DOCTORAL DISSERTATION

Gábor Murai

Pécs, 2022

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, 2022

To Cintia

Contents

List of Tables

Abstract

Despite the role of product related information in new product launches, our knowledge about its demand and supply in 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. The 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 is diminishing over time.

Then, we were able to formulate hypotheses 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 endogenized the overall interest towards the topic into a current state of satiation and topic awareness. The results indicate that both measures have a significant relationship with the performance of the videos, having positive and negative coefficients, 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. We found that the size of the channels has a significant positive impact on the performance of the videos, while it has a negative effect on the above-defined satiation and topic awareness. Our results suggest that the unobserved factors related to the image of the brand also significantly differentiate 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. The main hypothesis regarding these models was whether the performance of the videos translates into subscriber counts. We found that the performance positively influences the subscriber gaining process, and outstanding videos provide extra effects for this growth. In addition, we tested if the video-level reactions from the audiences can be related to the process and found that the average ratio of likes to views and dislikes to views are significant predictors of the subscriber count changes over time.

Acknowledgments

First and foremost, I would like to thank my entire family for always supporting and believing in me, which was a powerful motivation throughout these years. I would like to thank my Mom for all her effort to make sure I can follow this path, and I am particularly thankful to my wonderful and supportive wife, Cinti, who always gave me strength and advice during these years, which made it possible to complete this dissertation.

Second, I would like to thank all the members of the Faculty of Business and Economics, University of Pécs. I would like to especially thank János Barancsuk for being my mentor during the time I spent in Pécs and for raising my interest in microeconomics, which was the first stepping stone on this path. I am also grateful to my friends from the doctoral school, Olivér Kovács, Erik Braun, András Gyimesi, Zita Iloskics, Kristóf Németh, and Miklós Váry, for all the great memories I am keeping from these years and all the good advice they gave me regarding my research. I owe special thanks to László Szerb and Attila Varga for guiding me through the Ph.D. process and helping me achieve this destination.

I would like to also thank all the members of the Rotterdam School of Management for welcoming me into an open and kind environment during the time I visited them, which made it possible to start this research. I would like to especially thank Maciej Szymanowski for all the inspiration and both professional and personal advice he provided me.

Finally, I am also grateful to my two referees, Ildikó Kemény and Péter Németh. Their highly valuable reviews helped me to greatly improve this dissertation.

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 types 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 and experts. While the firms essentially have a very high-level control over 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 performance 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 to 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 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. The information coming from peers who have already bought the item or service now can be reached by almost anyone in the world in various forms. Nowadays, most online ecommerce platforms have a segment for user feedbacks, but there are also websites dedicated completely to such reviews.

However, information about the product can come not only from peers but also from third-party professionals or experts. In the case of this type of review, we can observe that the domain has changed just as much as that of peer reviews. First, the traditional magazine or newspaper segments of product reviews have moved to blogs and websites. Then, with the emergence of organized online attention platforms, such as YouTube, the professional or expert reviews evolved into the complex ecosystem that we can observe today. While the role of blogs and websites remained meaningful, the system, 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 the traditional newspaper or magazine model more. The centralized supply provides easier access to information from more sources for consumers. Meanwhile, the properties of the platform make the entry into the market accessible for anyone who aims 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 peer reviews and their effect on the consumers and firms is well-documented in the marketing domain. However, our understanding of the expert review ecosystem is very limited in general, while the knowledge regarding the modern shared platform reviewer market is especially scarce. Therefore, in this dissertation, we aim to fill this gap in the literature and shed light on the main drivers of this complex market. In addition, we chose 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 serve as a background for our research.

eWOM

Given that the content of the videos is essentially product reviews, the dissertation connects to the literature examining electronic word-of-mouth (eWOM), which is a piece of online information about a given product or firm from a company independent source (Hennig-Thurau et al., 2003).

The main literature streams in this domain that relate to our focal topic are examining the expertise, credibility, trustworthiness, and motivation of the eWOM poster and the impact of the information on the economic performance of the product or firm. Based on the literature on expertise, we can divide product reviews into peer and expert reviews. This helps us to understand the position and special features of our focus, expert reviews on YouTube.

The findings of this domain had many implications for the dissertation. For instance, it describes the evolution of the individual and the 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. Moreover, the findings in this domain (both peer and expert reviews) have shown that reviews, in general, play a crucial role in the consumers' quality perception and expectation about goods with uncertain properties.

Demand for Media Content

Expert reviews distributed on YouTube are different from the more traditional eWOM categories as the demand for these reviews could be driven by similar motivations to the ones behind media content watching.

Therefore, the examination of the demand for reviews in this market essentially leads us to the literature on the demand for media content. Combining these motives with the demand for product related information described above, we can identify product and non-product related motivations to watch the content creator generated product review videos. Among others, non-product related motives include the entertainment, social utility, or mood management needs of the audience.

We rely on this domain when we extend our spectrum with the examination of the supply and suppliers of the reviews. We keep in mind that the reviewer who aims to maximize the viewership of the videos is facing a demand that is not only related to the presented product, in which case the informativeness and credibility are the main defining factors, but it is also driven by non-product related needs, such as entertainment or social needs to interact with other people.

Personal Branding

Online personal branding has been one of the trending topics in the marketing literature in recent years. The main connection point here is the argument that YouTube product review channels are creating, building, and managing their own brands as it is defined by this domain. The unique, defining aspect of these brands is that it is built around an individual whose face is the brand itself. For instance, we can mention brands around popular figures such as Gordon Ramsey, LeBron James, or Calvin Klein.

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 consider being connected to the profession of the brand. To be successful in the market, channels need to appropriately merge different persona facets, features into a brand image and narrative.

Therefore, in the dissertation, we consider these YouTube channels similar to brands in other industries and develop our model to account for the differences among the channels' brand images in attracting product and non-product related demand.

Behavior of the media

In contrast to the studies in the field of eWOM and product reviews, the dissertation aims to model the market of reviews itself, and not only a specific aspect of the information, such as its credibility or impact on the consumers or the firms. Besides the demand for reviews described above, the supply side of the market consists of the information suppliers with their own incentives to post the reviews. While there is a literature stream in the eWOM domain that examines the motives behind the posting decision of peer reviews, expert reviewers on YouTube have different, financial incentives to provide these reviews.

From this perspective, the incentives of the reviewers resemble the goals of media firms and agents who aim to maximize their viewership. In this way, the dissertation connects to the theoretical literature on the behavior of the media. However, the framework of these studies is different compared 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 applicable to our model. However, perhaps the most important difference lies in 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 still 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 Soberman (2014). In their framework, they assume that news providers try to maximize the ex ante expected audience size to achieve the optimum revenue. This also means that they have a fixed rate per viewer advertising and content revenue. This domain also serves as a background to the dissertation's final set of models, which examines the long term growth of the channels through their incentives to maximize their revenue coming from the viewership of all the videos.

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 apply 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 videos when we estimate our model. Here, we can build on the studies that have already examined the evolution of the views of the videos on YouTube.

The other important aspect of our chosen segment is the video format. The studies in this domain pointed out that consumers 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

The first building block of our approach to model the product reviewer economy is the product related information aspect of the reviews. Thus, our model development process starts with the definition and identification of the information markets on the platform that can be linked to new products on the market. Based on the volume of both new products and product review videos, we selected the smartphone industry to estimate our models. We define an information market on YouTube as the collection of 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 contents. Therefore, we can measure the overall demand for information by the number of views the information content received in the past.

Using these definitions, we first examine the relationship between the aspect of the videos that it contains product related information, and the demand coming from the audience. Thus, we hypothesize that the above described segmentation of the platform to different information markets is indeed significant, which also establishes the baseline model for the product review economy in the dissertation.

More specifically, we investigate whether the topic of the video, which is the reviewed product, significantly impacts the demand for the video, denoted by the view count changes from one period to another. However, from the studies on new product diffusion processes and consumer learning, we also know that as the uncertainty of the consumers decreases, the demand for information decreases as well. Hence, we not only test the presence of the effect on the topic but also expect it to decrease over time.

H1:

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

B: The product's effect on the video's performance is decreasing over time.

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

Given the satiation of consumers, we can also infer that the YouTube channels on the supply side of the market may face a limited demand from the audience, which could induce competition among them. Overall, this effect would suggest a negative relationship between the performance of the videos reviewing the same product. However, we also argue that the increasing viewership of competitor video(s) may also increase the performance of the focal video. First, competitor product reviewers may bring new audience members to the market, which induces a spill-over effect for other videos on the same topic. Second, the potential viewers may use the aggregate demand and/or the number of videos on the topic as a signal of topic attractiveness. Third, as the viewership of the topic increases, the potential viewers may enter the market to have a sense of belonging to the social network and don't miss out on something important. Either way, these effects essentially raise the pool of audience that is not satiated yet. Instead, they are still aware and following up on the topic.

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 is already satiated. With this function, we are able to introduce the current probabilistic state of satiation and topic awareness into the model. Therefore, we hypothesize the following statement regarding the endogenized topic interest:

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 level 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 modeled the relationship between the video and topic level, but we did not account for the channel level. Thus, in the following, we shift our focus from the demand corresponding to a specific topic to the supply side of the market and examine the reviewers' role in the market.

As we outlined in the previous section, the personal branding literature shows us that we should not handle the supply on the market as a set of homogenous actors, as there could be heterogeneity in the channels' capability to attract product and non-product related demand as well. First, corresponding to the non-product related demand, we can approach the reviewers' brands as a buffer in terms of the performance of their videos. In other words, the channels with more attractive brand images have a competitive advantage compared to other channels. This means that consumers may have a perception about the channels' capability to fulfill their non-product related needs, and they take this consideration into account when they choose a video to watch. Second, corresponding to the product related demand, we also test the possibility that the brand image is not independent of the topic effects in the model. Meaning, that we test whether a channel with a more attractive brand image has a different relation to the topic information market than channels with a 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 differentiate the topic effects for the channels*

The other differentiating factor among channels that we account for to resolve the homogenous supply assumption on the market is the aspect that they have different sizes. Here, we are building on studies examining size dependent market power and the possibilities of being a "niche" topic creator. Similarly to the previous differentiation, we test the size's effect from two perspectives. It may be a separate buffer to the performance of the videos, but it can also alter the relations that are already present in the model. As an example, we may expect that bigger channels can facilitate the topic awareness effect better and benefit more from a trending topic. Since we denote the size of the channels with the number of subscribers they have at the moment, we outlined the following hypotheses:

H4:

A: The channel's subscriber count has a significant impact on the performance of the videos.

B: The channel's subscriber count has a significant interaction effect with the topic effects in the model.

The result of Hypothesis 4A could have another important implication for channels as it could lead to a multiplicative relationship between the size of the channels and the performance of the videos of the channel. The process in which this can work relies on the argument that the relationship outlined in this hypothesis could be present in both directions.

If we find evidence that not only the channel size affects the views of the videos, but the views could also translate into subscribers, the channel size has a multiplicative effect on the revenue of the channels. In this process, the channel size affects the number of views the videos receive, then the views translate into subscribers, which causes an even higher number of views in long term. Due to this potential connection and long-term incentives of the channel to maximize their revenue, our second set of models is 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 in 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 attempt to explain this process by using the reactions from the audience towards the videos to have a better understanding of the role of valence and audience engagement in the growth of the YouTubers. Here, we use two different methodologies, representing different consideration processes 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 audience reactions across the 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 considers the role of valence and audience engagement in the growth process on a video contribution level. This implies a more direct effect between a better-perceived video and the subscriber number of the channel, meaning that if a new video is perceived better by the audience than the average perception of other videos, the channel will experience a spike in the subscriber count, while with the previous approach, the impact is much more conservative.

Finally, 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 have a significant positive effect on the subscriber number change of the channel.

H6: The videos with outstanding view counts compared to the channel's other videos have a significant additional positive 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 video contribution audience reaction metrics.

We summarized the hypotheses 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 background theories from the related disciplines. This includes the literature on earned media and eWOM, the demand for media content, para-social interaction, personal branding, and finally, the domain of modeling the behavior of news firms and agents.

The third chapter presents the technical and methodological approach of the dissertation. First, this includes the description of the data collection procedure. Second, it consists of the definitions of different time dimensions in the obtained dataset. Finally, the chapter describes the methodology of hierarchical modeling, which will serve as a baseline model approach to estimate the model in the dissertation.

The fourth chapter starts with exploring the dataset with the goal of developing the baseline model for the following chapters. Then, we derive the first hypothesis about the presence of information markets in the market. The role of this question is crucial for the dissertation as it makes the ground for all other hypotheses regarding the performance of the videos.

The fifth chapter emerges from the question of whether we can endogenize the topic interest of the audience. To achieve this, we use arguments building on the satiation of the consumers, the competition among product reviewers, and possible topic interest increasing ability of the content creators.

The sixth chapter is the last chapter modeling the view counts of the videos. The main motivation of this chapter mainly comes from the consideration 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 product reviews are essentially different brands on the market. Second, we account for the fact that the channels are different in their size. In addition, this chapter also establishes the base question for the next chapter as it describes how the channel size affects the performance of the videos, which relationship could be present the other way around as well.

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 deepen the understanding of the connection between the subscription and the view counts and examine the possibility of multiplicative growth in the subscription gathering process of the channels. The main question in the chapter is whether we can explain the growth of the channels by the performance of the videos. Notwithstanding, we also aim to extend this framework with the audience reactions to the videos to explore the role of valence and audience engagement in the YouTube channels' growth process.

Table 1: Structure of the dissertation

Category	Chap-ter	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	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. Background

The second main chapter of the dissertation aims to explore the background literature behind our research. From this perspective, we can divide the following chapters into three parts and describe the background theories for each part separately. First, in chapter 4, we build our baseline model and define the premise of the dissertation, that information markets corresponding to a specific reviewed product can be identified on YouTube. Then, in chapter 5, we extend the baseline model and endogenize the interest coming from the audience towards the shared topic on the information market. Thus, connecting to these chapters, we first outline the main literature streams in the domain examining product related information. Chapter 6 shifts the focus from the demand for information on the market to the supply of information and examines the creators of the product review videos as personal brands providing media content. Hence, after the demand for product related information, we also outline the background for this chapter. This includes the literature on demand for media content, para-social interaction, and personal branding. Finally, chapter 7 examines the market from the perspective of the (long term) incentives of the reviewers, which is the revenue coming from the sum of all product review videos across multiple topics and their long-term follower count growth process. Therefore, the final part of this chapter outlines the background for this chapter by examining studies on the incentives of the reviewers.

2.1. *Demand for Product Related Information*

2.1.1. Role in Decision-making

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 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 consumers' choice 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, 1954). 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 implies that they cannot be sure about the optimality of their decision prior to the decision-making. For instance, such case can arise in 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 do not have enough (trusted) information about its quality. Other examples 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 example, 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 the decision. Consumers may have prior expectations about the unknown products before being exposed to any kind of information regarding the given product from a variety of sources, for instance, from 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 as 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 if the product is unknown for the consumers, their uncertainty cannot be reduced to zero until they gain first-hand experience. Thus, the additional information has decreasing benefits for the consumers, which property is closely related to the information search literature (Nelson, 1970; Stigler, 1961; Roos et al., 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 the information pieces into their expectations with different weights. The weights could be the manifestations of the consumers' preference over the difference in the information they are receiving. For example, it could be induced by the difference in the consumers' trust and credibility (Hu et al., 2011a, 2011b, 2012; Zhao et al., 2013) towards a product review, discussed in chapter 2.1.2.1.

One of the most important aspects 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 firm's website, 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 information senders or distributors, such as reviews, mentions, or ratings from the users of the products.

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 relevant earned media literature on which the dissertation is built, including the expertise, credibility and trust of the information source and the posting decision of the content creator as well.

2.1.2. eWOM and Expert Reviews

In the previous chapter, we narrowed down the focus on the product related information aspect of the expert review market to the information sources that are earned for the focal firm or product. Thus, in this chapter, we extend the background of the dissertation with studies from this domain. More specifically, starting with the broad category of Word-of-Mouth (WOM) and electronic Word-of-Mouth (eWOM), we are going to outline the position of online expert reviews in the literature of earned media and present the main literature streams from this domain that relates to our research.

In terms of the definition of WOM, one of the first interpretations comes from Arndt (1967), who described it as an *"oral, person-to-person communication between a receiver and a communicator whom the receiver perceives as noncommercial, regarding a brand, a product, or a service."* (Arndt, 1967, p.1967). In the following 50 years, many iterations, extensions, and modification have been proposed to this definition, regarding for instance, the specification of the content of the message or the relationship between the sender and the receiver. A detailed summary of these terminology proposals can be found in Markos-Kujbus's (2017) and Thao and Shurong's (2020) research.

Along the changing environment around WOM during these years, the brand and product related communication continuously changed as well. However, the most radical change started with the appearance of web 2.0, which completely reformed the way people communicate and interact (Thao and Shurong, 2020). As a result, these changes greatly impacted the brand and product related communication as well, shifting it to multiple different online platforms. The special type of communication format about products or brands which emerged this way is called the electronic Word-of-Mouth or eWOM, which consists of the focal topic in the dissertation, online product reviews.

With the emergence of eWOM, Hennig-Thurau et al. (2003) proposed a (new) definition, which became the most widespread definition for this type of communication. According to their approach, eWOM is *"any positive or negative statement made by potential, actual, or former customers about a product or company, which is made available to a multitude of people and institutions via the Internet."* (Hennig-Thurau et al., 2003, p.39). This definition shows that eWOM is essentially a broader category than product reviews, including every statement about a given product or company which does not necessary have reviewing purposes. Thus, in the following sections, we are going to further narrow our focus to studies from this domain where the subject is some kind of information aimed to review a certain product.

eWOM and product reviews can be categorized from multiple perspectives (Markos-Kujbus, 2017), however, from the perspective of the expert review market, the most relevant categorization is the one that distinguishes the information based on the expertise of the sender.

Based on the expertise, we can distinguish two types of product reviews: peer and professional or expert reviews (Markos-Kujbus, 2017, Smith et al., 2005, Keh and Sun, 2018, Parikh et al., 2016, Choi and Ok, 2011, Naujoks and Benkenstein, 2020). Parikh et al. 2016 also define a third, semi-professional category, however, in most studies, professional and semi-professional reviewers are studied together as opposed to peer reviews.

In terms of the criteria for being considered as an *"expert"*, there are multiple approaches in the literature. Naujoks and Benkenstein (2020) define experts as the reviewers who write their reviews as part of their profession, such as editors or critics. Similarly, Keh and Sun's (2018) study refers to expert reviews if it is professionally made by a third party and highlight that the difference between peer and expert reviews stems from the difference in domain expertise (Spence and Brucks 1997).

