Essays in Media Economics

Inauguraldissertation
zur
Erlangung des Doktorgrades
der
Wirtschafts- und Sozialwissenschaftlichen Fakultät
der
Universität zu Köln
2020

vorgelegt von
M.Sc. Anna-Kristina Kerkhof
aus Kleve

Referent: Prof. Dr. Johannes Münster

Korreferent: Prof. Dr. Felix Bierbrauer

Tag der Promotion: 19.05.2020

Acknowledgments

I am deeply indebted to my main advisor Johannes Münster for his support and guidance since my early undergraduate studies. Without you, I would not be where I am. You always took me seriously and I could always trust you. Thank you so much.

Many thanks also to my second advisor Felix Bierbrauer, who helped me see the bigger pictures.

I would also like to thank Jonas Loebbing and Sebastian Mertesacker for countless scientific and non-scientific discussions over lunch and dinner. You do not know how much I learned from you.

Annika Bohnhorst, Julian Freitag, Marina Happ, Sebastian Mertesacker, Sarah Recht, Leja Riedl, Marius Vogel, and Youpeng Zhang provided excellent research assistance. Thanks for your help!

I also thank Maja Adena, Pauline Affeldt, Ruediger Bachmann, Kai Barron, Davide Cantoni, Joerg Claussen, Gregory Crawford, Markus Dertwinkel-Kalt, Milena Djourelova, Ruben Durante, Marc Egger, Matthew Ellman, Florian Englmaier, Ruben Enikolopov, Thomas Fackler, Oliver Falck, Razi Farukh, Tobias Föll, Marcel Garz, Fabrizio Germano, Paul Gorny, Anna Gumpert, Oliver Gürtler, Emanuel Hansen, Julian Harke, Paul Heidhues, Matthias Heinz, Marit Hinnosaar, Erik Hornung, Steffen Huck, David Jaeger, Andre Kaiser, Ulrich Kaiser, Reinhold Kesler, Tobias Kretschmer, Ulrich Laitenberger, Johannes Loh, Leonardo Madio, Gregory Martin, Simon Martin, Frank May, Felix Mindl, Florian Morath, Axel Ockenfels, Martin Peitz, Christian Peukert, Pia Pinger, Susanne Prantl, Andrea Prat, Sven-Oliver Proksch, Riccardo Puglisi, Martin Quinn, Alexander Rasch, Markus Reisinger, Maximilian Schaefer, Paul Schempp, Olga Slivko, Robert Somogyi, David Strömberg, Daniel Sturm, Frederic Thenee, Kevin Tran, Marius Vogel, Van Anh Vuong, Fabian Waldinger, Alexander Westkamp, Christoph Wigger, Ilya Zarubin, and Dainis Zegners for helpful comments and suggestions.

Finally, I would like to thank my entire family for their support and their patience during difficult times, when I was moody, stressed out, and lost sight for what is really important in life. You always believe in me. Thank you so much.

Contents

List of Tables

List of Figures

1 Introduction

1.1 Overview of the thesis

Media markets play an essential role for modern societies. A free, unbiased, and diverse media landscape sustains democracy, freedom, and the public discourse, as the media inform, set the agenda, and influence both voters and politicians (Downs, 1957; Coase, 1974; Gentzkow et al., 2016; Puglisi and Snyder, 2016). Indeed, state-of-the-art research demonstrates that media content has a causal effect on the economic and political choices of individuals (DellaVigna and La Ferrara, 2015). It is therefore crucial to understand what drives the individual content choices of media outlets and how potential biases in a media market can be assessed and mitigated.

This thesis contributes to the understanding of media markets by studying the determinants of media outlets' content choice, by developing novel techniques to assess media bias, and by analyzing the welfare effects of potential regulations. To this end, the thesis is organized around the following research questions:

1. **How can we achieve a diverse media landscape?**
2. **How can we measure media bias?**
3. **How can we mitigate media bias?**

Chapter 2, **"Advertising and Content Differentiation: Evidence from YouTube"**, addresses the first research question and demonstrates that advertising has a causal positive effect on content differentiation on YouTube. In particular, I show that an exogenous increase in the technically feasible advertising quantity – in other words, the lift of an advertising cap – reduces the YouTubers' probability to duplicate mainstream content. A likely driver of this result is that consumers perceive advertising as a nuisance and thereby similar to a price they have to pay. Moreover, mainstream content – i.e., content in high demand by the consumers – is provided by many competing YouTubers. Hence, if a YouTuber who provides mainstream content increased her advertising quantity, her consumers could easily switch to a competitor. Losing consumers is less likely, however, if the YouTuber differentiates her content from her competitors and thereby makes her videos less substitutable.

My results make an important contribution to the existing literature; to the best of my knowledge, "Advertising and Content Differentiation: Evidence from YouTube" is the first paper that provides evidence of a causal positive effect of advertising on content differentiation. The paper thereby adds to the persistent and unresolved debate about the consequences of advertising on content differentiation in media markets (e.g., Herman and McChesney, 1997; Hamilton, 2004; McChesney, 2004; Gabszewicz et al., 2001; Huang et al., 2018; Anderson and Jullien, 2016) and contributes to recent discussions about the effect of digitization on content differentiation and diversity in media markets (Anderson, 2006; Brynjolfsson et al., 2003, 2011; Waldfogel, 2017, 2018; Goldfarb and Tucker, 2019).

The media are repeatedly accused of being biased towards the political left or right (Puglisi and Snyder, 2016). Thus, to address my second research question, Chapters 3, 4, and 5 develop novel techniques and applications to measure political media bias. Chapter 3, **"Incumbency Dominance in Letters to the Editor: Field Experimental Evidence"**, presents the results of a randomized field experiment three weeks before the 2017 federal elections in Germany. We wrote four different versions of a letter to the editor that differed in the subject the letter was about (Chancellor Merkel versus her main challenger Schulz) and in the evaluation of this subject (positive versus negative). We sent one randomly drawn version to each of over 200 German daily newspapers and observed whether the letter was published or not. The experimental design allows us to test three hypotheses on the (systematic) selection of media content on behalf of news outlets: political bias, negativity bias, and incumbency dominance. We find no political bias in the selection of letters to the editor and no statistically significant negativity bias. We do observe incumbency dominance, though: letters about the Chancellor were 40% more likely to be printed.

In Chapter 4, **"Coverage Bias on Wikipedia? Evidence from Biographies of German Members of Parliament"**, we study if the individual contributions of Wikipedia users lead to the unbalanced coverage of otherwise comparable MPs from different political parties. The central challenge of this analysis is to disentangle the effect of party affiliation on Wikipedia coverage from the effect of an MP's characteristics. We address this issue in two steps. First, we study a sample of relatively homogeneous observations. We consider the 18th German Bundestag (2013 to 2017) and focus on MPs from Germany's two biggest political parties, the center-right CDU/CSU and the center-left SPD, who jointly comprised more than three quarters of all MPs and formed a coalition government. Second, we compare the length of German and English Wikipedia biographies in a difference-in-differences framework. Partisan contributors are less likely to amplify the English biographies, since German voters are unlikely to read them. Assuming that unobserved MP characteristics affect the German and the English biography length equivalently, a difference-in-differences estimation using language as a first, and party affiliation as a second difference, yields unconfounded estimates of the effects of party affiliation on biography length. Our analysis reveals a small to medium size coverage bias against MPs from the SPD. The analysis of authorship patterns and talk pages shows that partisan contributions to Wikipedia are a plausible explanation for our results.

Chapter 5, **"Selective Sharing of News Items and the Political Position of News Outlets"**, presents a novel approach to measure the political position of online news outlets that is based on the selective sharing of news items by politicians on social media. Our central argument is that politicians predominantly share news items that are in line with their own political position, i.e., left-wing politicians prefer to share news items from left-wing news outlets, while right-wing politicians prefer to share news items from right-wing news outlets. Consequently, we can utilize the politicians' revealed preferences over news items to infer the political position of the news outlets. We apply our measure to twelve major German media outlets by analyzing tweets of German Members of Parliament on Twitter. For each news outlet under consideration, we compute the correlation between the political position of the seven parties in the 19th German Bundestag and their MPs' relative number of Twitter referrals to that outlet. We find that three outlets are positioned on the left, and two of them are positioned on the right.

Chapters 3, 4, and 5 make two major contributions to the literature on political media bias. First, measuring media bias is a difficult task, since researchers must find ways to overcome problems of subjectivity and the absence of suitable baselines against which to assess bias (e.g., Groeling, 2013). The chapters present three original approaches to measure political media bias that are applicable beyond German newspapers, German Members of Parliament, and German online news outlets. Second, knowledge on political media bias in online media outlets is scarce. Chapter 4 contributes to closing this gap by documenting coverage bias on Wikipedia, the world's largest online encyclopedia; relatedly, Chapter 5 exploits politicians' selective sharing of online news items via Twitter to determine the political position of online news outlets.

The neutrality of media outlets can also be compromised by the interference of advertisers on whom they financially depend (e.g., Ellman and Germano, 2009; Germano and Meier, 2013). For instance, advertisers prefer genres that put consumers in a more ad receptive mood (Wilbur, 2008) and may prefer the media not to report critically about their products (Blasco and Sobbrio, 2012). Indeed, commercial media bias is an increasingly important issue: given their severe financial situation, media outlets may become more and more susceptible to advertisers' pressures (FCC, 2011). Yet, not much is known about the determinants of commercial media bias and the mechanics of media markets in the presence of it.

To contribute to the understanding of this topic, Chapter 6, **"Quantity Restrictions on Advertising, Commercial Media Bias, and Welfare"**, studies commercial media bias that arises out of a conflict of interest between consumers and advertisers over media content. We analyze the welfare effect of a quantity restriction on advertising in a free-to-air television market in the presence of commercial media bias and thereby address research questions one and three. In our model, broadcasters face a trade-off between increasing the number of viewers by producing content that is highly valued by viewers, and increasing the price of advertising by choosing advertiser friendly content. A cap on advertising drives the per-viewer price of ads up; thus, content improves for viewers. Therefore, the cap can be welfare enhancing, even when viewers are not ad averse. Competition among broadcasters makes it more likely that a cap on advertising improves welfare. Thus, there is a complementarity between regulation and competition in this market.

Beyond contributing to the three broad research questions from above, the individual chapters of the thesis are linked through and contribute to three recurrent themes from the literature on media economics.

First, media markets are two-sided markets, which means that they finance their operations all or in part by advertising revenue instead of charging their consumers a monetary price (Anderson and Jullien, 2016). Yet, the effect of advertising on the individual media outlets' content choice is not well understood and provokes persistent debates about the consequences of advertising on content differentiation in media markets. In addition to that, the media's dependency on advertising revenue raises concerns about commercial media bias, i.e., about advertisers influencing the content choice of media outlets (e.g., Ellman and Germano, 2009; Germano and Meier, 2013; Blasco and Sobbrio, 2012). Chapter 2 of my thesis adds to this strand of literature by demonstrating that advertising has a causal effect on individual YouTubers' content choice; Chapter 6 contributes by analyzing a model of a two-sided media market with commercial media bias, where advertisers can exert pressure on the media outlets' content decisions.

Second, media content can in many circumstances be characterized as a public good. Moreover, media markets are of general interest since the working of these markets affects not only their active participants, but also generates important externalities, for example by helping citizens to take well-informed political decisions (see Batina and Ihori, 2005, for a review on the provision of public goods and Anderson and Coate, 2005, for the provision of public goods via advertising). While this applies to all media markets that are covered in this thesis, Chapters 4 and 6 emphasize this feature of media content.

Finally, media consumption is moving online; in particular, user-generated content is becoming more and more important (Luca, 2016b). Indeed, three out of the top five websites by Internet traffic – YouTube, Facebook, and Wikipedia – are based on user-generated content.[1] My thesis

[1]See https://www.alexa.com/topsites (Dec 2018).

takes this important development into account and focuses on online media markets. Chapter 2, for instance, studies the effect of advertising on content differentiation on YouTube, which is the world's most important user-generated content platform. Chapter 4 presents a measure of political media bias on Wikipedia and Chapter 5 uses the Tweets of German Members of Parliament Twitter to assess the political position of online media outlets.

