteristics grouping variable and topic size (x_i) independent variable on a simple hierarchical regression system. Then, in the following chapters, we introduce how can we derive the main objective function from the literature.

First, we can assume that we can build up our regression from two levels, the level of the grouping variable (channels) and the level of the response variable in the following way:

$$y_i = \beta_{0j} + \beta_1 x_i + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_{00} + \varepsilon_j \ ,$$

with:

$$\varepsilon_{ij} \sim N(0; \delta_{ij}^2)$$

$$\varepsilon_j \sim N\left(0; \delta_j^2\right) \ ,$$

where β_{0j} is the channel random effect δ_j is the channel j's expected squared deviation from the grand mean across all the channels (β_{00}) and δ_{ij} is the expected squared deviation of the response variable in case of channel j, given $\beta_{0j} + \beta_1 x_i$.

This regression equation shows that every channel has different random distribution, which shows the probabilities of the increase or decrease level on the value of the response variable. However, this extra effect is not related to the independent variable in the regression. In our example, it means, that while the channel characteristic is a significant predictor of the response variable by modifying the grand mean with a certain amount with some probability, it does not affect the relationship between the dependent variable and the topic size (x_i). Therefore, in the next chapters, we call this term as the random intercept.

In contrast, we can also assume, that the channel characteristics modifies how the topic size affects the dependent variable. Hence, our regression hierarchy changes to the following:

$$y_i = \beta_{0j} + \beta_{1j} x_i + \varepsilon_{ij}$$

$$\beta_{0j} = \beta_{00} + \varepsilon_{0j}$$

$$\beta_{1j} = \beta_{10} + \varepsilon_{1j} \ ,$$

with:

$$\varepsilon_{ij} \sim N(0; \delta_{ij}^2)$$

$$\varepsilon_{0j} \sim N(0; \delta_{0j}^2)$$

$$\varepsilon_{1j} \sim N\left(0; \delta_{1j}^2\right) \ ,$$

where β_{1j} is the channel random effect regarding the estimated effect between the independent variable and the response variable, δ_{1j}^2 is the expected squared deviation of the effect's grand mean (β_{10}), and the effect in case of channel j.

In this specification, we assume that the channel characteristics both affects the intercept and the effect of topic size on the response variable. Note, we do not assume a model with independent variable effect but without random intercept thorough the dissertation. The reason behind this simply comes from theoretical consideration similarly to the most basic setup of the linear regression: $y = \beta_0 + \beta_1 x_i$. It is possible that the most probable value for β_0 will be zero in our model, but assuming it before the estimation could make the model intrinsically biased.

Trivially, in our specifications we use more than one independent variable and more than one random effect as well. Hence, the regression equation system will be more complex. Despite its complexity, the foundation will be similar to the models in chapters 3.5.2-3.5.4. The details about our implementation of this model can be found in chapter 3.5.5.

3.3.3. Estimation

3.3.3.1. Drawing from Densities

In this chapter we derive the objective function that can be estimated using nonlinear optimization. First, a trivial solution could be the usage of simulation techniques to estimate the parameters of the unknown random effect distributions. We could do so by first, drawing random numbers from the probability distribution(s) with some set parameters. Then, we can calculate the mean log-likelihood across the draws. Finally, we can then iterate the set model parameters to achieve the best model fit by maximizing this calculated mean log-likelihood, using nonlinear optimizer(s) (Chapter 3.5.4.) (Train, 2009). The problem with this approach arises from the exponential properties of the computational resource requirements for the optimization process. The increase of the number of random effects and the corresponding possible levels for each grouping variable would make the task so computationally heavy that we would require to simplify the model. Therefore, we decided to choose an alternative approach to estimate our model that needs less resource in exchange for a few, but acceptable assumption about the model specification.

3.3.3.2. Variance Components

In the description of the formula, derived in this chapter, we are following Bates et al.'s (2004) study on comparing the formula to that of linear regression. Hence, our starting point is the equation of the standard linear regression:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \varepsilon\ ,$$

where $\boldsymbol{y}$ and $\boldsymbol{X}$ are the vector of dependent and independent variable, respectively, each having n elements. $\boldsymbol{\beta}$ is the coefficient vector for the independent variables with p elements. Hence, the response variable follows a normal distribution:

$$\boldsymbol{y}\ \sim\ \mathcal{N}(\boldsymbol{X\beta};\ \sigma^2\boldsymbol{I})$$

In this setting, we can introduce q number of random effects, modifying regression equation to:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{Z\mathcal{B}} + \varepsilon\ ,$$

where $\boldsymbol{\mathcal{B}}$ is the vector of random coefficient terms, and its elements can be both random intercept and random slope(s). Since these are random variables, to express the distribution of the response variable, we need its conditional, fixing the value of the random terms:

$$(\boldsymbol{y}|\ \boldsymbol{\mathcal{B}} = \boldsymbol{b}) \sim\ \mathcal{N}(\boldsymbol{X\beta}\ +\ \boldsymbol{Zb};\ \sigma^2\boldsymbol{I})\ ,$$

where we assume that the $\boldsymbol{\mathcal{B}}$ random effect vector follows a multivariate normal distribution with the following specification:

$$\boldsymbol{\mathcal{B}} \sim\ \mathcal{N}(0;\ \sigma^2\boldsymbol{\Sigma})\ .$$

Here, $\sigma^2\boldsymbol{\Sigma}$ is the variance-covariance matrix with σ^2 being the scaling factor. We can also observe that the expected values of the random effects are zero. However, when we calculate the overall effect between the independent variable with random slope or the random intercept and the response variable, we should add up the corresponding fixed effects $\boldsymbol{X\beta}$ and random effects $\boldsymbol{Zb}$. The detailed derivation of this calculation can be found in Chapter 4 for both random intercept and random slope.

Then, we can express the variance of $\boldsymbol{\mathcal{B}}$ distribution to be dependent on the introduced scaling factor (σ^2) and a vector of variance-component parameters ($\boldsymbol{\theta}$) by using the Cholesky decomposition (cites):

$$\boldsymbol{\Sigma_\theta} = \boldsymbol{\Lambda_\theta \Lambda_\theta^T} .$$

Therefore, we can derive the regression's log-likelihood function to be dependent only on the usual parameters in case of the non-hierarchical linear regression ($\boldsymbol{\beta}, \sigma^2$), plus the variance-component parameters ($\boldsymbol{\theta}$). The detailed derivation of the likelihood function (formula 1) can be found in Bates et al. (2004):

$$L(\beta, \sigma^2, \theta | y) = \int \frac{\sqrt{|\Sigma|}}{(2\pi\sigma^2)^{\frac{n+q}{2}}} exp\left(\frac{||y - X\beta - Z\mathcal{B}||^2 + \mathcal{B}^T \Sigma \mathcal{B}}{-2\sigma^2}\right) d\mathcal{B} \qquad (1)$$

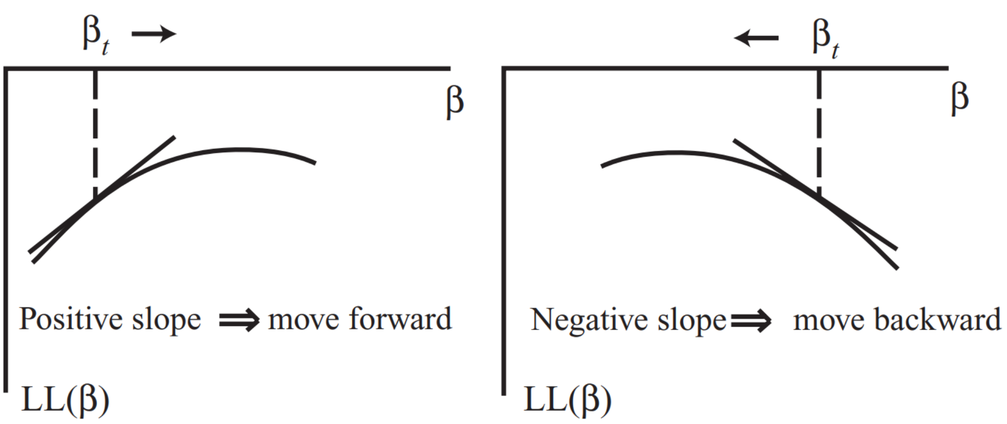
The main benefit of this approach that due to the shorter length of the variance-component parameters, the model optimizes much less parameters compare to the those of in Chapter 3.3.2. More specifically, the length of $\boldsymbol{\theta}$ vector is equals $\binom{p+1}{2}$.

3.3.4. Numerical Maximization

As Train (2009) describes the role of numerical maximization procedures in the research conducted nowadays by comparing it to the preceding times: "*In the past, researchers adapted their specifications to the few convenient models that were available. These models were included in commercially available estimation packages, so that the researcher could estimate the models without knowing the details of how the estimation was actually performed from a numerical perspective." - Train (2009).* However, with the emergence and widespread of the usage of simulation and numerical maximization procedures, researchers often specify models that can be tailor-made to the specific situations and issues. However, in this case, they need to write their own program code for the model (Train, 2009). One driver behind this phenomenon is caused by the boundary that if researchers define more and more complicated models, the derivation of the optimal parameters from the maximum (log-)likelihood function values becomes increasingly harder. Therefore, in some cases, the researchers will face an inability of the derivation of these parameters. The solution to this issue is the usage of numerical maximization procedures, that is often capable of finding the parameters corresponding to the optimal function values, when manual derivation would fail.

In our case, we are facing similar obstacles as we need to find the optimal parameter (vectors) $(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$ in case of formula 1, described in the previous chapter. Fortunately, nowadays there is a wide spectrum of available algorithms that we use in our estimations.

Figure 5: Parameter iterations in a numerical maximization method; deciding the direction of the change



Source: Train (2009)

Generally, these procedures mean, that we use an algorithm that finds the maximum objective function value with iterating the parameter values based on the following information (Train, 2009). Let

$$LL(\rho) = \ln(L(\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta}|\boldsymbol{y}))$$

denote the log-likelihood function, where ρ is a vector, containing all the parameters of the likelihood function. Then, the gradient vector of the function, the first derivatives shows the direction in which the algorithm should change the parameter values from the current iteration (i) to the next one $(i+1)$ (Figure 5).

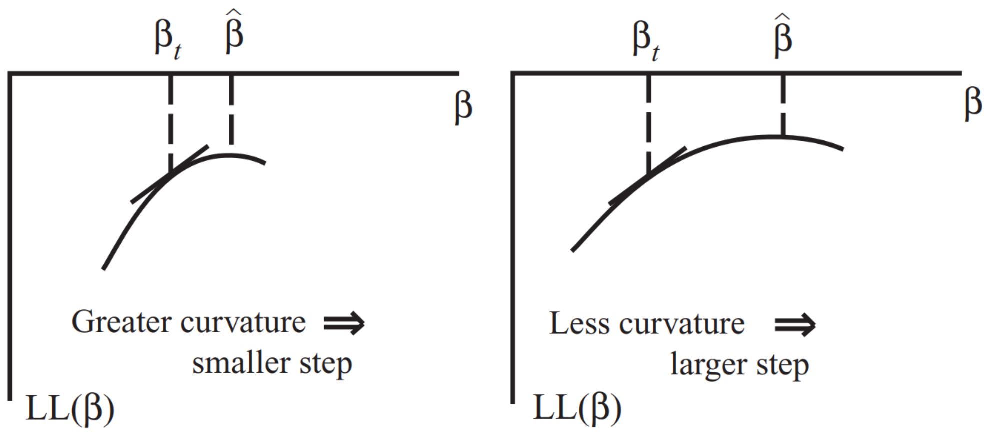
$$g_i = \left(\frac{\partial LL(\rho)}{\partial \rho}\right)_{\rho_i}$$

While the second derivative matrix, the Hessian of the function shows the step size in which the parameters should be changed (Figure 6).

$$H_i = \left(\frac{\partial g_i}{\rho'}\right)_{\rho_i} = \left(\frac{\partial^2 LL(\rho)}{\partial\rho\,\partial\rho'}\right)_{\rho_i}$$

Graphically it means, that optimal parameter values can be achieved by "walking up" on the objective function as long as an increase can be observed in the objective function value. (Train, 2009). Issues can arise with this solution if there are multiple local maximums of the functions, but these problems can be overcome with multiple run of the algorithm from different starting points.

Figure 6: Parameter iterations in a numerical maximization method; deciding the step size



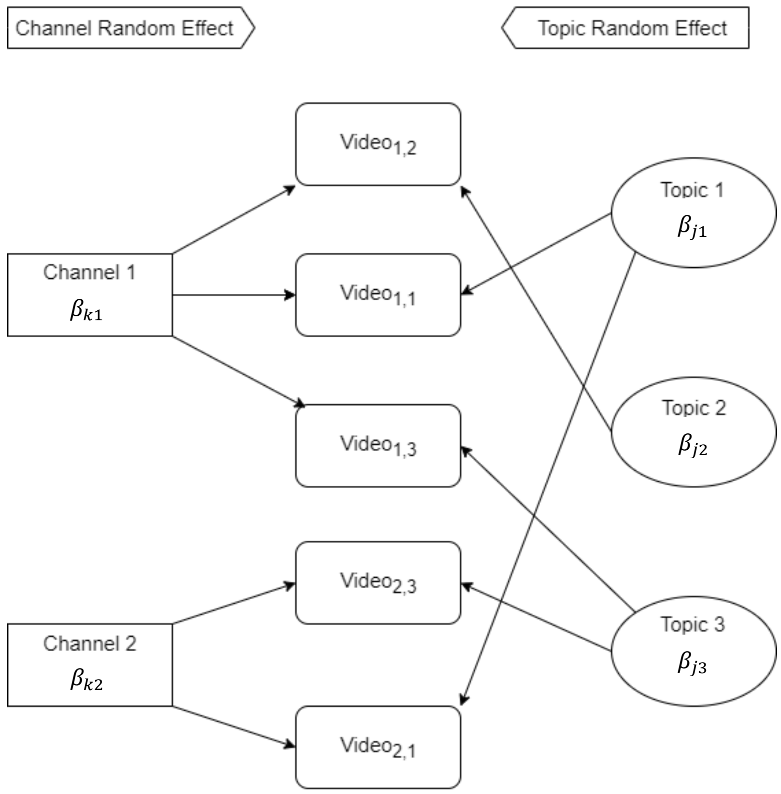
Source: Train (2009)

Finally, the differences among the multiple available algorithms can be described by the function form differences that determines the new iteration of parameter values from the previous objective function values. In the dissertation, we use multiple approach, including the "*nlminb*" (Fox et al. (1978); Fox (1997)), the Broyden–Fletcher–Goldfarb–Shanno ("*BFGS"*) (Shanno, 1970; Fletcher, 2013) and the "Nelder–Mead" (Nelder and Mead, 1965) algorithm. Based on the results presented in the upcoming chapters, regarding the performance of these algorithms, we can conclude, that in cases when every algorithm found the optimum, the results of the estimation was not significantly different from each other. However, there were cases when some of the algorithms did not found the global optimum. Overall, from this point-of-view, the "*nlminb*" algorithm proved to be the best performing one, as it found the optimum in every model specification.

3.3.5. Implementation

Given the complexity of our data, we define multiple grouping variables, motivated by different literature streams. Then, we test which identification is valid during the model development for our hypothesizes. The two main nests in our hierarchical system are the grouping the videos by the channels' and the products' level. The first is based on the characteristics of the persona of the reviewer channels, described in Chapter 2.2, while the second relies on the assumption that products creates their own information markets of reviewer videos on YouTube, outlined in Chapter 2.1. These two broad categories create a cross-classification of the observations. (Figure 7)

Figure 7: Illustration of the nested structure in the data



Source: own elaboration

In addition, as it was mentioned above, random effect estimation also provides a great tool to control for effects that are unobserved for the researcher. Hence, we also test the estimation of random intercept for the age of the topic and the age of video. The main goal here arises from the consideration that “only” using the time dimension as a dependent variable is too strict, and there are other aspects that should be controlled. Detailed description of this methodology can be found in Chapter 4.1 and 4.2. In conclusion, as Figure 7 and the above mentioned time controls highlight, we assume that behind the decisions of the channels in regard to the product information market they post a product review and the time when they post it, we can find complicated hierarchical system where the characteristics of channels and topics have a crucial role. One of the goals of this dissertation is to explore this system more thoroughly.

3.3.6. Hierarchical structure significance

Finally, since our model development and ultimately a considerable number of our hypothesizes rely on whether the grouping of the variable is significant or not, we need a test that is capable of assigning a p value for the presence of the random effects. Moreover, the hierarchical structure also changes the calculation of the p values corresponding to the fixed effects as well. This test of significance can be done by applying likelihood ratio test to that coefficient. The likelihood ratio test calculates the log-likelihood value of the nested model, and the that of model without the fixed/random effects. Then calculates the test statistic based on the difference between the two log-likelihood values. From this test statistic, we can decide whether the fixed/random effect is significant, or not. The likelihood ratio tests of our estimation were performed by using the “*lmerTest*” library in the R programming environment. Detailed description of the test (both fixed and random effects), and the program codes can be found in Kuznetsova et al.’s (2017) study.

4. Model Development

In this chapter the underlying baseline model will be developed. This will serve as an initial framework for the models in the following chapters, aimed to answer the formulated hypothesizes and research questions.

In Chapter 3, we already discussed the observed structure of our data and the main methodology aimed to address our questions and hypothesizes thorough the dissertation. Hence, in this chapter we focus on the economic phenomena behind this structure and the motivation regarding the baseline model. As we outlined in the introduction, the dissertation aims to examine the demand and supply of information on YouTube. In this platform, these aggregate measures build up from the individual demand and supply of the information *product*, which is the product review videos. Therefore, in case of YouTube, we can examine how much demand generated for that given information *product* in the past by observing how much views that given video has in the moment. In consequence, we first aim to model the view counts, more specifically, the changes in the view counts of the videos. Then, in the final chapter, we extend this framework of the product related information economy to include the suppliers long time incentives and growth dynamics (Chapter 7).

The models derived in the dissertation are designed to always answer only the focal question regrading to one specific relationship. Based on this approach, our goal is to get as reliable answers as we can get for these questions, and not to maximize R squared by adding as many significant independent variables as we can. This goal motivated the usage of the hierarchical random effect estimation, but it also requires the precise definition of the controls, that would lead to spurious relationship among our main variables in case we would not represent them.

4.1. Controls in the model

4.1.1. Controlling for the channel characteristics

The final model of the view counts of the videos consist of two main categories of independent variables, the topic, and the channel related effects. As half of the hypothesizes in the category of channel characteristics are related to the already present topic effects, the dissertation prioritizes the discussion of topic effects first (Chapter 4.2 and 5) before the channel related effects (Chapter 6). However, despite this distinction, it is important to control for the latter category in the first part of the dissertation as well. The reason behind the importance of representing these controls rely on the assumption that these effects are one of the main drivers of the view counts of the videos and they could be related to the topic effects and the channels' decisions of: *"Which product do they choose to review?"*, *"When do they review that product?"* as well.

The included channel characteristic controls are the subscriber count of the channel, denoted as the channel size, and the channel random intercept. While the detailed discussion of the motivation behind the role of channel size can be found in Chapter 6.1 and 7.1, it is worth to briefly note that it relies on the phenomenon of *"big-gets-more"* and *"big-gets-bigger"* type of multiplicative growth. In contrast, the representation of the channel random effect as a control relies on the channel characteristics or parts of the personas of the channels (Chapter 2.2.) that cannot be fully grabbed by the size of the channel. An extra information about the channel besides its size, that may be correlated with the view count, and would lead to spurious relationship if we do not represent it in the regression. Based on the assumption that this extra information does not change on the short term, we can adequately control for it with a time constant unique channel intercept. Such unique information could be the tendencies to use attractive thumbnails and titles for the videos the channel makes or the right usage of search tags (Li 2016; Trzcinski 2016; Diwanji 2014) but it can be any other aspect of the persona such as the entertainment style, or the objectivity of the content.