Another popular approach in the literature is to use the number of reviews and/or different expertise badges on the reviewer platforms to identify the expert reviewers from the overall population of reviewers. (Tan et al. 2008, Racherla and Friske, 2012, Park and Nicolau, 2015, Lo and Yao, 2019, Liu and Park, 2015, Filieri et al. 2018)

Parikh et al. (2016) define a separate category for semi-professional reviews. However, they do not define clear distinction criteria for the three groups. Instead, they give specific examples for each category, such as Consumer Reports magazines and professional restaurant critics to professional, Zagat reviews, containing numerical and textual information about restaurants for semi-professional and Yelp.com for peer reviews.

However, important to note that the changing environment, especially the widespread usage of the internet, not only reformed the peer review ecosystem by the emergence of online platforms collecting the reviews and by the integration of reviewer functionalities to most shopping websites, but also greatly impacted the expert review market. Thus, previous examples for the different categories become increasingly less relevant. In the offline era, professional reviews were first either a separate or part of printed media. Then, TV and radio stations had segments dedicated to these professionals. Examples of this kind of professional reviews could be book, movie, or museum review sections in the magazine “The World Today”[1], “It Its Innovation (i3)”[2] magazine by the Consumer Technology Association (CTA) or the popular TV show ”Top Gear”[3], focusing on reviewing primarily motor vehicles.

The offline publishing or broadcasting also meant, that becoming a professional reviewer had high entry costs, and it was not something that anyone can immediately start to pursue. This barrier has changed with the internet. Some of the offline media containing expert reviews has launched an online extension, while others fully moved to an online format. However, 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 previous printed media, such as “goodhousekeeping.com”, reviewing housekeeping appliances, or “expertreviews.co.uk”, which is a collection of reviews on a few different product categories. An example of a website that did not have a prior offline media could be “GSMArena.com”, which will be one of our information sources during the data collection procedure as well.

The professional review market has developed even further in the recent decade with the widespread usage of social media and organized online attention platforms, such

[1] https://www.chathamhouse.org/publications/the-world-today
[2] https://cta.tech/Resources/i3-Magazine
[3] https://www.topgear.com/

as YouTube (Smith, 2020). Essentially, these websites 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, which makes the entry into the market even easier for anyone aiming to pursue a carrier in this expertise. However, it could also be 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 information sources (e.g., user rating, advertisements, etc.) 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 on 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 information, and what are the incentives of the reviewers on the market.

Therefore, in the following sub-sections, we are going to present the most relevant available research streams from the domain of eWOM and product reviews, discussing questions that serve as a background for the research. First, we describe the literature on credibility and trust towards reviews and reviewers, which literature aims to answer the following questions: 1. “Do consumers find expert reviews more credible, and trustworthy than peer reviews?” 2. “How credibility and trust affect the consumers' decision?”. Then, we briefly discuss the incentives behind posting a peer review. Important to note that our focus in this dissertation is the market of expert reviews, where the reviewers have financial motives to post the reviews. Due to this nature, we leave the discussion of the related literature to the incentives of the reviewers examined in the dissertation to chapter 2.3.2. The last stream of literature in this chapter is the richest in this domain and examines the impact of the reviews with also answering why examining earned media is important for the firms and how consumers use product related information in their perception about the focal product.

Throughout chapter, we are discussing papers focusing on any type of eWOM. However, where it is possible, we make a distinction between the findings of papers about peer and expert reviews and in the case of credibility, we specifically highlight the importance of expertise in the literature.

2.1.2.1. Credibility and Trust

In this chapter, we describe the literature on one of the most important differentiating factors of expert reviews, its credibility. Essentially, we can divide the chapter into the examination of two research questions. While these two examined relationships are different in terms of cause and effect, if we take them together, we can understand why credibility and trust are especially important for the expert review market.

First, we should discuss the findings on how reviewer expertise affects the perceived credibility and trustworthiness of the product reviews. Second, we need to confirm whether the credibility and trustworthiness of the reviews actually affect the consumers' decision making process regarding the reviewed product. We can see that by combining the two directions, due to a potential difference in the credibility, expertise, and trust between peer and expert reviews, we may observe a difference in the impact of the review on the consumers as well, which is a key aspect for the firms.

Examining the difference between peer and expert reviews, we can see that the literature is contradictory in this question (and also very dependent on the settings of the study from the perspective of the definition of *"expert"* and on the platform where the expert reviews are posted).

Huang and Chen (2006) conducted an experiment where participants had to select books from an online bookstore to investigate the difference in recommendation sources. They found that recommendations from other consumers influenced the participants' decisions more than the recommendation from an expert.

In contrast, Racherla and Friske (2012) examined Yelp reviews and found that expert reviews are more useful for consumers than peer reviews. In their study, they used the number of posted reviews as an indicator of expertise.

There are also studies showing the relationship to both directions. Smith et al. (2005) conducted a survey-based research, where expert recommendations and peer reviews were also presented, both in a textual format. They found that the relationship is dependent on the goal of the subject. If it was utilitarian in nature, then expert, while in the case of hedonic underlying motivations, peer reviews were more trustworthy. Similarly to Racherla and Friske (2012)'s study, Keh and Sun (2018) also examined Yelp reviews and defined experts with the same approach, using the number of reviews as an indicator. However, they collected the data about reviews on different services. They found that if the underlying reviewed service was experience service, then peer reviews, while if it was a credence service, expert reviews turned out to be more important for consumers.

In addition to this research question, it is also important to confirm if credibility and trustworthiness indeed affect the consumers' decision-making. While Cheung et al. (2008) did not find a connection, Filieri (2016), Filieri et al. (2018), and Sussman and Siegal (2003) confirmed a significant positive relationship between the two aspects of product reviews. Thus, we can infer that as credibility and trust increase regarding a given review, it becomes more and more useful and valuable for the consumer.

2.1.2.2. Posting Decision

The research on the incentives behind the posting decisions of peer reviewers is a relatively small domain. Only a handful of studies examined the users' motives behind their expression to share their opinion and perception about the product (Westbrook 1987; Nardi et al. 2004a, 2004b; Mackiewicz, 2008, 2010).

Westbrook (1987) describes the motivation to spread WOM as a result of an inner tension due to the positive and negative feelings regarding the post-purchase experience of a product. Mackiewicz (2008, 2010) discusses three potential drivers behind the consumers' decision to take an effort and express their opinion about the product. First, consumers may see the reviews as beneficial for them because they seek a sense of efficacy. According to this reasoning, they may write the 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 to make better decisions, and driven by this goal, it is worth it for them to take the time and effort to write these reviews. Finally, a possible explanation could be the human tendency to crave attention and need to be heard.

In conclusion, we can assume that users usually buy the products for their own usage, and then they share their opinion about them offline by WOM or online by eWOM in forms such as mentions, recommendations, or 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 or the feeling that they have affected the world somehow.

In contrast to the incentives of peer review posting, in this dissertation, we are focusing on professional reviewers with monetary incentives. Hence, we model a market where the providers of the reviews are essentially posting them in a similar fashion as a firm providing a product or service. From this perspective, the motivations and incentives of the reviewers are more similar to the motivations of different media firms.

The most closely related studies from this aspect explore the behavior of media firms, news providers, and other media agents aiming to attract the attention of the audience. As this aspect of the reviewers is not connected to product related information, we leave the discussion of it to chapter 2.3.2, where we approach the supply on the information market as media content providers. However, important to note that these studies are examining the behavior of the media agents with theoretical concepts, while this thesis has different objectives. We aim to use empirical data to look for evidence for the questions and hypotheses arising from the identified gap in the literature regarding the demand and supply of product related information.

2.1.2.3. Impact of the Information

Product related information generated by peers has a wide range of forms, 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 various information pieces from peers could have an immense effect on the perception of the consumers who are still uncertain about the product(s) on the market. In addition, studies on this domain also outline how crucial it is for firms to understand the nature of this information market (Dellarocas et al., 2007; Chevalier and Mayzlin, 2006; Zhao et al., 2013; Wu et al., 2015), which also implicates the importance of examining the expert review market. We describe the most important findings related to these aspects of our research more thoroughly below.

Zhao et al. (2013) modeled consumer learning from both their own experience with the same genre (books) and learning from online reviews. Their results show 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 have 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 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.

Finally, reflecting on these results, He and Chen (2018) derived an optimal pricing strategy for the firms assuming 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 researchers to examine how uncertainty regarding a product evolves over time on an individual level by incorporating additional information pieces from various sources. Moreover, it also enables to investigate the already discussed credibility (Chapter 2.1.2.1) corresponding to the reviews (Zhao et al., 2013). For instance, this can be examined with a question if consumers indeed acknowledge the fact that there could be fake reviews among the total population of available reviews.

Essentially, the learning process happens by consumers updating their expectation about the product and uncertainty regarding this expectation after using more and more information about the products from others. While the learning process itself is not part of the core background literature of the dissertation, we are going to still leverage on these studies in Chapter 5, where we form assumptions regarding the audience's topic interest.

In contrast to peer reviews, the literature on professional or expert reviews is a relatively small in the marketing domain and it examines the reviews using 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; Hennig-Thurau et al., 2012; Terry et al., 2011), while Cox and Kaimann (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 the 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.

Concluding the literature on eWOM, in previous sections, we highlighted the most relevant available literature streams discussing important aspects of the product related demand in the information market. Among others, these studies focus on the credibility and trust towards the reviews, the incentives behind peer reviews, and the economic impact of the reviews (such as sales, market value, or purchase intention). However, crucially for us, these studies do not examine the demand and 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 and the incentives of the experts providing this information. Finally, we estimate the model using data collected from YouTube, which is currently an emerging platform for product related information.

2.2. Demand for Media Content

2.2.1. Uses and Gratifications of Expert Reviews

In the previous chapter, we discussed how the expert reviews shared on YouTube are related to product related information literature, including both peer and expert reviews. However, important to notice that the reviews are not only serving product related information needs, but they are also popular media content.

Hence, the fact that the reviews are devoted to a particular product does not imply that learning about the product is the only reason why viewers watch them, as media consumers can also watch it due to various other motivations.

Katz et al. (1973a) conducted a survey-based study to examine the underlying motives of media usage. They uncovered 35 needs, which they then grouped into five categories:

- Cognitive needs, such as information seeking
- Affective needs, such as emotional need
- Integrative needs, such as credibility
- Social needs to make connections
- Relaxing needs

Similarly, McQuail (1983) described four main reasons to watch media content: information, personal identity, integration and social interaction, and entertainment.

As the nature of the above categories suggests, nowadays, the examination of media content usage is not only an interest of economic and business domain but also an interesting literature stream in psychology, socialpsychology, sociology, and decision theory domain (Gálik-Csordás, 2020).

The needs and motives, in general, could be dependent on 1. the type of media in which the content is distributed or 2. on the format of the content. As we discussed earlier, in the past decades, the media had undergone immense changes, which emerged or changed the importance of certain motives, such as the ones corresponding to social media. Therefore, in the following sections, we will go through the literature that could be directly or indirectly linked with non-product related motives of YouTube expert review watching.

Studies about motivations of YouTube watching found that, besides information seeking, video viewing is driven by the audience's need for entertainment (Haridakis and Hanson, 2009; Khan, 2017) and social interaction (Haridakis and Hanson, 2009). More broadly, regarding user-generated media, Shao (2009) distinguishes information, entertainment, and mood management needs of social media content consumers.

While there are motivations that became more prominent with the emergence of social media and participatory culture (Gálik-Csordás, 2020), many of these motivations resemble the motivations to watch other, more traditional media content. Indeed, from multiple perspectives, the content which is posted on YouTube shares elements of the traditional entertainment sources of television, music, and movie (Shao, 2009) or even magazines and shows (Haridakis and Hanson, 2009; Khan, 2017).

One of the earliest and most widespread theories about the understanding of the traditional media viewership is the Uses and Gratifications Theory (Klapper, 1963; Katz et al., 1973b, 1974; Rosengren, 1974), which – among others – assumes that media watching is a goal-directed, motivated behavior in which people select the content to satisfy their different needs. Even though the theory became prominent when the dominant media sources were completely different from that of nowadays, studies building on it have uncovered many motivations that are just as relevant today as it was decades ago.

Besides information seeking motives (Ebersole, 2000; Kaye and Johnson, 2002; Papacharissi and Rubin, 2000; Sjoberg, 1999; Wolfradt and Doll, 2001), which in this sense can be linked to the product related literature outlined in chapter 2.1, one of the most crucial motives that could drive the demand for media content is the essential human need to be entertained (Ebersole, 2000; Kaye and Johnson, 2002; Papacharissi and Rubin, 2000; Wolfradt and Doll, 2001). However, in addition to this motive, media content is often used to just pass time and avoid boredom (Ebersole, 2000; Papacharissi and Rubin, 2000).

Expert Reviews shared on YouTube are not only competing with traditional media, but as part of social media, it could also provide social (Kaye and Johnson, 2002) or interpersonal (Papacharissi and Rubin, 2000) utility for the potential viewer. This utility is not coming from the product related information or the entertainment factor of the content. Instead, it stems from the social group or network which consists of the review itself.

Shao (2009) also links the motivations to features of social media environment, such as interactivity, formation of virtual communities, ease of producing own content (e.g., comments), and sharing it, features which are absent from traditional media (Neuberger and Nuernbergk, 2010) but are present in the case of content creator generated reviews.

Finally, while not all above studies were conducted by examining content on YouTube, Haridakis and Hanson (2009) and Khan (2017) conducted a survey-based study to examine whether prior findings regarding entertainment (and other) motives behind media content viewership apply to YouTube content as well. Haridakis and Hanson (2009) found that social activity, sensation seeking, entertainment, convenient information-seeking, co-viewing, and social interaction motives significantly predicted viewership of the videos. In addition, Khan (2017) found that behind the viewership, the strongest predictor was the relaxing entertainment, while information seeking motive also predicted the comment reading behavior of the viewers.

2.2.2. Para-Social Interaction

In previous chapters, through multiple literature streams, we outlined that demand for content creator generated product reviews could be driven by many different consumer needs, for instance, information seeking about the reviewed product or entertainment. We also highlighted that behind the viewership of the videos, there could also be some kind of social utility, which stems from the human need to interact with other people. However, there is an important aspect of this literature that needs further discussion as it highlights how content creators can further utilize the motives outlined above. This effect is called the para-social interaction (PSI) and describes the relationship between media personalities and media consumers (Frederick et al., 2012; Horton and Wohl, 1956; Lee and Watkins, 2016; Sokolova and Kefi, 2020).

Regarding the explanation of this relationship, Perse and Rubin (1989) describes that media users look at the media personality as a “friend” and they “feel that they know and understand the persona in the same intimate way they know and understand flesh and blood friends” - (p. 60, Perse and Rubin 1989). Ballantine and Martin (2005) describe that the characteristics of para-social interaction resemble interpersonal friendships. Thus, media users also seek advice from the personalities the similar way as they would do it with a friend (Rubin et al., 1985). The relationship itself evolves similarly to interpersonal relationship as well, in which the media users' uncertainty decreases and they eventually perceive similarity between them (Eyal and Rubin, 2003).

Given the history of examining the relationship between media users and personalities, the early studies were conducted by investigating the attitude towards personalities or celebrities on traditional media, such as radio (Horton and Wohl, 1956; Rubin and Step, 2000) or television (Horton and Wohl, 1956; Rubin et al., 1985; Eyal and Rubin, 2003). However, the theory can be applied to modern social environment, such as Instagram or YouTube as well (Lee and Watkins, 2016; Sokolova and Kefi, 2020).

Lee and Watkins (2016) conducted a survey and examined 1. the drivers of PSI, 2. whether PSI affects brand perception for brands presented in vloggers' YouTube videos. They found that physical attractiveness and homophily (which describes the level at which the users find the content creator similar to them) are significant predictors of PSI. In addition, PSI has a positive impact on the users' brand perception.

Sokolova and Kefi (2020) conducted a similar online survey for Instagram and YouTube. They found that homophily and social attractiveness positively influence PSI. However, they found a negative connection between physical attractiveness and PSI. Finally, they confirmed Lee and Watkins's (2016) study and found a positive relationship between PSI and the purchase intention towards the presented product.

Concluding, the audience of the content creator generated reviews can have either, or both types of motivations, interest in a product or/and be driven by the non-product related motivations. The non-product related audience motivations are something setting content creator generated reviews apart from reviews generated by peers. In the case of peer reviews, the audience has not been found to develop preferences for particular reviewers. For example, data of Banerjee et al. (2017) indicate that the feature allowing the audience to follow Yelp reviewers is seldomly used. Besides highlighting the main literature streams behind the demand for content creator generated reviews, we have also shown that reviewers can utilize the social interaction need through the para-social interaction between the media personality and the media users. Thus, reviewers who aim to maximize their audience sizes are facing with viewers with many different needs and motivations. This is an argument which will be key when we extend our framework with the investigation of the supply side of the market (Chapter 6.). Therefore, in the following chapter, we are looking at the content creators in a similar way as media firms or entities. First, we show the main literature on how the reviewer can utilize the consumers' needs and build their YouTube channels over time to create and strengthen their personal brands. Then, assuming that big YouTube reviewer brands have similar motives to media firms regarding their popular channels, programs, or shows, we outline the background on the incentives of the reviewers by using the theoretical literature on the behavior of media firms and agents.

2.3. *Reviewers as Media Brands*

2.3.1. Personal Branding

As we established, the last building block in the dissertation considers the expert reviews from the content creators' perspective, starting with 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; Varga and Sujbert, 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 also be cases when the supply only consists of individual brands. We can observe examples of these types in various traditional industries. For instance, Hewer and Brownlie (2013) and Dion and Arnould (2016) studied chefs Joël Robuchon, Gordon Ramsay, and 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 other literature streams, the emergence of the internet, social media, and organized online attention platforms opened new directions in the field of personal branding as well. On these platforms, individuals have the possibility to build their follower- or fanbase. Whether intentionally or not, this fanbase building often leads to similar personal brands to the traditional figures, discussed above. We can mention examples of such brand building processes in the case of bloggers (Duffy and Hund, 2015; Delisle and Parmentier, 2016; McQuarrie et al., 2013) or like in our case, YouTube channels (Chen, 2013; Tarnovskaya, 2017; Varga and Sujbert, 2018).

It is worth to mention another domain that also builds on the personal branding and therefore shows similarities to our focal topic. This body of literature examines influencers and celebrity endorsement (e.g., Lee and Watkins, 2016; Sokolova and Kefi, 2020; Burke, 2017; Varga and Panyi, 2018; Marchis and Markos-Kujbus, 2019;

Munnukka et al., 2019). Among others, this domain can highlight similarities in terms of the follower and subscriber gathering incentives of the content creators.

As we established earlier, the product reviewers in our focus aim to achieve profit by providing product reviews. Thus, the dissertation builds on the aspect of the domain of personal branding that through brand building, reviewers are incentivized to differentiate their content in order to grab a bigger share in the product review market. Therefore, in the following, we will more thoroughly discuss the studies that relate to this differentiation aspect.

Shaker and Hafiz (2014) explore various disciplines to identify different features of personal branding. They highlight the need for brand image creation and positioning while among others, they also describe the personality dimension (such as sincerity, excitement, competence, sophistication, and ruggedness) and both the core (including for example the brand's vision to differentiate itself) and the extended brand identity of the personal brand, which can include the personality dimensions to provide completeness.