1.2 Contribution to the co-authored chapters

Following the graduation regulations of the Faculty of Economics, Management, and Social Sciences at University of Cologne, this section explicates how I contributed to the co-authored chapters of this thesis.

The chapter "Advertising and Content Differentiation: Evidence from YouTube" is single-authored.

The chapter "Incumbency Dominance in Letters to the Editor: Field Experimental Evidence", published in *Political Communication (2019), Vol. 36:3, 337-356*, is joint work with Markus Dertwinkel-Kalt and Johannes Münster. The research idea was developed jointly after we taught the paper by Butler and Schofield (2010) in a course on media economics. Moreover, the research design and the hypotheses resulted from joint discussions. The implementation of the experiment was mostly done by me. The data collection and the data analysis were entirely done by myself.

The chapter "Coverage Bias on Wikipedia? Evidence from the Biographies of German Members of Parliament" is joint work with Johannes Münster. The research idea and the empirical strategy were developed jointly. Furthermore, the data collection was done by me and two research assistants. The data analysis was conducted entirely by myself. Finally, the current version of the paper was written by myself.

The chapter "Selective Sharing of News Items and the Political Position of News Outlets" is joint work with Julian Freitag and Johannes Münster. The paper builds on the research idea and data collection of Julian Freitag. My contributions to this project include pinning down the current version of the research question and the precise contribution to the existing literature, extending the data analysis, and re-writing the entire draft.

Finally, the chapter "Quantity restrictions on Advertising, Commercial Media Bias, and Welfare", published in *Journal of Public Economics (2015), Vol. 131, 124-141*, is joint work with Johannes Münster and outgrew an early idea that we developed together. Both of us contributed to the theoretical model and to writing down the final version of the paper.

2 Advertising and Content Differentiation: Evidence from YouTube

Awarded the Reinhard-Selten-Preis 2019 and the Best Paper Award at the 3rd Doctoral Workshop on the Economics of Digitization.

2.1 Introduction

Media diversity and content differentiation between media outlets are important for modern societies. A diverse media landscape sustains democracy, freedom, and the public discourse (Downs, 1957; Coase, 1974; Gentzkow et al., 2016; Puglisi and Snyder, 2016). Moreover, preferences over media content differ substantially across different groups of consumers (e.g., men and women prefer different types of media content) and the more differentiated the content in media markets, the more likely it is that all consumers' preferences are served (Waldfogel, 2007). Hence, it is no surprise that policy makers undertake great efforts to achieve and maintain diverse media markets.[1]

The media outlets' typical business model – charging low prices to attract consumers and generate revenue via advertising – provokes a persistent debate about the consequences of advertising on content differentiation, however. While media scholars argue that advertising revenue induces media outlets to duplicate mainstream content – i.e., content in high demand by the audience – to sell a maximum number of eyeballs to advertisers (e.g., Herman and McChesney, 1997; Hamilton, 2004; McChesney, 2004), predictions from economic theory are ambiguous. On the one hand, pioneering models on media outlets' content choice by Steiner (1952), Beebe (1977), and Gabszewicz et al. (2001) follow the above argumentation and predict that advertising leads to minimum differentiation à la Hotelling (1929). More recent models, on the other hand, acknowledge that many consumers perceive advertising as nuisance and thereby as a "price" they have to pay (Wilbur, 2008; Huang et al., 2018; Anderson and Jullien, 2016). Taking this into account leads to the opposite prediction: when incentivized by ad revenue, media outlets prefer to differentiate

[1] The European Council, for instance, has recently passed official guidelines for the protection of media diversity in the EU (CM/Rec(2018)1).

their content from each other to soften competition in the ad "price." Does advertising increase or diminish content differentiation in media markets? Empirical evidence on this question is scarce.

This paper studies the effect of advertising on content differentiation on YouTube – the world's most visited user-generated content platform[2] – to resolve the open question. I exploit two features of YouTube's monetization policy to identify the causal effect of advertising on the YouTubers' probability to duplicate mainstream content, where I define mainstream content as the content that attracts the largest number of views. First, I make use of the "ten minutes trick", which is a discontinuity in YouTube's mapping from video duration to the technically feasible number of ad breaks per video. If a video is shorter than ten minutes, YouTubers can permit for exactly one ad break in it. If the video is ten minutes or longer, YouTubers face no such limitation. Second, the ten minutes trick was unknown to the majority of YouTubers until Oct 2015, when YouTube launched a new ad break tool that made its existence prominent to the community.

To identify the effect of advertising on the YouTubers' content choice, I focus on YouTubers who produced short videos before Oct 2015, because those YouTubers were likely to be unaware of the ten minutes trick before the new ad break tool was launched. I classify a YouTuber as "treated" if she could increase her feasible number of ad breaks by increasing her share of videos that are ten minutes or longer *after* Oct 2015, and as control otherwise. Then, I compare the change in the probability to duplicate mainstream content before vs. after Oct 2015 for YouTubers in the treatment vs. the control group in a difference-in-differences framework.

Since the YouTubers have perfect control over their videos' duration, they might self-select into the treatment group. To account for self-selection in the YouTubers' treatment status, I use their *median video duration before Oct 2015* – in a sense, their "closeness" to the ten minutes threshold – as an instrument for being treated. The YouTubers in the sample did not choose their videos' duration *before* Oct 2015 bearing the ten minutes trick in mind, because they were unaware of the feature. As a result, a YouTuber's median video duration before Oct 2015 is uncorrelated to omitted variables that drive self-selection into the treatment group (e.g., commercial interests). On the other hand, extending their videos' duration to ten minutes or more is easier for YouTubers who were "closer" to the threshold *before* Oct 2015, i.e., median video duration before Oct 2015 is correlated to the YouTubers' (potentially endogenous) treatment status. A broad range of validity checks supports the identification strategy.

The analysis of around one million YouTube videos shows that an increase in the feasible number of ad breaks per video leads to a twenty percentage point reduction in the YouTubers' probability to duplicate mainstream content. The effect size is considerable: it corresponds to around 40% of a standard deviation in the dependent variable and to around 50% of its baseline value. The

[2] See www.alexa.com/topsites (July 2019)

large sample size allows me to conduct several subgroup analyses to study effect heterogeneity. I find that the positive effect of advertising on content differentiation is driven by the YouTubers who have at least $1,000$ subscribers, i.e., the YouTubers whose additional ad revenue is likely to exceed the costs from adapting their videos' content. In addition, I find heterogeneity along video categories: some categories are more flexible in terms of their typical video duration than others, hence, exploiting the ten minutes trick is more easy (e.g., a music clip is typically between three and five minutes long and cannot be easily extended).

Recent economic models on content choice in media markets acknowledge that consumers perceive advertising as a nuisance and similar to a "price" they have to pay; media outlets differentiate from each other to avoid competition in the ad "price" as a consequence (see Anderson and Jullien, 2016, for a survey). I show that the avoidance of ad "price" competition is a plausible economic mechanism behind my main results. First, I demonstrate that mainstream content – i.e., content in high demand – is also *supplied* by many YouTubers. Thus, viewers could easily switch to a competitor if a YouTuber increased her ad "price." Switching becomes less likely, however, when the YouTuber uploads content that is less mainstream and thereby covered by fewer competitors. Indeed, I find that an increase in the feasible number of ad breaks leads to a twenty percentage point reduction in the YouTubers' probability to upload content that is covered by many other YouTubers, too. Finally, I support this result by demonstrating that the audience of YouTubers who could increase the feasible number of ad breaks per video becomes more stable, i.e., the viewers become less likely to switch to competitors. I find no evidence for other economic mechanisms behind my results.

The paper makes at least two important contributions. First, I advance the knowledge on the effect of advertising on content differentiation in media markets. To my knowledge, this is the first paper that provides evidence of a causal *positive* effect of advertising on content differentiation, whereby it challenges the widespread opinion that the media inefficiently duplicate mainstream content when incentivized by ad revenue. This is a major insight, especially because the media's options to generate ad revenue are often subject to external regulation.[3]

Second, my results contribute to recent discussions about the effect of digitization on content differentiation and diversity in media markets (Waldfogel, 2017, 2018). The traditional cost structure of media markets – fixed costs are high and marginal costs are low – impedes media diversity, because the number of outlets that can co-exist is limited. Goldfarb and Tucker (2019), however, point out that digital technology has "reduced the cost of storage, computation, and transmission of data" (p.3). As a result, online media outlets can afford to provide niche content,

[3]The Audiovisual Media Services Directive, for instance, requires that the proportion of television advertising and teleshopping spots within a given clock hour shall not exceed 20% (Article 23 §1).

while enhanced search technologies enable consumers to find it – a phenomenon that Anderson (2006) summarizes as "the long tail."[4] YouTube serves as a point in case to study the determinants of content differentiation in digital media markets in general when fixed costs are low (e.g., online news markets or alternative user-generated content platforms). In particular, technology alone may not ensure a more diverse media landscape: although a large number of media outlets can co-exist, they might duplicate the most mainstream content, while niche preferences remain unserved. My paper shows that advertising provides additional incentives for media outlets to differentiate their content that – when falling on the fertile ground of digitization – can help to increase media diversity.

I contribute to two additional strands of literature. First, my paper adds to the extensive literature on horizontal product differentiation (see, e.g., Graitson, 1982; Gabszewicz and Thisse, 1986; Lancaster, 1990; Anderson et al., 1992), which shows that firms' degree of product differentiation is determined by two contrasting effects. On the one hand, a direct effect induces firms to move closer to their competitors to increase their consumer base, leading to *minimum differentiation* (Hotelling, 1929). On the other hand, a strategic effect prompts firms to move away from their competitors to soften price competition, which leads to *maximum differentiation* (d'Aspremont et al., 1979; Economides, 1986).[5] Accordingly, models on content differentiation in media markets that ignore consumers' ad aversion find that advertising leads to minimum content differentiation (Steiner, 1952; Beebe, 1977; Gabszewicz et al., 2001, 2002; Garcia Pires, 2014; Behringer and Filistrucchi, 2015). Models that acknowledge the conceptual equivalence between direct prices and consumers' nuisance costs from advertising, in contrast, predict that media outlets prefer to differentiate from each other to avoid ruinous competition in the ad "price" (Bourreau, 2003; Dukes, 2004; Gabszewicz et al., 2004; Peitz and Valletti, 2008; Anderson and Jullien, 2016).[6] My paper provides causal empirical evidence for the theoretical considerations from this literature. While a related paper by Seamans and Zhu (2014) shows that an increase in subscription prices is correlated to a higher degree of content differentiation, I demonstrate that an increase in the feasible number of ad breaks per video leads to content differentiation, because YouTubers want to soften competition in the ad "price." Most closely related to my work is Sun and Zhu (2013), who study the introduction of an ad-revenue-sharing program on a major Chinese online platform and find that advertising

[4]See also Brynjolfsson et al. (2003, 2011) for a discussion on the long tail and how consumer surplus benefits from increased product variety.

[5]De Palma et al. (1985) show that if the consumers are sufficiently heterogeneous in terms of their taste parameter, the direct effect prevails.

[6]Gal-Or and Dukes (2003) find minimal differentiation even if consumers are ad averse, but the result is driven by the assumption of informative advertising. When the outlets minimally differentiate their content, advertisers choose lower levels of advertising, because the consumers are ad averse. This implies lower levels of product information to consumers, whereby the advertisers gain higher margins on their products. As a result, the media outlets can set higher prices for advertisers (p.292).

leads to the duplication of mainstream content. Our results do not contradict each other, though. While ad breaks before or during YouTube videos are a true nuisance to viewers, Sun and Zhu (2013) explicitly state that the ads appearing on the bloggers' posts are not intrusive (p. 2317), which means that only a direct, but no strategic effect operates in their setting. The papers can therefore be seen as complements supporting the plausibility of each other's results.