4.1.2. The lifetime of the videos

Besides the change of the view count of the videos, Chapter 7 aims to model suppliers long time incentives and growth dynamics using the subscriber count changes of the channels as dependent variable. A key difference between these two variables is that the videos - on average - have relatively short lifetime, in which they are gathering most of their views (Crane and Sornette 2008; Cha et al., 2009; Yang et al. 2011; Figueiredo et al., 2011, 2014; Figueiredo 2103; Ahmed et al. 2013; Li, 2016), while the subscription number of the channels do not have such a lifetime. Usually the goal of the channels is to keep gaining subscribers over time. Their videos on the other hand - on average - tend to fall in terms of their new view count as the time increases. Therefore, it is important to control for the age of the videos when we model the changes in the view counts.

In conclusion, while it is true that the audience can pick up old videos, making them actively gain views again, the product review videos - on average - gain most of their views after they were posted, and then they usually slow down. (Crane and Sornette 2008; Cha et al., 2009; Yang et al. 2011; Figueiredo et al. 2011a, 2011b, 2014; Ahmed et al. 2013; Li et al., 2016) Hence, we can address this issue in two steps. First, we can introduce the logarithmic transformation of the number of days passed since the video was posted as a dependent variable to the regression. In this way, - since the changes in the view count is also on a logarithmic scale - we get the following function form:

$$\Delta Views_{i,t} = e^{\beta_0} * videoage_{i,t}{}^{\beta_1} . \qquad (2)$$

We expect a negative coefficient ($\beta_1 < 0$) for the age of the video, which transforms formula 2 into multiplicative inverse function:

$$\Delta Views_{i,t} = e^{\beta_0} \frac{1}{age\ of\ the\ video_{i,t}^{|\beta_1|}} + \varepsilon_{i,t} \ , \qquad (3)$$

which is supported by Cheng et al. (2007), Szabo and Huberman (2008) and Cha et al. (2009).

However, there is a chance that the multiplicative inverse relation defined by formula 3 may not result the best fit for our time control. A reason behind this possibility could be that we can apply different function forms or specifications at different time horizons (Li et al., 2016, Cha et al., 2009). Li et al. (2016) shown that the evolution patterns of videos could follow changing dynamics, such as the *burst-slow-burst-slow* or the *slow-burst-slow-burst-slow* process.

We can test this possibility and achieve a model that can contain non-continuous effect for the age of the video variable if we handle the age (by the number of days) as a categorical variable. Hence, we can estimate an adjustment for the effect for age of the topic compare to formula 2 for each day. This can be achieved by estimating a random intercept using the age as the factor for defining a hierarchical model, discussed in Chapter 3.3.2. Then, the posterior modes can be can retrieved for each day from the video is posted. Combining formula 2 and the adjustment by the posterior modes, we can calculate the overall effect for the age of the video. By estimating a hierarchical model instead of a linear regression with assuming random intercept for the ages of the videos, the formula defined above transforms into:

$$\Delta Views_{i,t} = e^{\beta_{0,j}} * videoage_{i,t}{}^{\beta_1} .$$

The duration since the video was posted is not the only time dimension in our model, hence we will also apply the same methodology when we are testing the effect of the age of the topic on the view count of the video.

4.2. *Information Market Identification*

4.2.1. Motivation

After discussing the represented controls in the model, this chapter lays down the foundations of our motivation, definition, and implementation of modeling product related information markets on YouTube.

In this market, we defined the supply as a set of third-party product reviewer channels, building on the literature on theoretical models of news providers, attention seekers and online personal branding (Chapter 2.1.2-2.2). The demand for information on the market comes from the audience that is interested in the topics of these videos, based on the literature on consumer learning, since these topics are essentially the products the channels are reviewing (Chapter 2.1.1).

Hence, we define the product review information market corresponding to a specific product on YouTube with the collection of the videos whose content is centered around the focal product, the YouTube channels that created these videos and the audience that watched these videos. Since every information market consists of videos that review different products from each other, both the size and the structure of the demand and supply vary across the markets. Thus, these differences may translate into different performances for the videos on the market, meaning that the choice of the topic may be reflected on the view count of the videos. Based on this premise, in this chapter, we argue that we can observe significant differences in the view counts of the videos by categorizing them into their corresponding product information market, because the topics of the videos had different effects on their performances. We denote this phenomenon as topic interest effect, since it shows how the topic's overall activity or engagement, coming from both the audience and the channels, is affecting the videos. In other words, we are exploring whether our differentiations of the information markets are viable, so whether the topic of a product review video on YouTube actually matters in terms of the view count it will gather in the future. Important to note, that in this chapter we are considering this topic interest as an exogenous factor for the videos that are posted on the topic, however, we extend this approach in the upcoming chapters to enable endogenous determination as well.

The defined effect of the topic is essentially dependent on the audience's and the reviewer channels' interest towards a given topic. Considering the literature of consumer learning (e.g. Erdem and Keane, 1996; Szymanowski and Gijsbrechts, 2012, 2013; Zhao et al., 2013; Wu et al., 2015) and diffusion of new products (e.g. Kalish 1985; Roberts and Urban 1988; Oren and Schwartz 1988; Mahajan et al., 1990; Peres et al., 2010), more information about this interest is available to us. Based on the findings of these studies, we can identify that the greatest number of consumers that are uncertain about a product is at the point in time when that product is launched, meaning that the demand for information is the highest when the product is launched. After this first period, the uncertainty, and the interest towards the topic decreases over time. Consequently, the topic interest effect in our model may also has a lifetime, such that it is the highest at the first periods of the age of the topic and then decreasing while it becomes irrelevant eventually. Therefore, we argue that not only the topic itself, but the age of the topic also matters for the videos on the market in terms of their view counts changes. Moreover, we expect a negative connection between the change in the view count number over time and the age of the topic. In conclusion we formulate the following hypothesis for the identification of the topic information markets.

H1:

A: The product reviewed in the video has significant effect on the performance of the video.

B: The effect of the product on the video is decreasing over time.

As it was discussed above, this approach for identifying market effects relied on an exogenous topic interest effect for the videos and can be served as a proof that the product related information markets indeed exist. In the following chapters (Chapter 5 and 6) we extend this approach with a more realistic view on how this market might work by enabling the actors on the market to influence each other's performance.

4.2.2. Information markets in the model

To model the information market, defined in the previous chapter, we can implement the method we derived in Chapter 3.5. Note, we already used this approach when we used random effect estimation to control for channel characteristics and the age of the video. In contrast to the controlling variables, the hierarchical model defined by the topic information market will be the base framework for most of the hypothesizes and research questions we formulate. In addition, we also interested in how the age of the topic affects the performance of the videos. Hence, we can use the age of the topic variable, derived in Chapter 3.4. Then, similarly to the age of the topic, we also have the possibility to adjust the time related coefficient by defining it as random in the model. However, there are multiple ways to build this model depending on our perception about the possible topic interest over time function.

In the first method, we use the age of the topic variable as an independent variable, which sets the shape of the topic interest over time function for all the topics, then we can estimate a unique topic interest intercept for the topic information markets to set a unique scale of the function for each topic. Model 3 in Table 7 was estimated following this methodology. In this model, all the topic has a unique number of topic interest at each point in time, but the relative differences of these interest between two point in time is fixed across topics. Then, the log-log specification of the regression leads to the following equation:

$$\Delta Views_{i,t} = e^{\beta_0} age\ of\ the\ topic_{i,t}^{\beta_1} + \varepsilon_{i,t}\ ,$$

which similarly to the formula that models the effect for the age of the video, becomes a multiplicative inverse function if $\beta_1 < 0$.

In the following model, we can raise similar arguments to that of age of the video, that the function form described above may not represent the best fit between the age of the topic and the new view counts between two periods. The reason behind this possibility could be technical, meaning that the assumed function form is correct, but there is a change in the parameters over time. However, it could be also driven by the nature of product diffusion processes (Kalish, 1985; Roberts and Urban, 1988; Oren and Swartz 1988; Mahajan et al., 1990; Peres et al., 2010). In these models, product diffusions often described by epidemic models, where the adoption of the new products follows a process where we can observe a slow increase in the product adoption, followed by a sharp increase, then a fast and finally a slow decrease, until the changes become irrelevant. Our variable meant to represent the interest for a topic and each topic corresponds to a certain new product. Thus, it is reasonable to assume that our regression may need adjustments, since it can only grab two segments of the adoption function.

Finally, we also derive a model, where we estimate unique topic interest over time function for each topic by estimating random slopes for the age of the topic variable for each topic information market besides the already present estimated random intercepts. With this method, we not only get unique scales of topic interest over time for each information market, but we also get unique shapes, so the effect of the relative lifetime of the topics can differ from each other.

4.3. Results

4.3.1. Represented controls

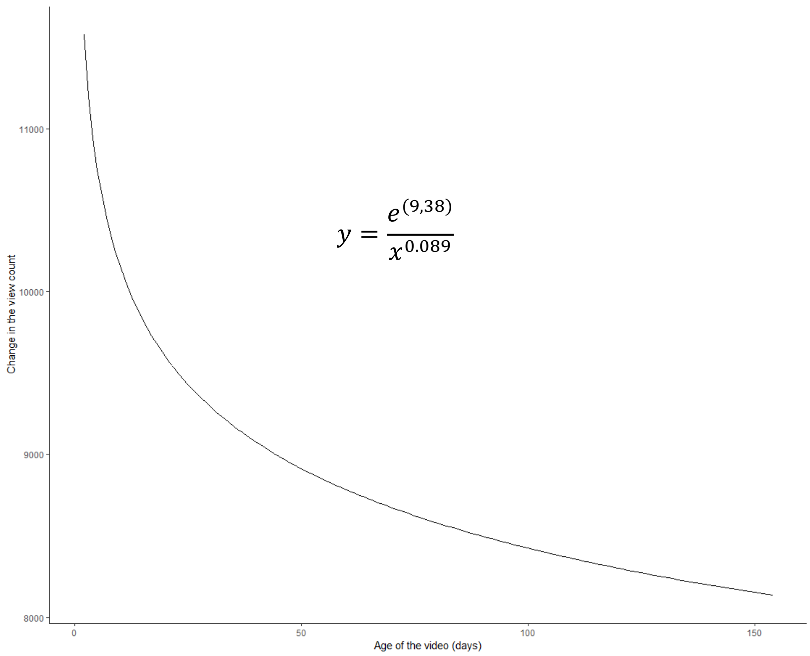
4.3.1.1. Channel Characteristics

We can observe that in consistent with our initial expectation, individual differences across channels play an important role on the view gathering process of the videos as both the size of the channel (defined by its subscription number) and the channel specific random intercept are significant. (Table 6)

4.3.1.2. Age of the video

The other important control, we accounted for in the model is the age of the video, defined by the number of days passed since the video was posted. We had two options to control for this effect, first, using the logarithmic transformation of the number of days passed since the video was posted as a dependent variable. We found that this effect is significant predictor of the view count changes of the videos, and we observe that there is a negative but diminishing connection between the two variables.

Figure 8: The effect of the age of the video without random effect



Source: own elaboration

Second, we can estimate a model with video age specific random intercepts to let the simulation readjust the defined logarithmic connection for a better fitting model to the data and essentially control for the age of the video better. This method gives us a unique opportunity to visualize random coefficients, as our grouping variable is a scale, that can be represented on the x-axis in a standard two-dimensional coordinate system.

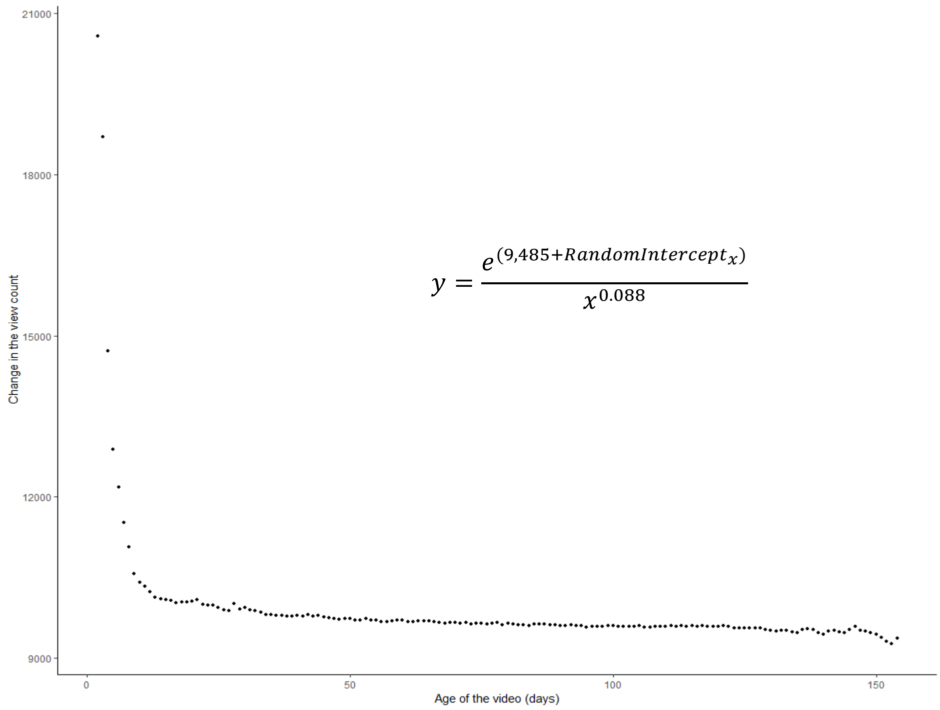
However, while the grouping variable can be represented on the x-axis easily, for the introduction of the estimated random effects to the y-axis, we need to calculate a central value for the estimated distributions. Hence, we simulated the posterior modes to represent the typical value of the random intercept and calculated the value of the independent variable. Using these values, we could calculate the overall effect of the age according to formula 3. Figure 8 shows the estimated connection between the age of the video independent variable and its view count, ceteris paribus, based on Model 1, while Figure 9 shows the readjustment connection in Model 2, applying random effects to the age of the video, which modifies the previous multiplicative inverse function.

Table 6: Estimated posterior modes for the age of the video

Number of Days	Random Intercept	Number of Days	Random Intercept	Number of Days	Random Intercept	Number of Days	Random Intercept
1	0,467165	16	-0,08172	31	-0,04734	46	-0,03137
2	0,403744	17	-0,07647	32	-0,04767	47	-0,03162
3	0,187338	18	-0,07188	33	-0,04891	48	-0,02718
4	0,0721	19	-0,06683	34	-0,04721	49	-0,0269
5	0,029911	20	-0,0595	35	-0,04574	50	-0,02816
6	-0,0133	21	-0,06466	36	-0,04385	51	-0,02546
7	-0,04367	22	-0,06203	37	-0,04409	52	-0,02227
8	-0,07967	23	-0,05867	38	-0,04175	53	-0,02249
9	-0,08681	24	-0,06102	39	-0,038	54	-0,02199
10	-0,08687	25	-0,06189	40	-0,03681	55	-0,02337
11	-0,08959	26	-0,05979	41	-0,03286	56	-0,02134
12	-0,09268	27	-0,04348	42	-0,03387	57	-0,01961
13	-0,09035	28	-0,05107	43	-0,03073	58	-0,01662
14	-0,08659	29	-0,04571	44	-0,03141	59	-0,01484
15	-0,08244	30	-0,04782	45	-0,03079	60	-0,01694

Source: own elaboration

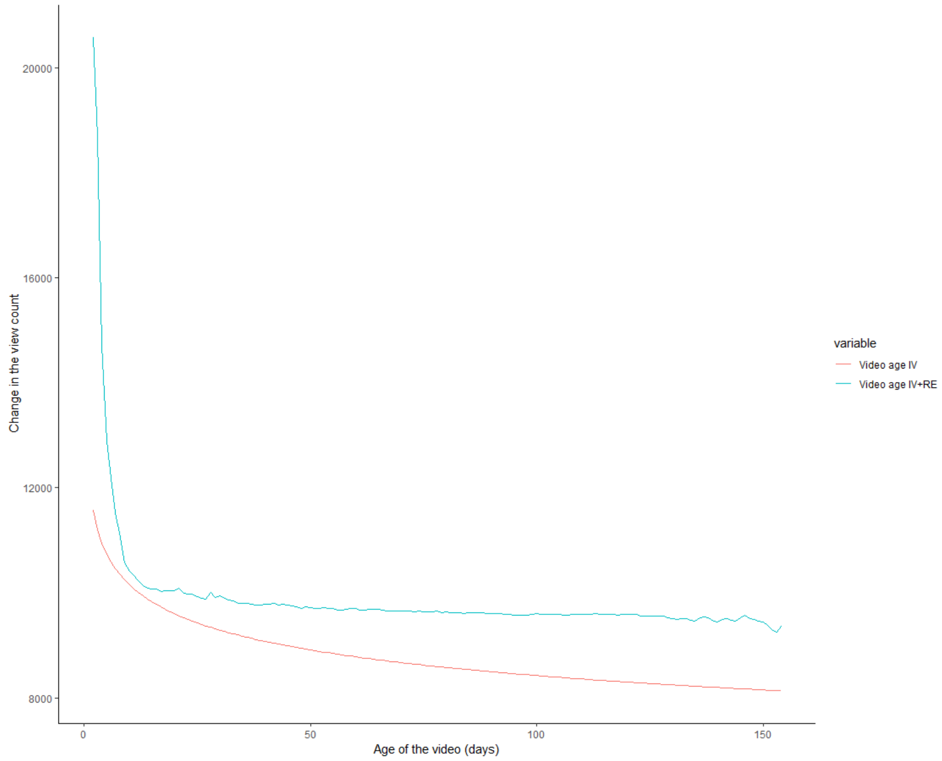
Figure 9: The effect of the age of the video with random effect



Source: own elaboration

If we create a curve from the distinct coefficients estimated for each period by assuming continuity along the time horizon, we can see that Model 2 prefers a connection which has a turning point around 2 weeks (Figure 10). The model shows that after this day the age has much less negative effect on the view count changes than it had before.

Figure 10: Comparison of the effect of the age of the video with and without random effect



Source: own elaboration

4.3.2. Topic Interest

Our results show that the hierarchical model defined to explore the effect the topic has on the videos' view counts (Model 3) performs better than previous models. The random effect estimated to the product of the video is significant. Moreover, the age of the topic variable also has a significant and negative coefficient. Combining these two findings of Model 3, we can conclude that there is a significant unique topic interest for each topic at each point in time, confirming our hypothesis regarding to identifying product information markets on YouTube. The estimated negative coefficient of the age of the topic variable confirms hypothesis (H1), stating that the topic has a significant and diminishing effect over time on the performance of the videos.

Then, we investigated whether we could make significant adjustments on the topic interest over time function across topics by assigning random intercepts for each day of the topic. The results of Model 4 indicate that such adjustments are not supported. In other words, we do not found evidence that a significant deviation from the multiplicative inverse function of $f(x) = 1/x^{0.013}$ would be present across the topics after we control for the age of the video and the scale of the topic effects with a random intercepts for the topic.