Tarnovskaya (2017) examines the major stages and elements of personal branding on YouTube. They revealed three major stages, overlapping over time: the loyalty towards the brand, the promotion of multiple social media accounts, and the encouraging co-creation. Regarding the elements of the brand, they pointed out the personality of the YouTuber, the typical topics, the tone of voice, and finally, the environment. In addition, they highlighted that the key features, such as the clarity, consistency, and authenticity of the content creator, are similar to the ones in the case of traditional brands.

Fournier and Eckhardt (2019) examine 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 by creating a perception in the consumers about intimacy and authenticity.

Varga and Sujbert (2018) build on the elaboration likelihood persuasion model, described by Cacioppo and Petty (1984). This model differentiates two types of persuasion, the central route, which relies on a person's careful and thoughtful consideration of the presented information, and the peripheral route, where the persuasion happens due to a cue in the context of the situation, such as the physical attractiveness of the message sender. Based on the assumption of the model that the persuasion which happens through the central route is stronger, Varga and Sujbert (2018) examine which route is more often used by YouTubers. They found that as opposed to the Cacioppo and Petty (1984) model, content creators use peripheral routes more frequently.

Dion and Arnould (2016) analyze the persona-fied brands, which term is summarized the best by the following description: *“[…] 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 their brand narrative.

Finally, Duffy and Hund (2015) show how fashion bloggers think about the ideal persona built by them; what is their “having it all” perception about this profession, while Scolere et al. (2018) highlight the platform dependency of the developed personal brands by the individuals and key elements such as 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 corresponding to the supply side of the market (Chapter 6.), we allow for effects that relate to this aspect of the channels.

2.3.2. Behavior of Media Firms

In this chapter, we highlight the last literature stream that can be linked to the content creators in our sample. However, the connection here is somewhat weak and only indirect. The studies in this domain examine the behavior of the information mediators with multiple different assumptions regarding the goals and incentives of the entities modeled by them. Hence, we can also observe that the decision variables of the information mediators, derived from these assumptions, are different across the different approaches.

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), the programming variety (Gal-Or and Dukes, 2003) or the 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) assume only advertising revenue, Godes et al. (2009) assume content and advertising revenues as well. Our approach in this regard is most closely related to Falkinger's (2007) and Xiang and Soberman's (2014) study, assuming that news providers try to maximize the ex ante expected audience size to achieve the optimal amount of revenue. This means that agents have a fixed rate per viewer advertising and content revenue. Important to note that through the dissertation, we derive that YouTube channels could essentially have two objectives: maximizing the size of the audience that watches their content and to maximize the size of the audience that become subscriber for the channel. However, out of these two potential goals, only one provides direct revenue for the channel, the audience that watches its content. The other objective only contributes to the revenue indirectly through the first objective.

The last segment we are highlighting from this literature stream and what we are building on during the development of our models consists of the studies concerning attention economies partly (Smith, 2020) or entirely (Falkinger, 2007). These studies highlighted how different these markets are from traditional markets where the demand and supply are clearly defined, based on the approach that YouTube channels, media firms, or similar information mediation entities are trying to attract the attention of the audience. Falkinger (2007) was able to derive the equilibrium audience sizes, assuming 1. audience members with different attention capacities, 2. information signal sellers with a decision to choose the strength of the signal. His findings rely on 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 of 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, will further motivate us to represent effects that relate to the information signal sellers' spill-over effect on the market (Chapter 5.) or to model their capability to build a follower base for long-term benefits (Chapter 7.).

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 hypotheses 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.

3. Data and Methodology

3.1. Data Collection Procedure

3.1.1. Identifying Product Reviewers on YouTube

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

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

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

- Technology
- Tech
- Smartphone
- Phone

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

- Product Review *(counted as one keyword)*
- Unboxing
- Review

To have more reliable search results, instead of pairing every product category to a specific channel type keyword, we used every combination of at least one phrase from each of the two keyword categories, with limiting the length of the search term to 3 words. In addition, we sent the channel search request to YouTube API by restricting the channel language option to English only.

These searches resulted in 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 count. Based on this table, we decided that the threshold value for channels to include 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 screening 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 labeled as English language channels. The reason behind this observation could be that

    a. the channels incorrectly state that they are making English content, or
    b. Google’s API regarding the filtering according to the language of the channels 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 our final goal is to bind the consumers' uncertainty due to new product launches to 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.

3.1.2. Observing the Reviewer Market

The main objective of the dissertation is to examine the demand and supply of product related information in 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 a daily basis to form a panel structure.

However, to understand our data gathering process, we also need to understand the structure of our chosen platform, YouTube. First, as we already outlined in the previous section, the set of information suppliers on the market translates to the set of product reviewer channels in the platform. Second, the information content on the market, which contains the product related information on a smartphone, is essentially the videos the above-described channels are posting regularly on the platform. Finally, the aggregate demand coming from the audience can be identified with the information on how many members were interested in the given video. 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 received views since the last observation could contain valuable information for us.

As we already have the list of YouTube channels, the next step is to gather the information contents they posted on the platform, which could be done by collecting all the video IDs the given channel posted from a given date. We had chosen to start collecting the video IDs from 01 May 2020, which resulted in a 47-day time window between the date when the first videos in the dataset were posted, and the day when the daily observation process began (16 June 2020).

Our motivation behind the chosen date relies on the goal to model the demand and supply of information 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, as we have both the channel and video IDs, we can collect information about them. Regarding the information content, we observe the number of views the video received up until that point in time. This is our most important variable throughout the dissertation since it shows the demand for product related information. Besides this information, we also have the possibility to collect aggregate number of reaction measures, such as the number of likes, dislikes, and comments, 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 download the title and description corresponding to the videos.

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

As we mentioned before, in contrast to the one-time channel ID collection (Chapter 3.1), we acquired data regarding the list of videos and data about each of the channel IDs and video IDs on a daily basis. Hence, every day, we checked whether new video(s) was/were posted on the market compared to the previous observation day. If there was/were, we added it/them to the list of videos, then repeated the downloading process for every channel IDs and video IDs in the list. 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 occurred with the data collection during this time window due to technical difficulties regarding the automatized handling of the continuously growing list of video IDs, leading to a gap 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 observations were collected for the video dataset and 8320 for the channel dataset over the course of the 108-day observation period. Descriptive statistics about these datasets 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

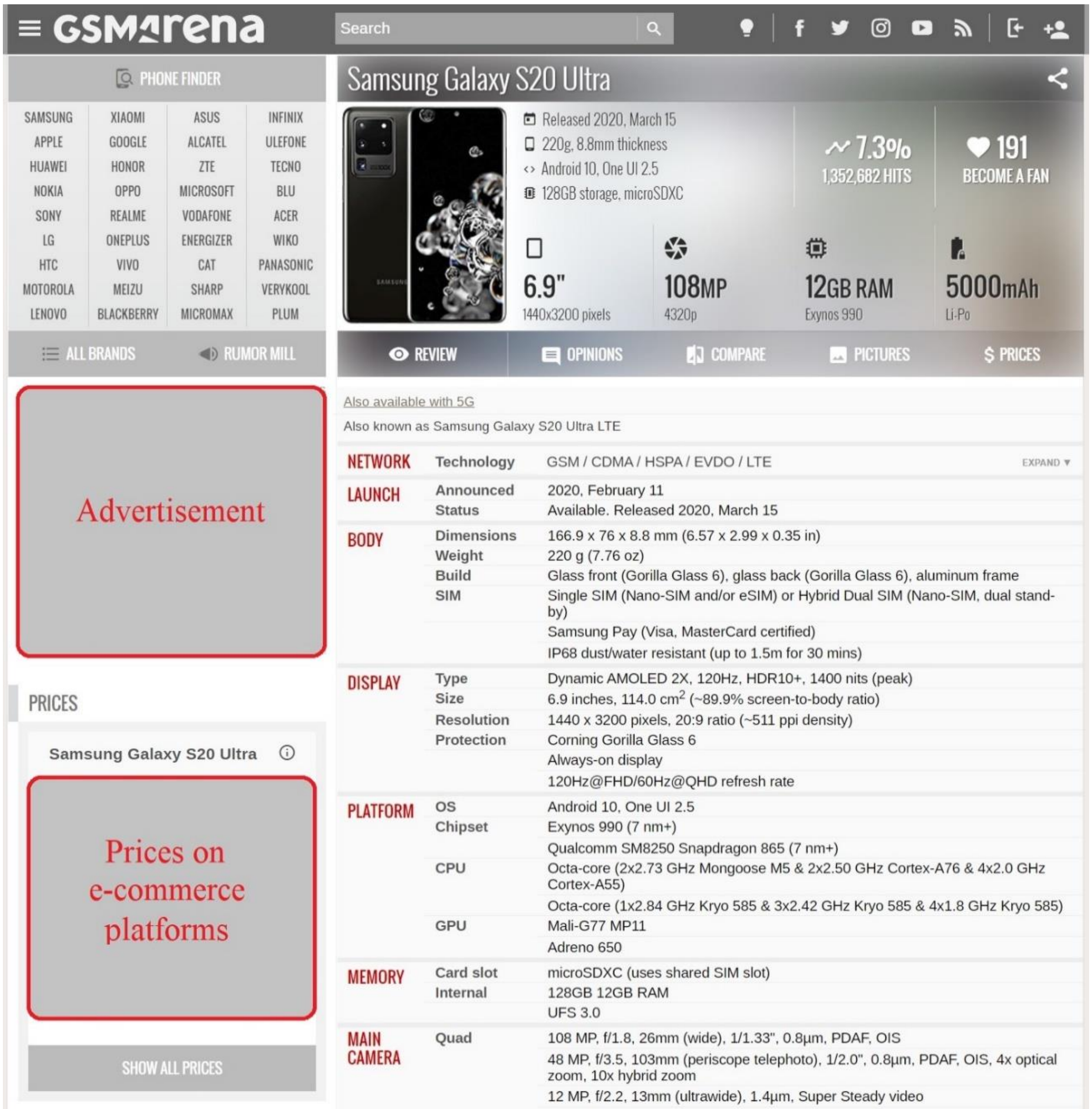
3.1.3. Collecting the List of New Products

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 should be filtered. The reason behind this comes from the aspect that we aim to examine the demand and supply of information about new products on the market. Hence, in this chapter, we aim to collect the list of new products in the smartphone industry during our specified data collection time window. Then, we can use this list to decide if the videos in the dataset can be matched to a new product or not.

The new products examined in the dissertation are 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[4] written in R language.

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

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 prices were hidden to avoid any unfair representation of products and e-commerce platforms

Following the successful collection of the release dates for each product, the next task is to match the ones with recent launch dates (after 1 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 not be used by the product reviewers. For instance, despite the official name being Apple iPhone SE (2020), product reviewers could simply use 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 the following other version or extra specification declaration words: *“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 new products list.
- 2. If the title 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 products on the list, screen the description of the video.
    - 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 dataset.
    - B. If the description does not contain any of the products on the list, remove the video from the dataset.

The reason behind the removal following the multiple product matches in the 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 channel is 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

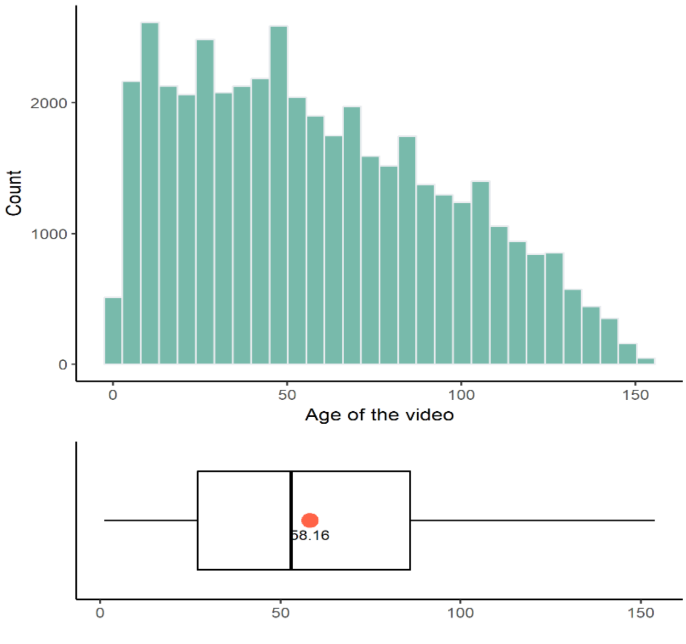
Nevertheless, our final product related video dataset contains 44 015 observations. Descriptive statistics of the dataset can be found in Table 5.

3.2. The Construction of Time Related Variables

Our constructed dataset is especially peculiar in the sense that multiple time dimensions can be identified in it, which could be correlated with the observed variables. However, before the distinction and definition of these dimensions, we should define a universal measure of time in the data. Even though we know the exact posting time of the videos to the nearest second, our measures are gathered on a daily basis. Hence, we define one period as one calendar day in the dataset, 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 the last day, 01 October 2020. Since we tracked every video, the focal channels are posted after the predefined starting date, the number of videos in the dataset increases 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



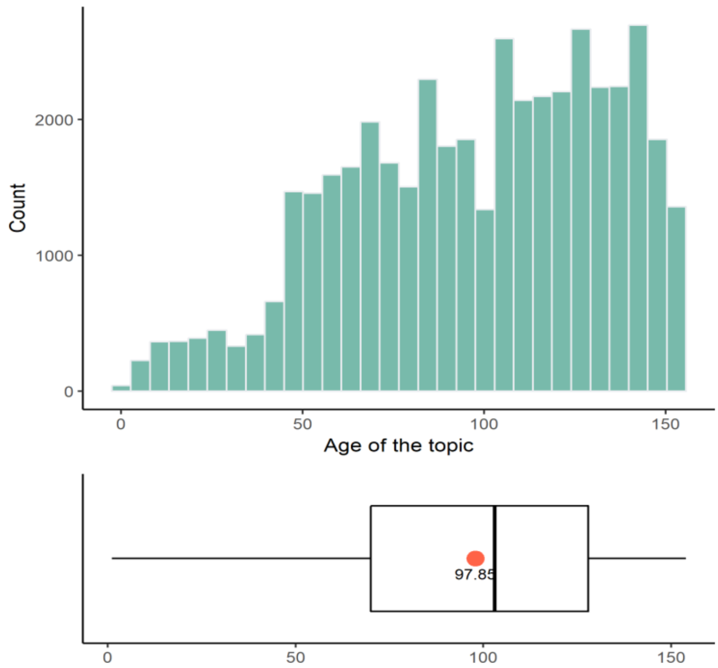
Source: own elaboration based on data from YouTube API

The second time dimension is video specific, and for each video, it starts on the corresponding posting day. Trivially, this dimension will only be the same for videos 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 each video. For instance, we can mention the important controlling role during the model development, but we can also introduce a video specific unique shocks to the viewership for the video's first day on the market, regardless of the absolute time.

The third dimension is motivated by the goal to identify product information markets on YouTube using the shared topics of the videos on the platform reviewing the same product. As one information market could contain multiple videos, the lifetime of a topic will be different from the previous time dimensions. Hence, all information markets could have their own unique lifetime, which creates our final time dimension. There are multiple possible ways to determine 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 information on that day. It could also be 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 the reviewers as a strategic tool and send them the products before the launch date. Our method to determine the starting day relies on the consideration that we are examining the YouTube information market of a given product. Thus, we define the first day of the topic as the posting date of the first video that was posted on this topic. This approach builds on the idea that even if we define an earlier date for the starting date of the topic, that would only lead to empty slices in the dataset until the first video appears. In contrast, if we define a later starting point, we 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 age 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

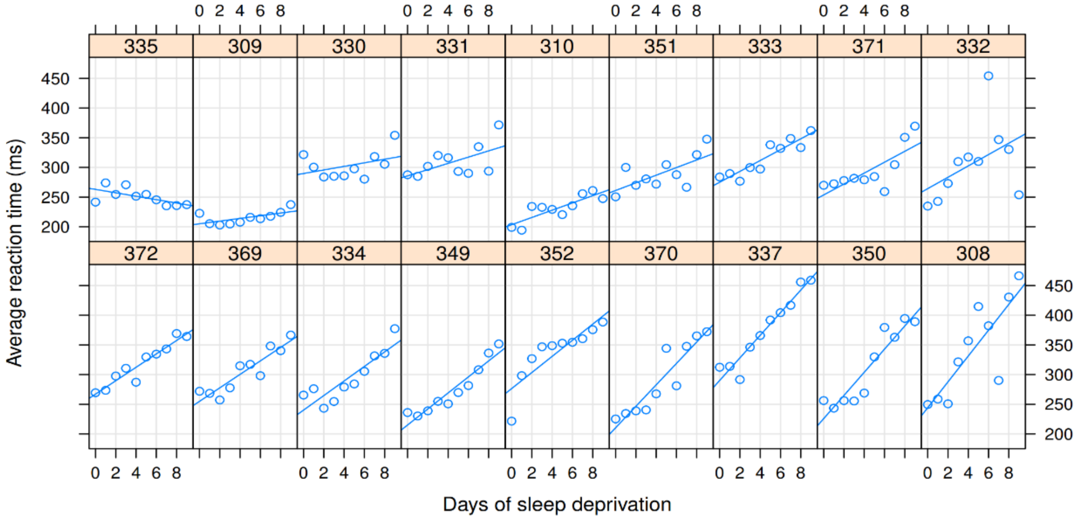
Based on the background theories, highlighted in Chapter 2, we can assume that there is an underlying hierarchical or nested structure(s) in the dataset. For instance, one nesting factor could be the topic of the videos, but we can also mention the channels as another grouping factor.

From the perspective of building and estimating regression equations, the nesting structure, in general, is caused by unobserved characteristics, which sort our examined variables into separate groups with significantly different regression coefficients.

A classic example of such hierarchical structure could be the frog-pond theory (Hox, 1995), where unobserved environmental characteristics of different ponds provide significantly different sizes of frogs. Another example could be Belenky et al.'s (2003) sleep deprivation study, where unobserved biological characteristics of the individuals create different regression equations for each group (Figure 4). Notice, significantly different regression equations could be present due to different intercepts, different slope parameters for the independent variable(s), or both. In this dissertation, we identify two potential nesting structures. First, the videos are nested in a product related information market. In this case, the characteristics of the given topic could contain the products' and brands' exogenous popularity or historical perception. Second, the videos are also nested by the corresponding content creators. Here, the unobserved factors could be the channels' presentation or title giving style, but we can list all the factors that are part of the channels' persona (Chapter 2.3.) and we do not measure them. However, we also take advantage of this methodology when we estimate the effect of the different time dimensions on our response variable.

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 hypotheses regarding the underlying structures in the product information market on YouTube. Moreover, the benefits of this method cannot be fully grabbed by being able to define complicated hierarchical regression. It also provides us a tool to control for unobserved factors that would cause a spurious relationship in the model.