In addition, my work makes three contributions to the literature on user-generated content (see Luca, 2016b, for a survey). First, I present a novel empirical strategy to identify causal effects on a user-generated content platform. While existing approaches use variation in institutional features *across* platforms (e.g., Chevalier and Mayzlin, 2006; Mayzlin et al., 2014), *within* platforms (Anderson and Magruder, 2012; Luca, 2016a), or conduct randomized experiments (Bond et al., 2012; Aral and Walker, 2012), I exploit two distinctive features of YouTube's monetization policy to identify the causal effect of advertising on the YouTubers' content choice. Second, I apply this novel identification strategy to a unique dataset of newly collected data on several thousand German YouTubers with more than a million videos that have not been investigated before. Third, my paper explores how monetization affects user-generated content. Since many other user-generated content platforms such as Wikipedia, TripAdvisor or Twitter do not allow their contributors to earn money, YouTube offers a unique environment to study this question. Previous analyses show that users contribute to user-generated content platforms for two main reasons: reputation (Wang, 2010; Anderson et al., 2013; Easley and Ghosh, 2013) and beliefs about a high impact of their contributions (Zhang and Zhu, 2011). These motives are non-pecuniary and render it unclear whether the YouTubers react to economic incentives at all. My results demonstrate that economic considerations matter. In particular, when incentivized by ad revenue, the YouTubers are willing to deviate from the content they provided before. Moreover, I show that ad revenue does not necessarily improve the YouTubers' video quality. Although the number of views goes up when a video has more ad breaks, the relative number of likes decreases.

The remainder of the paper is organized as follows. Section 2.2 provides background information on YouTube, its monetization policy, and the institutional features that the empirical strategy builds on. A stylized example introduces the central ideas of identification in Section 2.3, before I illustrate the data collection process and how I construct a dataset that is suitable for the analysis in Section 2.4. Section 2.5 discusses the details of the empirical strategy; the results are presented in Section 2.6. Next, in Section 2.7, I explore the economic mechanism that drives these results. Section 2.8 studies content differentiation in the aggregate; Section 2.9 investigates changes in video quality. Section 2.10 concludes.

2.2 YouTube: Background

2.2.1 Platform, audience, and contributors

YouTube is a video sharing platform founded in 2005 and acquired by Google in 2006. Its reach is tremendous: with 800 million unique users and 15 billion visits per month, it is the second-most popular website in the world (after google.com).[7] As of Oct 2018, several billion hours of video content from YouTube are watched every day.[8]

YouTube is based on user-generated content. While unregistered users are limited to watching, registered users can upload, share, and comment on videos. Registered users who upload videos on a regular basis are called *YouTubers*; YouTubers, in turn, operate a YouTube *channel* under their user name to distribute their videos.[9]

2.2.2 Monetization

YouTubers have the option to monetize their content; in particular, they can generate advertising revenue by permitting YouTube to show ads to viewers before or during their videos. However, while YouTubers can permit that ads *may* be shown, YouTube's algorithm determines *if* and *which* ad is displayed to a particular viewer. Thus, there is no direct relationship between YouTubers and advertisers.[10] According to anecdotal evidence – official statistics do not exist – YouTubers earn about three to five USD per 1,000 views per ad per video.[11]

Monetization via ad breaks is not open to all YouTubers, though. First, a YouTuber's content must be advertiser-friendly, i.e., free of violence, sex, and crime.[12] In early 2017, YouTube introduced a new policy of automated demonetization of non-advertiser-friendly content (also known as "adpocalypse") that aims at videos on sensitive social issues, tragedy, or conflict; many YouTubers reported losing more than half of their income as a result.[13] Second, while not bounded to a subscriber threshold before, YouTube disabled the monetization option for YouTubers with fewer than 1,000 subscribers in Feb 2018. This policy, too, is a reaction to advertisers' complaints about their products appearing next to dubious video content.[14] The subscriber threshold, YouTube

[7]See `www.alexa.com/siteinfo/youtube.com` (Oct 2018).
[8]See `www.youtube.com/yt/about/press/` (Oct 2018).
[9]I use the terms "YouTuber" and "channel" synonymously; cases where one YouTuber operates several channels are rare.
[10]In addition to permitting for ad breaks in their videos, YouTubers might also earn money through product placement and affiliate links. In this case, there exists a contractual basis with the advertiser.
[11]See `influencermarketinghub.com/how-much-do-youtubers-make/` (Dec 2018).
[12]See `support.google.com/youtube/answer/6162278?hl=en` (Dec 2018).
[13]See `nymag.com/intelligencer/2017/12/can-youtube-survive-the-adpocalypse.html` (Dec 2018).
[14]See `turbofuture.com/internet/YouTube-Screwed-Small-YouTube-Channels-With-`

argues, gives them enough information to determine the validity of a YouTuber's channel and to confirm that it is following the YouTube community guidelines and advertiser policies.[15]

2.2.3 The ten minutes trick

YouTube's monetization policy exhibits one distinctive feature, which is known as the "ten minutes trick." The ten minutes trick refers to a discontinuity in YouTube's mapping from a video's duration to the technically feasible number of ad breaks that the YouTuber can permit. If a video is shorter than ten minutes, YouTubers can permit for exactly one ad break in it. If, on the other hand, the video is ten minutes or longer, YouTubers face no technical restriction on the number of ad breaks.[16] Hence, the ten minutes trick can be summarized as

$$feasible\ number\ of\ ad\ breaks = \begin{cases} 1 & if\ video\ duration < 10\ min \\ \infty & if\ video\ duration \geq 10\ min. \end{cases} \quad (2.1)$$

While the ten minutes trick had long been a hidden feature, it gained sudden prominence in Oct 2015, when YouTube launched a new ad break tool for YouTubers.[17] The tool had two effects. First and foremost, it made the ten minutes trick apparent. In its old version, only a small additional input box would appear for videos exhibiting the ten minutes threshold (A in Figure 2.1). In contrast to that, the option to embed additional ad breaks is now permanently visible and points YouTubers to its existence (B in Figure 2.2). Second, editing additional ad breaks became less cumbersome. The new tool allows YouTubers to drag ad breaks back and forth on their video time line and it also offers a preview option to check whether an ad appears at an appropriate point in time during the video (C and D in Figure 2.2). The old version, in contrast, required typing and re-typing the point in time where the ad breaks were supposed to appear (A in Figure 2.1).

Their-New-Memorization-Policy (Dec 2018).
[15] support.google.com/youtube/answer/72857?hl=en&ref_topic=6029709 (Dec 2018).
[16] support.google.com/youtube/answer/6175006?hl=en (Oct 2018).
[17] See www.youtube.com/watch?v=z58Ed6q6xQg (Oct 2018).

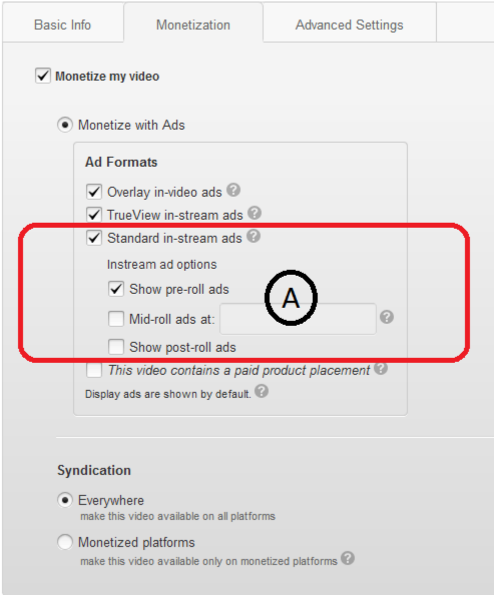

Figure 2.1: Old ad break tool (before Oct 2015).

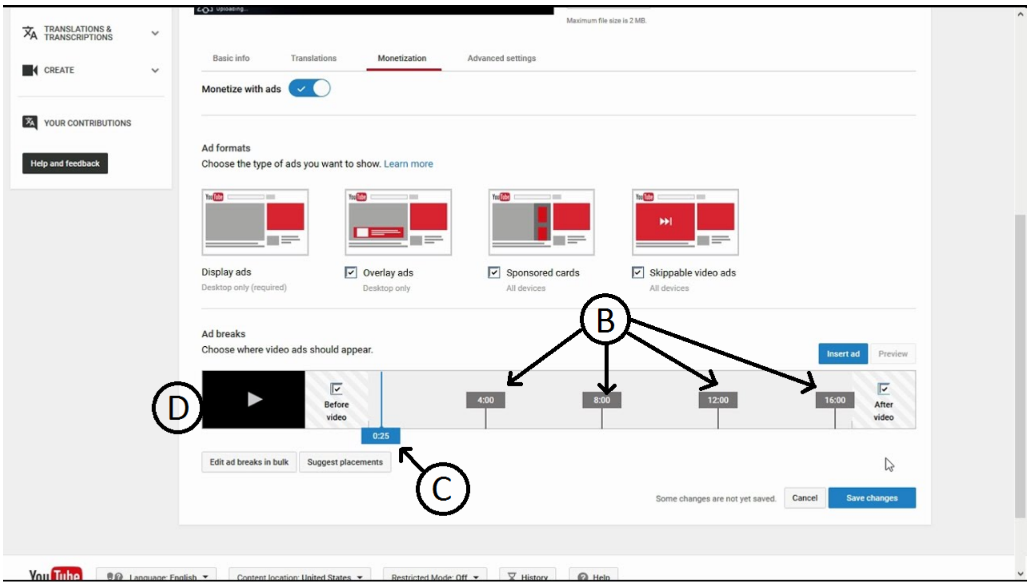

Figure 2.2: New ad break tool (after Oct 2015).

2.3 Identification: Stylized example

An ideal experiment would randomly assign some YouTubers to the option of showing just one, and others to the option of showing several ads per video to their viewers, and then compare the groups' probabilities to upload mainstream content. Given that the YouTubers' real life monetization settings are endogenous, however, the identification of a causal link from advertising to content choice requires a thoughtful empirical strategy. Though highly stylized, this section illustrates how combining the ten minutes trick with the launch of the new ad break tool yields variation in the YouTubers' feasible number of ad breaks per video that I exploit to identify the causal effect of interest.

Figure 2.3 illustrates YouTube's mapping from video duration to the technically feasible number of ad breaks per video as described in Section 2.2. Consider three hypothetical YouTubers A, B, and C *before* Oct 2015, where A's videos are very short, B's videos are close to but still below the ten minutes threshold, and C's videos are longer than that. Hence, while A and B may only permit for one ad break per video, C faces no such limitation. Note that this is no regression discontinuity setting, because the YouTubers have perfect control over their videos' duration. In particular, C could have chosen her videos' duration strategically to benefit from the jump in the feasible number of ad breaks per video.

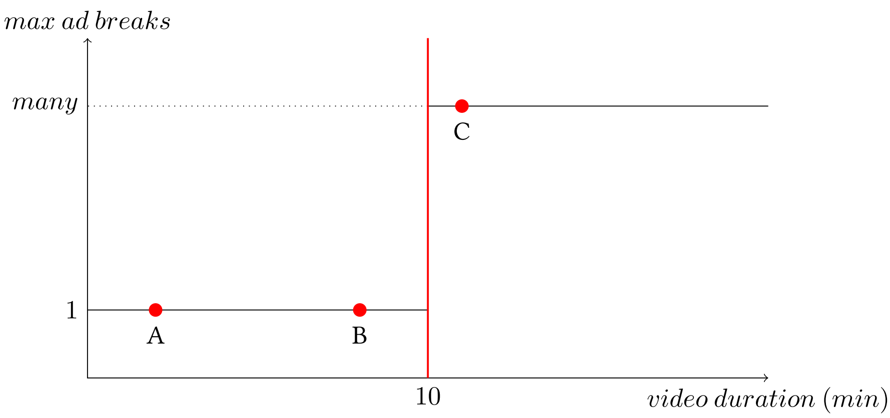

Figure 2.3: Stylized example of the identification strategy.