These results were obtained by using a model specification with a limitation that even though there is a unique topic interest at each point in time, the relative differences of the interests of two periods are the same for all topics. We aimed to resolve this limitation in Model 5 by estimating random slope for the age of the topic variable grouped by the topics. In this way we assumed a hierarchical structure not only for the intercept, but also for the slope regarding the age of the topic. The results of Model 5 show, that this hierarchical structure performs better than previous models. This indicates that we can achieve better fit for our model if we not only use different scales (Model 3) but we also estimate different shapes (Model 5) for the topic interest over time function.

These findings have multiple implications towards the creators of product reviewers on YouTube. First, it shows that the division of the videos by their corresponding product is significant, meaning that there is an observable product related differences in the view count of the video, so that the decision of which product should the channels review is important in terms of their revenue. The negative coefficient for the age of the topic also highlights that not only the product decision, but the timing of the review also matters. Moreover, the unique topic interest function for each product in Model 3-5 may serve as proofs that the product information market indeed exists and motivate our efforts to move towards a more complex but also more realistic model of the product information market on YouTube.

Model 5 also points out that the effect of the topic and the age of the topic are interrelated, meaning that different topics not only brings more views to the video, they also has different topic interest lifetimes, which can imply that the decision of the YouTubers to *"Which product should they choose to review?"* and *"When should they make the review?"* are cannot be separated from each other, although this implication needs more clarification by more findings.

Table 7: Regression results for market identification

Regression Results (1)

	Dependent variable:				
	ln ΔViews				
	(1)	(2)	(3)	(4)	(5)
ln channel subscriber count	0.023***	0.015***	0.015***	0.015***	0.015***
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
ln age of the video	-0.089***	-0.088***	-0.076***	-0.076***	-0.078***
	(0.001)	(0.006)	(0.006)	(0.006)	(0.006)
ln age of the topic			-0.012***	-0.013***	-0.045***
			(0.003)	(0.003)	(0.009)
Constant	9.380***	9.485***	9.520***	9.522***	9.625***
	(0.038)	(0.046)	(0.051)	(0.051)	(0.062)
Random Effects					
Intercept/Channel					
Standard Deviation	0.2076	0.219	0.24	0.2399	0.2343
Likelihood ratio	9936.786***	11227.172***	10529.683***	10530.647***	10647.202***
Intercept/Age of the video					
Standard Deviation		0.0659	0.0661	0.066	0.0642
Likelihood ratio		4658.871***	4827.786***	4287.911***	4346.003***
Intercept/Topic					
Standard Deviation			0.1605	0.1604	0.262
Likelihood ratio			2612.037***	2567.596***	3560.719***;1
Intercept/Age of the topic					
Standard Deviation				0.0046	
Likelihood ratio				1.601	
Age of the topic/Topic					
Standard Deviation					0.0095
Likelihood ratio					948.682***;2
Observations	41,670	41,670	41,670	41,670	41,670
Log Likelihood	11,334.290	13,663.730	15,096.830	15,097.630	15,571.170
Akaike Inf. Crit.	-22,658.590	-27,315.460	-30,177.660	-30,177.260	-31,122.340
Bayesian Inf. Crit.	-22,615.400	-27,263.630	-30,108.560	-30,099.520	-31,035.960

Note: *p<0.1; **p<0.05; ***p<0.01

[1] Calculated by dropping Age of the topic/Topic term

[2] Calculated by reducing Age of the topic/Topic term to Intercept/Topic

Source: own elaboration

5. Demand for product related information

5.1. Endogenous topic interest

So far, we denoted topic interest as the overall effect the topic has on the videos posted on it, including - among others - the effect of the overall activity, engagement or popularity in the topic's information market on and outside of YouTube. We then estimated a dynamic hierarchical model where every channel has a unique rate of and also uniquely evolving topic interest over time. However, the topic interest effect estimated in the previous chapter is a collective concept not differentiating between endogenous and exogenous effects from the perspective of the actors in our chosen platform of reviews.

Hence, in the following sections we aim to extend our approach of modeling the YouTube product review market with this direction in our minds. First, we explore how the properties of the individuals' demand for information affects the aggregate effect of the topic on the video, relying on the information search and consumer learning.

Second, we examine two potential manifestation in which the embeddedness of the channels in the YouTube reviewer economy affects the performance of other channels on the market. More specifically, on one hand, we base arguments that relies on the aspect that channels are competing on the supply side of the market for the pool of views from the audience. On the other hand, we also argue that in the case of such information platforms like YouTube, the channels (personal brands on the platform) are more connected to each other than brands in the traditional markets. Relying on the attention economy literature, we conclude that this connection through the platform and the topic means, that it is also easier to direct and redirect the audience's interest and attention. This leads to the conclusion that channels may as well have a positive impact on each other, and similarly to influencers, they can also the overall interest for a given topic and essentially for others' videos with their content.

After the theoretical background, we describe the methodology to represent these effects in our model. Here, we are building on the probabilistic properties of the satiation, competition, and topic awareness, representing the effects outlined above, to derive measurable metrics from the aggregate topic views over time, separating this variable into both positive and negative effects.

5.1.1. Satiation effect on the market

Our first and most important building block for the chapter and arguably for the dissertation is the reasoning about properties of the individuals' demand for information as time progresses. This argument is mostly relying on the results of information economics (Nelson, 1970; Stigler, 1961), but due to the same understanding about the agents, similar phenomena can be found in the literature of consuming learning from product reviews as well (e.g. Erdem and Keane, 1996; Szymanowski and Gijsbrechts, 2012, 2013; Zhao et al., 2013; Wu et al., 2015). However, many economists consider Herbert Simon's (1959) superb study as the main starting reference point for this theory. From this paper we can infer that the classical way of identifying humans as homo oeconomicus, in a way that they are or at least they aim to be fully informed is inherently wrong. On the contrary, humans are satisfied with *"only"* satisfactory solutions and they *"only"* have bounded rationality. Then, this theory was formalized by Stigler (1961), showing that we can model this result if we assume some costs to the search of information and diminishing returns to the benefits the information provides. This cost can be monetary in nature, but most importantly for the dissertation, it can be also time and/or (mental) effort. Then, the diminishing returns induces that the same amount of benefit of the information becoming more and more costly until it is not worth for the consumer to search or reach for more information. Therefore, the theory predicts that there will be a point in time in which the consumers have become satiated with information. This argument can be translated to our model. From a theoretical point of view, the satiation point means, that the viewer will not watch more videos on the given topic, she/he will not follow up on upcoming videos on the same topic. From the perspective of the content creators on the platform, who aim to post videos on the focal topic in the future, this phenomenon may implicate a potential *"missing-out"* element of decision making. Meaning, that as the number of satiated consumers grow in the market, channels are missing-out viewers, and therefore revenue, despite the potential that those viewers may would have watched the creators' videos if he/she would posted it earlier. Henceforth we assume a negative relationship between satiation and the performance of the videos in terms of new views from one period to another.

These consequences were derived with the help of the argument about the consumers' decision to stop seeking for more information when it is not worth for them anymore. While this narrative certainly can be a driver of the satiation effect, important to note that most of the research papers in this area use this models assuming that we can model the information search process as if it would come from the conscious decision of the consumers. In reality, the phenomenon is more likely connected to unconsciousness cognitive motives and boundaries.

Similarly, a recent literature stream highlights that the limited attention of the information signal receivers can also result similar phenomenon. (Davenport and Beck, 2001; Falkinger 2007; Smith, 2020) Then, as Smith (2020) argues, the limited attention can be more prominent in case of organized online attention platforms, such as YouTube, where consumers are exposed to an enormous number of stimulus. Here, consumers simply have a cognitive boundary, their limited attention, resulting a situation when they must decide – consciously or unconsciously – which video should they prioritize. However, this argument will be further developed in the following chapter, as it introduces the role of competition between the channels on the market.

5.1.2. Competition among channels

In this dissertation and especially in the literature review we highly emphasized the special nature of the product reviewer information market on YouTube as it consists and resembles elements from multiple literature streams, such as the consumer learning, personal branding, behavior of the media firms or information search literature. In this chapter we reach back to a more traditional way of thinking about the economy, while we investigate the role and manifestations of the competition in our market of product review videos. We highlight the understanding of the supply side of the market as a set of competing brands in this approach.

In Chapter 4.2 we outlined our base model as we separated the videos posted by the channels into different information markets based on the topic of the video. From this initial framework, a reasonable assumption could be raised could be raised that the channels in the same information market thus direct competitors of each other, which may induce a negative relationship between the creators that are already posted a video on the channel. However, this argument does not immediately result a negative relationship between the videos on the market, when one can be successful at the cost of decreasing others' market share. We need the information described in Chapter 5.1.1. that the demand for information on the market is limited to assume a negative competition effect between the channels. The reason behind this requirement is comes from the fundamentals of economics, that the scarcity of the resources is what leads to the competition among the actors (Robbins, 1932).

In the previous chapter we derived how consumers can be satiated with the topic after watching a certain number of videos. We also mentioned that there are other considerations coming from the literature on attention economies (e.g. Falkinger, 2007), such as the consumers limited attention that leads to limited pool of views on the information market. Hence, we can assume, that the average audience member indeed cannot watch all the video on the topic, leading to a limited number of views over the topic lifetime horizon. With this argument, we can derive the argument that channels on the information market can be considered as direct competitors of each other, attempting to grab as big of a share from the pool of views available as much they can.

Therefore, we expect, that there is – on average – a negative relationship between the performance of two competitor videos. Moreover, as we described in this chapter, that satiation and competition are linked together from the channels point of view, we handle the satiation-competition effect together in the following chapters.

5.1.3. Topic Awareness

So far, during our attempt to derive internal topic interest dynamics, which extends our exogenous approach in Chapter 4.2, we derived how the topic interest could contain a missing-out element to it and how the topics are competing each other. We expect that through satiation effect, both the missing-out element and the competition of the channels -on average- harm the creators on the platforms. Hence, we anticipate that we will find negative relationship between this phenomenon and the performance of the videos.

Nevertheless, topic interest still can provide positive extra effects for the videos posted on it through the part of the audience that are still aware of the topic and still eager to follow-up on it, providing extra views for the videos. Hence, the total positive effect that the topic provides over the total topic lifetime horizon should be restricted by moderating it with the share of topic interest that represents the satiation of the consumers.

In this sense, the topic awareness in our model similar to the role of topic popularity, but while we assume topic popularity as the exogenous effect of the topic on the videos, topic awareness is going to be determined from the internal (YouTube) popularity of the topic. Thus, the effect represents how the actors on the platforms are relate to the topic at a given time period, showing the current state of the trend of the topic through the engagement of both the audience and the channels. Trivially, we assume a positive relationship between the topic awareness and the performance of the videos posted on this topic.

One of the most interesting aspect of the topic awareness we defined above is that its dynamics is not purely dependent on the satiation of consumers. That case would mean, that for instance the topic awareness is determined by a topic interest trend curve over time moderated by the satiation of consumers. In this case as satiation increases, ceteris paribus, topic awareness decreases. In contrast, the dynamics of the topic awareness is also dependent on the activity of the channels. On one hand, as they are joining to the market, they may bring new viewers. On the other hand, their content may also affect the viewers demand for information on the topic. Whether it is due to an informativeness, an entertainment or the controversiality of the videos, it may incentivize the viewers to watch more videos and become topic followers. From this point of view, this property of content creators in the product review area resemble that of opinion leaders, expert reviewers, online personalities, and influencers, as their content attracts attention towards a topic.

Either from the new joiner audience members or the topic interest buff effect, we can see that the posted video increases the pool of aware viewers. We can imagine an event that newly posted appears on the market. Then, it generates waves in the topic information market and essentially raising the total views of the topic in a multiplicative manner by increasing the overall topic interest and motivating the audience members to watch more other videos on the same topic as well.

5.2. Probabilistic properties of the Satiated-Interested audience

It can be easily seen that the arguments presented in the previous sections of the chapters are cannot be directly measured from the data available to us. Hence, in this chapter, we reformalize our reasoning and link it to measurable variables, relying on probabilistic assumptions about the main points of the arguments, namely the satiation and topic awareness effects. However, before doing so, we are also formulating a research question regarding the resultant of the outlined of the effects in the dataset.

RQ1: What is the resultant of the potential positive and negative endogenous topic effects on the YouTube product reviewer market?

The first objective in linking these effects to measurable variables is to define a variable that accounts for the total interest for a given topic. Then, we can derive the probabilistic distribution of satiation and topic awareness from this metric. Ultimately, by definition, the manifestation of the audience' interest can be measured by the view count of the topic. Hence, we aggregate the video level views to the level of topics to attain a metrics that shows us the total interest that a video received. Since, our argument relies on the dynamics of the topic interest and the probability distribution of satiation and topic awareness over time, instead of aggregating the total views at any time period, we sum up the changes in the view counts over the topic lifetime. The timing here is key to the model. We only sum up the view count changes of the videos on the topic until the observation day, in this way we model the total past interest for a topic from a perspective of the channels at that time period. However, the aim to model the effect of topic interest to the videos posted on the topic requires one more modification of this definition, as the current description would contain the focal video's view count changes as well. This specification would lead to an effect that the views of the videos could affect itself directly through the topic interest, resulting spurious correlation between the independent and the dependent variable. Hence, we will not include the focal video's view count changes during the calculation. Therefore, the calculation of total past views of topic j at time $\bar{T}$ ($1 \leq \bar{T} \leq T$) is calculated according to the formula:

$$TPV_{i,\bar{T}} = \sum_{l=1}^{N} \sum_{t=1}^{\bar{T}} \Delta Views_{l,t} \quad i \in \ topic_j \ \& \ l \neq i \qquad (4)$$

Note, that if we investigate the effect of the total past views on the performance of the videos, we do not differentiate between the views that happened close to the focal period and the ones that happened in the past, yet. We are only examining if there is a connection between all the views of other videos that was on the topic and the views of the focal video, regardless of the posting dates. Hence, this variable is not suitable to examine the dynamics of satiation and topic awareness, but it answers our research question about the resultant of these effects. (RQ1)

Then, our next objective is to derive satiation and topic awareness effects from the variable defined with formula 4. First, as we described in Chapter 5.1.1, the satiation effect shows how the audience can gradually lose interest towards a topic over time as they are watching more and more videos about it. Intuitively, we may derive that we can find the most number of people that are still interested towards a topic with the highest probability among the viewers that joined most recently, and that probability gradually decreases as we are looking at the audience that joined earlier in the topic lifetime. However, that would be only true, if the number of new joiners to the market over time can be described by a uniform distribution. Hence, our argument is linked to the share of audience that joined at that period instead. Meaning, the highest probability to find the highest share of still interested people that joined in a given period from all the people that are joined to the market on that date, corresponds to the day of observation. Then, this probability gradually decreases as we are examining earlier dates, while we can find the lowest share with the highest chance on the first day of the topic. As we can define the audience as either being satiated or aware towards a topic, we can use this distinction to define the share of audience that is interested towards a topic from the total audience and the share of audience that is already satiated. Notice, with this approach, we can also include the new videos' topic interest buff effect which works through the new views generated them. Since the new video always generates new views in the day of observation, regardless of the question whether it comes from a new audience member or from an old one, it strengthens the argument about the probability distribution of the interested viewers among all the viewers.

Let $TA_{j,T}$ denote the total audience of topic j at any time period $\bar{T}$ $(1 \leq \bar{T} \leq T)$, and the audience that joined at time t, where $t \leq \bar{T}$. Therefore, the total number of audience can be calculated as:

$$TA_{j,\bar{T}} = \sum_{t=1}^{\bar{T}} A_{j,t} \quad \forall\, \bar{T} \in T.$$

Denote the time period when the number of new joiners to the market is not significantly different from zero with T, then by definition we find a connection between the total past views and the total audience at time T as:

$$TPV_{j,T} = \varphi TA_{j,T} = \varphi \sum_{t=1}^{T} A_{j,t} \ ,$$

where φ is the average number of videos watched by one person. Then, as its discussed above, we can distinct this metric to the total number of satiated and interested audience:

$$TPV_{j,T} = \varphi(SA_{j,T} + IA_{j,T}) \ . \qquad (5)$$

Using our arguments about the distribution of satiated and interested viewers, we define a function $w_{\bar{T}}(t)$ that results the share of audience that is joined at time t and already satiated at time $\bar{T}$. Based on this function, we can derive the number of viewers that is satiated at time t as

$$SA_{j,\bar{T}} = \sum_{t=1}^{\bar{T}} w_{\bar{T}}(t)\, A_{j,t} \quad for\ \forall\, \bar{T} \in T,$$

and the number of viewers that is still interested towards the topic as

$$IA_{j,\bar{T}} = \sum_{t=1}^{\bar{T}} \big(1 - w_{\bar{T}}(t)\big) A_{j,t} \qquad for\ \forall\, \bar{T} \in T.$$

While equation 5 successfully connects the number of views and the number of views, the equation in this form only holds for $t = T$.

$$TPV_{j,T} = \varphi \left(\sum_{t=1}^{T} w_T(t)\, A_{j,t} \ + \sum_{t=1}^{T} \big(1 - w_T(t)\big) A_{j,t} \right)$$

However, for the channels, not especially the total number of these metrics that matters. Instead, we are interested in the satiation and topic awareness at time period $\bar{T}$, which is $1 \ \leq \bar{T} \leq T$. The problem here is that we do not know the volatility of φ at each period. We cannot be sure that the ratio of the number of viewers to the number of audience will be equal to the average ratio over the whole topic lifetime, or there is a deviation compare to it. Hence, similarly to our base arguments we only assume that our equations hold on a probabilistic level.

$$E\left(TPV_{j,\bar{T}}\right) = \varphi\left(\sum_{t=1}^{\bar{T}} w_{\bar{T}}(t)\, E\left(A_{j,t}\right) + \sum_{t=1}^{\bar{T}}\left(1 - w_{\bar{T}}(t)\right)E\left(A_{j,t}\right)\right) \quad for\ \forall\, \bar{T} \in T,$$

where φ becomes a scaling factor. Since we use hierarchical regression method in our model, a scaling factor does not influence the results in any way. Henceforth, we do not account for φ anymore. Finally, we rely on the property of random numbers that the expectation value of any drawn sample element from a probability distribution equals to the expected value of the probability distribution. Hence, our observation for the total topic views at time $\bar{T}$:

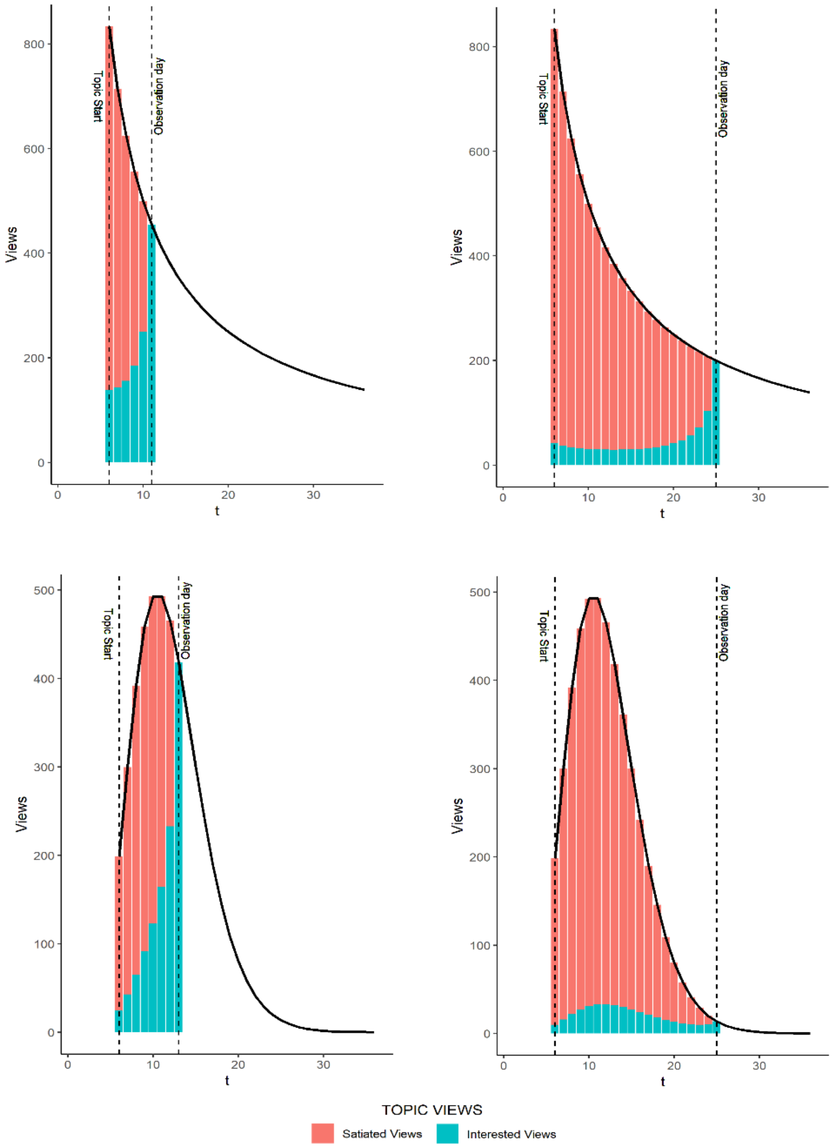
$$TPV_{j,\bar{T}}^{Obs} = \varphi\left(\sum_{t=1}^{\bar{T}} w_{\bar{T}}(t)\, E\left(A_{j,t}\right) + \sum_{t=1}^{\bar{T}}\left(1 - w_{\bar{T}}(t)\right)E\left(A_{j,t}\right)\right) + \varepsilon_{j,\bar{T}}$$

$$with\ \ E\left(\varepsilon_{j,\bar{T}}\right) = 0\ \ .$$

This equation then gives us a possibility that if we know $w_{\bar{T}}(t)$, we can also calculate the expected value of satiated $\left(E\left(SA_{j,t}\right)\right)$ and interested $\left(E\left(IA_{j,t}\right)\right)$ audience. We then use these expected values in our model to examine the effect of these metrics have on the performance of the video. To achieve this goal, we need a $w_{\bar{T}}(t)$ function.