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



Source: Bates et al. (2014)

3.3.2. Random Effects

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

First, we assume that the regression can be build up 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 distributions, showing the effect of the channel on the response variable in a probabilistic fashion. This effect is not related to the independent variable(s) in the regression. In our example, this has two implications, 1. the channel characteristics are a significant predictor of the response variable, modifying the grand mean with certain probabilities, and 2. it does not affect the relationship between the dependent variable and the topic size (x_i). Therefore, in the next chapters, we denote this term as the random intercept.

In contrast, we can also assume that the channel characteristics modify how the topic size affects the dependent variable. Hence, our hierarchical regression 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 corresponding to 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 affect the intercept and the topic size's effect on the response variable. Note, throughout the dissertation, we are not going to derive a model, where an independent variable effect is present, but a random intercept is not present. The reasoning behind this aspect is similar to the motivation of the intercept in the most basic setup of the linear regression: $y = \beta_0 + \beta_1 x_i$. In this equation, it is possible that the most probable value for β_0 is zero, but assuming it before the estimation, and therefore estimating the model without it could lead to a model, which is intrinsically biased.

Trivially, in our specifications, we use more than one independent variable and more than one random effect. Hence, the regression equation system will be more complex than the simple example shown in this chapter. However, despite its complexity, the foundation will still be similar. The details about our implementation of the methodology 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 with nonlinear optimization. First, a trivial solution could be the usage of different simulation techniques to estimate the parameters of the unknown random effect distributions. We can achieve this by first, drawing random numbers from the probability distribution(s) with some set of parameters. Then, we can calculate the mean log-likelihood across the draws. Finally, we can then iterate the model parameters to achieve the best fitting model by maximizing the 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 of the optimization process. The increase in 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 models that need less resource in exchange for a few, but acceptable assumptions about the model specification.

3.3.3.2. Variance Components

During the description of the estimation methodology, we are following Bates et al.'s (2014) study, which compares the formula to a 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:

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

We can introduce q number of random effects to this framework, which modifies the regression equation to:

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

where $\boldsymbol{\mathcal{B}}$ is the vector of random coefficient terms. The elements of the vector 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 $\boldsymbol{\mathcal{B}}$ 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, we need to 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:

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

Finally, 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. (2014):

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

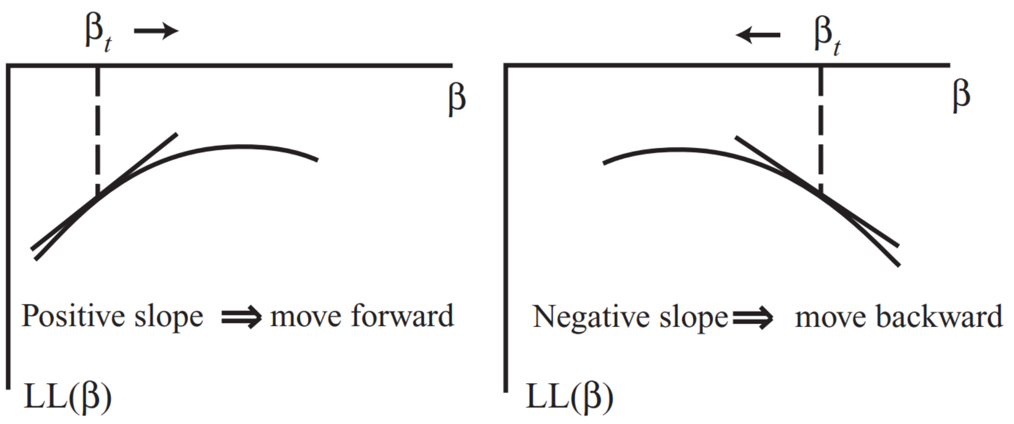
The main benefit of this approach is that due to the shorter length of the variance-component parameters, the model optimizes less parameters compared to the model in Chapter 3.3.2. More specifically, the length of $\boldsymbol{\theta}$ vector is equals $\binom{p+1}{2}$.

3.3.4. Numerical Maximization

Train (2009) describes the role of numerical maximization procedures in research conducted nowadays by comparing it to 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 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 to 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 in cases when the manual derivation fails.

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, we are facing similar obstacles. Fortunately, there are a wide spectrum of available algorithms that we can use during the 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 by iterating the parameter values based on the following information (Train, 2009).

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

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 show the direction in which the algorithm should change the parameter values from the current iteration (i) to the next one ($i + 1$) (Figure 5).

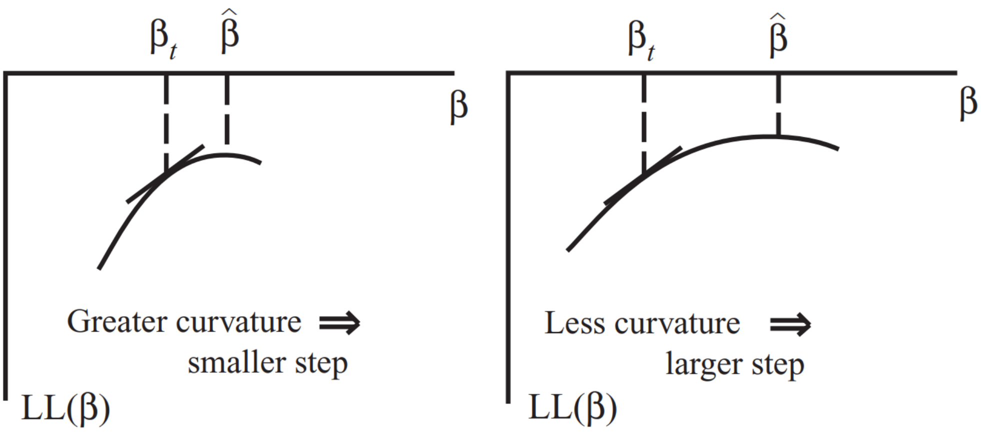
$$g_i = \left(\frac{\partial LL(\rho)}{\partial \rho}\right)_{\rho_i}$$

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 runs of the algorithm from different starting points.

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



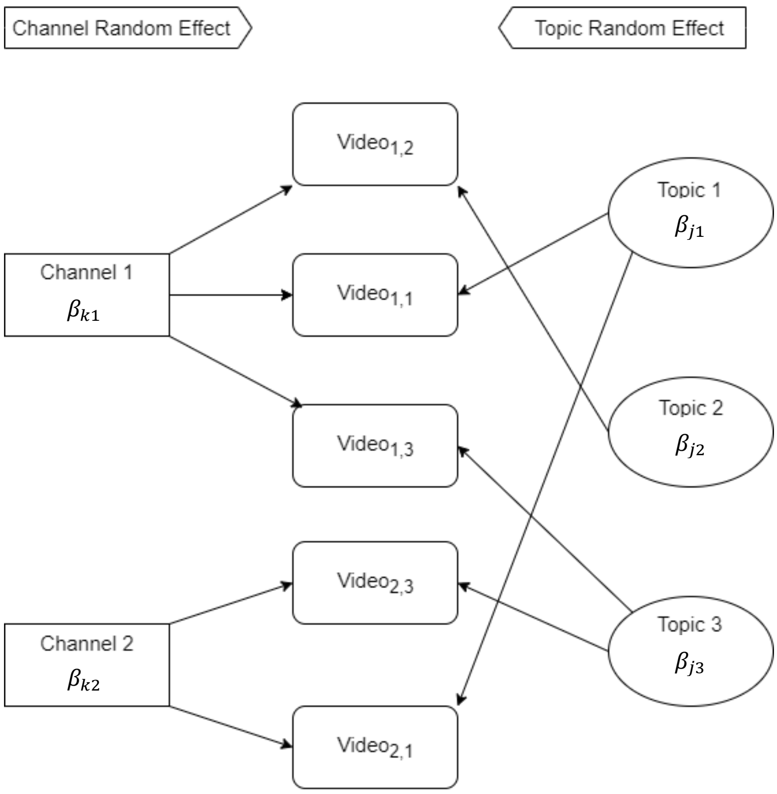
Source: Train (2009)

Finally, the difference among the multiple available algorithms can be described by the function form differences that determine the new iteration of parameter values from the previous objective function values. In the dissertation, we use multiple approaches, 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. Regarding the performance of these algorithms, based on the results presented in the upcoming chapters, we can conclude that in cases when every algorithm found the optimum, the results of the estimation were not significantly different from each other. However, there were cases when some of the algorithms did not find 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 section. The two main nests in the hierarchical system are the grouping of the videos by the corresponding channels and products. The first is based on the characteristics of the persona of the reviewer channels, described in Chapter 2.3, while the second relies on the assumption that products create their own information markets 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 by the researcher. Hence, we also test the estimation of random intercept for the age of the topic and the age of the 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.

A detailed description of this methodology can be found in Chapter 4.1 and 4.2. In conclusion, as Figure 7 and above mentioned time controls highlight, we assume that behind the decisions of the channels regarding the product information market they enter by posting a product review and regarding the time when they post it, we can find a 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. Significance of the Hierarchical Structure

Finally, since our model development and ultimately a considerable number of our hypotheses 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. In addition, the presence of the hierarchical structure also changes the calculation of the p values corresponding to the fixed effects. This significance test can be done by applying likelihood ratio test to the coefficient. The likelihood ratio test calculates the log-likelihood value of the nested model and the model without the fixed/random effects. Then it calculates the test statistic based on the difference between the two log-likelihood values. From this 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. A 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. In Chapter 3, we already discussed the observed structure of the data and the main methodology aimed to address the questions and hypotheses throughout the dissertation. Hence, in this chapter, we focus on the economic phenomena behind this structure and on the motivation regarding the baseline model. As we already established, 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 contents, which is the product review videos. Therefore, in the case of YouTube, we can examine how much demand was generated for a given information content in the past by observing how many views the video has at the moment. In consequence, our first dependent variable relies on the view counts of the videos. Then, in the final chapter, we extend this framework 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 regarding 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 controls to avoid spurious relationships among the main variables.

4.1. Controls in the Model

4.1.1. Controlling for the Channel Characteristics

The final model of the view counts of the videos consists of two main categories of independent variables, the topic and the channel related effects. As half of the hypotheses in the category of channel characteristics are related to the already present topic effects, the dissertation prioritizes the discussion of topic effects first (Chapters 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 relies on the possibility that there could be systematic differences among the channels regarding their decision of which products they are reviewing and when they post those reviews. This would lead to a biased estimation regarding the topic's effect over time on the videos corresponding to the topic.

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 6.1.2. and 7.1., it is worth to briefly note that it assumes *"big-gets-more"* and *"big-gets-bigger"* phenomena on the market. In contrast, the discussion about the role of the channel random effect is discussed in Chapter 6.1.1. In a nutshell, it assumes a unique characteristic for the channels on the market, which reflects the differences among the viewers to attract the demand coming from the audience. Among others, such unique characteristics could be the tendency to use attractive thumbnails and titles for the videos or the right usage of search tags (Li et al., 2016; Trzciński and Rokita, 2017; Diwanji et al., 2014), but it can be any other aspects of the persona such as the entertainment or presentation style.

4.1.2. The Lifetime of the Videos

Besides the change in the view count of the videos, Chapter 7 aims to model the suppliers' 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 a relatively short lifetime, in which they gather most of their views (Yang and Leskovec, 2011; Figueiredo et al., 2011, 2014; Figueiredo, 2103; Ahmed et al., 2013; Li et al., 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 time passes. 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 (Yang and Leskovec, 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:

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

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

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

which is supported by Cheng et al. (2007) and Szabo and Huberman (2010).

However, there is a chance that the multiplicative inverse relation defined by Formula 3 may not represent 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. For instance, Li et al. (2016) showed that the evolution patterns of videos could follow changing dynamics, such as *burst-slow-burst-slow* or *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, for each day, we can estimate an adjustment for the age of the topic effect. 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 retrieved for each day. Combining Formula 2 and the posterior modes of the adjustment, 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:

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

4.2. *Information Market Identification*

4.2.1. Main Hypothesis Development

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

First, we defined the supply in the market as a set of third-party product reviewer channels, building on the literature on theoretical models of news providers, attention seekers, para-social interaction, and online personal branding. We are going to focus more on these literature streams when we extend our framework with a more detailed examination of the supply side of the market in Chapters 6 and 7.

The demand for information on the market comes from the audience that is interested in the topics of the videos, based on the literature on consumer learning, as these topics are essentially the products the channels are reviewing.

Hence, we define the product review information market corresponding to a specific product on YouTube with the collection of 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 reviewing different products, 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 in 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 the topic interest effect since it shows how the overall activity or engagement corresponding to a topic, 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, 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 on 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), discussed in chapter 2.1., 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 the product is launched, meaning that the demand for information is the highest when the product is launched. After this first period, the uncertainty, and interest towards the topic decreases over time. Consequently, the topic interest effect in our model may also have a lifetime, such that it is the highest in the first period of the topic and then decreases while it becomes irrelevant eventually. Therefore, we argue that not only the topic itself but the age of the topic could also matter 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 reviewed product has a significant effect on the performance of the video.

B: The product's effect on the video's performance is decreasing over time.

As we 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 (Chapters 5 and 6), we extend this approach with a more realistic view on how this market might work by enabling the actors in the market to influence each other's performance.

4.2.2. Modeling Information Markets

To model the information markets, defined in the previous chapter, we can implement the method we derived in Chapter 3.5. Note, that we already used this approach when we used random effect estimation to control for channel characteristics and 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 hypotheses and research questions we formulate.

In addition, we are also interested in how the age of the topic affects the performance of the videos. Hence, we use the corresponding topic age variable we derived in Chapter 3.4. Then, similarly to the age of the video, we also have the possibility to adjust the time related coefficient by defining it as random in the model.

There are multiple ways to build this model depending on the assumption regarding the topic interest over time function. Our first approach is to 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 topic specific intercept to set a unique scale for the interest of each topic. Model 3 in Table 7 was estimated following this methodology. In this model, all topic has a unique number of topic interest at each point in time, but the relative differences in the interest between two points in time are fixed across the 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 $\beta_1 < 0$.

We can raise similar arguments to that of in case of the age of the video regarding the possibility 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, for instance, if the assumed function form is correct, but there is a change in the parameters over time. However, it could also be 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 are often described with epidemic models, where the adoption of new products follows a process where we can observe a slow increase in the product adoption at first, followed by a sharp increase, then a fast, and finally a slow decrease until the changes become irrelevant. Our variable is 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 test a model where we estimate unique topic interest over time function for each topic by estimating a random slope for the age of the topic besides the already present estimated random intercepts. With this method, we not only get unique scales for the topic interest over time function in each information market, but we also get unique shapes, so the relative effect of a change in the 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 in 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 7)

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, we used the logarithmic transformation of the number of days passed since the video was posted as a dependent variable (Model 1 – Table 7). We found that this effect is a 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 estimated 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 (Model 2 – Table 7). Table 6 shows the estimated adjustments for the age of the video coefficient. In addition, this method gives us a unique opportunity to visualize random coefficients, as our grouping variable can be represented on the x-axis in a standard two-dimensional coordinate system.

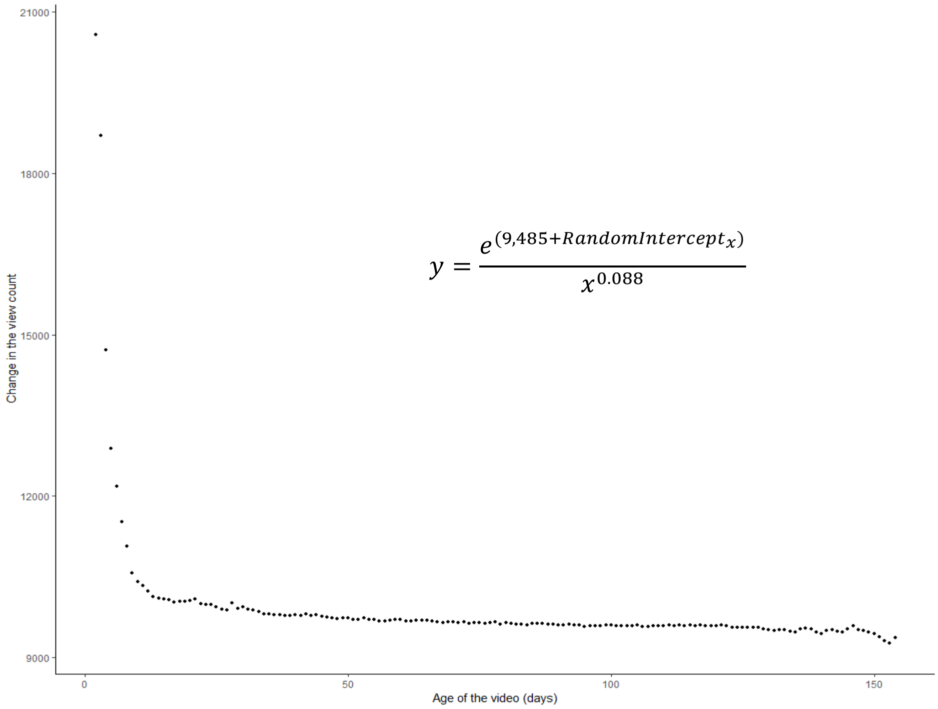
However, while the grouping variable can be represented on the x-axis easily, for the introduction of the estimated random effects to the y-axis, we need to calculate a central value for the estimated distributions. Hence, we simulated the posterior modes to represent the typical value of the random intercept and calculated the value of the independent variable. Using these values, we could calculate the overall effect of age according to Formula 3. Based on Model 1, Figure 8 shows the estimated connection between the age and view count of the video, ceteris paribus. In contrast, Figure 9 shows the readjusted relationship in Model 2, which applies 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

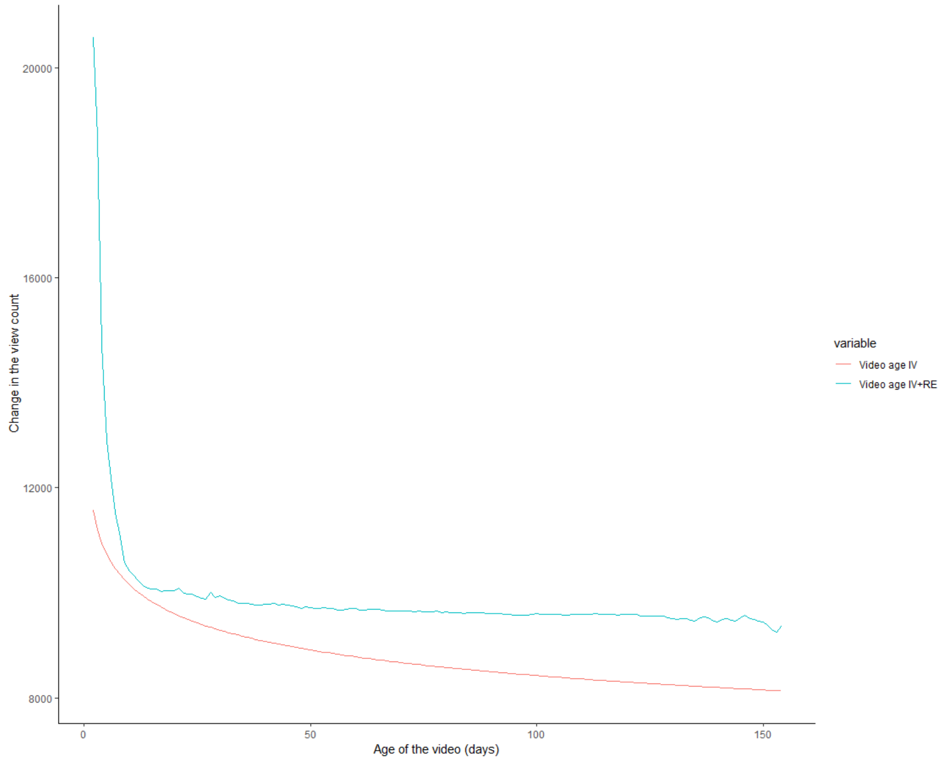
Figure 9: The effect of the age of the video with random effect



Source: own elaboration

If we attempt to 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 that has a turning point around two 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 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 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 to 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 did not find 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 scale of topic interest with random intercepts.