Next, consider the launch of the new ad break tool in Oct 2015. While C is unaffected, A and B realize that they can increase the feasible number of ad breaks per video by uploading videos that are ten minutes or longer. Pushing her video duration beyond the threshold, however, is easier to accomplish for B than for A. The key identifying assumption is that although a YouTuber has perfect control over her videos' duration, A and B, who were initially ignorant of the threshold's existence, did not choose their videos' *distance* to the ten minutes threshold having the discontinuity in mind. As a consequence, the cost of moving beyond the threshold *after* it became prominent – and thereby also the probability to actually do so – is orthogonal to unobserved characteristics such as, for instance, commercial incentives that may also drive a YouTuber's decision to increase her feasible number of ad breaks.[18]

I exploit the variation in the YouTubers' cost to move beyond the threshold as follows. First, I consider only YouTubers like A and B, i.e., YouTubers who were "left" to the ten minutes threshold before Oct 2015. Then, I compare the change in the probability to upload mainstream content before and after Oct 2015 of YouTubers who could increase the feasible number of ad breaks per video by uploading videos that are ten minutes or longer (treatment group) to YouTubers who did not do so (control group) in a difference-in-differences framework. Finally, I account for self-selection into the treatment group by using a YouTuber's "closeness" to the ten minutes threshold before Oct 2015 as an instrument for her treatment status. Thus, my empirical strategy boils down to exploit variation YouTubers who were close to the threshold before Oct 2015 to YouTubers who were further away from it (in contrast to comparing YouTubers just left to the threshold to YouTubers just right to it, as one would do in a regression discontinuity design). A detailed discussion of the empirical strategy follows in Section 2.5.

2.4 Data

2.4.1 Data collection

To carry out the above analysis, I collect data via the YouTube Data API and via HTML webscraping. First, I use the website `channelfinder.com` to compile a list of all active German YouTube channels as of Oct 2017. Based on this list, I collect data on the YouTuber level, including a full history of video uploads by each YouTuber, from the Data API. Finally, I retrieve data on the video level, including the date of upload, video duration, views, likes, dislikes, category, and keywords. Note that views, likes and dislikes are accumulative measures; thus, I retrieve these numbers as they are on the day of data collection.

Data on the YouTubers' monetization settings is, unfortunately, highly limited; the Data API,

[18] To be precise, A and B could correspond to three types of YouTubers: (i) those who did not know about the threshold, as discussed above, (ii) those who knew about the threshold, but found it too cumbersome to permit for additional ad breaks, and (iii) those who knew but did not want to increase their videos' duration. The logic that applies to YouTubers in group (i) holds for YouTubers in group (ii) as well. YouTubers in group (iii) can be interpreted as "never takers", see Section 2.5.2 for a discussion of IV heterogeneity.

for instance, does not provide any information regarding a video's number of ad breaks. Moreover, YouTube technically prohibits any automated program from collecting data "faster than a human could."[19] Hence, although the permitted ad breaks are detectable in a video's HTML code, a webscraper could not crawl each video in the dataset within a reasonable amount of time. Instead, I let a webscraper crawl twenty randomly drawn videos per YouTuber.[20] If it detects at least one ad break in at least one video, I classify the YouTuber as "advertising YouTuber", and as "non-advertising YouTuber" otherwise. This compromise allows me to collect monetization data on the YouTuber level for all YouTubers in my dataset, but forgoes more fine-grained information on the video level. Appendix A.2.1 discusses the consequences of a potential measurement error.

2.4.2 Definition of mainstream content

I use the number of video views and the videos' keywords to construct a measure for mainstream content. Each video is given illustrative keywords by its YouTuber – for instance, a funny cat video would be equipped with the keywords "funny" and "cat" – which help viewers to find them via YouTube's search engine.[21] For each month, for each video category, I compute how many views a certain keyword has attracted and rank them in descending order; the upper one percent of the keywords in this distribution is then classified as "mainstream."[22] Finally, I assign a dummy variable that is equal to one to all videos equipped with a mainstream keyword.[23] Note that it is important to consider each month and each video category separately. First, what is mainstream is likely to change over time, second, different video categories attract very different audiences whose preferences need to be considered separately. Moreover, it is crucial to define mainstream content based on the universe of *all* active YouTubers, i.e., before I exclude observations to construct the final dataset. Otherwise, I would compute the most mainstream keywords within the sample of YouTubers selected for the main analysis (see Section 2.4.3), which is conceptually different.

Take the category "Science & Technology" in April 2015 as an example. Videos are given $13,555$ different keywords; the three most viewed are "diy", "homemade", and "selfmade". Figure 2.4 shows that the distribution of views over keywords is heavily skewed: a small number of keywords accounts for a large part of the views. For instance, the upper one percent of the keywords accounts for 45.1%, while the lowest ten percent of the keywords account for just

[19] See `www.youtube.com/static?gl=de&template=terms&hl=en` (Oct 2018).

[20] The webscraper pauses for eight seconds before proceeding to the next video; crawling each video this way would take several years. Crawling twenty videos per YouTuber, in contrast, is feasible within three and four months.

[21] If a video is not given keywords, I generate keywords from its title.

[22] I ignore trivial keywords that appear in the video categories' titles. For instance, I ignore "people" and "blog" for videos in the category "People & Blogs" and "science" and "technology" for videos in the category "Science & Technology."

[23] In that, I follow the procedure by Sun and Zhu (2013), only that instead of blogs' hashtags I use videos' keywords.

0.02% of the views. The numbers are similar for other categories and other points in time.

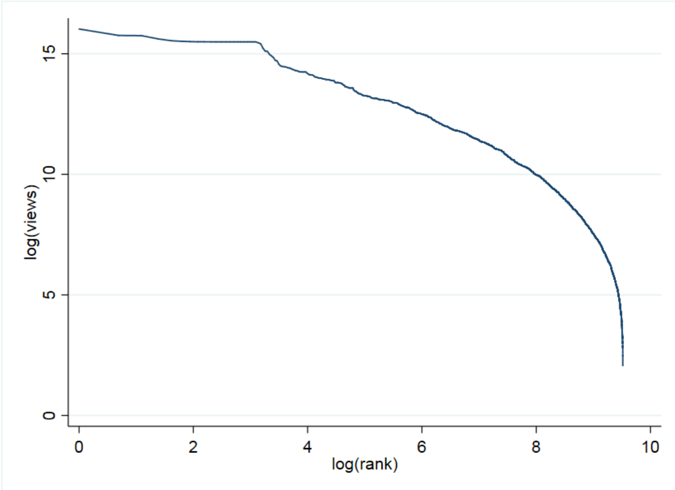

Figure 2.4: Log-log plot of the number of views a keyword attracts and its associated rank in the category "Science & Technology" in March 2015.

2.4.3 Final dataset

In a last step, I construct my final dataset. First, I define an appropriate observation period. The central event – the launch of the new ad break tool – took place in Oct 2015. Including videos uploaded between Jan 2013 and Jan 2017 into the final dataset yields a sufficient number of before and after observations. At the same time, this choice excludes both videos that are too old – and therefore not well comparable to more recent ones in terms of content or duration – as well as videos that were too "recent" on the date of data collection. By leaving at least nine months between the latest upload of a video and the data collection process (that started in Oct 2017) all videos in my dataset can be considered as "old", which minimizes any potential bias that may arise through the accumulative nature of some descriptive variables such as likes, dislikes, and views. Moreover, an observation period from Jan 2013 to Jan 2017 excludes the two big demonetization waves from 2017 and 2018 (see Section 2.2) that could have affected the YouTuber's content choice. Robustness checks on my main results using other observation periods and a summary of minor events that occurred between Jan 2013 and Jan 2017 are presented in Appendix A.1.1 and Appendix A.2.2.

Second, I determine which YouTubers to include. Following the outline from Section 2.3, I restrict the analysis to YouTubers "left" to the ten minutes threshold before Oct 2015 (YouTubers A and B in the example), where I use a YouTuber's median video duration before Oct 2015 to define her "position" on the x-axis in Figure 2.3. Thus, I include only YouTubers whose median video duration before Oct 2015 is smaller than ten minutes into the final dataset. In addition, I include only YouTubers who uploaded at least one video before and after Oct 2015. Finally, due to the "adpocalypse" (see Section 2.2), I exclude all videos from the category "News & Politics", since many of these videos were forcefully demonetized by YouTube. The final panel dataset includes $15,877$ YouTubers with $1,349,267$ videos over a time period of 49 months. Table 2.1 summarizes all variables used in the main analysis. Appendix A.1.2 shows several robustness checks based on different selections of YouTubers.

2.4.4 Illustrative evidence

Based on the final dataset, this section provides illustrative evidence of the two major arguments in Section 2.3. That is, I confirm that the launch of the new ad break tool made the ten minutes trick more apparent and that YouTubers who were closer to the ten minutes threshold before Oct 2015 are more likely to exploit it. In addition, I provide video level evidence for an increase in the actual (not the feasible) number of ad breaks per video.

First, Figure 2.5 demonstrates that the advertising YouTubers' share of videos between ten and fourteen minutes increases after Oct 2015. The non-advertising YouTubers, on the other hand, are unaffected. The diverging trends confirm that the launch of the new ad break tool in Oct 2015 made the ten minutes trick more apparent to the advertising YouTubers.

In what follows, I consider only advertising YouTubers. A further comparison of advertising and non-advertising YouTubers is problematic, since they act based on entirely different motives: non-advertising YouTubers had neither chance nor interest to exploit the ten minutes trick at any point in time. Thus, I exclude the non-advertising YouTubers from my main analysis, but come back to them for falsification checks in Section 2.6.2.

As a second step, I show that (advertising) YouTubers who were closer to the ten minutes threshold before Oct 2015 are more likely to exploit it. Since "closeness" – in terms of a YouTuber's median video duration (see Section 2.4.3) – is a continuous measure, I cannot compare the trends of distinct groups, though. Instead, Figure 2.6 illustrates that the increase in YouTubers' share of videos between ten and fourteen minutes after Oct 2015 is stronger for YouTubers whose "closeness" is around the 75^{th} percentile of its distribution than for YouTubers around the 25^{th} percentile.

Table 2.1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
$Mainstream_{vit}$	0.425	0.494	0	1	1,397,267
$Competitive_{vit}$	0.641	0.480	0	1	1,397,267
$Advertising_i$	0.764	0.425	0	1	15,877
$post_t$	0.475	0.499	0	1	1,397,267
D_i	0.226	0.418	0	1	15,877
$Duration_{vit}$	6.411	13.341	0	1440.033	1,397,267
$Subscribers_i$	18,234.506	138,282.229	0	6,581,640	15,877
$Film\&Animation_{vit}$	0.086	0.280	0	1	1,397,267
$Cars\&Vehicles_{vit}$	0.081	0.272	0	1	1,397,267
$Music_{vit}$	0.025	0.155	0	1	1,397,267
$Pets\&Animals_{vit}$	0.026	0.159	0	1	1,397,267
$Sports_{vit}$	0.085	0.278	0	1	1,397,267
$Travel\&Events_{vit}$	0.056	0.229	0	1	1,397,267
$Let'sPlay_{vit}$	0.085	0.278	0	1	1,397,267
$People\&Blogs_{vit}$	0.202	0.402	0	1	1,397,267
$Comedy_{vit}$	0.015	0.121	0	1	1,397,267
$Entertainment_{vit}$	0.201	0.401	0	1	1,397,267
$HowTo\&Style_{vit}$	0.064	0.245	0	1	1,397,267
$Education_{vit}$	0.046	0.210	0	1	1,397,267
$Science\&Technology_{vit}$	0.014	0.119	0	1	1,397,267
$Nonprofit\&Activism_{vit}$	0.015	0.120	0	1	1,397,267
$I(1^{st}to10^{th})_{vit}$	0.673	0.469	0	1	1,397,267
$I(10^{th}to25^{th})_{vit}$	0.581	0.493	0	1	1,397,267
$I(25^{th}to50^{th})_{vit}$	0.548	0.498	0	1	1,397,267
$I(50^{th}to75^{th})_{vit}$	0.390	0.488	0	1	1,397,267
$I(75^{th}to100^{th})_{vit}$	0.284	0.451	0	1	1,397,267
$SumAffiliations_{vit}$	2.472	1.160	0	5	1,397,267
$Likes_{vit}$	631.945	5,993.188	0	1,269,177	1,397,267
$Dislikes_{vit}$	33.452	532.98	0	149,614	1,397,267
$Views_{vit}$	35,098.614	564,233.77	0	337,832,408	1,397,267

Notes: This table presents the summary statistics of all variables used in the analysis. The variables $Mainstream_{vit}$, $Competitive_{vit}$, $Advertising_i$, $post_t$, D_i, all percentile indicators, and all category indicators are dummy variables. The variables $Advertising_i$, D_i, $close_i$, and $Subscribers_i$ are available only on the YouTuber level.