This function essentially shows how the satiated and interested audience distributes over time from a perspective of a specific point in time $\bar{T}$. Based on the value of the argument (t), the function answers the question: *"What is share of audience that is joined the market at time t and already satiated at time $\bar{T}$ compare to all the viewers that joined at time t?"*. We illustrated how can we imagine the effect of $w_{\bar{T}}(t)$ in Figure 11 assuming exponentially decreasing interest over time from the audience. In this graph, we used two type of new topic views over time function.

Figure 11: Illustration of the distributions of satiated and interested views



Source: own elaboration

First, we imagined an exponentially decreasing topic interest over time function as well. Then, we extended this idea with a *"building-up"* period at the beginning of the topic interest, resulting a gamma function overall. Moreover, the graph contains two function curves for each topic views function form, showing the differences between the values of $\bar{T}$. Important to note, that these functions only have illustration purposes, and we do not assume such topic views functions during the estimations of the topic.

As we mentioned, after we apply the $w_{\bar{T}}(t)$ function on the total topic views, we can calculate the variables to examine the satiation and topic awareness effect which can be represented in the regression. Regarding these models, we formulated the following hypothesis:

H2: Recent topic views have a positive, while the ones that happened earlier have a negative impact on the performance of the videos.

5.3. The model of endogenous topic interest

As suggested in the previous chapter, our main objective during the implementation of the derived metrics into the model is to find the $w_{\bar{T}}(t)$ function. Our approach to this task is the following:

1. Assume a function form for $w_{\bar{T}}(t)$ which describes the nature of the increase of the share of satiation as Δt compare to $\bar{T}$ increases, but not specifies the extent of the decrease.
2. Optimize the parameter of the function by running the model iteratively.
3. Choose the best fitting model based on a decision criterion, such as the R squared of the model.
4. Repeat the process with different function form.

Corresponding to this process we formulated the following research question:

RQ2: How can we separate the aggregate effect to represent the satiation and topic awareness of the consumers and the competition among channels?

We hypothesize two function forms for representing different types of topic awareness decrease over time. First, we use the function form:

$$w_{\bar{T}}(t)_{lin} = \mu(\bar{T} - t) \ ,$$

where we optimize the value of μ. Second, we also optimize a multiplicative inverse or reciprocal function, which represents nonlinear decrease over time. The function specification corresponding to this form is:

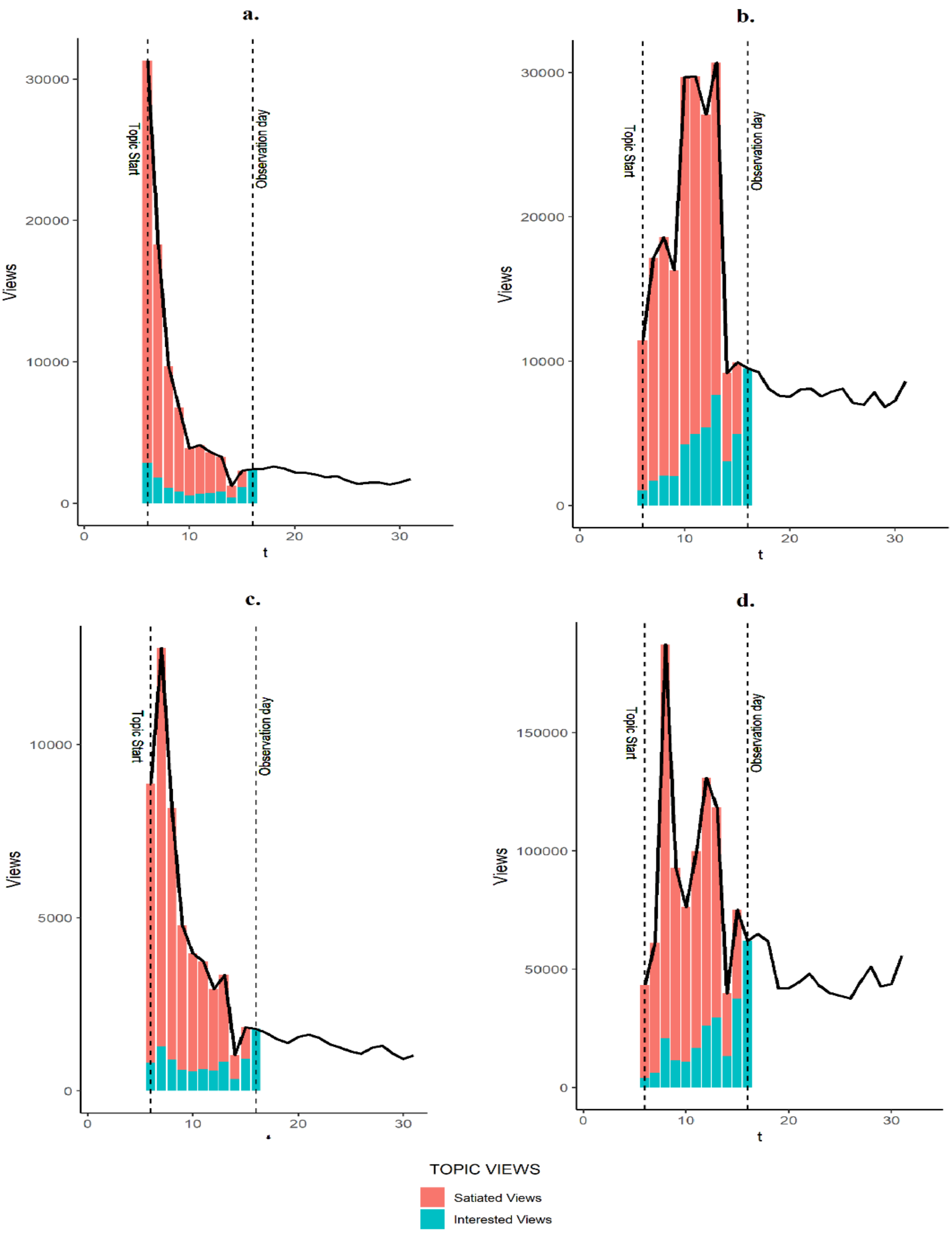
$$w_{\bar{T}}(t)_{rec} = 1 - \frac{1}{(\bar{T} - t + 1)^{\theta}} \ ,$$

where we optimize the value of θ.

In Figure 11 we have already shown how this model looks like on a theoretical level, with assumed distributions. Hence, in this chapter, concerning the implementation, we illustrated how the same effect looks like on our data with four topics from a point-of-view of different channels (Figure 12).

Our final model, answering the questions raised in this chapter, thus builds as follows. First, we aim to answer the question what the resultant of the possible positive and negative effects the topic has on the videos is by extending our previous model with the total past views of topic j. Then, we divide topic views into satiation and topic awareness with the method derived in Chapter 5.2. We illustrated our research question and hypothesizes in Figure 13, showing how the chapter extends our initial framework derived before.

Figure 12: Product related information market from the perspective of the first video poster

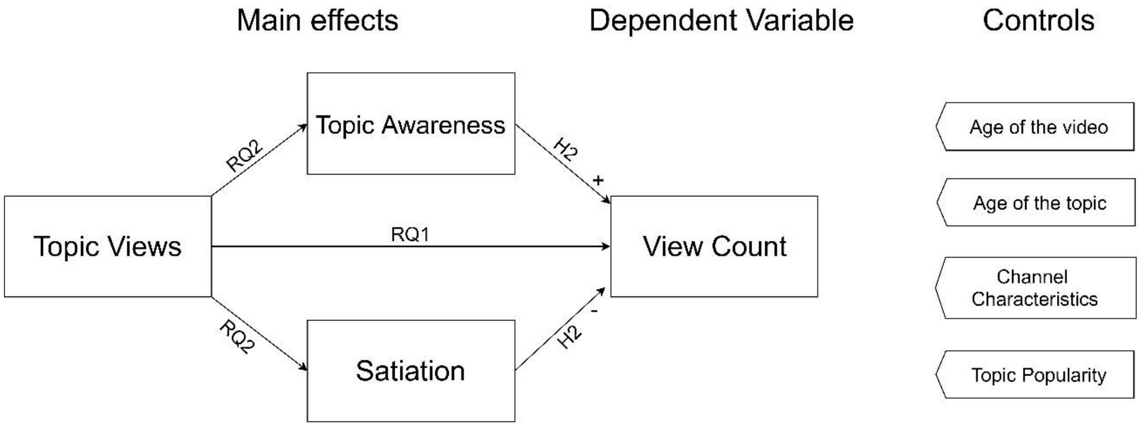


Source: own elaboration

Note: a: Motorola Edge+, b: Apple iPhone SE, c: Sony Xperia 10 II, d: OnePlus 8

In conclusion, in the previous chapters, we denoted the effect a given topic has on the videos on it as topic interest effect, and estimated it using hierarchical model approach. In this chapter, we can extend this model with the satiation and topic awareness buff effect within the market, which endogenizes the current state of topic interest. In addition, with the introduction of topic awareness buff effect through the new video posting, we also made it possible for the topic interest to grow over time in our model.

Figure 13: Conceptual model for the demand for product related information



Source: own elaboration

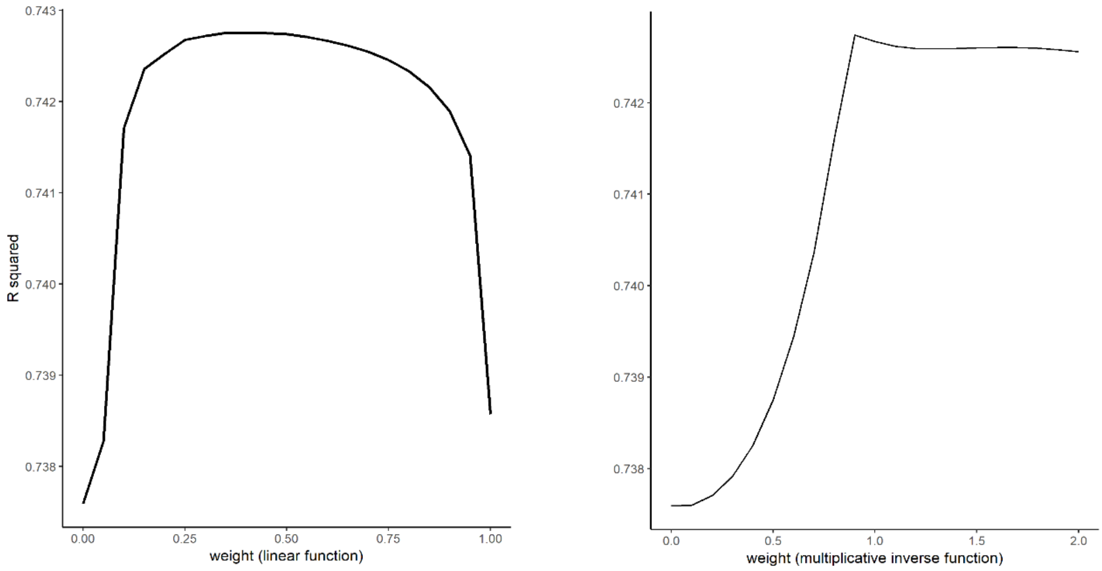
5.4. Results

Introducing the total topic views variable into the regression, we found little evidence that this variable would influence the view counts of the videos. The effect is very small (coeffient: 0.002) and only significant on 10% confidence level. However, the estimated coefficient is negative, which means, that the negative effects are slightly outweighing the positive effects, if there are meaningful positive effects on the market. This may also implicate that if we can observe satiation and topic awareness effect on the market, satiation effect is stronger.

Following the definition of the total views on the topic, we aimed to divide this variable into two separate parts with the goal of investigating the topic awareness and satiation effects. This division was made by two weighting function separately, whose parameter was optimized by iteratively calculating the values of the variables corresponding to satiation and topic awareness for each observation. Then, we estimated the model with those variables.

Hence, we repeated the estimation 42 times over the two function types with 21 different parameters for each function. First, the linear model, with the slope parameter having a value from 0 to 1 with a step of .05, and then an exponential model with the exponent having a value from 0 to 2 with a step of .1. We illustrated the achieved R squared values for these model estimations in Figure 14. In this figure, we assumed, that such a function curve can be made by eliminating the possibility of significant positive and negative spikes between two estimation results. We found that we can observe the best fitting model in the linear weighting function case at slope parameter .35. While, with the multiplicative inverse function form we get the best fit at the exponent having a value of .9. The corresponding model results for this specification can be found in Table 8.

Figure 14: Model performances by different weighting function forms and parameters



Source: own elaboration

Both models have similar results in terms of the estimated coefficients for the independent variables and their corresponding standard errors, resulting highly significant model parameters for each model. These results unambiguously suggest that the approach to divide the audience based on the distance of its corresponding period and the observation day with a weighting function results a better model than using the total past views alone.

Moreover, we can also observe that these coefficients have different signs. Based on these signs, we can confirm our expectation that satiation effect has negative while topic awareness has positive connection with the view count changes of the videos. Based on these results, we can accept our second hypothesizes.

Table 8: Regression results for the demand for product related information

Regression Results (2)

	Dependent variable:		
	ln ΔViews		
	(6)	(7)	(8)
ln channel subscriber count	0.016***	0.014***	0.012***
	(0.002)	(0.002)	(0.002)
ln age of the video	-0.076***	-0.076***	-0.077***
	(0.006)	(0.006)	(0.006)
ln age of the topic	-0.015***	0.003	0.001
	(0.003)	(0.003)	(0.003)
ln total past views of the topic	0.002*		
	(0.001)		
Linear weights (α=0.35)			
ln Topic Awareness		0.011***	
		(0.001)	
ln Satiation		-0.019***	
		(0.002)	
Exponential weights (γ=0.9)			
ln Topic Awareness			0.017***
			(0.002)
ln Satiation			-0.032***
			(0.003)
Constant	9.510***	9.593***	9.740***
	(0.051)	(0.054)	(0.057)
Random Effects			
Intercept/Channel			
Standard Deviation	0.2394	0.2417	0.2427
Likelihood ratio	10530.499***	10394.996***	10383.121***
Intercept/Age of the video			
Standard Deviation	0.0662	0.0656	0.0657
Likelihood ratio	4826.947***	4746.769***	4757.655***
Intercept/Topic			
Standard Deviation	0.1589	0.1691	0.1672
Likelihood ratio	2602.765***	2551.379***	2454.588***
Observations	41,670	41,670	41,670
Log Likelihood	15,092.380	15,145.280	15,143.370
Akaike Inf. Crit.	-30,166.760	-30,270.560	-30,266.750
Bayesian Inf. Crit.	-30,089.020	-30,184.190	-30,180.370
Note:			*p<0.1; **p<0.05; ***p<0.01

Source: own elaboration

6. Examining the information suppliers

As we highlighted thorough the dissertation, our understanding of the actors in the product information market builds from multiple literature streams as the properties of this unique market can be found in multiple discipline. In the previous chapters, we used this multidisciplinary view to show that the channels on the supply side of the market may compete but also help each other at the same time. The aim of this chapter is to move away from the previous, homogenous view about the competing information providers on the supply side of the market and explore whether and how the information suppliers differ from each other. Ultimately, our goal is to investigate the effect the channel differences have on the view count and similarly to the previous chapter, we use a multidisciplinary approach to achieve this goal.

The potential individual heterogeneities across YouTube channels explored in this chapter is sorted into two main categories. We are going to explore each main category through two types of effect. First, we explore a potential direct effect between the difference among channels and our response variable. Second, we also derive a more complex indirect effect to the model.

The first main category of channel differentiation relies on the literature of personal branding, more specifically on the role and nature of the persona of the self-brands on the YouTube platform. Hence, we first derive a model, where we account for these soft, unobserved variables for each channel. Then we extend these approaches by assuming that the persona of the channel not only results a distinct benefit for the channel, but it may change the structure and the dynamics of the relation between the videos and the information market.

Second, we also consider the aspect that the reviewers are different in their *"sizes"* on the platform that may result benefits for the channels outside of their channel characteristics. This dissertation is following the overall consensus in the YouTube platform, and lets the number of subscriber count of the channel denote the size of the channel. Hence, we first discuss the role of the subscriber count in the channel's performance. However, this argument leads outside of the boundaries of this chapter by motivating the model of subscriptions in Chapter 7. Finally, similarly to the persona of the channel, we consider the potential cross-effects between the topic information market and the size of the channel.

6.1. Brand related factors

As it was discussed in the introduction of this chapter, this section aims to consider the differences among channels completely from the perspective of personal branding, without accounting for the consequences of the differences in their sizes. Hence, the following arguments are aimed to translate the results, that we outlined in Chapter 2.2. into our approach of the product review information market. Then, extend this consideration by not only translate the implications of the personal branding literature, but also examining the potential cross effect to our previous results in the model.