These results were obtained by using a model specification that has a limitation. Even though there is a unique topic interest at each point in time, the relative differences in the interests of two periods are the same for all topics. We aimed to resolve this limitation with 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 a better fit for the model if we use not only different scales (Model 3) but 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 products is significant. This means that there are observable differences in the view counts of the videos reviewing different products, so the decision of which product the channels should review is important in terms of their revenue.

The negative coefficient for the age of the topic 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 proof that the product information market indeed exists and motivate our efforts to move towards a more complex 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 bring more views to the video, they also have 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. Hypothesis Development

5.1.1. 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 on the topic's information market on and outside of YouTube. We then estimated a dynamic hierarchical model where every channel has a 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.

Therefore, 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 affect the aggregate effect of the topic on the video. Second, we examine two potential manifestations in which the embeddedness of the channels in the YouTube reviewer economy affects the performance of other channels on the market. More specifically, on the one hand, we argue that the channels on the supply side of the market are competing for the pool of views coming from the audience. This argument would suggest a negative relationship between the performance of two competitor videos on the same topic. On the other hand, the suppliers on the market may also have a positive impact on each other by increasing the overall interest for a given topic. Thus, channels may not only compete but also complement each other.

After the theoretical background, we describe the methodology to represent these effects in our model. Here, we are building on the probabilistic properties of 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 positive and negative effects.

5.1.1.1. Satiation Effect on the Market

Our first and most crucial building block for the chapter and arguably for the dissertation is the reasoning about the properties of the individuals' demand for information as time progresses. This argument is mainly relying on the results of information economics (Nelson, 1970; Stigler, 1961), but due to the same understanding of the agents, similar phenomena can be found in the literature on 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), discussed in chapter 2.1. However, many economists consider Herbert Simon's (1959) extraordinary study as the primary starting reference point for this theory. From this paper, we can infer that the classical way of identifying humans as homo oeconomicus and assuming 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 outcome if we assume some costs to the search for 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 also be time and/or (mental) effort. Then, the diminishing return induces that the same amount of benefit of the information becoming more and more costly until it is not worth it 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 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 increases in the market, channels are missing out on viewers and therefore, revenue. In contrast, there was a potential that those viewers would have watched the creators' videos if he/she would have posted them 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 using the argument that consumers decide to stop seeking more information when it is not worth it for them anymore. This narrative certainly can be a driver of the satiation effect. However, important to note that most of the research papers in this area use these models, assuming that we can model the outcome of the information search process as if it were the result of a conscious decision of the consumers. In reality, the phenomenon is more likely connected to unconscious cognitive motives and boundaries.

Similarly, a recent literature stream highlights that the limited attention of the information signal receivers can also lead to a similar phenomenon (Davenport and Beck, 2001; Falkinger, 2007; Smith, 2020). Then, as Smith (2020) argues, the limited attention can be more prominent in the case of organized online attention platforms, such as YouTube, where consumers are exposed to an enormous number of stimuli. Here, consumers simply have a cognitive boundary, their limited attention, resulting in a situation when they must decide – consciously or unconsciously – which video they should 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.

5.1.1.2. 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 of 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 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 that the channels in the same information market thus direct competitors to each other. However, the argument does not immediately imply a negative relationship between the performances of the creators on the market when one can be successful at the cost of decreasing others' market share.

For this assumption, we also need the information described in Chapter 5.1.1.1., that the demand for information on the market is limited. The reason behind this requirement comes from the fundamentals of economics that the scarcity of resources is what leads to competition among the actors (Robbins, 2007).

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 which leads to limited pool of views on the information market. Hence, we can assume that the average audience member indeed cannot watch all videos on a topic, leading to a limited number of views over the topic lifetime horizon. With this argument, we can derive the argument that channels in the information market can be considered direct competitors of each other, attempting to grab as big a share from the pool of views as possible.

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, since satiation and competition are linked from the channels' point of view, we handle the satiation-competition effect together in the following chapters.

5.1.1.3. Topic Awareness

So far, during the 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 videos are competing with each other on the same information market. We expect that through the satiation effect, both the missing-out and the competition element – on average – harm the creators on the platforms. Hence, we would anticipate that we will find a negative relationship between this phenomenon and the performance of the videos. However, topic interest, in general, is still a positive phenomenon for the videos on the market. It can still provide extra views for them through the part of the audience that is still aware of the topic and eager to follow up on it.

Hence, the total positive effect that the topic provides over its total lifetime horizon should be restricted by moderating it with the share of interest that represents the satiation of the consumers. In contrast to the exogenous topic interest introduced in Chapter 4.2, topic awareness will be determined by the internal (YouTube) popularity of the topic.

Thus, the effect represents how the actors on the platforms relate to the topic at a given time period, showing the current state of interest and engagement from both the audience and channels. Trivially, we assume a positive relationship between topic awareness and the performance of the videos posted on this topic.

One of the most interesting aspects of the topic awareness we defined above is that its dynamics are not purely dependent on the satiation of consumers. That case would mean that as satiation increases, topic awareness necessarily decreases on the market. In contrast, the dynamics of the topic awareness are also dependent on the market participants.

Channels can affect the topic interest multiple ways. For instance, as they are joining the market by posting a video on the topic, they may bring new viewers to the market who will be interested in the topic. They can also make their content in a way that specifically motivates the viewers to demand more information on the topic, for example, due to its informativeness, entertainment, or controversiality. Hence, it may incentivize the viewers to watch more videos and become topic followers. Essentially, we can argue that the videos on the same topic may not only compete but also complement each other. This is also supported by various evidence from the literature. For example, this could be the case if the popularity of reviews, e.g., the number of views a review attracts, is observed by the audience, and this observation influences the audience's interest in the product. Prior research has documented consumer inferences of a similar nature. For example, Van Herpen et al. (2009) showed that consumers infer product attractiveness from cues about product popularity, such as stockout. Findings of Cui et al. (2012), Micheli and Gemser (2016), and Nguyen and Chaudhuri (2019) suggest that consumers make similar inferences from the volume of consumer reviews and media attention, respectively. In our empirical context, consumers can observe the popularity (the number of views) of reviews of a particular product. This means that product reviews on YouTube are not only a source of learning for consumers interested in a product but also can facilitate inference-making about which products are worthy of consumers' attention. Such a process implies a complementarity between the reviews of the same product.

Additional sources of complementarity between reviews can be as follows. The audience can develop a sense of belonging to a broader community spanning a set of creators and their audience (Neuberger and Nuernbergk, 2010; Neuberger et al., 2019; Jönsson and Örnebring, 2011). Viewers may perceive watching multiple reviews from different creators as an act signifying their membership in a community and, thus, derive social and emotional value from it.

In conclusion, these studies highlight that the overall attractiveness and interest of a topic can increase as the number of videos and the aggregate demand increase. This means that the pool of not satiated viewers, those who are still aware and following up on the topic, increases for the videos on the same topic. For example, we can imagine an event that a newly posted video appears on the market. Then, it generates waves in the topic information market and essentially raises 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.

5.1.2. Probabilistic Properties of the Satiated-Interested Audience

It can be easily seen that the arguments presented in the previous sections of the chapter cannot be directly measured from the data available to us. Hence, as a last step of the hypothesis development, we reformalize our reasoning and link it to measurable variables by relying on probabilistic assumptions regarding the main points of the arguments, namely the satiation and topic awareness effects.

The first objective of this process 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’s 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 metric that shows us the total interest that a topic 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 the channels’ perspective at that time period. However, we need one more modification to this methodology to be able to appropriately measure the effect of topic interest on the videos. This is due to the fact that the current aggregation contains 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 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 following 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. We are only examining if there is a connection between all the views of other videos on the topic and the views of the focal video, regardless of the posting dates. Hence, this variable is not suitable for examining the dynamics of satiation and topic awareness. However, it shows the resultant of these effects.

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 watch more and more videos about it. Intuitively, we may derive that we can find the highest number of people that are still interested towards a topic with the highest probability among the viewers that joined most recently. This probability then gradually decreases as we are looking at the audience that joined earlier along the topic lifetime. However, this would be only true if the number of new joiners to the market over time can be described by a uniform distribution. Hence, instead of absolute values, we link our arguments to the share of audience that joined at a given period. Meaning, we expect that from the total audience, we can find the highest share of still interested people among the viewers that joined the market most recently. In contrast, we expect the highest share of satiated people among the viewers that joined the market on the starting day of the topic. Important to emphasize that we only expect the share to decrease over time as we examine viewers that joined the market at earlier and earlier stages of the topic lifetime. It can still happen that a later period in time has a higher absolute number of satiated viewers if the total number of views on that period was higher enough to balance out the time difference in the satiation process.

As we can define the audience as either being satiated or aware of a topic, we can use this distinction to derive the share of audience that is interested towards a topic and the share of audience that is already satiated from the total audience. Notice, with this approach, we can also include the new videos' topic interest buff effect, which works through the new views generated by them. Since the new video always generates new views on the day of observation, regardless of 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.

Denote the total audience of topic j at a given $\bar{T}$ time period $(1 \leq \bar{T} \leq T)$ with $TA_{j,T}$ and the audience of topic j that joined at time t $(t \leq \bar{T})$ with $A_{j,t}$. Then, the total audience can be calculated as:

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

Denote a time period when the number of new joiners to the market is not significantly different from zero with T, then by definition, we can 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 shows 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 total number of viewers that is satiated at time $\bar{T}$ 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 total number of viewers that is still interested in the topic as

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

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 the total number of these metrics what matters. Instead, we are interested in the satiation and topic awareness at time period $\bar{T}$, which is $1 \ \le \bar{T} \le T$. 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 views to the audience will be equal to the average ratio over the whole topic lifetime, or if there is a deviation from it. Hence, similarly to our base arguments, we only assume that our equations hold on a probabilistic level.

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

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. 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 of total topic views at time $\bar{T}$:

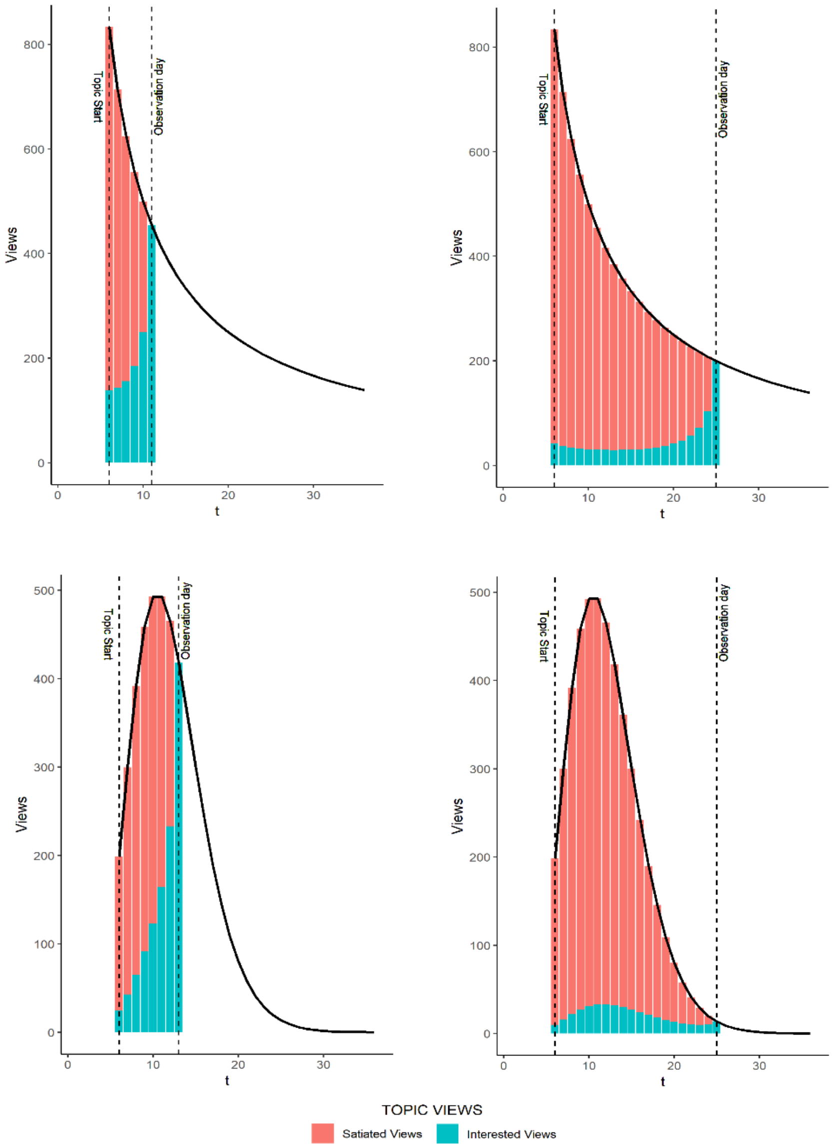
$$TPV_{j,\bar{T}}^{Obs} = \varphi\left(\sum_{t=1}^{\bar{T}} w_{\bar{T}}(t)\, E\big(A_{j,t}\big) + \sum_{t=1}^{\bar{T}} \big(1 - w_{\bar{T}}(t)\big) E\big(A_{j,t}\big)\right) + \varepsilon_{j,\bar{T}}$$

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

This equation then gives us a possibility that if we know $w_{\bar{T}}(t)$, we can also calculate the expected value of satiated $\left(E\big(SA_{j,t}\big)\right)$ and interested $\left(E\big(IA_{j,t}\big)\right)$ audience. We then use these expected values in our model as independent variables to examine the effect of these metrics to the performance of the video. However, 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 $\bar{T}$ time period. 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 $\bar{T}$ compared 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 types of new topic view counts function.

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



Source: own elaboration

First, we imagined an exponentially decreasing topic interest function. Then, we extended this idea with a *"building-up"* period at the beginning of the topic interest, resulting in a gamma function overall. The graph contains two 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 estimation of the model. There, we are going to rely on the observed amounts of topic views.

After we apply $w_{\bar{T}}(t)$ function on the total topic views, the variables to examine the satiation and topic awareness can be calculated, which finally enables us to formulate the hypothesis corresponding to the endogenized topic interest, which is the following:

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

5.2. The Model of Endogenous Topic Interest

As suggested in the previous chapter, our main objective during the introduction 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 compared 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.

We hypothesize two function forms for representing different types of topic awareness decrease over time. First, we use a linear 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 a nonlinear decrease over time. The function specification corresponding to this form is:

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

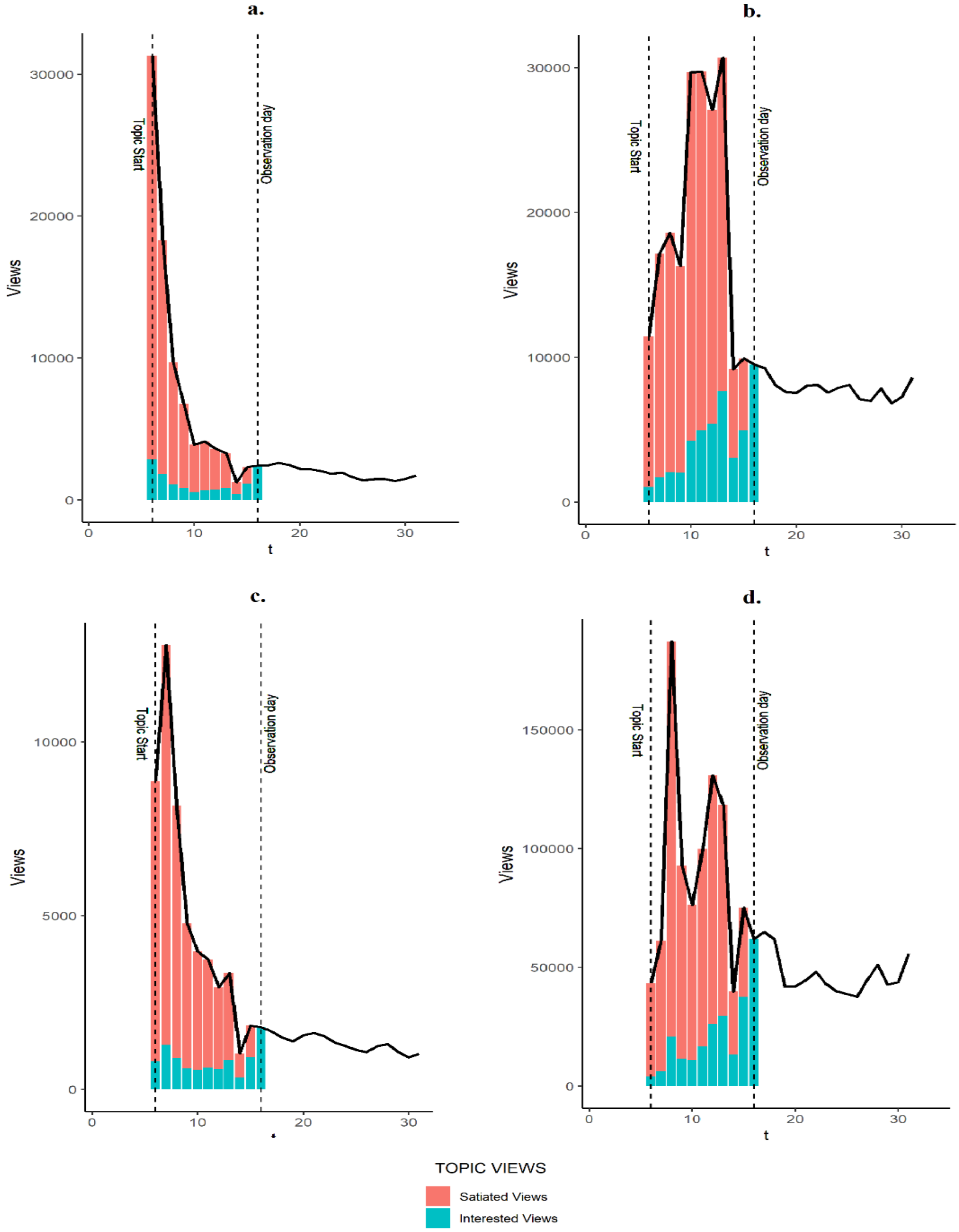
where we optimize the value of θ.