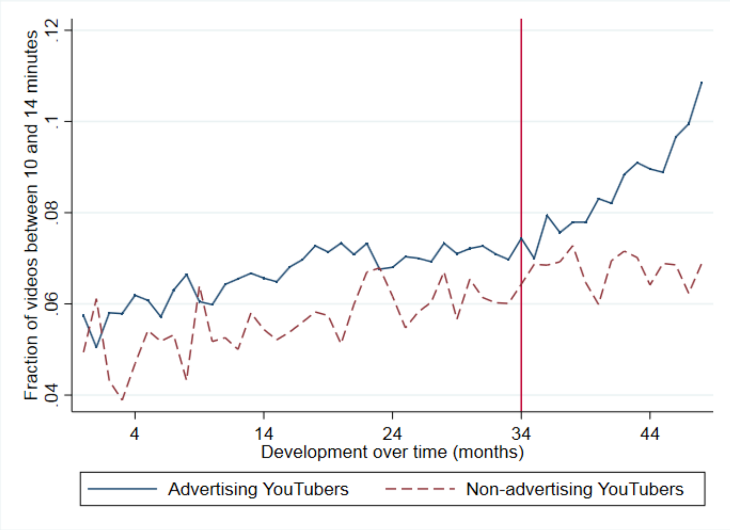

Figure 2.5: Trends in advertising vs. non-advertising YouTubers' share of videos between ten and fourteen minutes. The vertical line depicts Oct 2015.

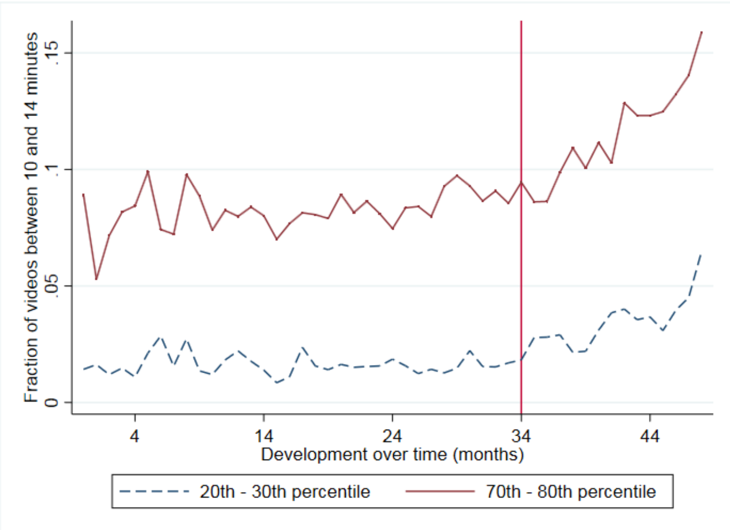

Figure 2.6: Trends in advertising YouTubers' share of videos between ten and fourteen minutes. The vertical line depicts Oct 2015.

In addition to that, I examine the distribution of video durations before and after Oct 2015 for the same two groups of YouTubers. Figures 2.7 to 2.10 illustrate three important facts. First, if the YouTubers increase their videos' duration after Oct 2015 to benefit from YouTube's discontinuous mapping from video duration to the feasible number of ad breaks, one should see bunching just behind the ten minutes threshold *after* Oct 2015. Figures 2.8 and 2.10 show that this is the case. In addition, Figures 2.8 and 2.10 illustrate that it is appropriate to focus on the share of videos between ten and fourteen minutes: if the YouTubers exploit the ten minutes trick after Oct 2015, they start to upload videos that *just* enable them to increase the feasible number of ad breaks per video. Second, if exploiting the ten minutes threshold it is less costly for YouTubers whose median video duration was closer to the ten minutes threshold before Oct 2015, the bunching should be more pronounced for YouTubers with a higher median video duration before Oct 2015; Figures 2.8 and 2.9 confirm that this is the case. Third, the distributions of video durations before Oct 2015 in Figures 2.7 and 2.9 document that the dataset is likely to be limited to YouTubers who were ignorant of the ten minutes trick before the launch of the new ad break tool, since – in contrast to Figures 2.8 and 2.10 – the distributions of video durations are smooth around the ten minutes threshold.

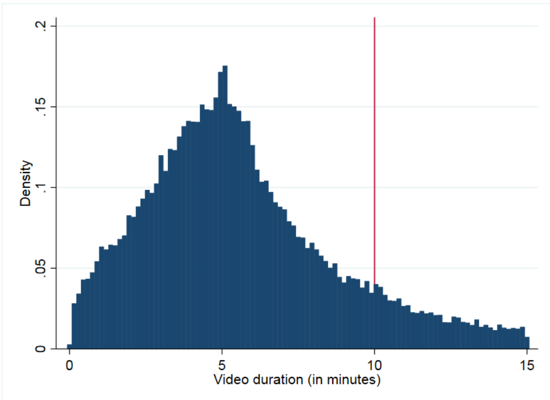

Figure 2.7: Histogram of the distribution of video durations *before* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

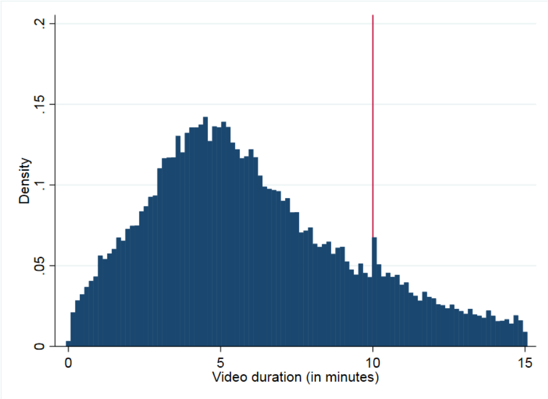

Figure 2.8: Histogram of the distribution of video durations *after* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration after Oct 2015. The vertical line depicts the ten minutes threshold.

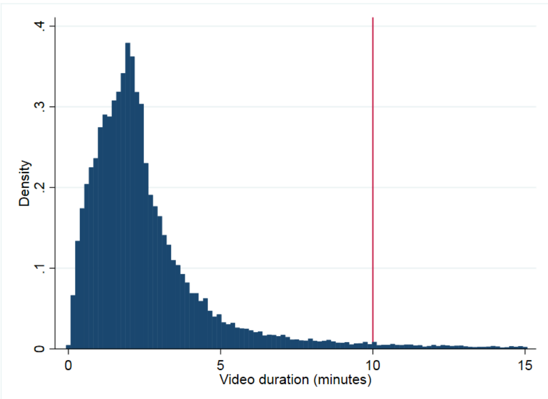

Figure 2.9: Histogram of the distribution of video durations *before* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

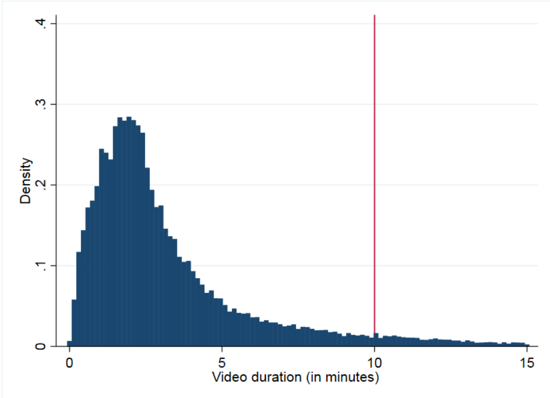

Figure 2.10: Histogram of the distribution of video durations *after* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration after Oct 2015. The vertical line depicts the ten minutes threshold.

I augment this illustrative evidence with the results from a formal McCrary test (McCrary, 2008), which is based on an estimator for the discontinuity at a given threshold in the density function of the running variable (H_0: no discontinuity). Here, I apply a McCrary test to obtain a measure for the discontinuity in the distributions of video durations in Figures 2.7 to 2.10. The results are displayed in Figures 2.11 to 2.14. Before Oct 2015, in Figures 2.11 and 2.13, the estimates for the discontinuities are small for both groups of YouTubers. In contrast to that, the estimate discontinuity after Oct 2015 is still small in Figure 2.14, but much more pronounced in Figure 2.12, where I consider the YouTubers whose median video duration was closer to the ten minutes threshold. Estimates and standard errors of the discontinuities can be found in Table 2.2.

Finally, I show that the *actual* number of permitted ad breaks in videos that are ten minutes or longer has increased. To this end, I draw a random subsample of 500 advertising YouTubers and collect *video level data* on their monetization settings ($52,462$ videos).[24] I consider only videos that are ten minutes or longer. I find that the average number of ad breaks in these videos has grown from 0.86 before Oct 2015 to 1.04 after Oct 2015, which corresponds to an increase of 20%. Moreover, the share of videos that has more than one ad break has increased from 17.7% to 20.7%. Finally, while 23 is the largest number of ad breaks in a single video before Oct 2015, this number has risen to 52 after Oct 2015. Thus, in the random subsample, the actual number of ad breaks has increased on the intensive as well as on the extensive margin.

[24]Collecting this fine grained data is only feasible for a small subsample of YouTubers; see Section 2.4.1 for details.