This literature stream investigates how the channels are creating a brand image, a persona for their channels, and not the actual person who is presented. Despite most of these studies did not put the examination of the different personas on the performance of the channel into the main objectives, important to notice, that this consideration is still the intrinsic driver of this literature and essentially the personas of the channels. The channels creating and developing their style and brand image over time to achieve better results in the information market, to be more successful. However, this development is based on their view of what factors could make a channel successful and what is the optimal implementation of the mix of these factors specifically for them, as Duffy and Hund (2015) shows in their study in this stream. The factor considerations can come from various sources, such as audience reactions received for the videos, performance, and audience reactions for other creators, merged with the overall worldview of the actual person. Hence, we can observe that while there are clear trends in the brand images on the platform, the strategies and the implementation of the different trends differ from channel to channel. Therefore, there will be relative winners and losers of the brand image development process, based on the creators conscious or unconscious decision about the persona. As it mentioned above, this good and bad or at least worse decision rely on various sources, and we assume that every channel uses the available information differently. There are multiple reasons why we do not assume rational behavior by allowing channels to infer the way of success incorrectly. First, it can be a cognitive boundary, that channels simply cannot imagine themselves as an average viewer or average audience member of their target group (anymore). Hence, they make false conclusions about what the audience expects from them. Second, there are known biases influencing such conclusion formation from channels. While there are dozens of such possible biases (e.g. Kahneman et al., 1982), we highlight the role of survivorship bias (Brown et al., 1992) in this decision to illustrate the potential flawed logic. According to this bias, channels may infer the wrong success characteristics simply because they do not see the channels that are stopped their activity due to the lack of success. Third, there could be cases, when there is an overall good understanding of the success factor, but channels choose a bad mix that may result a negative attitude from the audience towards the channel. These thought processes essentially lead to a range of different brand images on the market, compare to the previous homogenous view of the supply. Along this difference, using argument above, we hypothesize two scenarios in which the described brand image differences can affect the performance of the videos in our model.

First, we approach the benefit of a good brand image to define it as a buffer to the performance of the channel's videos compare to the worse brand images. However, here, we assume that the buffer's effect on the performance is only dependent on the persona of channel, or in other words, the bolstering effect is independent from other factors (such as the topic) in the model.

However, one can argue, that such independency may not hold in case of the relation to the topic information market. If we accept our hypothesis regarding the presence of unique self-brands in the market, we can conclude that the market-related economic consequences of the presence of brands may also play a role in our model. The economic literature has been examining these consequences for a long time. From the perspective of the objectives of the dissertation, the most important out of them is the effect of brands on the competition, consumer loyalty and price elasticity (e.g. Simon, 1979; Krishnamurti 1992; Delgado-Ballerster, 2001; Alnawas 2015). Based on these papers we can derive, that the economic consequences of the competition present in the market can be moderated if the firm builds his brand such that the price elasticity for its products is lower than its competitors. In this way, brands can build more resilience against competition. Therefore, we hypothesize that the competition related (Chapter 5.1.2) satiation effect in our model can be moderated by the persona of the channel. In addition, from the defined nature of good brand image, such that it is attractive to the audience, we also enable that possible mediator or bolster effect could be present in the positive side of the topic related effects, and a channel with better brand image may benefit more from the topic awareness effect.

In conclusion, motivated by the above arguments, we aim to answer the following hypothesis regarding the persona of the channel:

H3:

A: The unique channel characteristics have a significant effect on the performance of the videos.

B: The unique channel characteristics significantly differentiates the topic effects for the channels

6.2. The size of the channels

The differentiation the channels along their sizes is motivated from various perspectives and similarly to the persona of the channel, we are exploring the effect of differentiation on its own and also by its cross-effect with the topic information market. However, in contrast to the previous chapter we start our argument with the possibility that sizes may modify the channel's relation to the topic information market first, then we discuss the meaning of the effect second. The reasoning of this order is simply due to the fact that the topic cross-effects relates to the previous part of the dissertation, while the sheer size effect leads outside of the limits of our current framework, introducing the second set of models in the dissertation.

When we examined the effect of the channel characteristics on the topic's effect on the videos, we consciously omitted the size of the channel as a characteristic. The rationale behind this is based on the consideration that the size of the channel is not a chosen trait, rather it is a result of the channel's previous activity. The consequences of the differentiation based on the size of the entities of the supply on the market has a wide literature in the field of economics (e.g. Amato and Wilder, 1985; Amato and Amato, 2004; Lee, 2009; Niresh and Thirunavukkarasu, 2014). From this literature stream, we may infer, that as the size of the entities increases, usually they are more capable of capitalizing on their goods on the market, to the detriment of others' interest. Therefore, our prior expectation is that as channel size increases, channels can also capitalize on the present topic awareness better in the market by attracting more interested views to their videos compare to smaller channels.

Regarding the satiation/competition effect, although we anticipate the cross-effect to be significant, we do not have any expectations about the sign of the parameter. We can motivate the positivity by relying on the argument that channels can, similarly to the brand image, build resilience from the market effects by creating large enough loyal fanbase. However, we can also motivate the negativity by assuming that small channels are more capable of avoiding the competition and enter to niche topics, while big channels are the faces of the market and cannot avoid this effect.

Finally, we consider the effect of the channel sizes on its own to the performance of the videos. While we left this argument to the last regarding the views model, the question how the fanbase of the channels affects the views of the videos is probably one of the most important ones for the creators in the market.

First it seems trivial to assume that as the number of subscribers is growing, there would be a higher number of views (Yoganarasimhan, 2012; Diwanji et al., 2014; Liikkanen and Salovaara, 2015; Welbourne and Grant, 2016; Burgess and Green, 2018;). However, the order of the causation, whether more views are causing more subscribers or the higher fanbase will watch more videos is not trivial at all. Behind the issues of whether and how the subscriber number may affect the number of views, there is essentially one important question: With subscribing to a channel, will there be a higher probability for the representative audience member to watch the upcoming videos from the channel or she/he would watch those videos with the same probability either way. We can argue that besides other reasons, one may become a subscriber to get notifications if a new video is posted from the channels, she/he subscribed. Another reason could be to get faster path for the channel's videos. Therefore, we can assume that the subscriber count positively affects the number of views.

The other direction of this assumed positive connection is the number of views affecting the number of subscriptions of a given channel, suggesting that the subscription base can be both a reason and a result simultaneously. Due to this consideration, while we consider the size of a channel as an exogenous variable, in the next chapter we extend our approach with a second model, estimating the number of subscription number gain, which can be dependent on all the views a given channel received to its videos.

Based on these arguments about the sizes of the channels, we hypothesize the following:

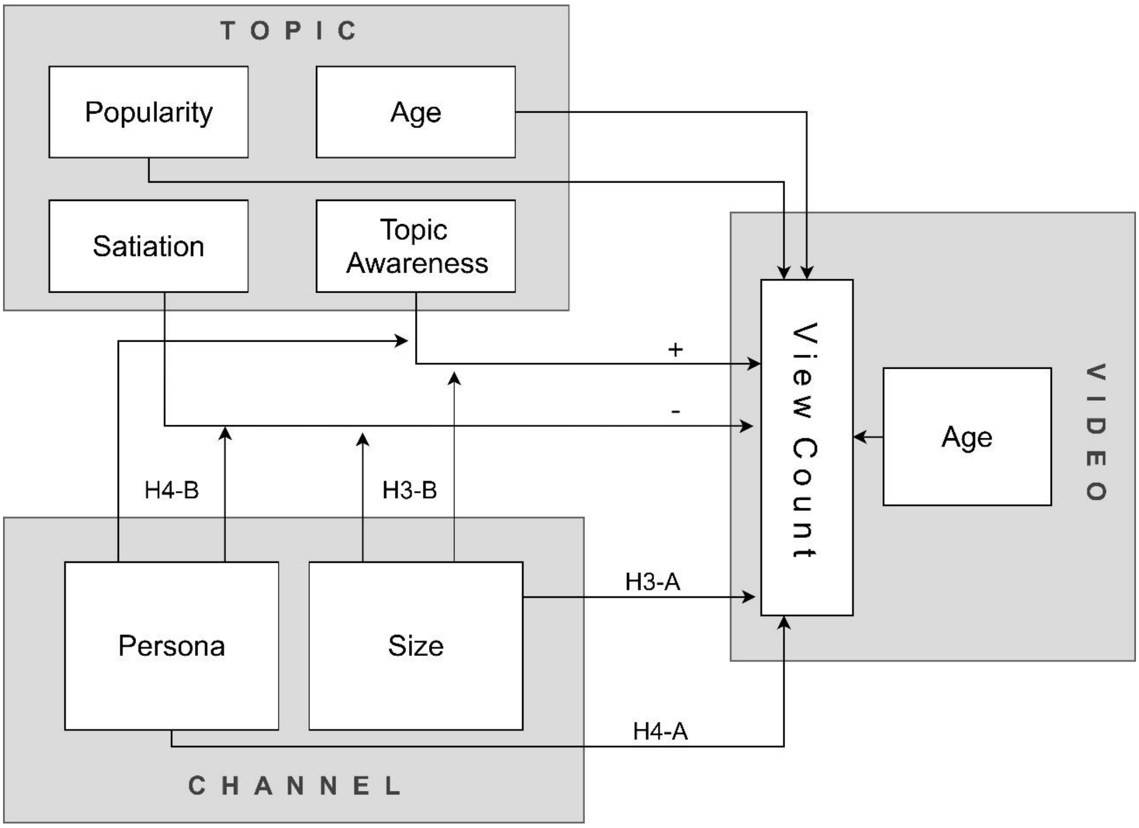
H4:

A: The number of subscribers of the channels has a significant impact on the performance of their videos.

B: The number of subscribers of the channel has a significant interaction effect with the topic effects in the model.

In conclusion, we illustrated the updated conceptual model with the introduced new elements in Figure 15, including our hypothesizes for the chapter. Here, we can see that compare to the previous models, now the channel characteristics and the topic variables are interconnected, creating a complex structure of relations in the model.

Figure 15: Conceptual model of the product information economy on YouTube,



Source: own elaboration

6.3. Methodology

Similarly to the topic effects in the model, we start the modelling of the channel specific elements of the model by defining a hierarchical regression equation system by the dimension of channels. In this way we can handle the channels as factors that create different groups with different structure compare to the base model or the model with topic information markets. The level in which these groups' equation structure differ compared to the base model is based on the level of complexity we assign for the effect of the channel differentiation. (Chapter 6.1)

First, we model the more straightforward performance buffer persona effect. Here, we define a random distribution for the intercept for each channel. Note, that we already used this approach by assigning a random distribution of intercepts for each topic. Hence, we extend the present vector of random intercepts into a matrix based on the channel and topic of the videos.

Then, we extend this model by assuming that the channel's persona can alter the already defined topic effects in the model, namely the satiation and topic awareness effect. Meaning, besides the matrix of intercepts, we also define a vector of topic effect in the model. In other words, we are modelling different, random slopes (curves) for the topic effects for each channel. This effect is illustrated by Figure 4 in Chapter 3.3.1.

Finally, we also described to model the effect of the channel sizes as the other differentiating factor among youtubers in our approach. However, for control purposes, we already represented the logarithmic transformation of the subscriber count of the channel in the model. The only term that is yet to be derived is the cross-effect between the topic of the channel and its size.

This cross-effect in case of the subscriber count can be modelled by introducing two interaction terms into the regression, a satiation-size effect, and a topic awareness-size effect. The significance value corresponding to these interaction terms shows whether such cross-effects are supported by our data.

6.4. Results

6.4.1. Brand effects

The summary of results of the model estimations can be found in Table 9. First, we investigated whether the persona of the channel provides any extra effect for the videos through a random intercept. Note, that this model assumption was also used in previous models (Table 7 and Table 8) for control purposes. Based on the likelihood ratio test (Chapter 3.5.6.), we can conclude that in every model, the usage of random intercepts performed better than a constant intercept across the channels. Moreover, the results of Model 11 and 12 show that the model with randomly defined coefficients corresponding to the satiation/competition and topic awareness effect, respectively, performs better than the constant slope model. The significance of the current model setup compared to previous frameworks highlight, that not every channel relates to the market in the same way. There are creators who enjoys more benefit from the topic awareness, and less exposed to the satiation/competition effect on the market. Hence, these models have crucial implications for the channel. Besides their decision to which topic should they choose to review and when should they post the review, they should carefully design their brand image, because it changes the whole structure in which their performance is dependent on factors other than the information presented in the video. Although, the analysis of the exact elements of the brand image reaches outside of the scope of this dissertation, it highlights a potential future research direction in this literature stream.

6.4.2. Channel size effects

Similarly to the previous section, let us examine the results of the models containing cross-effects between the channel sizes and the topic effects first. Our results implicate that both coefficients are significant, suggesting that the size indeed influences the channels connection to the market. Moreover, we can also infer, that both coefficients of the interaction terms are negative. Thus, given the opposite sign of the original effect, the two interaction terms have opposite consequences to the baseline (size independent) effect of the topic. In case of the satiation/competition effect, it means that the negative effect on the performance of the videos becomes stronger as the channel size increases. This result contradicts the argument that bigger channels may built up a resilience against this effect and indicates that smaller channels are less exposed to the satiation effect on the market. One explanation could be through the visibility of the channels due to their sizes in the eyes of the audience, that makes competition stronger in case of big channels.

There is another implication coming from the observation that the satiation on the market is more important for big channels. Since there is a higher satiation effect for them, the timing is more important for channels with big subscriber counts. Hence, they should pay attention to not wait too long for potentially big topics, since the growing satiation on the market damage their final view count they will receive for the video, resulting smaller revenue than the potential if they post the video earlier.

On the contrary, in case of the topic awareness effect we can observe an opposite relationship. Here, the negativity of the coefficient corresponding to the defined interaction effect means, that the role of topic awareness weakens as the positive effect on the performance of the videos becomes smaller as the channel size increases. We can conclude from this result that our prior expectation proved to be wrong, and big channels are not capable of capitalizing on the topic more than small channels. We observed that the opposite is true, and small channels are benefiting more from a topic that is *"trending"*. This consideration implicates, that it is worth for small channels to follow the trends on the market as they are receiving much more extra views from the trend (that is mostly determined by big channels) compared to big channels, which results extra revenue in our approach.

Calculating the overall effect of the channel sizes, we can see that based on the relative sizes of the effects, there is a trend in the model in which the two coefficient is approaching each other as the channel size increases. After a certain number of subscribers, there will be an overall negative effect of the topic. Overall, it means, that the topic awareness effect, according to our results, helps the small channels to gain subscribers and but less or not effective for big channels. A possible explanation could be that as the supply of videos are growing, it raises the topic awareness of the audience. The increased awareness means increasing demand for information as well, that can reach beyond the scope of the supply of the videos by big channels, that usually serve as the first videos to be watched. Then, this extended demand can highlight the small channels on the market, providing information about the same topic. Hence, as the topic interest rapidly grows due to big channels coming to the market, small channels may have a chance to get attention through recommendations or YouTube searches from viewers that are not familiar with these small channels yet.

Finally, we also examined the connection of the subscriber counts on their own on the performance on the videos without any interaction. Our results regarding the benefits of the fanbases are very robust, since all the model unambiguously shows a significant and positive relationship between the variables. Therefore, we can affirm that the fanbase is an important source of revenue corresponding to product review videos. However, for channels, the most important aspect of this relationship is the potential for the long-term benefits of building a fanbase. This aspect is based on the consideration that this effect is applicable for every video the channels have; it is going to continually have a positive impact on the view counts and essentially on the revenue of the channel.

Moreover, these results can highlight the possibility of even more meaningful long-term benefits if we account for the multiplication effect of the subscription number of the channel. This effect relies on the idea that a higher view count may translate into higher subscription number as well for the channels at later periods. Through this way, a bigger fanbase can result higher view counts, which results even bigger fanbase, leading to a multiplication effect in the model. Notice, that this multiplication effect, if it is indeed present, also applies for extra effects coming from the topic and persona, since as we already shown, it results higher view count, which in this theory may contributes to the fanbase building as well. Therefore, in the following sections, we test the hypothesizes that is motivated by the subscription multiplication effect, so at which extent does the view counts of the videos converts into subscriptions.

Table 9: Regression results for channel-topic cross effects

Regression Results (3)

	Dependent variable:			
	ln ΔViews			
	(9)	(10)	(11)	(12)
ln channel subscriber count	0.012***	0.012***	0.013***	0.011***
	(0.002)	(0.002)	(0.002)	(0.002)
ln age of the video	-0.077***	-0.076***	-0.076***	-0.072***
	(0.006)	(0.006)	(0.006)	(0.006)
ln age of the topic	0.001	-0.001	-0.002	-0.017***
	(0.003)	(0.003)	(0.003)	(0.003)
ln Topic Awareness	0.017***	0.050***	0.083***	0.052***
	(0.002)	(0.006)	(0.025)	(0.007)
ln Satiation	-0.032***	-0.008**	-0.001	-0.049
	(0.003)	(0.004)	(0.004)	(0.042)
ln Topic Awareness x sub.count		-0.034***	-0.039***	-0.041***
		(0.006)	(0.007)	(0.006)
ln Satiation x sub.count		-0.025***	-0.028***	0.005
		(0.003)	(0.004)	(0.004)
Constant	9.740***	10.584***	10.271***	10.653***
	(0.057)	(0.098)	(0.301)	(0.697)
Random Effects				
Intercept/Channel				
Standard Deviation	0.2427	0.2984	2.2049	5.4179
Likelihood ratio	10383.121***	9350.84***	11488.59***;2	12803.501***;4
Intercept/Age of the video				
Standard Deviation	0.0657	0.0661	0.0655	0.0647
Likelihood ratio	4757.655***	4800.616***	4857.988***	5017.936***
Intercept/Product				
Standard Deviation	0.1672	0.158	0.1641	0.1716
Likelihood ratio	2454.588***	2468.505***	1943.736***	2214.735***
Topic Interst Buff/Channel				
Standard Deviation			0.1891	
Likelihood ratio			2137.751***;4	
Satiation/Channel				
Standard Deviation				0.3319
Likelihood ratio				3452.661***;5
Observations	41,670	41,670	41,670	41,670
Log Likelihood	15,143.370	15,195.680	16,264.560	16,922.020
Akaike Inf. Crit.	-30,266.750	-30,367.370	-32,501.120	-33,816.030
Bayesian Inf. Crit.	-30,180.370	-30,263.720	-32,380.200	-33,695.110

Note: *p<0.1; **p<0.05; ***p<0.01

[1]Normalized value

[2,4]Calculated by dropping [2]Topic Interest Buff/Channel or [4]Satiation/Channel term

[3,5] Calculated by reducing [3]Topic Interest Buff/Channel or [5]Satiation/Channel term to Intercept/Channel

Source: own elaboration

7. The growth of the channels

In the previous chapters, we successfully modelled how the dynamics of the view counts of YouTube videos in the product reviewer market are evolving over time from the perspective of the creators on the market. Although, this framework highlighted key findings for youtubers, such as the role of topic and time decision, it only focused on the performance of one single video a given channel has, and not the performance of the channel in general. Nevertheless, a channel that aims to be successful on the market should aim to maximize the revenue coming from the sum of multiple sources of videos, not just only one. In Chapter 6 we derived how the sizes of the channels can affect the view count of the videos and discussed that it opens up the possibility of long-term benefits for the channel as the number of subscribers affects every video the channel has. Moreover, the channels' follower count could have even more benefits for the channel if the performances of the videos matter in the subscriber count gathering process. If the view count of the videos successfully turned into subscribers, the channel enjoys a multiplicative growing process in which the higher subscriber count result higher views, that translates into even higher follower counts.