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

Our final model, answering the questions raised in this chapter, builds as follows. First, we aim to examine the resultant of the possible positive and negative effects the topic has on the videos 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 hypotheses in Figure 13, showing how the chapter extends our initial framework derived before.

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

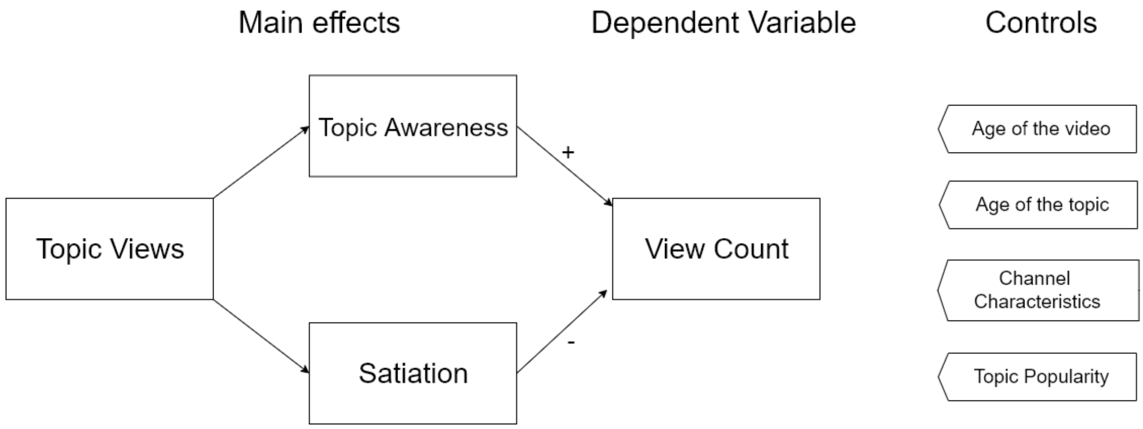
Figure 12: Product related information market from the perspective of the first video poster



Source: own elaboration

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

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



Source: own elaboration

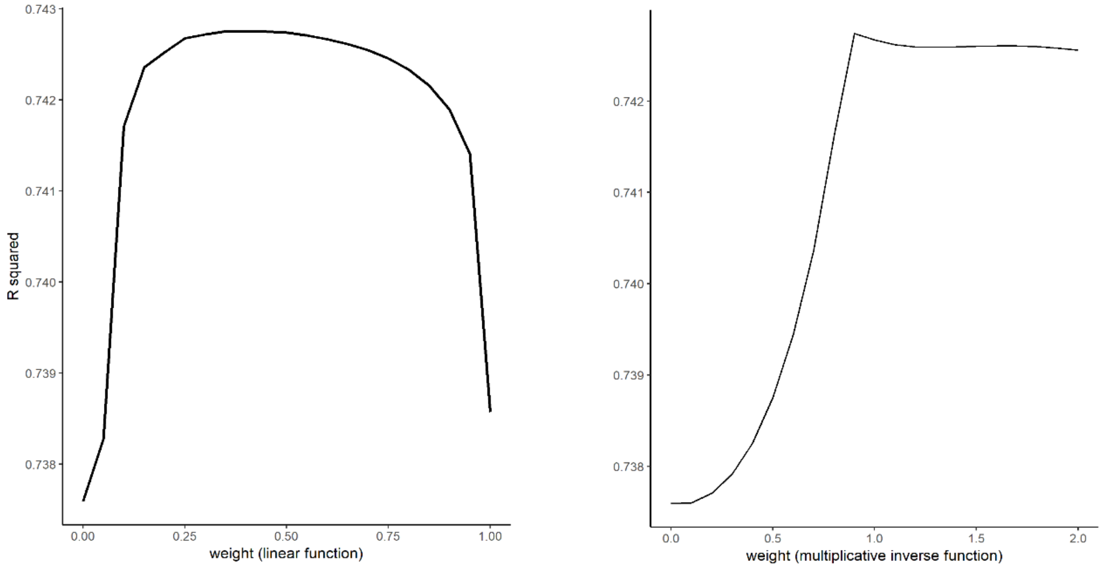
5.3. Results

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 (coefficient: 0.002) and only significant on a 10% confidence level.

Following the definition of total past views of the topic, we aimed to divide this variable into two separate parts with the goal of investigating the topic awareness and satiation effects. The division was made by two weighting functions separately, whose parameter was optimized by iteratively calculating the values of the variables corresponding to satiation and topic awareness and estimating the model with those variables.

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 could be made by eliminating the possibility of a significant positive and negative spikes between two estimation results. We found that in the case of the linear weighting function, we can observe the best fitting model at the slope parameter .35. Using the multiplicative inverse function form, we get the best fit with exponent .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 regarding the estimated coefficients for the independent variables and their corresponding standard errors, both being highly significant. Based on the information criteria, there is a slight favor for the multiplicative inverse function form.

These results unambiguously suggest that the approach to divide the audience based on the distance between its corresponding period and the observation day leads to 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 the satiation effect has a negative while topic awareness has a positive connection with the view count changes of the videos. Based on these results, we can accept our second hypothesis.

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 throughout the dissertation, our understanding of the actors in the product information market builds from multiple literature streams as the different features of this unique market can be linked to multiple disciplines. 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 complement each other at the same time. The aim of this chapter is to move away from the previous, homogenous view about the information providers and explore whether and how they can differ from each other. Ultimately, our goal is to investigate the effect of these differences on the view count of the video. Thus, 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 are sorted into two main categories. Moreover, we are going to examine two types of effects in the case of each main category. 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 mostly relies on the literature on personal branding (Chapter 2.3.1.), more specifically, the role and nature of the self-brands' persona on the platform. However, we also use the findings of the literature on the demand for product related information (Chapter 2.1.) and demand for media content (Chapter 2.2.). We first derive a model where we account for the unobserved channel characteristics for each channel to examine the brands' effect on the non-product related demand. Then we extend this framework by assuming that the persona of the channel affects the product related demand as well. This way, the brand's image could not only provide direct benefit for the channel, but it may change the structure and 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. The dissertation follows the overall consensus on 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 subscriptions model in Chapter 7. Finally, similarly to the persona of the channel, we consider potential cross-effects between the topic information market and the size of the channel.

6.1. Hypothesis Development

6.1.1. Brand Related Factors

As we discussed in previous chapters, the professional or expert reviewers on YouTube could face a demand that is not only related to the presented product (Chapter 2.1.), but it could also be driven by non-product related motives (Chapter 2.2.1.), such as entertainment or social needs to interact with other people. Then, we also highlighted that demand from the audience could come specifically towards the focal channel through the para-social interaction between the media user and the media personality (Chapter 2.2.2.). Finally, we discussed that channels might attempt to facilitate these sources of demand by building a personal brand (Chapter 2.3.1.), which could help the reviewer to realize higher view counts from the product and non-product related demand. In this chapter, we will examine both relationships by introducing effects to the framework corresponding to the relationship between the channels and the types of demand.

The personal branding literature investigates how channels are creating a brand image, a persona for their channels, which will be presented instead of the actual person. Even though the studies in this domain do not examine the effect of different personas on the performance of the channel, important to notice, that this consideration is still the intrinsic driver of the studies in this literature and essentially the creation of the personas. The channels establish and develop their style and brand image over time to achieve better results in the information market, to be more successful.

This literature also highlights key features of a good brand image or persona on the market, such as different personality dimensions, including sincerity, excitement, competence, sophistication, and ruggedness. As Dion and Arnould (2016) highlight, the content creator needs to appropriately integrate these features into the overall brand narrative for a successful brand management. The channel's perception about the optimal brand image mix can rely on various information sources available to them. For example, they can use the received audience reactions or the performance of the channel as feedback, but they can also monitor other brands on the market as well. Then, all this information is merged with the overall worldview, personality, and skill set of the actual person behind the persona. Thus, every channel could have a different perception about the optimal brand image on the market. As a result, we may observe that while there are trends in the brand images on the platform, the strategies and the implementations of the different trends differ from channel to channel, leading to relative winners and losers on the market, based on the creators conscious or unconscious decision about the persona.

There could be multiple reasons why we may observe channels to infer the way of success incorrectly. For instance, it can be a cognitive boundary that they simply cannot imagine themselves as an average viewer or average audience member of their target group. Hence, they make false conclusions about what the audience expects from them. There are also 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 the decision to illustrate a potentially flawed logic. According to this bias, channels may infer the wrong success characteristics simply because they do not see the channels that stopped their activity due to the lack of success. There could also be cases when there is an overall good understanding of the success factor, but channels choose a bad mix that may induce a negative attitude from the audience towards the channel.

In conclusion, we may observe that the supply on the market is not homogenous; instead, it contains a wide variety of personal brands. Given the different brand images, we can assume that there is a difference in their success in attracting views. This could relate to the product and non-product related demand as well. First, starting with the non-product related demand, we can approach the benefit of a good brand image as a buffer to the performance of the channel's videos compared to a worse brand image. Second, brands may affect the channels' capability to attract non-product related demand as well. If we accept our hypothesis regarding the presence of unique self-brands on the market, we can assume that similarly to traditional brands, it may have economic consequences to the market.

The 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 competition, consumer loyalty, and price elasticity (e.g., Simon, 1979; Krishnamurthi and Raj, 1991; Delgado-Ballester and Munuera-Alemán, 2001; Alnawas 2016). Based on these papers, we can derive that overall, the competition which is present on the market can be moderated if the firm builds the brand in a way that the price elasticity for its products will be lower than that of for the competitors. Therefore, 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, we defined a good brand image in a way that it is attractive to the audience. Based on this premise, we can also assume that a good brand image may be capable of attracting aware viewers of the topic better than a worse brand image.

In conclusion, motivated by 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 differentiate the topic effects for the channels.

6.1.2. The Size of the Channels

In this chapter, we examine the differentiation of the channels by their corresponding sizes. Similarly to the persona of the channel, we investigate the direct effect of differentiation on the performance of the videos, and we also explore potential cross-effects with the topic interest. However, in contrast to the previous chapter, we start our argument with the indirect relationship. The reasoning behind 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 channel characteristics on the relationship between the topic's effect and the performance of 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; it is the result of the channel's previous activity.

There is a wide literature on the economic consequences of the size of the entities in the supply (e.g., Amato and Wilder, 1985; Amato and Amato, 2004; Lee, 2009; Niresh and Thirunavukkarasu, 2014). From this body of literature, we can 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' interests. Therefore, our prior expectation is that as channel size increases, channels can benefit more from the topic awareness on the market by attracting more interested views to their videos compared 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 a large enough loyal fanbase. However, we can also motivate the negativity by assuming that small channels are more capable of avoiding the competition and engage to niche topics, while big channels are essentially the faces of the market, so they cannot escape the competition.

Finally, we consider the direct effect of the channel sizes on the performance of the videos. While we left this argument to the last in this chapter, the question of how the fanbase of the channel 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 increasing, there will be a higher number of views (Welbourne and Grant 2016, Hoiles et al. 2017). However, the order of the causation at all. It could be that the bigger fanbase induces more video views, however, more views may also cause more subscribers to the channel. 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 that the representative audience member is going to watch the upcoming videos from the channel as well 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 to. Another reason could be to get a 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 shows that the number of views can eventually turn into higher subscriber counts, suggesting that the subscription base can be both a reason and a result simultaneously. Due to this consideration, while so far, we consider the size of a channel as an exogenous variable, in Chapter 7, we extend our framework with a second set of models, estimating the change in the number of subscription count of the channel, which is going to be dependent on the views a given channel received to its videos.

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

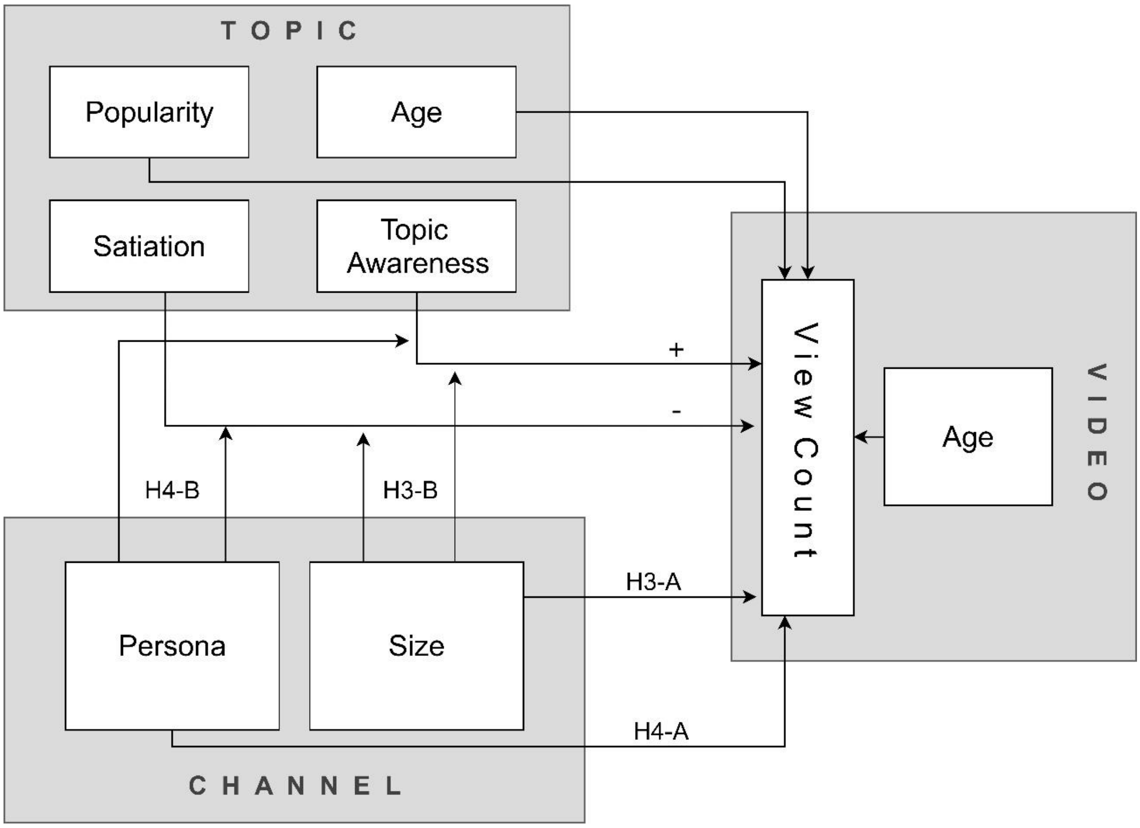
H4:

A: The channel's subscriber count has a significant impact on the performance of the videos.

B: The channel's subscriber count 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 compared to previous models, now the channel characteristics and the topic variables are interconnected, creating a complex structure of the relations in the model.

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



Source: own elaboration

6.2. Methodology

Similarly to the topic effects in the model, we start modeling channel-specific elements by defining a hierarchical regression equation system. With this methodology, we can approach the channels as a grouping variable that creates a different model structure compared to the baseline model or the model with topic information markets. The level at which these groups' equation structures differ from the base model is dependent on the level of complexity we assign to the channel related effects in the model (Chapter 6.1).

First, we model the more straightforward, direct effect of the persona. Here, we define a random distribution for the intercept using the channels as a grouping variable. Note that we already used this approach when we assigned a random distribution for the intercepts using the topic of the videos. Hence, we extend the present vector of random intercepts into a matrix based on the channel and topic of the videos.

Then, we further develop our 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 modeling different random slopes (curves) for the topic effects for each channel. This effect is illustrated in Figure 4 in Chapter 3.3.1.

Finally, we also aim to model the effect of the channel sizes on the performance of the videos. However, for control purposes, we already represented the logarithmic transformation of the subscriber count of the channel in the model. The only terms that are missing correspond to the cross-effects between the topic effects and the size of the channel.

These cross-effects in the case of the subscriber count can be modeled 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.

6.3. Results

6.3.1. Brand Effects

The summary of the model estimations can be found in Table 9. Based on the likelihood ratio test (Chapter 3.5.6.), we can conclude that the usage of random intercepts leads to a better model than the model with constant intercept across the channels. Moreover, the results of Models 11 and 12 show that the model with randomly defined satiation/competition and topic awareness coefficients performs better than the constant slope model. Overall, this shows that there is heterogeneity among the channels to attract product and non-product related demand as well.

The significance of the current model setup compared to previous frameworks highlights that not every channel relates to the information 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 they should choose to review and when they should 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.

6.3.2. Channel Size Effects

Similarly to the previous section, we first examine the results corresponding to cross-effects between the channel sizes and the topic effects (Model 10). The 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 for the interaction terms are negative. Thus, given the opposite signs of the original effects, the two interaction terms have opposite consequences on the baseline (size independent) effect of the topic.

In the 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 build up resilience against competition and indicates that smaller channels are less exposed to the satiation effect on the market. One explanation behind this result could be the visibility of the channels. Due to their size, these channels could be more focused by the audience of the topic, which makes the competition stronger for them.

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 the final view count they will receive for the video. Therefore, the waiting could lead to lower revenue for the channel.

On the contrary, in the case of the topic awareness effect, we can observe an opposite relationship. Here, the negativity of the coefficient means that the positive topic awareness effect on the performance of the videos weakens as the channel size increases.

Hence, we can conclude 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 benefit more from a topic that is *"trending"*. This consideration implies that it is worth it for small channels to follow the trends on the market as they are receiving much more extra views from a popular, attractive topic (that is mostly determined by big channels) than big channels, which provides extra revenue for them.

Calculating the overall effect of the topic given 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 coefficients approach each other as the channel size increases. After a certain number of subscribers, there will be an overall negative effect of the topic. It means that the topic awareness effect, according to our results, helps small channels to gain subscribers but is less or not effective for big channels. A possible explanation could be that as the supply of videos is growing, it raises the topic awareness of the audience. The increased awareness means increasing demand for videos as well, which can reach beyond the scope of big channels on the market. 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.

Overall, this result highlights a spill-over effect for small channels on the platform.

Finally, we also examined the direct connection of the subscriber counts on the performance of the videos without the interaction term. Our results regarding this effect are very robust since all 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 to every video the channels have; it will continuously have a positive impact on the viewership and essentially on the channel's income.