Table 2.2: McCrary test

Figure	Estimate
70th to 80th percentile pre Oct 2015 (Figure 2.11)	0.1654*** (0.0640)
70th to 80th percentile post Oct 2015 (Figure 2.12)	0.4049*** (0.0604)
20th to 30th percentile pre Oct 2015 (Figure 2.13)	0.0035 (0.1609)
20th to 30th percentile post Oct 2015 (Figure 2.14)	0.2659** (0.1356)

Notes: Standard errors in parentheses. The estimates depict discontinuity estimates (log difference in height) of a McCrary test with bin width 1 and band width 60. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

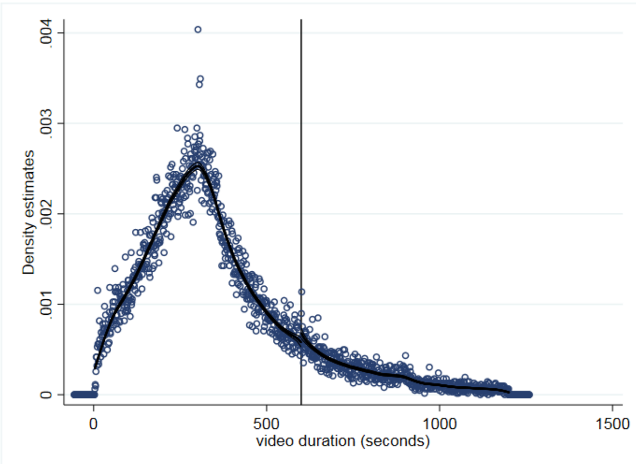

Figure 2.11: McCrary test of the distribution of video durations *before* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

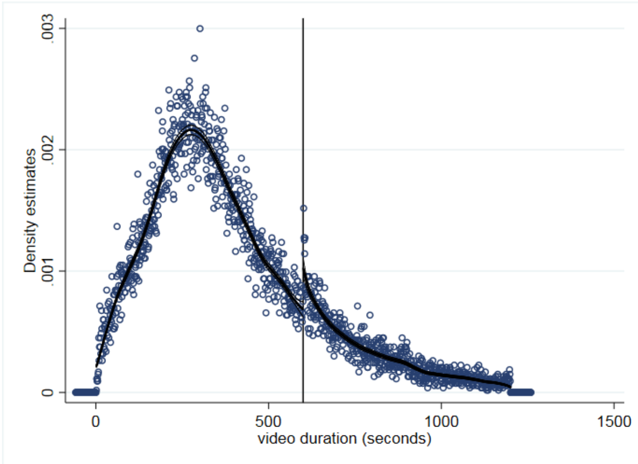

Figure 2.12: McCrary test of the distribution of video durations *after* Oct 2015 for YouTubers from the 70th to 80th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

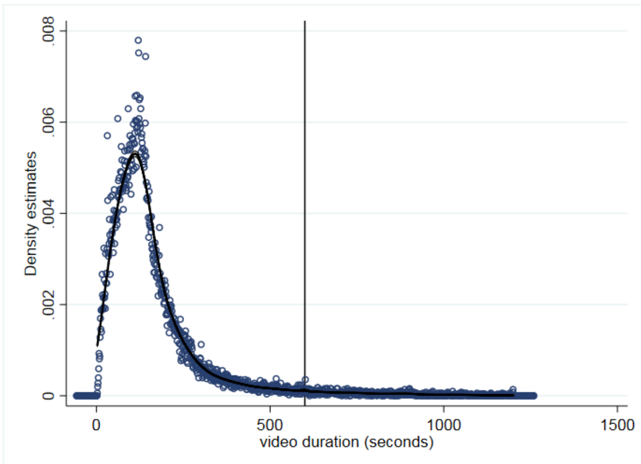

Figure 2.13: McCrary test of the distribution of video durations *before* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

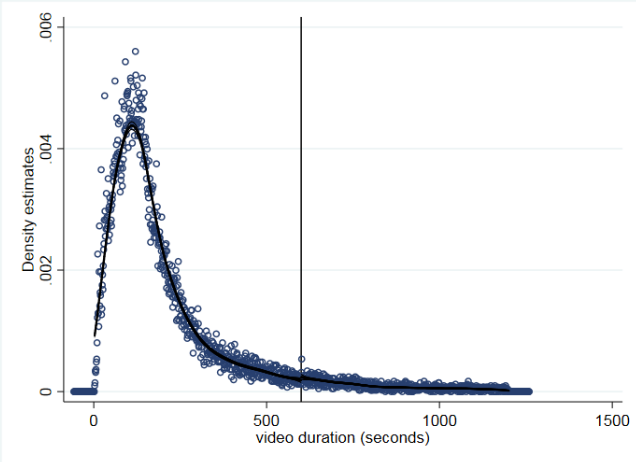

Figure 2.14: McCrary test of the distribution of video durations *after* Oct 2015 for YouTubers from the 20th to 30th percentile in median video duration before Oct 2015. The vertical line depicts the ten minutes threshold.

2.5 Empirical strategy

2.5.1 Baseline regression

This section formalizes the empirical strategy outlined in Section 2.3. As a first step in the empirical analysis, I define the treatment and the control group. Following the outline from Section 2.3, I classify a YouTuber as treated if she could increase the feasible number of ad breaks in her videos after Oct 2015. To this end, I compute each YouTuber's share of videos between ten and fourteen minutes before and after Oct 2015; if this share has increased by at least five percentage points, YouTuber i is assigned to the treatment group ($2,513$ YouTubers), and to the control group otherwise ($8,086$ YouTubers). See Appendix A.1.3 for robustness checks that use other classifications of the treatment and the control group.

The baseline difference-in-differences regression is given by

$$Mainstream_{vit} = \beta D_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \epsilon_{vit}, \tag{2.2}$$

where D_i indicates the treatment group, $post_t$ indicates all months after Oct 2015, X_{vit} controls for video categories, ϕ_i and ϕ_t are YouTuber and monthly fixed effects, respectively, and t_i is a YouTuber specific linear time trend. The dependent variable $Mainstream_{vit}$ is a dummy variable equal to one if video v of YouTuber i in month t is given a mainstream keyword, and zero otherwise (see Section 2.4.2 for details). Thus, I estimate a Linear Probability Model and the parameter β

measures the average percentage point change in the probability to upload mainstream content for YouTubers in the treatment relative to the control group.

2.5.2 IV regression

Model

An OLS estimation of equation (2.2) is unlikely to yield a causal estimate of the effect of advertising on the probability to upload mainstream content for three interrelated reasons. First, YouTubers can self-select into the treatment group. This applies, for instance, to particularly money-loving YouTubers. If these YouTubers are at the same time more likely to upload mainstream content, the OLS estimate for β would be upward biased. Second, omitted YouTuber specific time-varying factors that are neither captured in the YouTuber specific linear time trend nor in YouTuber or monthly fixed effects may drive $Mainstream_{vit}$ and D_i at the same time. To stick with the example, some YouTubers may develop a taste for money over time. If these YouTubers are more likely to upload mainstream content, the OLS estimate of β would, again, be upward biased. Finally, reverse causality may generate a spurious relationship between $Mainstream_{vit}$ and D_i. If, for instance, YouTubers who produce more mainstream content are more likely and more willing to increase their number of ad breaks per video, the OLS estimate for β would be upward biased, too.

To account for the endogeneity in a YouTuber's treatment status, I use YouTubers' *median video duration before Oct 2015* as an instrument for D_i. The first stage equation is given by

$$D_i * post_t = \pi close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + u_{vit} \tag{2.3}$$

where $close_i$ denotes the median video duration of YouTuber i before Oct 2015. The interpretation is as follows. If $close_i$ is a valid instrument (a discussion follows in Section 2.5.2), it initiates a causal chain. As good as random variation in $close_i$ generates as good as random variation in D_i, which is isolated by the first stage. Using this exogenous variation, I can consistently estimate β in equation (2.2) using Two Stage Least Squares (2SLS).

The reduced form of equations (2.2) and (2.3) is given by

$$Mainstream_{vit} = \gamma close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + v_{vit}. \tag{2.4}$$

The parameters β in equation (2.2) and γ in equation (2.4) answer different questions. The parameter γ is the average effect of an additional unit of $close_i$ – i.e., an additional minute – on the difference in the probability to upload mainstream content before and after Oct 2015. In other words, γ measures how a better chance to increase the feasible number of ad breaks per video affects the probability to upload mainstream content, whereby it is comparable to an intention-to-treatment effect. In contrast to that, β is the average effect of the actual treatment status D_i on the difference in the probability to upload mainstream content before and after Oct 2015.

Instrument heterogeneity

The instrument $close_i$ is likely to affect different YouTubers in different ways. In particular, some YouTubers' treatment status may be entirely unchanged. On the one hand, some YouTubers have no interest in increasing their feasible number of ad breaks per video; these YouTubers remain untreated, no matter how close to the ten minutes threshold they are. On the other hand, some YouTubers are desperate to increase the feasible number of ad breaks per video; these YouTubers pursue the treatment, no matter how far away from the ten minutes threshold they are. Thus, the 2SLS estimate for β measures a Local Average Treatment Effect (LATE, see Angrist and Imbens, 1995), i.e., a weighted average of the individual treatment effects, where the weights capture the individual magnitudes of π_i, i.e., the extent to which $close_i$ affects $Pr(D_i = 1)$.

Instrument validity

The validity of $close_i$ as instrument for D_i hinges on four requirements: Instrument relevance, the exclusion restriction, instrument independence, and monotonicity. These requirements are now discussed.

Instrument relevance First, the parameter π in the first stage equation (2.3) must be non-zero, which means that $close_i$ must be correlated to D_i. It is plausible that the closer a YouTuber's position to the ten minutes threshold before Oct 2015, the easier it is to produce videos that are ten minutes or longer after Oct 2015, for instance, because she does not have to deviate far from her former concepts or because she does not have to spend much additional effort. Illustrative evidence is provided by Figures 2.6, 2.8, and 2.10 in Section 2.4.4. Moreover, a bivariate regression of D_i on $close_i$ yields a t-statistic of around 15. Finally, the first stage diagnostics discussed in Section 2.6.1 confirm the instrument's relevance.

Exclusion restriction Second, $close_i$ must operate through the single, known channel $D_i * post_t$. In other words, the instrument must not be correlated to the dependent variable. This is a plausible assumption, too. A YouTuber's median video duration before Oct 2015 – when the YouTubers were ignorant of the ten minutes trick's existence – is most likely a result of her personal style, taste, or preferred level of effort and orthogonal to whether the video covers mainstream topics or not.

The panel structure of my dataset allows me to conduct an event study that confirms the plausibility of the exclusion restriction. Based on the reduced form equation (2.4), I interact $close_i$ with each monthly dummy, using Oct 2015 ($t = 34$) as the baseline. This specification allows me to treat the coefficients of the interaction terms as the effect of $close_i$ on $Mainstream_{vit}$ relative to a base month just before the YouTubers could start to adapt their content. The event study regression equation is given by

$$Mainstream_{vit} = \sum_{t=1}^{33} \gamma_t close_i * pre_t + \sum_{t=35}^{49} \gamma_t close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \upsilon_{vit}. \quad (2.5)$$

The interpretation of this approach is analogous to checking the validity of a parallel trends assumption. While the indirect impact of $close_i$ on $Mainstream_{vit}$ may accumulate over time, it should not begin before a YouTuber became aware of the new ad break tool, i.e., before the treatment status D_i was switched on. Thus, if the only way $close_i$ affects the dependent variable $Mainstream_{vit}$ is via $D_i * post_t$, then all estimates $\gamma_t, t \leq 33$, should be close to zero and be statistically insignificant. In contrast, the estimates $\gamma_t, t \geq 35$, should be unequal to zero and statistically significant.

Instrument independence In addition to the exclusion restriction, the instrument $close_i$ must be independent of potential outcomes and potential treatments. In other words, $close_i$ must be as good as randomly assigned such that the first stage captures the causal effect of $close_i$ on D_i. Note that reverse causality is of no concern here, because $close_i$ is by definition determined before, and D_i after Oct 2015. Yet, YouTuber specific time-varying factors that drive both $close_i$ and D_i as well as the potential manipulation of $close_i$ on behalf of the YouTubers – in the sense that they choose high values of $close_i$ to increase their treatment probability – may be an issue.

Four facts, however, argue against the manipulation of $close_i$. First, the ten minutes trick was unknown until Oct 2015. Second, YouTube did not announce the new ad break tool before its launch, so the knowledge of the ten minutes trick caught the YouTubers unprepared.[25] Third, YouTubers do not benefit from higher values of $close_i$ before Oct 2015, since the number of ad breaks per video is limited to one, irrespective of how close their are to the threshold. Finally, if a YouTuber chose a high value of $close_i$ to increase her treatment probability, she must know about the ten minutes trick; if she knew about the ten minutes trick, she would either exploit or ignore

[25] I searched through the YouTube creators blog (`https://youtube-creators.googleblog.com/`) and found no entries announcing the new ad break tool from before Oct 2015.

it, but she would not just move closer to the threshold.