Building on these arguments, our goal in Chapter 7 is to answer the question, how can we model the channels subscriber building process over time. The chapter builds up from the motivation, methodology and results of two main, distinct parts. First, we derive the main model, and examine how can we model the subscriber count gathering trends of youtubers over time with the possibility of both performance independent and dependent growth. Then, we extend the base framework to answer, whether we could explain a significant part of the growth by the audience reactions of the channels' videos.

7.1. Performance induced growth

In the first section of the chapter we aim to answer the most essential question regarding the presence of the translating effect of views into subscribers. Our model is based on a proportional process, in which a certain ratio of the new views for the videos of the channel becomes subscriber at every period.

However, besides the performance related growth, we also need to control for the performance independent elements of the model. Hence, besides representing the performances of the channels' videos, we also represent a unique channel specific trend when we model the growth processes.

Based on this model setup, we outline the following hypothesis regarding the performance related growth of the channel:

H5: The view count changes of the channels' videos has a significant positive effect on its subscriber number changes.

7.2. The reach of the channels

In the subscription gathering process, the main interest of the channels is to reach the viewers that lies outside of their fanbase. Moreover, they can also aim for reaching the audience that are even outside of the set of viewers that are already familiar with the them. Hence, from the channel's point-of-view, we can divide the audience into three segments:

1. The viewers that watched a video and already subscribed to the channel.
2. The share of audience that watched a video but decided not to subscribe.
3. Finally, the viewers that are not familiar with the focal channel, thus not considered the decision of subscribing yet.

In the first case, in terms of the subscribing gathering process, the channel's main goal is to keep these viewers in the follower base, and prevent potential unsubscribe. While, there is interesting line in the literature examining crises when the brand can rapidly lose reputation in multiple domain (e.g. Zhao et al. (2011), investigating product harm crises in the consumer learning literature), this consideration lies outside of the scope of this dissertation. For this segment, we are assuming that the channels are capable of keeping the quality level that the subscribers expect from them.

Therefore, this chapter of the dissertation focuses on the remaining two segments. In case on the second group of viewers, the channel can assume that there is a possibility that in the future they eventually become subscribers. This can happen, if they change their mind if they showed disinterest towards the channel in the past or the channel can provide enough evidence for the viewers that are uncertain about the channel. The channels therefore aim to provide this evidence through their videos which may incentivize them to finally commit to subscribe.

In case of the viewers, that are not familiar with the channels, we can argue along multiple consideration. First, since they have not seen any content from the focal channel in the past, they have not considered subscribing to the channel, yet. Hence, the group potentially contains viewers that would potentially subscribe immediately, and also the ones who will move to second group after watching the content of the channel. Therefore, we can argue, that the possibility of a viewer becoming subscriber from this group after watching a video is higher than that of viewer from the second segment. However, reaching these people is potentially harder compare to the members of the previous categories, since they have not seen a content the focal channel made, yet. Nevertheless, based on the higher chance corresponding to converting these viewers into subscribers, if the channel can reach them, we expect a higher growth.

Therefore, we define a measure of *"reach"* by how far the channel's videos can spread on the market beyond its regular viewership. This viewership defined by the "usual" view counts the channel's videos gets. Based on this measure, we hypothesize, that as the reach of the channel increases, we expect a boost, an increase in the subscriber gathering process.

H6: Outlier videos of the channel in terms of their view counts have significant positive extra effect on the subscriber number changes of the channel.

7.3. Audience reactions

Examining the growth mechanism of the channels, especially its performance related factors, one may ask, what is the underlying role of the valence of the audience towards the channels. Are the channels with positively rated content going faster? Or only the engagement from the audience is that what matters for them? Or simply, there is no such connection, and channels with low engagement can also grow fast if they are making content that is desirable for certain set of viewers.

For channels on the market, the answers to these questions could lead to multiple implications regarding their long-term strategy to create bigger market share. Besides this strategy, if theses metrics indeed matter, it also extends the list of indicators for the channel, that can help to find the strengths and weak spots of their current performance. Therefore, this chapter extends the previously defined baseline model with the reactions from the audience to the focal channel's content.

However, there are multiple theoretical and technical challenges to overcome if we aim to represent these effects in our model. In this section, we provide solutions for the theoretical questions, then, in the methodological section (Chapter 7.4), we show, how can we solve the technical issues.

From the theoretical standpoint, our main question lies in the nature of the aggregation of the audience reactions from the video level into an overall channel effect. Essentially with this consideration, we also form an assumption about the audience's mental model about the channel prior subscribing. We can model this mechanism such that we assume an aggregate valence value for the channel, which is an average view coming from the content of the channels. In this framework the videos are essentially not perfect manifestations of their creator's overall image. Thus, the image of the channel can be inferred by watching its videos and deriving an average view from them. From this perspective, channels that are not joined to the profession recently have robust view from the audience. It can be changed, but only slowly, so channels have to consistently create videos that are welcomed by the audience. Consequently, the image can also be worsened, but similarly, only slowly. Therefore, an outlier video in terms of the valence and engagement towards it cannot ruin the image immediately in respect to the subscriber gathering process. We denote this solution of aggregation by the *"Average Subscribing Image"* of the channel.

In contrast, we can also think about the connection between the audience reaction and the new subscriber counts as it is present on the video level. In this approach, the overall effect on the growth of the channel is the aggregation of the video level contributions. From this perspective, legacy performance, the content that was posted long time ago - on average - has little effect on follower base changes. It is mostly affected by the videos that were relatively recently posted. Obviously, this modelling approach leads to a more volatile process. However, this theory may be more capable of grabbing the effect of the current trends, about the perception of the channel and its effect on its growth. Moreover, it can show the effect, if there is any, of a sudden positive or negative burst in the perception about the creator, such as a sudden wave of dislikes after a controversial video.

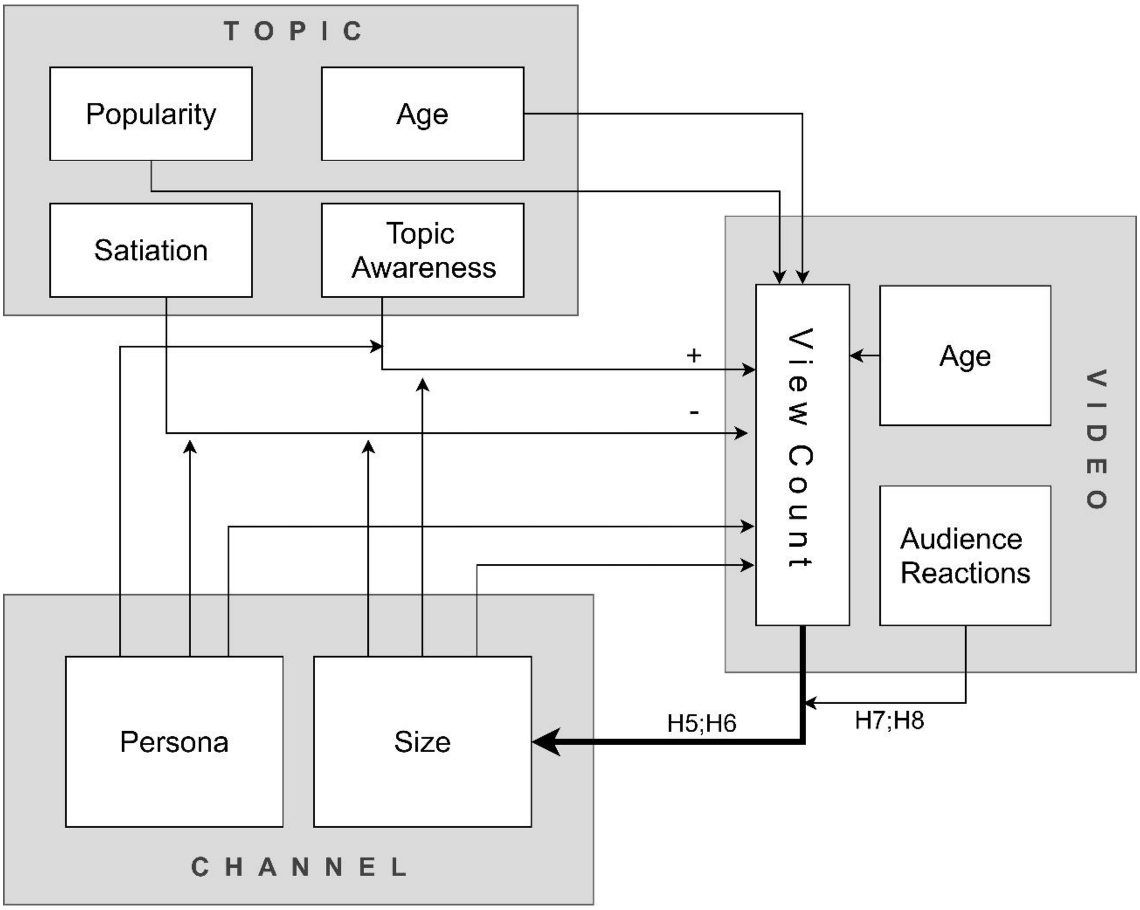
Based on the arguments discussed above, at this point of motivating the model development, we do not take a side which framework represents better the connection between the feedbacks and the growth. Instead, our solution to this question is to continue the modelling towards both direction and let the data provide the answer which approach represents better the relationship. Therefore, in the next section we derive two model, one for each consideration to be able to decide which approach fit better to the data, resulting the following hypothesizes:

H7: We can explain the channel growth better if we use the channels' average audience reaction metrics.

H8: We can explain the channel growth better if we use the video contribution audience reaction metrics.

In addition, we also updated the conceptual graph of the models in this dissertation to contain the final model extension, the growth of the channels (Figure 16).

Figure 16: Conceptual model of the product information economy on YouTube, including the growth of the channels



Source: own elaboration

7.4. Methodology

7.4.1. Representing the performance in the model

The main goal of this chapter is to describe the model of the channels' growth. As we discussed in previous chapters, we denote the channels' sizes at a given period by their measured subscriber counts at that period. Hence, our response variable in this chapter:

$$\Delta Subscribers_{k,t} = Subscribers_{k,t} - Subscribers_{k,t-1}\,,$$

Similarly to the views models, we assume that a nonlinear connection between the subscriber gaining process is more realistic than a linear relationship. Hence, we use the logarithmic transformation of our variables. Then, according to Chapter 7.1, we build the base model by assuming both performance independent, and dependent growth factors.

In consistent to previous chapters, where we denoted the performance of the videos at a given period as the number of views gained compared to the previous period, we define the performance of the channel as the sum of the performance of the videos (posted on any topic):

$$\sum_{i}^{N_{kt}} \Delta Views_{it} = \sum_{i}^{N_{kt}} (Views_{it} - Views_{i,t-1}) \text{ ,}$$

where N_{kt} is the number of videos channel k has at time t. Therefore, we define the base model with both performance dependent and independent factors as:

$$\Delta Subscription_{kt} = \beta_{0k} + \beta_1 \sum_{i}^{N_{kt}} \Delta Views_{it} + \varepsilon_{kt}$$

$$\beta_{0k} \sim N\left(E(\beta_{0k}), \delta_{\beta_0}^2\right)$$

where β_{0k} is the trend component of the model and β_1 is the rate in which the performance of the channels translates to subscribers. Thus, the trend component in the model is unique for the channels, but we are interested in modelling constant performance ratio across all the channels.

7.4.2. Deriving the reach effect

In the following section we derive a metric that is aimed to represent the effect of the reach, that we defined in chapter 7.2. This effect essentially was defined to show how far the channel's videos can spread on the market beyond its regular follower base. The underlying assumption of the effect is based on the argument that the channels may get more subscribers if they make a video that can reach outside of the usual audience of the channel compared to the number of subscribers that the number of views would suggest. Thus, we expect that we observe extra amount of growth, if the one or more videos of the channels are getting unusually high views, compare to its regular view counts.

Hence, our metric should rely on the performance of the videos. Thus, before defining the overall effect that can be represented in the regression, we derive a video level metric for the phenomenon first. However, based on our arguments, it should only show notable effect on the growth process, if the performance is outlier in the channel's videos in terms of its view counts. We can grab this effect if we derive our metric in a way that it attains exponentially higher values if the performance of the video is increasingly higher compared to the other videos.

Finally, we need to grab the property of the video that it only has to be outlier in the set of the channel's videos to have an extra effect on the growth. We can achieve this by normalizing the performances of the videos the channels have for each creator separately. In this way, every channel will have their own reference system of performances, while our metric in the regression will denote the same effect for every channel. Without the channel level normalization, this method would result a biased metric, led by the sizes of the channels across all the creators.

Therefore, we calculate the defined reach metric in the following way:

$$r_{it} = \Delta Views_{it}^{\overline{Views}_{it}} \ ,$$

where $\overline{Views}_{it}$ is the normalized value of the view counts of channel k (with videos i = 1…N_k) in the scale of the all the video of the channel.

Then, we can aggregate the reach metric for each channel across all the video to get the channel's total reach at time t, which can be represented in the regression equation in our models.

$$R_{kt} = \sum_{i}^{N_{kt}} \Delta Views_{it}^{\overline{Views}_{it}}$$

Important to note, that this is the only term in the regression, that is represented without the logarithmic transformation. The reason behind the lack of conversion is that we aimed to represent an exponential connection with the formula. If we would take the logarithm of it, we would lose some level of this exponentiality in the model, and it would not be capable of sufficiently denote our hypothesized connection.

7.4.3. Using audience reactions

Our last extension of the base model of the channels' growth aims to explore the connection between the audience reactions, and the subscriber gaining process. Modeling this relationship, we are asking, whether we can explain a significant part of the variance of the growth among channels by introducing the audience's reveled valence, opinion, or engagement in the model. Essentially this relationship may shed light on some of the underlying thought processes, viewers have on average prior to subscribing to a channel in this domain.

From the perspective of connecting the audience's opinion about a given content on the market and the growth of the channel that posted that video, the most valuable asset for us is the observations that reveals the audience valence towards the focal video. Therefore, we can use the information about the number of likes and dislikes a given video received as a good measure of revealed valence. The limitation here is obvious, as this measure classifies the underlying valence of the viewers that committed to express it in a binary fashion. However, since the name of this function of the platform clearly suggest the underlying valence of the viewer, we consider this metric as an appropriate measure on an aggregate level.

The implementation of these measures to truly show the valence towards the video facing an obstacle as simply introducing it to the regression would result a biased relationship. This is due to the positive connection between the number of views and the number audience reactions a given video receives. Therefore, to achieve an appropriate measure, we divide both the number of likes and dislikes at a given period with the number of views in that period.

Finally, one can also argue that these valence metrics still contain unfolded information, that can be examined if we handle them together. Meaning, the overall valence towards a video from the audience may lie in comparing the number of likes to the number of dislikes at any given period. Hence, in this way, we not only representing the absolute number of likes and dislikes, but also a relative measure expressed by the ratio of the audience who expressed them.

Our last audience reaction measure has a special role in our model, as it does not reveal us the valence of the audience. While one can argue that comments for the videos can contain information that can show both positive and negative valence (even at the same time) towards a video, retrieving this information would simply require too much resource in the model development process. Hence, the reason is only due to a technical limitation since it would require highly sophisticated natural language processing (NLP) and sentiment analysis techniques. Nevertheless, the number of comments can still provide extra information about the audience. Our underlying assumption that motivates the representation of this variable is based on the consideration that posting a comment requires more effort from the viewers than clicking on the like/dislike function of the platform. This is even more accurate, if we consider that a significant part of the comments are replies to other comments, which suggest that the viewer spent more time with the particular video. Thus, we argue, that under the number of comments, there is more engagement from the audience than the number of likes and dislikes. This argument holds regardless of the valence of the comment. Therefore, we are representing the number of comments as an extra measure of engagement from the audience.

In case of the number of comments, we can apply the same assumption regarding its correlation with the number of views as in case of the likes and dislikes. Meaning, we expect that as the viewership of the video grows, the number of comments is increasing as well. Hence, once again, we should divide the number of comments with the number of views before representing it in the regression. Finally, we summarized our main variables in this chapter grouped by their underlying driver and the relation to each other in Table 10.

To this point, we defined measures from the feedbacks that a channel receives to their videos and we do not described how these video level metrics can be aggregated to an overall channel metric at every period in time, that can extend our baseline model. Therefore, in the next sections, we discuss two methods of aggregation, each corresponding to different thought processes of the audience.

Table 10: Audience reaction categories

	Valence		Audience Engagement
Absolute terms	Likes/Views	Dislikes/Views	Comments/Views
Relative	Likes/Dislikes		

Source: own elaboration

7.4.3.1. Modeling the average subscribing image of YouTube channels

Our first defined method regarding the aggregation of the video level valence and audience engagement metrics is denoted by the average subscribing image the audience form about a given channel. From a theoretical standpoint it means that the audience looking at the videos a given channel has as all of them are the manifestations of the same channel image, quality, or other channel related properties. Hence, they treat every video as equal when they form their decision of subscribing. Becoming a subscriber, the viewer essentially commits to get notifications and easier access for all the future videos.

From a methodological point-of-view this translates to an aggregation where all the videos of the channel weighted equal. It also means, that we should not differentiate between videos in terms of the overall impact of one increment of likes and dislikes. In other words, one like or dislike worth the same for each video, regardless of the video’s other properties such as its size.

Therefore, accounting for the correlation between the view count of the video and our measures, we can aggregate the video level metrics to a channel level variable by dividing the sum of the videos’ measure of valence or audience engagement and the sum of the views of the channel’s videos. Then, consistently to our previous models, we take the logarithmic transformation of this variable to get our independent variables in the model:

$$
ln\ \Delta Subscription_{kt} = \beta_{0k} + \beta_1\ ln \sum_i^{N_{kt}} \Delta\, Views_{it} + \beta_2\ ln\ \frac{\sum_i^{N_{kt}} Likes_{it}}{\sum_i^{N_{kt}} Views_{it}} + \beta_3\ ln\ \frac{\sum_i^{N_{kt}} Dislikes_{it}}{\sum_i^{N_{kt}} Views_{it}} +
$$

$$
+\beta_4\ ln\ \frac{\sum_i^{N_{kt}} Comments_{it}}{\sum_i^{N_{kt}} Views_{it}} + \beta_5\ ln\ \frac{\sum_i^{N_{kt}} Likes_{it}}{\sum_i^{N_{kt}} Dislikes_{it}} + \varepsilon_{kt}\ ,
$$

$$
where\colon \beta_{0k} \sim N\left(E(\beta_{0k}), \delta^2_{\beta_{0,}}\right)
$$

7.4.3.2. Modeling video level subscriber contributions

Our final model extension represents a different thought process than that of corresponding to the average subscribing image. In the previous method, we hypothesized that the channels have an overall image based on the audience reaction metrics coming from all the video, and then, this image can explain the variance in the subscriber count gains across channels. In this model, the valence and audience engagement have an indirect underlying relationship with the subscriber count changes through the overall image of the channel. In contrast, we can think about the effect that a better perceived video has on the subscriber number compared to a video that is welcomed worse as having a direct relationship with the subscriber count changes.

Hence, this approach assumes an aggregation which rely on the individual contributions of the videos, similarly to the performance dependent elements of the model (Chapter 7.1. and 7.2). However, with extending the model in the direction of the audience reactions, we aim to explore whether we can explain the variance in our response variable if we account for the number of likes, dislikes and comments of the individual videos that caused the increase in the dependent variable in the first place.