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 a higher subscription count for the channels at later periods. This way, a bigger fanbase can induce higher view counts, which results in an even bigger fanbase, leading to a multiplication effect in the model. Notice that this multiplication effect, if indeed present, also applies to the extra effects coming from the topic and persona since, as we have already shown, they drive higher view counts, which in this theory may contribute to the fanbase building as well. Therefore, in the following sections, we test hypotheses regarding this subscription multiplication effect by exploring which extent the view counts of the videos convert 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 previous chapters, we successfully modeled how the dynamics of the view counts of YouTube videos are evolving in the product reviewer market 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, the chapter only focused on the performance of one single video from a given channel 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 all the videos, not just 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 for long-term benefits as the number of subscribers affects every video the channel has. Moreover, the follower base could provide even more benefits for the channels if the performances of the videos play a role in the subscriber count gathering process. If the views of the videos can be successfully translated into subscribers, the channel enjoys a multiplicative growing process in which the higher subscriber count leads to higher views, which translates into even higher follower counts.

Building on these arguments, our goal in Chapter 7 is to answer the question of how we can model the channels' subscriber building process over time. The chapters consist of the motivation, methodology, and results of two distinct sections. First, we derive the baseline model and examine the subscriber count gathering trends of YouTubers with the possibility of both performance independent and dependent growth. Then, we extend this framework with the intention to explore whether we could explain a significant part of the growth by the audience reactions of the channels’ videos.

7.1. Hypothesis Development

7.1.1. Performance Induced Growth

In the first section of the chapter, we aim to answer the most essential question in this chapter by examining if YouTubers can indeed successfully translate their viewers into subscribers. Our model is based on a proportional process, in which a certain ratio of the videos' new views becomes subscribers at every period. We denote this part of the model with the channel's performance related growth.

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

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 have a significant positive effect on the subscriber number change of the channel.*

7.1.2. The Reach of the Channels

The previous hypothesis was formulated by weighting every video equally in the growth process and only aimed to answer the performance's effect on average. Hence, resolving this limitation, we extend our previous approach with effects that differentiate the videos from different perspectives.

As a first step, we investigate whether the videos that reached a wider audience than the usual viewership of the channel act as a booster in the channels' growth process. Our arguments regarding this perspective rely on the categorization of the audience from the channel's point of view.

For a YouTube channel, there are three mutually exclusive viewer categories:

1) The viewers that have already subscribed to the channel.

2) The audience that watched at least one video but decided not to subscribe (yet).

3) Finally, the viewers that are not familiar with the channel, thus not considered the decision to subscribe yet.

In the first case, the channel's primary goal is to keep these viewers in the follower base and prevent a potential unsubscribe. While there is an exciting line in the literature examining crises when the brand can rapidly lose reputation (e.g., Zhao et al. (2011), investigating product harm crises in the consumer learning literature), this consideration lies outside of the scope of the dissertation. Therefore, we assume that the channels can keep the quality level their subscribers expect from them. Therefore, this chapter of the dissertation focuses on the remaining two segments. In the case of the second group, the channel can assume that there is a possibility that they will eventually become subscribers in the future. Hence, the channels aim to provide evidence through their videos to incentivize them to subscribe.

In the case of the viewers that are not familiar with the channels, we can argue along multiple considerations. The viewers in this group are not familiar with the content creator; they have not seen any content posted by the channel. Thus, they have not considered subscribing to the channel yet. This group could contain viewers who would subscribe immediately and viewers who would go to the second group after watching the channel's content. Therefore, the probability of a viewer becoming a subscriber is higher in the third group than in the second group.

The following hypothesis builds on this higher probability. Based on the higher chance of converting the viewers into subscribers, we expect higher growth if the channel reaches the third group. In other words, we assume that if channels can reach out from their usual audience, they may realize higher growth. Therefore, we hypothesize that the videos with a significantly higher view count than the usual view count of the channel's videos have an extra positive impact on the new subscriber count of the channel compared to the new subscriber count suggested by the view count of the video.

We define a measure of *“reach”*, showing how far the channel’s videos can spread on the market beyond its regular viewership. This viewership is defined by the *“usual”* view counts the channel’s videos get. 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: The videos with outstanding view counts compared to the channel’s other videos have a significant additional positive effect on the subscriber number changes of the channel.

7.1.3. 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 a certain set of viewers.

For the channels on the market, the answers to these questions could lead to multiple implications regarding their long-term strategy to create a bigger market share. Besides this strategy, if these 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, and then, in the methodological section (Chapter 7.4), we show how we can 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 to the subscribing decision. We can model this mechanism such that we assume an aggregate valence perception about the channel, which is an average view coming from the content of the channels. In this framework, the videos are essentially imperfect manifestations of their creator's overall image. Thus, the image of the channel can be inferred by watching its videos and forming an average view based on them. This approach also suggests that channels that have been on the market for a long time have a robust view from the audience. It can be changed, but only slowly. Thus, channels have to consistently create videos that are welcomed by the audience. In parallel, the image can also be worsened, but similarly, only slowly. We denote this approach by the *"Average Subscribing Image"* of the channel.

In contrast, we can also define the relationship between the audience reactions and the new subscriber counts of the channels on the video contribution level. With this approach, the overall effect on the growth of the channels is the aggregation of the video-level contributions. Given the videos' decreasing activity over time, this method also suggests that legacy performance, the content that was posted a long time ago - on average - has only marginal effect on the follower base changes. It is mostly affected by the videos that were relatively recently posted. Obviously, this modeling approach leads to a more volatile process. However, this theory may be more capable of grabbing current trends in the perception of the channel and its effect on the growth process. 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 the model development motivation, we do not take a side on which framework represents better the relationship between the feedbacks and the growth. Instead, our solution to this question is to continue the modeling in both directions and let the data provide the answer which approach performs better.

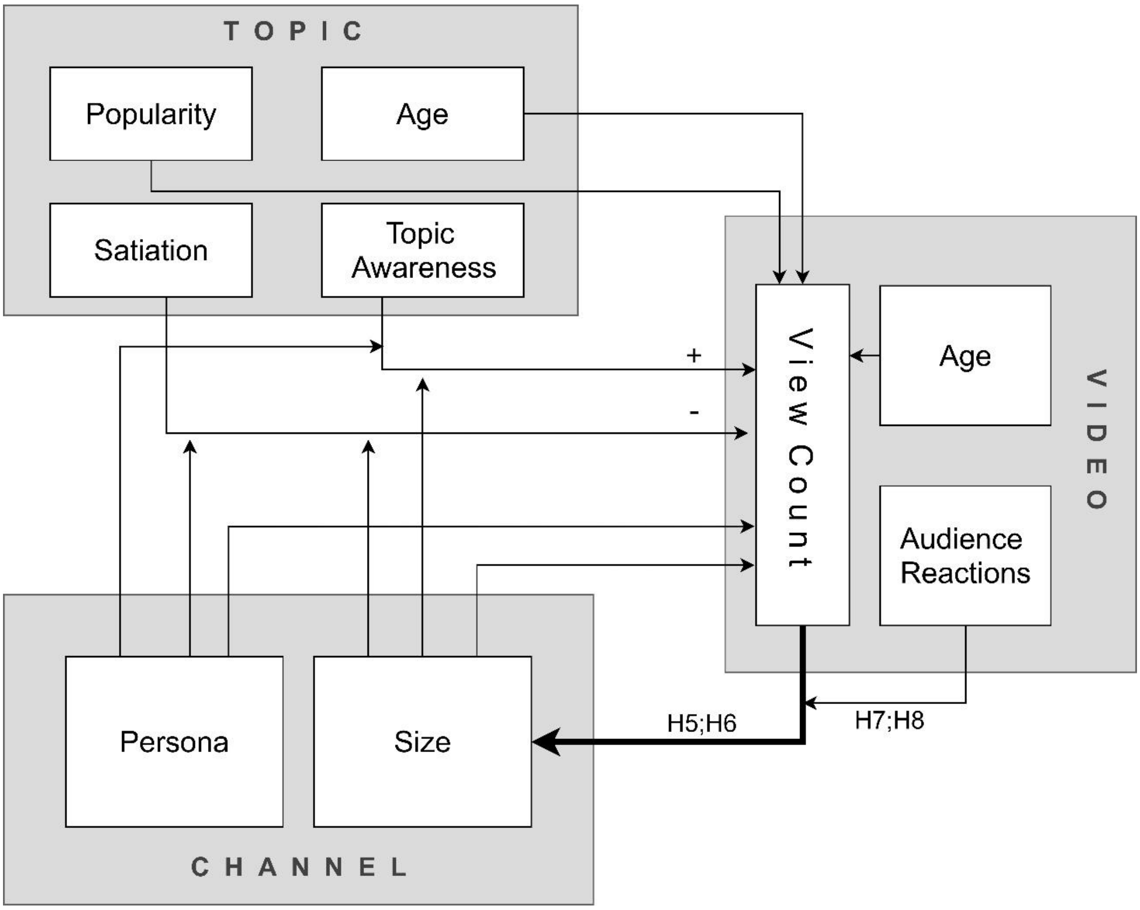
Therefore, in the next section, we derive two models, one for each consideration, to be able to decide which approach fits better to the data, resulting in the following hypothesizes:

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 video contribution audience reaction metrics.

In addition, we also updated the conceptual graph of the models in the 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

7.2. Methodology

7.2.1. 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 nonlinearity could be present in the connection between the subscriber gaining process and our independent variables. 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):

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

where N_{kt} is the number of videos channel k has at time t.

For the performance independent growth, we assume that every channel has a unique growth rate separate from the views of the videos. Then, we use hierarchical mixed-effects modeling to define a random intercept for the channels on the market and define the following model with both performance dependent and independent factors:

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

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

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 modeling constant performance ratio across all the channels.

7.2.2. Deriving the Reach Effect

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

However, before defining the overall effect represented in the regression, we should first derive the video level reach metric. Based on our arguments, the video's reach effect should only be notable if the performance is an outlier compared to the performances of the channel's other videos. This can be achieved if we derive the metric so that it attains exponentially higher values if the performance of the video stands out from the usual performances.

Finally, we need to grab the property of this effect that the video is only an outlier in the set of the channel's videos, it does not have to be an outlier in the full dataset. We can accomplish this by normalizing the performances of the videos the channels have for each content creator separately. In this way, every channel will have its 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 in 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 all channel videos.

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

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

The reason behind the lack of conversion is to keep the exponential. 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 denoting the hypothesized connection.

7.2.3. Using Audience Reactions

Our last model extension aims to explore the connection between the audience reactions and the subscriber gaining process. Modeling this relationship, we ask if we can explain a significant part of the variance of the growth among channels by introducing the audience's revealed valence, opinion, or engagement to 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 reveal the audience's 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.

However, simply introducing these measures to the regression would result in a biased relationship due to the positive connection between the number of views and the audience reactions a given video receives, so we divided 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 could still contain unfolded information if we do not handle them separately. Meaning the audience's overall valence towards a video may lie in comparing the number of likes to the number of dislikes at a given period.

Hence, we represent not only the absolute number of likes and dislikes but also a relative measure expressed by the ratio of these two variables. The last audience reaction measure has a unique role in the model, as it does not reveal the audience's valence. While one can argue that the comments of the videos may contain information that shows both positive and negative valence (even at the same time) towards a video, the resource requirement for retrieving reliable information from the comments (e.g., with sophisticated natural language processing (NLP) and sentiment analysis techniques) lies beyond the limits of the research.

Nevertheless, the number of comments can still provide extra information about the audience. The 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 is replied to other comments, suggesting that the viewer spent more time with the particular video. Thus, the number of comments may show higher engagement from the audience than the number of likes or dislikes. This argument holds regardless of the valence of the comment. Therefore, we represent the number of comments as an extra measure of engagement from the audience.

For the number of comments in the regression, we can apply the same assumption regarding its correlation with the number of views as in the case of the likes and dislikes in the model. Meaning, we expect that as the viewership of the video grows, the number of comments increases as well. Hence, once again, we divide the number of comments by 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 their relation to each other in Table 10.

To this point, we defined measures from the feedbacks that the channels are receiving to their videos and we did not describe how these video-level metrics can be aggregated to an overall channel metric at every period in time. Therefore, in the following sections, we discuss two aggregation methods, 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

7.2.3.1. Modeling the Average Subscribing Image of YouTube Channels

The 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 is looking at all the videos a given channel has as 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 about subscribing to the channel. By becoming a subscriber, the viewer essentially commits to receiving notifications and easier access to all future videos.

From a methodological point of view, this translates to an aggregation where all channel videos are weighted equally. 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 is worth the same for each video, regardless of the video's other properties.

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 by the sum of the views.

Then, consistently to the previous models, we take the logarithmic transformation of this variable to achieve the independent variable in the following model.

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

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

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

7.2.3.2. Modeling Video-level Subscriber Contributions

The final model extension represents a different thought process than the one 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 videos. Then, this image can explain the variance in the subscriber count gains across channels. With this approach, the valence and audience engagement have an indirect relationship with the subscriber count changes through the overall image of the channel. In contrast, we can hypothesize a more direct effect between a better-perceived video and the subscriber number of the channel.

Hence, this approach assumes an aggregation that relies on the individual contributions of the videos, similarly to the performance-dependent elements of the model (Chapter 7.1. and 7.2). However, by extending the model in the direction of the audience reactions, we aim to explore if 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 metrics are weighted by the number of views the videos received compared 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 biases coming from the video size effect, we divide these variables by the views of the video at the given period. Then, we can aggregate these video-level metrics to one aggregate measure that can be introduced to the model. Worth noting that the weighting with the view count changes also assures that we avoid other biases in the model. It would come from the fact that a channel with higher number of videos would have - on average - higher number of audience reactions as the number of likes, dislikes, and views are always nonnegative numbers.

Finally, taking the log-transformation of these variables, the 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

7.3. Results

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

Analyzing the first model, we can observe that the coefficient corresponding to the performance of the channels is significant. Therefore, we found evidence that the aggregated number of view count changes has a significant positive impact on the channel's growth. In other words, we should reject the hypothesis that the coefficient is zero, and we can accept hypothesis 5. This means that besides a unique performance-independent element, we could also observe performance-dependent effects in the model. The implication of this result is crucial for channels on the market. With the evidence on performance dependent growth, we can confirm the performance's multiplicative effect on the channel's revenue. This process essentially shows that 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 may provide multiplicative long-term benefits for the channels.

The second model aimed to explore if we can observe extra growth for channels that have videos with outstanding viewership compared to the viewership of the channel's other videos. Our results suggest that the presence of a video with exceptional viewership is a significant predictor of the channel growth and implicate that the reach of the videos is an important growth potential for the channels. Thus, we accept hypothesis 6. As the channels have outstanding videos, they – on average – receive an extra number of subscribers compared to what our previous model would have suggested. As a result, the channels on the market, especially the small ones that have not had explosive videos yet, may derive the implication that it is worth experimenting 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 impact on future revenues. However, important to keep in mind that the valence of the videos could also matter, which may prevent the overall positive resultant of the experimenting process.

The two follow-up models aimed to explore the connection between the audience reactions and the subscription growth of the channels. The two models tested two different approaches about the possible relationship between the variables. In consistent to the previous sections, we discuss the results corresponding to the *average subscriber* *image* first. This approach of the process argues that the channel's videos 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 overall number of 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 videos of the channels, we reformulated this average to an average feedback ratio. We denoted this method as the *average subscribing image* of the channel as it shows the unweighted mean from the feedbacks towards the channel. Our results indicate that we can explain a significant portion of the variance in the growth process among the channels with the usage of likes to views and 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 terms 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 tested the relationship between the audience reactions and the growth of the channels from the video contribution perspective. The 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 the 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 by representing the valence-related variables, the likes and dislike ratios, we can achieve a better performing model than our previous models. 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: $\ln \sum_{i=1}^{N_k} \Delta Views_{it}$	0.121*** (0.010)	0.138*** (0.016)	0.115*** (0.011)	0.136** (0.063)
Reach: $\Delta Views_{it}^{\overline{Views}}$		0.828*** (0.158)	0.820*** (0.158)	0.825*** (0.159)
METHOD 1:				
Likes: $\ln \frac{\sum_{i=1}^{N_k} Likes_{it}}{\sum_{i=1}^{N_k} Views_{it}}$			3.012** (1.487)	
Dislikes: $\ln \frac{\sum_{i=1}^{N_k} Dislikes_{it}}{\sum_{i=1}^{N_k} Views_{it}}$			-33.722** (14.088)	
Comments: $\ln \frac{\sum_{i=1}^{N_k} Comments_{it}}{\sum_{i=1}^{N_k} Views_{it}}$			1.072 (2.385)	
Like Ratio: $\ln \frac{\sum_{i=1}^{N_k} Likes_{it}}{\sum_{i=1}^{N_k} Dislikes_{it}}$			-0.028 (0.019)	
METHOD 2:				
Likes: $\ln \sum_{i=1}^{N_k} \frac{Likes_{it}}{Views_{it}} \Delta Views_{it}$				0.055 (0.053)
Dislikes: $\ln \sum_{i=1}^{N_k} \frac{Dislikes_{it}}{Views_{it}} \Delta Views_{it}$				-0.036 (0.045)
Comments: $\ln \sum_{i=1}^{N_k} \frac{Comments_{it}}{Views_{it}} \Delta Views_{it}$				-0.031 (0.020)
Like Ratio: $\ln \sum_{i=1}^{N_k} \frac{Likes_{it}}{Dislikes_{it}} \Delta Views_{it}$				-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 is aimed to model the product review economy on YouTube. This model provides valuable information about the audience watching these reviews for firms that launched or intend to launch a product and for product reviewers on the information market.

The main direction of the dissertation can be connected to multiple domains in the literature. However, most of these literature streams have not approached this market from the perspective of the demand and supply of information. The eWOM, product review, and consumer learning literature mostly focus on the product related aspect of the reviews, such as the credibility, expertise, or impact of the presented information.

Another reference point is the literature on para-social interaction and personal branding, 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 channels' brand image is built up 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 investigating 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 the 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 aimed to define and identify the markets corresponding to information about different products on YouTube. We denoted the collection of 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 that posted the content, while the demand comes from the audience that seeks information.

Building on this denotation, we were able to examine the baseline effect of the topic on the videos. In addition, from the studies on new product diffusion processes and consumer learning, we can infer that the demand for information is the highest when the product is launched, and then it decreases over time. Thus, we expect that the topic's effect is decreasing over time as well. Hence, we hypothesized that *H1-A: the reviewed product has a significant effect on the performance of the video, and H1-B: the product's effect on the video's performance is decreasing over time.* Based on the model estimations, we found that we can accept both H1-A and B on every common significance level. This implies that our framework of information segmentation on the platform was supported by the data, which made it 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's 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 coming outside of the platform.

To endogenize this effect, on the one hand, we relied on the information economics 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 in the topic. Finally, we derived a weighting function in the model that can separate the views of 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 parameters of the function and estimated a model with the variables created by the function. Based on this setup, 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*. Regarding the function form, both the linear and the multiplicative inverse functions resulted in significant model specifications. 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 has shown that both the satiation and the topic awareness effects are significant, having negative and positive coefficients, 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 agents. However, motivated by multiple literature streams, we may observe differences among the channels in their capability to attract product and non-product related demand. Overall, we considered two aspects that differentiate channels in terms of the performances of their videos posted on different information markets. These are the brand and the size of the channels.