It remains to rule out that unobserved YouTuber specific time-varying factors drive both $close_i$ and D_i. Three arguments speak against such concerns. First, t_i in equation (2.3) controls for YouTuber specific linear time trends; in Appendix A.2.3, I also include higher order polynomials of t_i into equation (2.3). Second, while commercial interests are a plausible driver of D_i, they are unlikely to affect $close_i$, as argued above. Third, YouTubers with a strong commercial interest might self-select into particular video categories that, in turn, require a certain video duration. The vector X_{vit} in equation (2.3), however, captures category specific characteristics and therefore prohibits that $close_i$ is indirectly driven by a YouTuber's commercial interest.

Monotonicity Finally, while $close_i$ may have no effect on some YouTubers (see Section 2.5.2), those who are affected must be affected in the same direction, i.e., $\pi_i \geq 0 \,\forall\, i$. Again, this is a plausible assumption: it is hard to believe that a high value of $close_i$ prohibits treatment from YouTubers who would have been treated if $close_i$ was low. Figure 2.15 provides illustrative evidence. It plots all values of $close_i$ against the corresponding probability of treatment, $Pr(D_i = 1)$. With the exception of some outliers at the upper left and the lower right corner, the relationship between $close_i$ and $Pr(D_i = 1)$ is monotone.

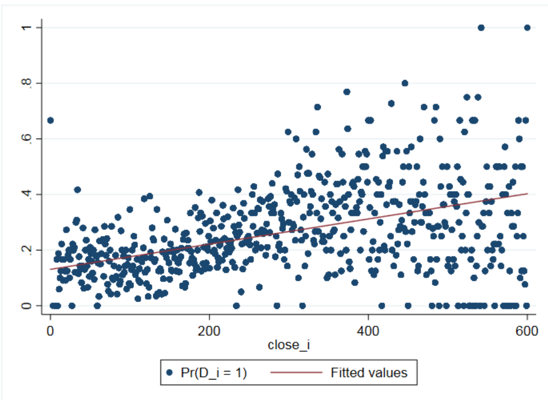

Figure 2.15: Plot of all values of $close_i$ on the associated average probability to be treated, $Pr(D_i = 1)$.

Note that I might violate the monotonicity assumption if I used a continuous measure of treatment intensity – i.e., the extent to which a YouTuber increases her share of videos that are ten minutes or longer – instead of the binary treatment status D_i. As argued, YouTubers with high values of $close_i$ have a higher *probability* to increase their share of videos that are ten minutes or longer. At the same time, however, they have *less scope* to do so, because their initial share of videos that are ten minutes or longer is already high. Hence, while the impact of $close_i$ on the *extensive margin* of treatment is monotone and increasing – as shown above – it might follow an inverted U-shape on the *intensive margin.*

Additional requirements

In addition to the validity of the instrument, two further requirements must be fulfilled. First, to be consistent with the idea of the identification strategy, the effect of an increase in the feasible number of ad breaks per video on the probability to upload mainstream content must be driven by the videos that are ten minutes or longer. Second, video duration as such must not have a direct impact on the probability to upload mainstream content. These additional requirements are now discussed.

Evidence from the video level The parameter β in equation (2.2) aggregates the effect of an increase in the feasible number of ad breaks on the probability to upload mainstream content on the YouTuber level. The aggregation is coherent with my empirical strategy: the instrument provides as good as random variation on the YouTuber level, too. Yet, to check if the aggregate effect is driven by videos that are ten minutes or longer, I augment the first stage regression equation (2.3) to

$$I(\geq 10)_{vit} = \alpha close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \upsilon_{vit} | Mainstream_{vit}, \tag{2.6}$$

where $I(\geq 10)_{vit}$ indicates if video v of YouTuber i in month t is ten minutes or longer. Then, I estimate equation (2.6) by OLS for mainstream and for non-mainstream content separately.

The interpretation is as follows. The parameter α measures the effect of an additional unit of $close_i$ on the probability that a video is ten minutes or longer, conditional on whether the video is mainstream or not. Suppose β in equation (2.2) is negative, i.e., an (aggregate) increase in the feasible number of ad breaks per videos reduces the probability to upload mainstream content. If the aggregate effect is driven by the videos that are ten minutes or longer, the OLS estimate for α should be positive and statistically significant when I condition on non-mainstream content, because the probability that a video is ten minutes or longer increases within this subsample. In contrast to that, the OLS estimate for α should be close to zero and not statistically significant when I condition on mainstream content. Note that reverse causality concerns prohibit an interaction of the term $close_i * post_t$ in equation (2.6) with a dummy that indicates mainstream content and a corresponding regression based on the entire sample. If, for instance, an increase in the feasible number of ad breaks led to a reduction in the probability to upload mainstream content, the estimate for the triple interaction would be downward biased.

Video duration and mainstream content Finally, when a YouTuber is treated, not only her treatment status D_i changes, but – by construction – her videos' duration increases, too. Hence, I must also ensure that video duration as such does not affect the dependent variable $Mainstream_{vit}$.

The difficulty resembles a regression discontinuity design: when comparing observations left and right to a cutoff, not only the treatment status, but also the value of the assignment variable determining the treatment status changes. Standard regression discontinuity designs would include the assignment variable as a control. Simply controlling for video duration may, however, be problematic in my application, since the videos' duration after Oct 2015 may be manipulated to exploit the ten minutes trick. In contrast to that, the videos' duration before Oct 2015 is – similar to the instrument $close_i$ – likely to be as good as randomly assigned. Hence, I run the following regression

$$Mainstream_{vit} = \delta duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \tag{2.7}$$

including only the time period before Oct 2015, where I expect δ to be close to zero and statistically insignificant.

2.6 Results

2.6.1 Main results

Table 2.3 presents the main results. Columns 1 to 3 show the results from the potentially biased OLS estimation of equation (2.2). The estimates are close to zero and not statistically significant despite the large sample size. In contrast to that, the estimates obtained by a 2SLS estimation of equations (2.2) and (2.3), displayed in columns 4 to 6, are negative and statistically significant at the 1%-level. According to these estimates, an increase in the feasible number of ad breaks per video decreases the probability to duplicate mainstream content by about twenty percentage points. The effect size is considerable: it corresponds to 40% of a standard deviation in the dependent variable $Mainstream_{vit}$ and to around 50% of its baseline value 0.448. The large difference between the OLS and the 2SLS estimates confirms the endogeneity concerns expressed earlier: YouTubers' self-selection into treatment, omitted YouTuber specific time-varying factors as well as reverse causality may lead to an upward bias in the estimate for β when not taken into account.

The first stage diagnostics in columns 4 to 6 confirm the validity of my empirical strategy. Having been closer to the ten minutes threshold before Oct 2015 leads to a higher treatment probability: an additional unit of $close_i$ (i.e., an additional minute) increases the treatment probability by about 2.9 percentage points. The estimate is highly statistically significant. Moreover, an F-statistic between 144 and 151 demonstrates the strength of the instrument (Stock and Yogo, 2002; Kleibergen and Paap, 2006).

Finally, columns 7 to 9 show the reduced form estimates of equation (2.4). As argued in Section 2.5, these estimates measure the effect of an additional unit of $close_i$ on the probability to duplicate mainstream content. Consistent with the results from the 2SLS regression, the estimates are negative: a one unit increase in $close_i$ leads to a 0.6 percentage point reduction in the probability to duplicate mainstream content. Though small, the estimates are statistically significant at the 1%-level.

In sum, the results presented in Table 2.3 lead to the conclusion that the exogenous increase in the feasible number of ad breaks per video causes a considerable reduction in the probability to duplicate mainstream content. In other words, I find evidence that advertising has a causal positive effect on content differentiation. The results match the theoretical considerations discussed in Section 2.1: when the YouTubers increase the number of ad breaks in their videos, they raise the ad "price" that their viewers have to pay. A higher ad "price", in turn, goes along with higher content differentiation. A detailed discussion of the economic mechanism follows in Section 2.7. Appendix A.1.5 shows that the main results are robust to alternative measures of mainstream content.

2.6.2 Validity checks

Exclusion restriction

This section confirms the plausibility of the exclusion restriction as discussed in Section 2.5.2. Figure 2.16 presents the results of the event study. The dots connected by the solid line display the estimates γ_t from a regression of equation (2.5), the dashed lines depict a 95% confidence interval. The estimates for γ_t, $t \in [1, 33]$, fluctuate around zero without a visible trend. The lion's share of the estimates is not statistically significant at the 5%-level. In contrast to that, the estimates for γ_t, $t \in [35, 49]$, are negative and downward trending. Moreover, most estimates are statistically significant at the 5%-level. Hence, $close_i$ had no clear and statistically significant effect on the dependent variable $Mainstream_{vit}$ before Oct 2015, but a clear negative and increasingly strong effect after Oct 2015. See Appendix A.1.6 for a series of placebo regressions that supports the plausibility of the exclusion restriction, too.

Table 2.3: Main results

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red.Form (9)
$D_i * post_t$	0.009	0.006	0.006	-0.209***	-0.200***	-0.192***			
	(0.0078)	(0.0077)	(0.0077)	(0.0505)	(0.0491)	(0.0480)			
$close_i * post_t$							-0.006***	-0.006***	-0.006***
							(0.0013)	(0.0013)	(0.0013)
First stage				0.0286***	0.0286***	0.0290***			
				(0.0023)	(0.0023)	(0.0024)			
F-test of excluded instruments				144.13	143.85	151.32			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

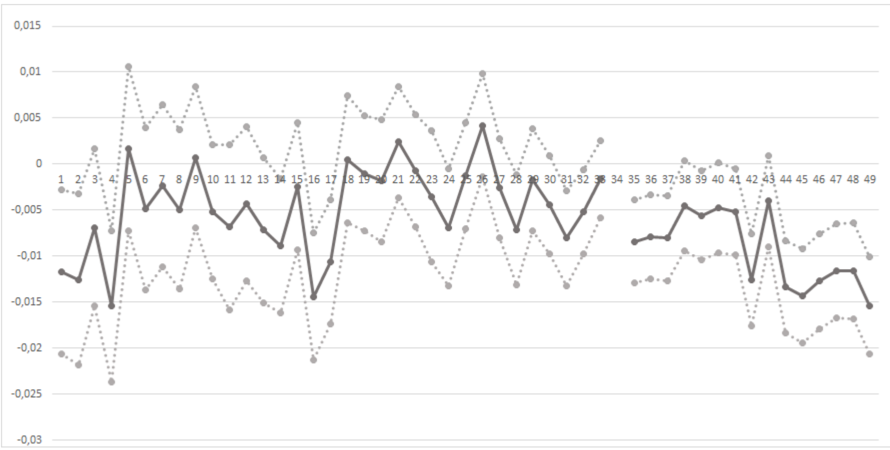

Figure 2.16: Event study mainstream content (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t$ from equation (2.5). The dashed line depicts a 95%-confidence interval.

Additional requirements

Next, I show that the additional requirements from Section 2.5.2 are fulfilled. First, Table 2.4 displays the results from an OLS regression of equation (2.6) on the subsample of mainstream videos (columns 1 to 3) and on the subsample of non-mainstream videos (columns 4 to 6). While the OLS estimate of α is small and not statistically significant in columns 1 to 3, it is around six times larger and statistically significant at the 5%-level in columns 4 to 6. These results are consistent with the ideas from Section 2.5.2. If the estimate for β is negative and if this aggregate effect is driven by the videos that are ten minutes or longer, the estimate for α should be close to zero when considering only mainstream, and positive when considering only non-mainstream content.

Table 2.4: Mainstream content – Evidence from the video level

	Mainstream (1)	Mainstream (2)	Mainstream (3)	Non-main. (4)	Non-main. (5)	Non-main. (6)
$close_i * post_t$	0.0005	0.0006	0.0005	0.003**	0.003**	0.003**
	(0.0012)	(0.0011)	(0.0011)	(0.0012)	(0.0012)	(0.0011)
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	9,855	9,855	9,855	10,248	10,248	10,248
Videos	477,532	477,532	477,532	589,468	589,468	589,468

Notes: Robust standard errors in parentheses. The dependent variable is $I(\geq 10)_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is ten minutes or longer, and 0 otherwise. All estimates are obtained by OLS and based on using the advertising YouTubers only. In addition, the estimates in Columns 1 to 3 are based on videos classified as mainstream. The estimates in Columns 4 to 6 are based on videos classified as non-mainstream. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Second, I consider the regression results from equation (2.7). Table 2.5 shows that the estimate for δ in equation (2.7) is very small and statistically insignificant. Thus, I find no evidence in my data that video duration as such directly affects $Mainstream_{vit}$.