Therefore, following the logic of the performance dependent growth, we derive a metric where our video level metrices are weighted by the number of views the videos received compare to the previous period. In this way, our variables show the valence effect of the video weighted by the number of views the video received. Similarly to the previous methodology, to avoid the biasness coming from the video size effect, we divide these variables with the views of the video at the given period. Then, we can aggregate this video level metrics to one aggregate measure that can be introduced to the model. Worth to note, that the weighting with the view count changes also assures, that we avoid another biasness in the model. It would come from the fact that a channel with higher number of videos would have - on average - higher values for these metrics as the number of likes, dislikes and views are always nonnegative numbers.

Finally, after taking the log-transformation of these variables, our final model of the subscriber gathering process is the following:

$$ln\ \Delta Subscription_{kt} = \beta_{0k} + \beta_{0k}\ ln \sum_{i}^{N_{kt}} \Delta\, Views_{it} + \beta_1\ ln \left(\sum_{i}^{N_{kt}} \frac{Likes_{it}}{Views_{it}} \Delta Views_{it}\right) +$$
$$\beta_2\ ln \left(\sum_{i}^{N_{kt}} \frac{Dislikes_{it}}{Views_{it}} \Delta Views_{it}\right) + \beta_3\ ln \left(\sum_{i}^{N_{kt}} \frac{Comments_{it}}{Views_{it}} \Delta Views_{it}\right) + \beta_4\ ln \left(\sum_{i}^{N_{kt}} \frac{Likes_{it}}{Dislikes_{it}} \Delta Views_{it}\right)$$
$$+ \varepsilon_{kt}$$

In consistent to Table 10, we extended the table containing our measures of feedbacks, that are going to be tested in the model estimations in Table 11.

Table 11: Audience reaction metrics in the model

		Valence		Audience Engagement
Method 1	Absolute terms	$\sum_{i}^{N_{kt}} \frac{Likes_{it}}{Views_{it}}$	$\sum_{i}^{N_{kt}} \frac{Dislikes_{it}}{Views_{it}}$	$\sum_{i}^{N_{kt}} \frac{Comments_{it}}{Views_{it}}$
	Relative terms	$\sum_{i}^{N_{kt}} \frac{Likes_{it}}{Dislikes_{it}}$		
Method 2	Absolute terms	$\sum_{i}^{N_{kt}} \left(\frac{Likes_{it}}{Views_{it}} \Delta Views_{it}\right)$	$\sum_{i}^{N_{kt}} \left(\frac{Dislikes_{it}}{Views_{it}} \Delta Views_{it}\right)$	$\sum_{i}^{N_{kt}} \left(\frac{Comments_{it}}{Views_{it}} \Delta Views_{it}\right)$
	Relative terms	$\sum_{i}^{N_{kt}} \left(\frac{Likes_{it}}{Dislikes_{it}} \Delta Views_{it}\right)$		

Source: own elaboration

7.5. Results

Based on the objectives we set to this chapter, and the methodology to achieve these goals we estimated four models. The results of these models examined the channels growth from different perspectives to answer our four main questions. First, we estimated the base model to find out the role of the channels' performance in their growths. Then, we extended this approach by the reach of the channels, to investigate the effect of the videos that have outstanding performances compare to the other videos of the channel. Finally, we extended this model into two directions, motivated by the different approaches for the same objective, examining the explanatory power of the audience reactions in the models. We summarized results of these estimated models in Table 12.

Analyzing the results of the first model, we can observe that the coefficient corresponding to the performance of the channels is significant. Therefore, based on the methodology behind this independent variable, we found evidence that the aggregated number of view count changes has a significant, positive impact on the growth of the channel. In other words, as the model indicates, we should reject the hypothesis of the coefficient being zero, and we can accept hypothesis 5, meaning, besides a unique performance independent element, we can observe performance dependent effects in the model.

The implication of this result is crucial for channels on this market. With the evidence of the presence of a performance dependent growth, we can also confirm the multiplication effect of the performance on the revenue of the channel. This process essentially shows that a higher performance leads to even higher performances through the follower base building of the channel. Moreover, since we accepted the hypothesis that the topic of the video has a positive effect on the performance of the video, product review channels should consider choosing topics that have high potential and can result multiplicative long-term benefits for the channels.

The second model was aimed to explore whether we can observe extra effects for channels that reach far on the market compared to the usual videos. Our results suggest that the top performing videos of the channels in fact have extra subscriber effect that can implicate that the reach of the videos are an important growth potential for channels. Thus, we accept our hypothesis, that as the channels has outstanding videos than their usual view counts, -on average- receive an extra number of subscribers compare to what our previous model would have suggested. The channels on the market and especially the small ones that did not have explosive videos yet, may derive the implication that it is worth to experiment with the content of the video, since a groundbreaking video's effect can outweigh the ones with poor performances. Hence, it could have an immense multiplicative effect on the future revenue. However, important to keep in mind that the valence of the videos could also matter in terms of the growth, which may prevent the overall positive resultant of the experimenting process.

Our last two models aimed to explore the connection between the audience reactions and the subscription growth of the channels. The two models are aimed to test two different approach about the possible relationship between the two variables. In consistent to the previous sections, we discuss the results corresponding to the average subscriber image approach first. This understanding of the process argues that the videos of the channel are the manifestations of the underlying properties of the channel. Hence, the framework behind this model assumes that the channels can be evaluated on the number of the likes, dislikes, and comments, without differentiating between the videos. Since we aimed to avoid spurious effects from the performance of the videos and the number of the videos of the channels, we reformulated this average number to an average feedback ratio. We denoted this method as the average subscribing image of the channel as it shows the unweighted average from the feedbacks towards the channel.

Our results indicate that we can explain a significant part of the variances in the growth process of the channels with the usage of the likes to views and the dislikes to views ratio on a 5% significance level. However, we have not found evidence that the number of comments or the like to dislike ratio would be related to our response variable. In term of the directions of the effects, we can conclude that the results meet our prior expectations, as we can observe a positive regression coefficient corresponding to the overall like ratio of the channel, while there is a negative coefficient for the overall dislike ratio.

Finally, we also tested the relationship between the audience reactions and the growth of the channels from a different perspective. In this approach we modelled the effect on the video contribution level. Our previous method explored the relationship between the variables using an indirect relationship through the image of the channels. In contrast, this approach assumes a direct relationship between the two variables by weighting the audience reaction metrics of the videos with their new view counts they received compared to the previous period.

After estimating the model, we did not find any evidence that this model extension would further explain the growth of the channels. Therefore, we can conclude that the average subscribing image approach proved to be better in estimating the connection between the audience reactions and the subscriber gathering process. More specifically, based on the information criteria of the models , it is also suggested, that representing the valence related variables, the likes and dislike ratios, we can achieve a better performing model than that of derived in Chapter 7.2. Hence, our final model regarding the new subscriber count of the channels for the next period contains a performance independent unique intercept, and independent variables of the performances of the videos, the reach of the videos, the like and the dislike ratio of the channel.

Table 12: Regression results for the growth of the channels

Regression Results
(4)

	Dependent variable:			
	ln ΔSubscriptions			
	(13)	(14)	(15)	(16)
Performance: $\ln \sum_{i=1}^{N_k} \Delta \text{Views}_{it}$	0.121*** (0.010)	0.138*** (0.016)	0.115*** (0.011)	0.136** (0.063)
Reach: $\Delta \text{Views}_{it}^{\overline{\text{Views}}}$		0.828*** (0.158)	0.820*** (0.158)	0.825*** (0.159)
METHOD 1:				
Likes: $\ln \frac{\sum_{i=1}^{N_k} \text{Likes}_{it}}{\sum_{i=1}^{N_k} \text{Views}_{it}}$			3.012** (1.487)	
Dislikes: $\ln \frac{\sum_{i=1}^{N_k} \text{Dislikes}_{it}}{\sum_{i=1}^{N_k} \text{Views}_{it}}$			-33.722** (14.088)	
Comments: $\ln \frac{\sum_{i=1}^{N_k} \text{Comments}_{it}}{\sum_{i=1}^{N_k} \text{Views}_{it}}$			1.072 (2.385)	
Like Ratio: $\ln \frac{\sum_{i=1}^{N_k} \text{Likes}_{it}}{\sum_{i=1}^{N_k} \text{Dislikes}_{it}}$			-0.028 (0.019)	
METHOD 2:				
Likes: $\ln \sum_{i=1}^{N_k} \frac{\text{Likes}_{it}}{\text{Views}_{it}} \Delta \text{Views}_{it}$				0.055 (0.053)
Dislikes: $\ln \sum_{i=1}^{N_k} \frac{\text{Dislikes}_{it}}{\text{Views}_{it}} \Delta \text{Views}_{it}$				-0.036 (0.045)
Comments: $\ln \sum_{i=1}^{N_k} \frac{\text{Comments}_{it}}{\text{Views}_{it}} \Delta \text{Views}_{it}$				-0.031 (0.020)
Like Ratio: $\ln \sum_{i=1}^{N_k} \frac{\text{Likes}_{it}}{\text{Dislikes}_{it}} \Delta \text{Views}_{it}$				-0.010 (0.057)
Constant	5.820*** (0.108)	5.634*** (0.152)	5.907*** (0.108)	5.696*** (0.205)
Random Effects				
Intercept/Channel				
Standard Deviation	0.1984	0.9727	0.1742	0.1745
Likelihood ratio	820.096***	958.137***	481.108***	454.778***
Observations	7,928	7,928	7,928	7,928
Log Likelihood	-6,146.188	-6,077.167	-6,128.066	-6,143.165
Akaike Inf. Crit.	12,300.380	12,166.330	12,274.130	12,304.330
Bayesian Inf. Crit.	12,328.290	12,208.200	12,336.930	12,367.130
Note:				*p<0.1; **p<0.05; ***p<0.01

Source: own elaboration

8. Conclusion

The dissertation was aimed to model the product review economy on YouTube. This model then provides valuable information about key elements about earned media for the firms that launched or intent to launch a product on the market.

The main direction of the dissertation can be connected to multiple domains in the literature. However, most of these literature streams has not approached this market from the perspective of the demand and supply of information.

The product review and consumer learning literature mostly investigate the effect of reviews on the demand of the consumers or the performance of the firm that launched the product. Another reference point for the dissertation is the personal branding literature, based on the argument that the supply of the market is essentially a set of individual reviewer self-brands. This domain thoroughly describes how the brand image is built up of these channels and explains how this image may translate into success. The only domain that examines the market around the information that an agent mediates for the audience is the literature examining the behavior of media firms and agents. However, these studies examine the decision of the agents with purely theoretical models, while our study aims to do so with empirically tested models. Moreover, these studies generally examine other aspects of the decision making of the information mediator, such as the objectivity, accuracy, political orientation, price, or programming variety of their content.

Thus, our framework is the first to empirically model the economy of product related information (on YouTube) in the marketing domain. Our broad objectives prior to the research were the following:

1. Explore the role of product related information in the reviewer market.
2. Identify the key characteristics of the demand and supply in the market.
3. Examine the relationship between these characteristics and the information "product", which is the video containing the information

Along these goals, the first part of the dissertation was aimed to define and identify the markets corresponding to information about different products on YouTube. We denoted the collection of the videos posted on the same topic, which is a new product on the market as an information market. Relating to this, the supply of the market consists of the channels who posted the content, while the demand comes from the audience that seeks for information. Building on this denotation, we were able to examine the baseline effect of the topic on the videos. Hence, we hypothesized that *H1-A: the product reviewed in the video has significant effect on the performance of the video, H1-B: this effect is decreasing over time.* Based on the model estimations, we found that we can accept both H1-A and B hypothesis on every common significance level. This implicates that our framework of information segmentation on the platform was supported by the data, which made us possible to further develop the model.

Hence, in the next chapters we examined the demand and supply on the market more thoroughly. In the previous part of the thesis we argued that the estimated effect of the topic could highlight the overall topic interest towards the demand for information about a certain product. However, this effect was estimated in a way that it represents the topic interest in an exogenous fashion. Thus, we argued that if we aim to examine the dynamics of the demand and supply of information on the market, we need to endogenize a part of this effect, while we should also keep an exogenous part, accounting for effects that comes from outside of the platform.

To endogenize this effect, on one hand, we relied on the information search and consumption literature, which showed how the individual information need evolves, how the audience becomes satiated over time. On the other hand, we also used arguments regarding the competition among channels and the topic awareness of the audience that is still interested towards the topic. Finally, we derived a weighting function in the model which can separate the views in the topic according to its recency. Based on the properties of this function, the most recent views represented a certain share of interested views while the views that happened earlier showed us a certain share of satiated views. Then, we optimized the properties of the function by iteratively changing both the form and the parameters of the function and estimating the model with the variables created by the function. Based on this setup, we had the following research questions and hypothesizes for this segment: *RQ1: What is the resultant of the potential positive and negative endogenous topic effects on the YouTube product reviewer market? RQ2: How can we separate the aggregate effect to represent the satiation and topic awareness of the consumers and the competition among channels? H2: Recent topic views have a positive, while the ones that happened earlier have a negative impact on the performance of the videos.* For RQ1, we found that the resultant of the effects is only significantly different from zero at a 10% confidence level. Regarding the function form, we observed both function form, the linear and the multiplicative inverse, to be significant. Based on the slight favor towards the multiplicative inverse function, we found that the optimal exponent of this function is 0.9. Then, using this weighting function, the estimated model have shown that both the satiation and the topic awareness effect are significant, having negative and positive coefficient, respectively. Thus, we accepted H2.

So far in the dissertations, our approach to the suppliers of the information could be described by a set of uniform, homogenous agent. In contrast, in the next parts, based on the personal branding literature, we aimed to resolve this assumption and account for individual heterogeneity across YouTubers. This also defined us the third level of our overall model. First, the video level measures, second, the product level, consisting the videos posted on the same topic by different creators and finally the channel level, which includes the videos posted by the same channel on various topics.

We considered two aspect which differentiates channels in terms of the performances of their videos, posted on the different information markets. First, the personal branding literature, second, the sizes of the channels. We tested the brand related elements of the model first. From the available literature we inferred that the brand images of the channels may have multiple different roles in the model of product review economy. First, it can act as a buffer for the performance of the videos of the channel. Meaning, it provides a fixed amount of views for the channels over time, so it is not dependent on other aspects of the model. In contrast, our second approach resolves this assumption and enables cross-effects with the topic effects defined before. Hence, the following hypothesizes were formulated: *H3-A: The unique channel characteristics have a significant effect on the performance of the videos. H3-B: The unique channel characteristics significantly differentiates the topic effects for the channels.* Our results supported both the buffer and the topic cross-effects, thus we accepted H3-A and B.

The other channel differentiating factor we examined is the sizes of the channels. This aspect relied on the literature of size dependent market power across firms or brands. To investigate the effect of this aspect on the performance of the videos we followed similar logic to that of brand images. Hence, we first assumed a direct relationship, representing the effect as an independent variable in the model. Second, we tested cross-effects between the topic effect as well. Here, we assumed that based on the subscriber counts, channels may moderate or boost the positive or negative effect of the current state in the market in terms of the demand or supply of information. From these arguments we tested the following hypothesizes: *H4-A: The number of subscribers of the channels has a significant impact on the performance of their videos. H4-B: The number of subscribers* *of the channel has a significant interaction effect with the topic effects in the model.* Our results have shown that both approaches are significant in the model, hence we can accept hypothesis 4-A and B.

From these effects, we obtained a model with the above described three layers: videos, topics, channels. However, channels when making decisions not only interested in short term benefits but rather to maximize their revenue on the long term as well. From this perspective channels may be more interested in building their follower base. This consideration also arises when we examine the correlation between the two aspects, the performances of the videos and the growth of the channel. In the previous segment we described the effect of subscriber count on the performance of the videos, here we consider the relationship in the other direction. In other words, we assume that the two aspects can be related through a process in which viewers eventually become subscribers. Therefore, the performances of the videos could translate to the growth of the channel, resulting long term benefits. If this connection is proven to be right, it has important implications for the channel as it highlights a potentially multiplicative benefits for channels. From this process, we can infer that as channel size grows, it positively impacts all the videos of the channel, which result higher growth rate, resulting a multiplicative process. These considerations motivated our second set of models, modelling the growth of the channels.

Besides our main objective in this segment, which is to examine the effect between the performance and the growth, we also aimed to investigate other drivers that can have important implications for the channels in terms of their growth. This extends our baseline framework into multiple directions. First, we argue that channels may achieve higher growth if they can reach the audience that is not familiar with their content. Motivated by this consideration, we derived a metric which was aimed to show whether outstanding videos of the channel provides extra benefits for them. Second, we also aimed to explain the phenomenon better by assuming that valence and audience engagement can be connected to the channels' growth. Here, we assumed and tested two different approaches. First, we tested the average subscribing image, which assumes that in the eyes of the audience the properties of the videos are the manifestations of the overall image of the brand. Therefore, we can aggregate the available feedback metric of the videos into an *average subscribing image* of the channel. These metrics are the likes to views, dislikes to views, comments to views and likes to dislikes. The other approach took a different path and rather than handling all the videos equal, it tried to explain the growth on the video contribution level. Hence, the main driver of this methodology was the number of new views the certain videos received compared to the previous period, weighted by the audience reaction metrics mentioned above. In conclusion, the hypothesizes outlined to this set of models were the following: *H5: The view count changes of the channels' videos has a significant positive effect on its subscriber number changes. H6: Outlier videos of the channel in terms of their view counts have significant positive extra effect on the subscriber number changes of the channel. H7: We can explain the channel growth better if we use the channels' average audience reaction metrics. H8: We can explain the channel growth better if we use the video contribution audience reaction metrics.* Our results unambiguously suggest that we can accept both hypothesis 5 and 6. However we can only partly accept hypothesis 7, as only the average likes per views and average dislikes per views have proven to be significant. Based on the results, we found no evidence, that the framework derived for hypothesis 8 would be appropriate to model the connection between the audience reactions coming to the videos and the growth of the channels.

References

Ahmed, M., Spagna, S., Huici, F., & Niccolini, S. (2013). A peek into the future: Predicting the evolution of popularity in user generated content. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 607-616).

Ahn, D. S., & Sarver, T. (2013). Preference for flexibility and random choice. *Econometrica*, *81*(1), 341-361.

Alnawas, I., & Altarifi, S. (2016). Exploring the role of brand identification and brand love in generating higher levels of brand loyalty. *Journal of Vacation Marketing*, *22*(2), 111-128.

Amato, L. H., & Amato, C. H. (2004). Firm size, strategic advantage, and profit rates in US retailing. *Journal of Retailing and Consumer Services*, *11*(3), 181-193.

Amato, L., & Wilder, R. P. (1985). The effects of firm size on profit rates in US manufacturing. *Southern Economic Journal*, 181-190.

Arrow, K. J. (1959). Rational choice functions and orderings. *Economica*, *26*(102), 121-127.

Barone, M. J., Miyazaki, A. D., & Taylor, K. A. (2000). The influence of cause-related marketing on consumer choice: does one good turn deserve another?. *Journal of the academy of marketing Science*, *28*(2), 248-262.

Basuroy S, Chatterjee S, & Ravid S.A. (2003). How critical are critical reviews? The box office effects of film critics, star power and budgets. Journal of Marketing 67(4): 103–117.

Basuroy, S., & Chatterjee, S. (2008). Fast and frequent: Investigating box office revenues of motion picture sequels. Journal of Business Research, 61(7), 798-803.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2014). Fitting linear mixed-effects models using lme4. *arXiv preprint arXiv:1406.5823*.