We tested the brand related elements of the model first. From the personal branding, we inferred that the brand images of the channels might have multiple different roles in the model of product review economy. First, corresponding to the non-product related demand of the audience, it can act as a buffer for the performance of the videos. Meaning, it can directly provide extra views for the channels over time, independently from other aspects in the model. Second, corresponding to the product related demand, it could be connected to the topic effects in the model as well.

Hence, the following hypotheses were formulated: *H3-A: The unique channel characteristics have a significant effect on the performance of the videos, and H3-B: the unique channel characteristics significantly differentiate the topic effects for the channels.* Our results supported both the buffer and topic cross-effects, thus we accepted H3-A and B.

The other channel differentiating factor we examined is the size 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 that in case of the brand images. Hence, we assumed a direct relationship, representing the effect as an independent variable in the model, and we also tested cross-effects with the topic effect as well. Here, we assumed that based on the subscriber counts, channels might moderate or boost the positive or negative effect of the current state of topic interest on the market. From these arguments, we investigated the following hypotheses: *H4-A: the channel's subscriber count has a significant impact on the performance of the videos,* and *H4-B: the channel's subscriber count 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 three layers: videos, topics, and channels. However, channels are not only interested in the short-term benefits but also in maximizing their revenue in the long term. 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 have shown 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 a process in which viewers can 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 potentially multiplicative benefits for channels. From this process, we can infer that as channel size increases, it positively impacts all the videos of the channel, which leads to a higher growth rate, indicating a multiplicative process. These considerations motivated our second set of models, modeling the growth of the channels.

Besides our main objective in this segment, which is to examine the effect between performance and 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 in 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 that was aimed to show whether outstanding videos of the channel provide 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 growth of the channels. 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 metrics 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 instead of handling all 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 videos received compared to the previous period, weighted by the audience reaction metrics mentioned above. In conclusion, the hypotheses outlined to this set of models were the following: *H5: the view count changes of the channels' videos have a significant positive effect on the subscriber number change of the channel, H6: the videos with outstanding view counts compared to the channel's other videos have a significant additional positive 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, and H8: we can explain the channel growth better if we use video contribution audience reaction metrics.* Our results unambiguously suggest that we can accept both hypotheses 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. In addition, based on the results, we found no evidence that the framework derived for hypothesis 8 would be appropriate to model the relationship between the audience reactions coming to the videos and the growth of the channels.

Concluding our findings, we found that the demand for content creator generated reviews is driven by both product and non-product related needs of the audience. Our finding that the reviewed product is a significant driver of its audience size is fundamentally important. It establishes a clear link between the reviews and consumers' demand for product information. A prominent stream of research on earned media focuses on the link between earned media and sales (e.g., Chevalier and Mayzlin, 2006; Moon and Kamakura, 2017; Marchand et al., 2017). These studies do not study information consumption by consumers. Our study provides evidence regarding earned media consumption, thereby shedding light on what, based on experimental data (Kostyra et al., 2016), appears to be a causal link between earned media and brand sales.

Our next set of findings refers to the nature of competition between reviews of the same product. We found that the reviews of the same products are predominantly complements in the short run and predominantly competitors in the long run. These findings shed light on the dynamic nature of the product review market. A creator can opt to post their review early to capture the information demand before the other reviews appear. However, such a strategy can involve risk that the product will not be picked by other reviewers and, as a consequence, will not garner much attention. On the other hand, posting a review late exposes the video to the negative, competitive effects.

The evidence regarding the complementarity and competition between content creators has implications for creators and brand managers. Audience's attention to the product is a common good which is of value to brands and content creators. Our findings suggest that prior consumer attention to the product contributes to subsequent attention, but only up to a point. For content creators, this has implications for review topic choice and publication timing. The dissertation does not derive optimal decisions of the actors involved. It could be that the incentives for creators are to publish their content as early as possible. This could maximize the positive effects of complementarity and minimize the effects of competition. Such incentives for creators could lead to mixed outcomes for marketers. On the one hand, they could lead to buzz right after product launch. Moreover, they could also give marketers a tool to influence creators, for example, they could select which creators receive review units before product launch. On the other hand, such incentives could also shorten the burst of the public's product attention. This could, ultimately, lead to a smaller reached audience compared with a scenario when product attention is stimulated over a longer period of time.

We also find that there is heterogeneity across YouTube channels in their capability to attract the product and non-product related sources of demand for videos. Thus, the views of the review depend on the creator of the review, underscoring the importance of the creators and their characteristics. This finding is consistent with the literature on para-social interaction and personal branding, implicating that the audience can develop a relationship with the creator. This finding sets content creator generated reviews apart from reviews coming from peers. In the case of peer reviews, the audience does not seem to develop such a relationship but instead, relies on extrinsic cues such as review helpfulness rating to assess message credibility (e.g., Forman et al. 2008). In contrast, we find that for the reviews on YouTube, the creator's identity is important to the audience.

Finally, we also identified a multiplicative process in the long-term growth of the reviewers on the market. This implies that big channels get even bigger over time. However, we have also found that smaller channels still have a chance to step on the path that leads to catching up with large channels if they make videos that reach outside of their usual audience. Moreover, the growth of the channels has a strong positive connection with the average revealed valence towards their content, which can be a signal for both small and big channels about the long-term growth potential of their current content. The unveiled trajectories on the market structure highlight potential threats for the firm whose product is being reviewed. The growing concentration could essentially mean that the economic performance of the product will be largely dependent on a small number of reviewers. Thus, marketers need to identify the key figures on the market and use this information during the product's marketing strategy.

Our research can be considered a novel attempt to model the market of product reviews. However, our approach is not comprehensive nor without limitations. First, we estimated our models on data collected from product reviewers in the tech genre on YouTube. As a natural extension, follow-up research is needed to validate our findings on other product categories or other platforms. Second, our data is aggregated across consumers. Such aggregate data allow us to include a broad set of creators, products, and a long sample period. However, we do not observe video watching histories and click streams of individual audience members. While individual data on YouTube watching is not in the public domain, future research should seek to access such data to produce a more granular picture of drivers of demand for, and competition between the reviews. Third, our data does not include information about the platform behavior, in particular, platform's content choices. Such choices are driven by the platform's recommender system. Future research should seek to include additional data capturing key aspects of the platform's behavior. Such research could shed light on how the drivers of the demand and competition emerge from an interaction of viewer preferences, social interaction, and platform behavior. Regarding the model on the growth of the reviewers, while we considered the importance of representing the revealed valence of the audience in the model, due to the limitations of the scope of this research, the usage of these measures could be improved. One can argue that a more sophisticated approach could be achieved by mining the audience's comments on the channels' content. This highlights a research direction of extending our framework with the application of natural language processing (NLP) and sentiment analysis on the audience's comments.

Finally, future research should further our understanding of the link between earned media consumption and sales. Prior research has explored the direct link between properties of earned media, such as the valence of reviews and sales. We document that the demand for earned media can be associated not just with product interest but also for entertainment or social reasons. We also document that audiences' interest in the product depends on earned media popularity. Taken together, this implies that the relationship between earned media consumption and sales is complex. Future research should study earned media, information consumption, and sales jointly.

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.

Arndt, J. (1967). Word of Mouth Advertising: A Review of the Literature, *Advertising Research Foundation*, Inc., New York, NY.

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

Ballantine, P. W., & Martin, B. A. S. (2005). Forming para-social relationships in online communities. *Advances in Consumer Research*, 32, 197–201.

Banerjee, S., Bhattacharyya, S. & Bose, I., (2017). Whose online reviews to trust? Understanding reviewer trustworthiness and its impact on business. *Decision Support Systems*, 96, pp.17-26.

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.

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.

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

Cacioppo, J. T. – Petty, R. E. (1984). The Elaboration Likelihood Model of Persuasion. *Advances in Consumer Research,* 11(1), 673–675.

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.

Chen, Y., Liu, Y., & Zhang, J. (2012). When do third-party product reviews affect firm value and what can firms do? The case of media critics and professional movie reviews. *Journal of Marketing,* 76(2), 116-134.

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*.

Cheung, C. M. K., Lee, M. K. O., & Rabjohn, N. (2008). The impact of electronic word-ofmouth: The adoption of online opinions in online customer communities. *Internet Research,* 18(3), 229–247

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.

Choi, J. W., & Ok, C. (2011). The Effect of Online Restaurant Reviews on Diners' Visit Intention: A Comparative Analysis of Expert vs. Peer Reviews.

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.

Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce,* 17(1), 39-58.

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

Davenport, T. H., & Beck, J. C. (2001). The attention economy. *Ubiquity, 2001*(May), 1-es.

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. (2011). Retail luxury strategy: assembling charisma through art and magic. *Journal of retailing,* 87(4), 502-520.

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). 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

Ebersole, S. (2000). Uses and gratifications of the web among students. *Journal of ComputerMediated Communication,* 6(1), 1–17.

Eliashberg, J., Shugan, S. M. (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.

Eyal, K., & Rubin, A. M. (2003). Viewer aggression and homophily, identification, and para-social relationships with television characters. *Journal of Broadcasting & Electronic Media,* 47(1), 77–98.

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).

Filieri, R. (2016). What makes an online consumer review trustworthy? *Annals of Tourism Research,* 58, 46–64.

Filieri, R., Hofacker, C. F., & Alguezaui, S. (2018). What makes information in online consumer reviews diagnostic over time? The role of review relevancy, factuality, currency, source credibility and ranking score. *Computers in Human Behavior*, 80, 122–131.

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

Forman, C., Ghose, A., & Wiesenfeld, B. (2008). Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information systems research*, 19(3), 291-313.

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.

Frederick, E. L., Lim, C. H., Clavio, G., & Walsh, P. (2012). Why we follow: An examination of para-social interaction and fan motivations for following athlete archetypes on Twitter. *International Journal of Sport Communication,* 5, 481–502.

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

Gálik, M., & Csordás, T. (2020). A média gazdaságtanának kézikönyve. (M. Gálik & T. Csordás, Eds.). Budapest: Médiatudományi Intézet.

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*.

Haridakis, P., & Hanson, G. (2009). Social Interaction and Co-Viewing With YouTube: Blending Mass Communication Reception and Social Connection. *Journal of Broadcasting & Electronic Media,* 53(2), 317–335.

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.

Hennig-Thurau, T., Walsh, G., & Walsh, G. (2003). Electronic word-of-mouth: Motives for and consequences of reading customer articulations on the Internet. *International journal of electronic commerce,* 8(2), 51-74.

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.

Hoiles, W., Aprem, A., and Krishnamurthy, V. (2017). "Engagement and popularity dynamics of YouTube videos and sensitivity to meta-data", *IEEE Transactions on Knowledge and Data Engineering*, 29(7), 1426-1437.

Horton, D., & Wohl, R. R. (1956). Mass communication and para-social interaction: Observations on intimacy at a distance. *Psychiatry,* 19, 215–229.

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., 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.

Huang, J. H., & Chen, Y. F. (2006). Herding in online product choice. *Psychology & Marketing,* 23(5), 413–428.

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.

Jönsson, A. M., & Örnebring, H. (2011). User-generated content and the news: Empowerment of citizens or interactive illusion?. *Journalism Practice,* 5(2), 127-144.

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.

Katz, E. – Blumler, J. G. – Gurevitch, M. (1973b). Uses and Gratifications Research. *The Public Opinion Quarterly,* 37(4), 509-523.

Katz, E. – Gurevitch, M. – Haas, H. (1973a). On the use of the mass media for important things. *American Sociological Review,* 38(2), 164-181.

Katz, E., Blumler, J. G., & Gurevitch, M. (1974). Utilization of mass communication by the individual. In *J. G. Blumler & E. Katz (Eds.), The uses of mass communications: Current perspectives on gratifications research* (pp. 19–32). Beverly Hills, CA: Sage.

Kaye, B. K., & Johnson, T. J. (2002). Online and in the Know: Uses and Gratifications of the Web for Political Information. *Journal of Broadcasting & Electronic Media,* 46(1), 54–71.

Keh, H. T., & Sun, J. (2018). The Differential Effects of Online Peer Review and Expert Review on Service Evaluations. *Journal of Service Research,* 109467051877945.

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.

Khan, M. L. (2017). Social media engagement: What motivates user participation and consumption on YouTube? *Computers in Human Behavior,* 66, 236–247.

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

Klapper, J. T. (1963). Mass communication research: An old road resurveyed. *The Public Opinion Quarterly,* 27(4), 515–527.

Kostyra, D. S., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. International *Journal of Research in Marketing*, 33(1), 11-26.

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.

Liu, Z. and Park, S. (2015). "What makes a useful online review? Implication for travel product websites", *Tourism Management*, Vol. 47, pp. 140-151.

Lo, A.S. and Yao, S.S. (2019). "What makes hotel online reviews credible? An investigation of the roles of reviewer expertise, review rating consistency and review valence", *International Journal of Contemporary Hospitality Management*, Vol. 31 No. 1, pp. 41-60.

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.

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. International *Journal of Research in Marketing*, 34(2), 336-354.

Marchis, A. and Markos-Kujbus, É. (2019). Minden jó influencer…" – Avagy hogyan lehet hiteles egy Instagram-mikroinfluencer az ügynökségek és az influencerek szerint?

In *DMMD ADAPTER Tanulmányok a digitális marketing, média és design területéről. Budapesti Corvinus Egyetem, Budapest,* (pp. 19-33.). ISBN 978-963-503-798-8

Markos-Kujbus, É. (2017). Az online szájreklám (e-WOM) mint marketingkommunikációs eszköz. Az online fogyasztói vélemények információs szerepe a TripAdvisor példáján keresztül (Ph.D. értekezés, Budapesti Corvinus Egyetem).

McQuail, D. (1987). Mass communication theory: An introduction. *Sage Publications,* Inc.

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*.

Micheli, P., & Gemser, G. (2016). Signaling strategies for innovative design: A study on design tradition and expert attention. *Journal of Product Innovation Management,* 33(5), 613-627.

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

Moon, S., & Kamakura, W. A. (2017). A picture is worth a thousand words: Translating product reviews into a product positioning map. *International Journal of Research in Marketing*, 34(1), 265-285

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.

Naujoks, A., & Benkenstein, M. (2020). Expert cues: how expert reviewers are perceived online. *Journal of Service Theory and Practice.*

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.

Neuberger, C., & Nuernbergk, C. (2010). Competition, complementarity or integration? The relationship between professional and participatory media. *Journalism practice*, 4(3), 319-332.

Neuberger, C., Nuernbergk, C., & Langenohl, S. (2019). Journalism as Multichannel Communication: A newsroom survey on the multiple uses of social media. *Journalism Studies*, 20(9), 1260-1280.

Newman D. (2014). [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

Nguyen, H. T., & Chaudhuri, M. (2019). Making new products go viral and succeed. *International journal of research in marketing*, 36(1), 39-62.

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).

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.

Papacharissi, Z., & Rubin, A. M. (2000). Predictors of Internet use. *Journal of Broadcasting & Electronic Media,* 44, 175–196.

Parikh, A. A., Behnke, C., Almanza, B., Nelson, D., & Vorvoreanu, M. (2016). Comparative content analysis of professional, semi-professional, and user-generated restaurant reviews. *Journal of Foodservice Business Research,* 20(5), 497–511.

Park, S. and Nicolau, J.L. (2015). "Asymmetric effects of online consumer reviews". *Annals of Tourism Research,* Vol. 50, pp. 67-83.

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.

Perse, E. M., & Rubin, R. R. (1989). Attribution in social and para-social relationships. *Communication Research,* 19, 59–77

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.

Racherla, P., & Friske, W. (2012). Perceived "usefulness" of online consumer reviews: An exploratory investigation across three services categories. *Electronic Commerce Research and Applications,* 11(6), 548–559.

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.*

Rosengren, K. E. (1974). Uses and gratifications: A paradigm outlined. In J. G. Blumler & E. Katz (Eds.). *The uses of mass communications: Current perspectives on gratifications research* (pp. 269–286). Beverly Hills, CA: Sage.

Rubin, A. M., & Step, M. M. (2000). Impact of Motivation, Attraction, and Parasocial Interaction on Talk Radio listening. *Journal of Broadcasting & Electronic Media,* 44(4), 635–654.

Rubin, A. M., Perse, E. M., & Powell, R. A. (1985). Loneliness, para-social interaction, and local television news viewing. *Human Communication Research,* 12(2), 155–180.

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.

Shaker, F. – Hafiz, R. (2014). Personal Branding in Online Platform. *Global Disclosure of Economics and Business*, 3(2), 109–120.

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

Shao, G. (2009). Understanding the appeal of user-generated media: a uses and gratification perspective. *Internet research.*

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.

Sjoberg, U. (1999). The rise of the electronic individual: A study of how young Swedish teenagers use and perceive the internet. *Telematics & Informatics,* 16(3), 113–133.

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.

Smith, D., Menon, S., and Sivakumar, K. (2005). Online peer and editorial recommendations, trust, and choice in virtual markets. *Journal of Interactive Marketing,* 19, 3, 15–37

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.

Spence, Mark T. and Merrie Brucks (1997). “The Moderating Effects of Problem Characteristics on Experts’ and Novices’ Judgments*,” Journal of Marketing Research,* 34 (2), 233-47

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.

Sussman, S. W., & Siegal, W. S. (2003). Informational Influence in Organizations: An Integrated Approach to Knowledge Adoption. *Information Systems Research,* 14(1), 47–65.

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.

Tan, K.W.P., Swee, D., Lim, C., Detenber, B.H. and Alsagoff, L. (2008). “The impact of language variety and expertise on perceptions of online political discussions”. *Journal of Computer-Mediated Communication*, Vol. 13 No. 1, pp. 76-99.

Tarnovskaya, V. (2017). Successful Personal Branding on Social Media Building a Personal Brand through Content on YouTube. *Presented at the 12th Global Brand Conference Kalmar*.

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.

Thao, T., & Shurong, T. (2020). Is it possible for" electronic word-of-mouth" and" usergenerated content" to be used interchangeably. *Journal of Marketing and Consumer Research*, *65*, 41-48.

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.

Van Herpen, E., Pieters, R., & Zeelenberg, M. (2009). When demand accelerates demand: Trailing the bandwagon. *Journal of Consumer Psychology,* 19(3), 302-312.

Varga, Á. – Panyi, K. (2018). Híres leszek! - A magyar YouTube piac influencer központú vizsgálata. *Vezetéstudomány / Budapest Management Review,* 49(12), 24-30. doi: 10.14267/VEZTUD.2018.12.03

Varga, Á., & Sujbert, V. (2018). Énmárkázás online: A Youtuberek márkaépítése során használt tartalomelemek analízise. *A hatékony marketing: EMOK.*

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.

Westbrook, Robert A. (1987). "Product/Consumption-Based Affective Responses and Postpurchase Processes". *Journal of Marketing Research*, 24 (3), 258-70.

Wolfradt, U., & Doll, J. (2001). Motives of adolescents to use the Internet as a function of personality traits, personal and social factors. *Journal of Educational Computing Research,* 24(1), 13–27.

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

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.