Battaggion, M. R., & Vaglio, A. (2015). Pin-ups and Journalists: A Model of Media Market with News and Entertainment. Journal of Media Economics, 28(4), 217–245. doi:10.1080/08997764.2015.1094078

Belenky, G., Wesensten, N. J., Thorne, D. R., Thomas, M. L., Sing, H. C., Redmond, D. P., ... & Balkin, T. J. (2003). Patterns of performance degradation and restoration during sleep restriction and subsequent recovery: A sleep dose-response study. *Journal of sleep research*, *12*(1), 1-12.

Bendisch, F., Larsen, G., & Trueman, M. (2013). Fame and fortune: a conceptual model of CEO brands. *European Journal of Marketing*.

Berger, J., Draganska, M., & Simonson, I. (2007). The influence of product variety on brand perception and choice. *Marketing Science*, *26*(4), 460-472.

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.

Bode, M. (2010). Showing doing. The art-science debate in a performative perspective. Journal of Consumer Behaviour, 9(2), 139–155. doi:10.1002/cb.310

Brown, S. J., Goetzmann, W., Ibbotson, R. G., & Ross, S. A. (1992). Survivorship bias in performance studies. *The Review of Financial Studies*, *5*(4), 553-580.

Burgess, J., & Green, J. (2018). *YouTube: Online video and participatory culture*. John Wiley & Sons.

Burke, K. E. (2017). Social Butterflies-How Social Media Influencers are the New Celebrity Endorsement (Doctoral dissertation, Virginia Tech).

Cha, M., Kwak, H., Rodriguez, P., Ahn, Y. Y., & Moon, S. (2009). Analyzing the video popularity characteristics of large-scale user generated content systems. *IEEE/ACM Transactions on networking*, *17*(5), 1357-1370.

Chen, Chih-Ping (2013) Exploring Personal Branding on YouTube, Journal of Internet Commerce, 12:4, 332-347

Chen, Y., & Xie, J. (2005). Third-party product review and firm marketing strategy. *Marketing science*, *24*(2), 218-240.

Cheng, X., Dale, C., & Liu, J. (2007). Understanding the characteristics of internet short video sharing: YouTube as a case study. *arXiv preprint arXiv:0707.3670*.

Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of marketing research*, *43*(3), 345-354.

Colicev, A., Malshe, A., Pauwels, K., & O'Connor, P. (2018). Improving consumer mindset metrics and shareholder value through social media: The different roles of owned and earned media. *Journal of Marketing*, *82*(1), 37-56.

Cox, J., & Kaimann, D. (2015). How do reviews from professional critics interact with other signals of product quality? Evidence from the video game industry. Journal of Consumer Behaviour, 14(6), 366-377.

Cox, J., & Kaimann, D. (2015). How do reviews from professional critics interact with other signals of product quality? Evidence from the video game industry. Journal of Consumer Behaviour, 14(6), 366-377.

Crane, R., & Sornette, D. (2008). Robust dynamic classes revealed by measuring the response function of a social system. *Proceedings of the National Academy of Sciences*, *105*(41), 15649-15653.

D'Adderio, L. (2008). The performativity of routines: Theorising the influence of artefacts and distributed agencies on routines dynamics. Research Policy, 37(5), 769–789. doi:10.1016/j. respol.2007.12.012

Debreu, G. (1954). Representation of a preference ordering by a numerical function. *Decision processes*, *3*, 159-165.

Dekel, E., Lipman, B. L., & Rustichini, A. (2001). Representing preferences with a unique subjective state space. *Econometrica*, *69*(4), 891-934.

Delgado-Ballester, E., & Munuera-Alemán, J. L. (2001). Brand trust in the context of consumer loyalty. *European Journal of marketing*.

Delisle, M. P., & Parmentier, M. A. (2016). Navigating person-branding in the fashion blogosphere. *Journal of Global Fashion Marketing*, *7*(3), 211-224.

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

Dion, D., & Arnould, E. (2016). Persona-fied brands: managing branded persons through persona. *Journal of Marketing Management*, *32*(1-2), 121-148.

Diwanji, P., Simon, B. P., Märki, M., Korkut, S., & Dornberger, R. (2014, November). Success factors of online learning videos. In *2014 International Conference on Interactive Mobile Communication Technologies and Learning (IMCL2014)* (pp. 125-132). IEEE.

Duffy, B. E., & Hund, E. (2015). "Having it all" on social media: Entrepreneurial femininity and self-branding among fashion bloggers. *Social Media+ Society*, *1*(2), 2056305115604337.

Durand, R., Rao, H., & Monin, P. (2007). Code and conduct in French cuisine: Impact of code changes on external evaluations. Strategic Management Journal, 28(5), 455–472. doi:10.1002/ smj.583

Eliashberg J, Shugan SM (1997) Film critics: Influencers or predictors? J. Marketing 61(2):68–78.

Erdem, T., & Keane, M. P. (1996). Decision-making under uncertainty: Capturing dynamic brand choice processes in turbulent consumer goods markets. *Marketing science*, *15*(1), 1-20.

Falkinger, J. (2007). *Attention economies. Journal of Economic Theory, 133(1), 266–294.* doi:10.1016/j.jet.2005.12.001

Feldman, M. S., & Pentland, B. T. (2003). Reconceptualizing organizational routines as a source of flexibility and change. Administrative Science Quarterly, 48(1), 94–118. doi:10.2307/3556620

Figueiredo, F. (2013, February). On the prediction of popularity of trends and hits for user generated videos. In *Proceedings of the sixth ACM international conference on Web search and data mining* (pp. 741-746).

Figueiredo, F., Almeida, J. M., Gonçalves, M. A., & Benevenuto, F. (2014). On the dynamics of social media popularity: A YouTube case study. *ACM Transactions on Internet Technology (TOIT)*, *14*(4), 1-23.

Figueiredo, F., Benevenuto, F., & Almeida, J. M. (2011, February). The tube over time: characterizing popularity growth of youtube videos. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 745-754).

Fletcher, R. (2013). *Practical methods of optimization*. John Wiley & Sons.

Fournier, S., & Eckhardt, G. M. (2019). Putting the person back in person-brands: understanding and managing the two-bodied brand. *Journal of Marketing Research*, *56*(4), 602-619.

Fox P (1997). The Port Mathematical Subroutine Library, Version 3. AT&T Bell Laboratories, Murray Hill, NJ.

Fox, P. A., Hall, A. P., & Schryer, N. L. (1978). The PORT mathematical subroutine library. *ACM Transactions on Mathematical Software (TOMS)*, *4*(2), 104-126.

Friedman, M., & Savage, L. J. (1948). The utility analysis of choices involving risk. *Journal of political Economy*, *56*(4), 279-304.

Gabszewicz, J. J., Laussel, D., & Sonnac, N. (2001). *Press advertising and the ascent of the "Pensée Unique." European Economic Review, 45(4-6), 641–651*. doi:10.1016/s0014-2921(01)00139-8

Gabszewicz, J. J., Laussel, D., & Sonnac, N. (2002). Press Advertising and the Political Differentiation of Newspapers. Journal of Public Economic Theory, 4(3), 317–334. doi:10.1111/1467-9779.00100

Gabszewicz, J. J., Laussel, D., & Sonnac, N. (2004). Programming and Advertising Competition in the Broadcasting Industry. Journal of Economics Management Strategy, 13(4), 657–669. doi:10.1111/j.1430-9134.2004.00027.x

Gal-Or, E., & Dukes, A. (2003). Minimum Differentiation in Commercial Media Markets. Journal of Economics Management Strategy, 12(3), 291–325. doi:10.1111/j.1430-9134.2003.00291.x

Gemser, G., Van Oostrum, M., & Leenders, M. A. (2007). 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.

Godes D, Ofek E, Sarvary M (2009) Content vs. advertising: The impact of competition on media firm strategy. Marketing Sci. 28(1):20–35.

Hall, J., Lockshin, L., & O'Mahony, G. B. (2001). Exploring the links between wine choice and dining occasions: Factors of influence. *International journal of wine marketing*.

He, Q. C., & Chen, Y. J. (2018). Dynamic pricing of electronic products with consumer reviews. *Omega*, *80*, 123-134.

Hennig-Thurau, T., Marchand, A., & Hiller, B. (2012). The relationship between reviewer judgments and motion picture success: re-analysis and extension. Journal of Cultural Economics, 36(3), 249-283.

Hewer, P., & Brownlie, D. (2013). Spaces of hope, enlivenment and entanglement: Explorations in the spatial logic of celebrity culinary brands. *Journal of Consumer Culture*, *13*(1), 46-63.

Hilger, J., Rafert, G., & Villas-Boas, S. (2011). Expert opinion and the demand for experience goods: an experimental approach in the retail wine market. *Review of Economics and Statistics*, *93*(4), 1289-1296.

Hox, J. J. (1995). *Applied multilevel analysis*. TT-publikaties.

Hu, N., Bose, I., Gao, Y., & Liu, L. (2011a). Manipulation in digital word-of-mouth: A reality check for book reviews. *Decision Support Systems*, *50*(3), 627-635.

Hu, N., Bose, I., Koh, N. S., & Liu, L. (2012). Manipulation of online reviews: An analysis of ratings, readability, and sentiments. *Decision support systems*, *52*(3), 674-684.

Hu, N., Liu, L., & Sambamurthy, V. (2011b). Fraud detection in online consumer reviews. *Decision Support Systems*, *50*(3), 614-626.

Iyengar, R., Ansari, A., & Gupta, S. (2007). A model of consumer learning for service quality and usage. *Journal of Marketing Research*, *44*(4), 529-544.

Kahneman, D., Slovic, S. P., Slovic, P., & Tversky, A. (Eds.). (1982). *Judgment under uncertainty: Heuristics and biases*. Cambridge university press.

Kalish, S. (1985). A new product adoption model with price, advertising, and uncertainty. *Management science*, *31*(12), 1569-1585.

Kerrigan, F., Brownlie, D., Hewer, P., & Daza-LeTouze, C. (2011). 'Spinning'Warhol: Celebrity brand theoretics and the logic of the celebrity brand. *Journal of Marketing Management*, *27*(13-14), 1504-1524.

Kim, K., Kevin Chung, and Noah Lim (2019), "Third-Party Reviews and Quality Provision," Management Science.

Kjellberg, H., & Helgesson, C.-F. (2006). Multiple versions of markets: Multiplicity and performativity in market practice. Industrial Marketing Management, 35, 839–855. doi:10.1016/j. indmarman.2006.05.011

Kreps, D. M. (1979). A representation theorem for" preference for flexibility". *Econometrica: Journal of the Econometric Society*, 565-577.

Krishnamurthi, L., & Raj, S. P. (1991). An empirical analysis of the relationship between brand loyalty and consumer price elasticity. *Marketing science*, *10*(2), 172-183.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. (2017). lmerTest package: tests in linear mixed effects models. *Journal of statistical software*, *82*(13), 1-26.

Lee, J. (2009). Does size matter in firm performance? Evidence from US public firms. *international Journal of the economics of Business*, *16*(2), 189-203.

Lee, J. E., & Watkins, B. (2016). YouTube vloggers' influence on consumer luxury brand perceptions and intentions. Journal of Business Research, 69(12), 5753-5760.

Li, C., Liu, J., & Ouyang, S. (2016). Characterizing and predicting the popularity of online videos. *IEEE Access*, *4*, 1630-1641.

Li, F., & Du, T. C. (2017). Maximizing micro-blog influence in online promotion. *Expert Systems with Applications*, *70*, 52-66.

Liikkanen, L. A., & Salovaara, A. (2015). Music on YouTube: User engagement with traditional, user-appropriated and derivative videos. *Computers in Human Behavior*, *50*, 108-124.

Lovett, M. J., & Staelin, R. (2016). The role of paid, earned, and owned media in building entertainment brands: Reminding, informing, and enhancing enjoyment. *Marketing Science*, *35*(1), 142-157.

Mackiewicz, J. (2008). Reviewer motivations, bias, and credibility in online reviews. In *Handbook of research on computer mediated communication* (pp. 252-266). IGI Global.

Mackiewicz, J. (2010). Assertions of expertise in online product reviews. *Journal of Business and Technical Communication*, *24*(1), 3-28.

Mahajan, V., Muller, E., & Bass, F. M. (1990). New product diffusion models in marketing: A review and directions for research. *Journal of marketing*, *54*(1), 1-26.

McQuarrie, E. F., Miller, J., & Phillips, B. J. (2013). The megaphone effect: Taste and audience in fashion blogging. *Journal of Consumer Research*, *40*(1), 136-158.

Melewar, T. C., Lim, L. L., & Petruzzellis, L. (2010). Mobile phone choice: technology versus marketing. The brand effect in the Italian market. *European Journal of marketing*.

Miller, R. A. (1984). Job matching and occupational choice. *Journal of Political economy*, *92*(6), 1086-1120.

Moulard, J. G., Garrity, C. P., & Rice, D. H. (2015). What makes a human brand authentic? Identifying the antecedents of celebrity authenticity. *Psychology & Marketing*, *32*(2), 173-186.

Mullainathan, Sendhil, and Andrei Shleifer. 2005. "The Market for News." American Economic Review 95 (4) (September): 1031–1053. doi:10.1257/0002828054825619.

Munnukka, J., Maity, D., Reinikainen, H., & Luoma-aho, V. (2019). "Thanks for watching". The effectiveness of YouTube vlogendorsements. *Computers in human behavior*, *93*, 226-234.

Murphy, P. (2010). The intractability of reputation: Media coverage as a complex system in the case of Martha Stewart. *Journal of Public Relations Research*, *22*(2), 209-237.

Narayanan, S., & Manchanda, P. (2009). Heterogeneous learning and the targeting of marketing communication for new products. *Marketing science*, *28*(3), 424-441.

Narayanan, S., Manchanda, P., & Chintagunta, P. K. (2005). Temporal differences in the role of marketing communication in new product categories. *Journal of Marketing Research*, *42*(3), 278-290.

Nardi, B. A., Schiano, D. J., & Gumbrecht, M. (2004b). Blogging as social activity, or, would you let 900 million people read your diary?. In *Proceedings of the 2004 ACM conference on Computer supported cooperative work* (pp. 222-231).

Nardi, B. A., Schiano, D. J., Gumbrecht, M., & Swartz, L. (2004a). Why we blog. *Communications of the ACM*, *47*(12), 41-46.

Nelder, J. A., & Mead, R. (1965). A simplex method for function minimization. *The computer journal*, *7*(4), 308-313.

Nelson, P. (1970). Information and consumer behavior. *Journal of political economy*, *78*(2), 311-329.

Newman D. (2004). [online] The Role Of Paid, Owned And Earned Media In Your Marketing Strategy: https://www.forbes.com/sites/danielnewman/2014/12/03/the-role-of-paid-owned-and-earned-media-in-your-marketing-strategy/?sh=6e93544b28bf

Niresh, A., & Thirunavukkarasu, V. (2014). Firm size and profitability: A study of listed manufacturing firms in Sri Lanka. *International journal of business and management*, *9*(4).

Oren, S. S., & Schwartz, R. G. (1988). Diffusion of new products in risk-sensitive markets. *Journal of Forecasting*, *7*(4), 273-287.

Parmentier, M. A., Fischer, E., & Reuber, A. R. (2013). Positioning person brands in established organizational fields. *Journal of the Academy of Marketing Science*, *41*(3), 373-387.

Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International journal of research in marketing*, *27*(2), 91-106.

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.

Reinstein, D.A., Snyder, C.M., 2005. The influence of expert reviews on consumer demand for experience goods: a case study of movie critics. J. Ind. Econ. 53 (1), 27–51.

Robbins, L. (2007). An essay on the nature and significance of economic science. Ludwig von Mises Institute.

Roberts, J. H., & Urban, G. L. (1988). Modeling multiattribute utility, risk, and belief dynamics for new consumer durable brand choice. *Management Science, 34*(2), 167-185.

Roos, J. M., Mela, C. F., & Shachar, R. (2013). Hyper-media search and consumption. Available at SSRN 2286000.

Scolere, L., Pruchniewska, U., & Duffy, B. E. (2018). Constructing the platform-specific self-brand: The labor of social media promotion. *Social Media+ Society*, *4*(3), 2056305118784768.

Shanno, D. F. (1970). Conditioning of quasi-Newton methods for function minimization. *Mathematics of computation*, *24*(111), 647-656.

Simon, H. (1979). Dynamics of price elasticity and brand life cycles: An empirical study. *Journal of Marketing Research*, *16*(4), 439-452.

Simon, H. A. (1959). Theories of decision-making in economics and behavioral science. *The American economic review*, *49*(3), 253-283.

Smith, A. N., & Fischer, E. (2020). Pay attention, please! Person brand building in organized online attention economies. Journal of the Academy of Marketing Science, 1-22.

Sokolova, K., & Kefi, H. (2020). Instagram and YouTube bloggers promote it, why should I buy? How credibility and parasocial interaction influence purchase intentions. Journal of Retailing and Consumer Services, 53.

Stephen, A. T., & Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of marketing research*, *49*(5), 624-639.

Stigler, G. J. (1961). The economics of information. *Journal of political economy*, *69*(3), 213-225.

Szabo, G., & Huberman, B. A. (2010). Predicting the popularity of online content. *Communications of the ACM*, *53*(8), 80-88.

Szymanowski, M., & Gijsbrechts, E. (2012). Consumption-based cross-brand learning: are private labels really private?. *Journal of Marketing Research*, *49*(2), 231-246.

Szymanowski, M., & Gijsbrechts, E. (2013). Patterns in consumption-based learning about brand quality for consumer packaged goods. *International Journal of Research in Marketing*, *30*(3), 219-235.

Tellis, Gerard and Joseph Johnson (2007), "The Value of Quality," Marketing Science, 26 (6), 758–73.

Terry, N., Butler, M., & De'Armond, D. A. (2011). The determinants of domestic box office performance in the motion picture industry. Southwestern Economic Review, 32, 137-148.

Thomson, M. (2006). Human brands: Investigating antecedents to consumers' strong attachments to celebrities. *Journal of marketing*, *70*(3), 104-119.

Tirunillai, S., & Tellis, G. J. (2012). Does chatter really matter? Dynamics of user-generated content and stock performance. *Marketing Science*, *31*(2), 198-215.

Train, K. E. (2009). *Discrete choice methods with simulation*. Cambridge university press.

Trzciński, T., & Rokita, P. (2017). Predicting popularity of online videos using support vector regression. *IEEE Transactions on Multimedia*, *19*(11), 2561-2570.

Welbourne, D. J., & Grant, W. J. (2016). Science communication on YouTube: Factors that affect channel and video popularity. *Public understanding of science*, *25*(6), 706-718.

Wu, C., Che, H., Chan, T. Y., & Lu, X. (2015). The economic value of online reviews. *Marketing Science*, *34*(5), 739-754.

Xiang Y, Sarvary M (2007) News consumption and media bias. Marketing Sci. 26(5):611–628

Xiang, Y., & Soberman, D. (2014). Consumer favorites and the design of news. *Management Science*, *60*(1), 188-205.

Yang, J., & Leskovec, J. (2011, February). Patterns of temporal variation in online media. In *Proceedings of the fourth ACM international conference on Web search and data mining* (pp. 177-186).

Yoganarasimhan, H. (2012). Impact of social network structure on content propagation: A study using YouTube data. *Quantitative Marketing and Economics*, *10*(1), 111-150.

Zhao, Y., Yang, S., Narayan, V., & Zhao, Y. (2013). Modeling consumer learning from online product reviews. *Marketing Science*, *32*(1), 153-169.

Zhao, Y., Zhao, Y., & Helsen, K. (2011). Consumer learning in a turbulent market environment: Modeling consumer choice dynamics after a product-harm crisis. *Journal of Marketing Research*, *48*(2), 255-267.