Non-advertising YouTubers

The non-advertising YouTubers, whom I do not consider in the main analysis, allow me to conduct two additional validity checks. The non-advertising YouTubers' content choices are not driven by commercial considerations. As a consequence, their probability to upload mainstream content cannot be affected by the launch of the new ad break tool in Oct 2015. Hence, non-advertising

Table 2.5: Video duration and mainstream content

	OLS (1)	OLS (2)	OLS (3)
video duration	0.0000760 (0.000103)	0.0000125 (0.000101)	0.0000119 (0.000101)
Time FE	X	X	X
YouTuber FE	X	X	X
Controls		X	X
YouTuber Time Trend			X
YouTubers	10,113	10,113	10,113
Videos	566,079	566,079	566,079

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, the estimates are based on a regression that excludes all months $t \geq 34$. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

YouTubers who are classified as "treated" must have increased their share of videos between ten and fourteen minutes for reasons other than exploiting the ten minutes trick. As a consequence, the estimate for β should be close to zero and statistically insignificant when I estimate equations (2.2) and (2.3) by 2SLS on the non-advertising YouTubers only.

The regression results in Table 2.6 support these considerations. While the potentially biased OLS estimates in columns 1 to 3 are positive and significant at the 5%-level, both the IV estimates (columns 4 to 6), and the reduced form estimates (columns 7 to 9) are close to zero and statistically insignificant.

Figure 2.17 provides an additional plausibility check of the exclusion restriction. If the only way the instrument $close_i$ affects the dependent variable $Mainstream_{vit}$ is via the increase of the feasible number of ad breaks per video, then *all* estimates γ_t obtained when estimating equation (2.5) on the subsample of non-advertising YouTubers should be close to zero and be statistically insignificant. Figure 2.17 demonstrates that this is the case.

Table 2.6: Main results non-advertising YouTubers

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.033**	0.0364**	0.0364**	0.0227	-0.0121	-0.0125			
	(0.0146)	(0.0143)	(0.0142)	(0.0938)	(0.0894)	(0.0886)			
$close_i * post_t$							0.0005	-0.0003	-0.0003
							(0.0022)	(0.0021)	(0.0021)
First stage				0.0237***	0.0237***	0.0239***			
				(0.0035)	(0.0035)	(0.0035)			
F-test of excluded instruments				45.74	45.81	47.70			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278	5,278
Videos	329,725	329,725	329,725	329,725	329,725	329,725	329,725	329,725	329,725

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. The estimates are based on using the non-advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

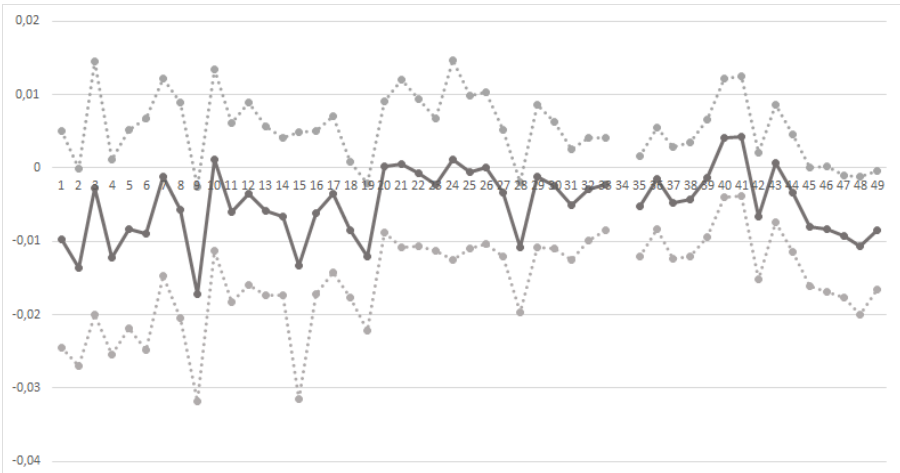

Figure 2.17: Event study mainstream content (non-advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t$ from equation (2.5). The dashed line depicts a 95%-confidence interval.

2.6.3 Effect heterogeneity

One particular strength of my dataset is its size, which allows me to conduct a series of subgroup analyses. To this end, this sections illustrates effect heterogeneity along two dimensions. First, I show that the average effect from Section 2.6.1 is driven by YouTubers with many subscribers. Second, I document that some video categories are more flexible regarding their typical video duration, which leads to heterogeneity on the first stage.

Heterogeneity along the subscriber count

The adaption of video content entails costs. The YouTubers must deviate from the content they were producing before, which may force them to focus on topics that they are less intrinsically motivated to cover. The larger a YouTuber's audience, however, the higher is her benefit from additional ad breaks and therefore also the probability that the additional ad revenue covers the costs. To confirm that the effect of an increase in the feasible number of ad breaks on the probability to upload mainstream content is stronger for YouTubers with a high subscriber count, I split my sample at the $1,000$ subscriber threshold – which roughly corresponds to the median number of subscribers – and consider YouTubers with at least $1,000$, and YouTubers with fewer than $1,000$ subscribers separately.[26] Note that reverse causality prohibits including the subscriber count as an interaction term. If, for instance, YouTubers who upload much mainstream content have a larger audience, I would overestimate the effect of a YouTuber's subscriber count.

Tables 2.7 and 2.8 show the results from regressing equations (2.2) and (2.3) on the two subsamples. The potentially biased OLS estimates in columns 1 to 3 are close to zero and statistically insignificant in both tables. The 2SLS estimates, however, are larger than the average effect in Table 2.3 when considering the YouTubers with at least $1,000$ subscribers, but close to zero for the other subsample. The first stage estimates follow a similar pattern: they are around 15% smaller than in Table 2.3 when considering YouTubers with fewer than $1,000$ subscribers, but statistically significant at the 1%-level in both cases. Finally, consistent with the 2SLS results, the reduced form estimates in columns 7 to 9 are larger than the average effect and statistically significant for the YouTubers with at least $1,000$ subscribers in Table 2.7, but close to zero for the YouTubers with fewer than $1,000$ subscribers in Table 2.8. Thus, while an increase in the feasible number of ad breaks leads to an increase in content differentiation for YouTubers with a relatively large audience, YouTubers with a low subscriber count refrain from adapting their content.

[26] YouTube has also recently disabled all YouTube channels with fewer than $1,000$ subscribers from monetization, arguing that this is a meaningful threshold for a channel to be considered "eligible" for ad revenues (see Section 2.2 for details).

Table 2.7: Main regression, subscribers $\geq 1,000$

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.009	0.008	0.007	-0.259***	-0.238***	-0.230***			
	(0.0099)	(0.0097)	(0.0097)	(0.0689)	(0.0662)	(0.0650)			
$close_i * post_t$							-0.007***	-0.007***	-0.007***
							(0.002)	(0.002)	(0.002)
First stage				0.028***	0.026***	0.029***			
				(0.003)	(0.003)	(0.003)			
F-test of excluded instruments				83.08	83.23	86.28			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	5,183	5,183	5,183	5,183	5,183	5,183	5,183	5,183	5,183
Videos	677,590	677,590	677,590	677,590	677,590	677,590	677,590	677,590	677,590

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers with at least $1,000$ subscribers are included. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.8: Main regression, subscribers $< 1,000$

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.015	0.008	0.009	-0.050	-0.069	-0.070			
	(0.012)	(0.012)	(0.012)	(0.082)	(0.082)				
$close_i * post_t$							-0.001	-0.002	-0.002
							(0.002)	(0.002)	(0.002)
First stage				0.0247***	0.0245***	0.0248***			
				(0.0035)	(0.0036)	(0.0035)			
F-test of excluded instruments				48.26	47.79	48.81			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	5,416	5,416	5,416	5,416	5,416	5,416	5,416	5,416	5,416
Videos	389,952	389,952	389,952	389,952	389,952	389,952	389,952	389,952	389,952

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers with fewer than $1,000$ subscribers are included. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Heterogeneity along video categories

Next, I demonstrate that some video categories are more flexible regarding their typical video duration. For instance, a music clip typically takes between three and five minutes and cannot be easily extended to ten minutes. Similarly, a comedy video becomes boring if it does not get the gag across. To illustrate heterogeneity between the fourteen video categories considered in the analysis, I estimate equations (2.2) and (2.3) on fourteen subsamples that include all videos from a particular video category.

The results in Table 2.9 reveal effect heterogeneity in terms of the first and also in terms of the second stage. The first stage estimate is close to zero for the categories "Music", "Comedy", and "Let's Play." Let's Play videos are often based on how YouTubers finish video game levels, many of which include a time constraint. The first stage estimate is largest for the categories "Cars & Vehicles", "Pets & Animals", and "Sports", hence, videos from these categories can either be most easily extended to ten minutes or more, or YouTubers who have the strongest desire to increase their feasible number of ad breaks self-select into these categories. The first stage estimate for the remaining categories is similar to the results from Section 2.6.1.

For the discussion of the second stage estimates, I focus on the categories with a first stage F-statistic above 10. Consistent with the main results from Section 2.6.1, all estimates are negative; their size ranges from -0.0762 ("Cars & Vehicles") to -0.922 ("Film & Animation"). The estimates are statistically significant for the categories "Film & Animation", "People & Blogs", and "Entertainment", which are also the categories with the highest number of observations. Hence, in addition to heterogeneity on the first stage, the video categories differ in the extent to which the video content is adapted. There are, again, two plausible explanations. First, it could be easier to create videos that cover non-mainstream content for some categories; in other words, the effect heterogeneity is driven by category specific differences (that are not captured by X_{vit}). Second, YouTubers who are more creative or more willing to try out something new might self-select into the video categories "Film & Animation", "People & Blogs", and "Entertainment" whose second stage effect is strongest.

2.6.4 Differentiation along the tail

Up to this point, I have considered the effect of an increase in the feasible number of ad breaks per video on the *most* mainstream content only. In this section, I study content differentiation along the "tail." In particular, I show that the effect I document in Section 2.6.1 diminishes for less mainstream content until it eventually switches its sign. To this end, I generate five dummy variables that indicate alternative percentiles of the distribution of most-viewed keywords (see

Table 2.9: Heterogeneity along video categories

	Film and Animation (1)	Cars and Vehicles (2)	Music (3)	Pets and Animals (4)	Sports (5)	Travel and Events (6)	Let's Play (7)	People and Blogs (8)	Comedy (9)	Enter-tainment (10)	How To and Style (11)	Edu-cation (12)	Science and Techn. (13)	Nonprofit and Activ. (14)
$D_i * post_t$	-0.922***	-0.0762	2.712	0.170	-0.105	0.127	-0.658	-0.243**	-0.548	-0.255**	-0.236	-0.100	-0.651	-0.227
	(0.3405)	(0.0953)	(12.376)	(0.179)	(0.134)	(0.139)	(0.852)	(0.109)	(0.936)	(0.122)	(0.188)	(0.258)	(0.761)	(0.197)
First stage	0.0220***	0.044***	-0.003	0.037**	0.035***	0.030***	-0.009	0.031***	0.020	0.028***	0.028***	0.026**	0.021	0.047
	(0.006)	(0.010)	(0.017)	(0.017)	(0.009)	(0.010)	(0.009)	(0.006)	(0.023)	(0.005)	(0.008)	(0.012)	(0.017)	(0.031)
F-test of excluded instruments	13.49	21.08	0.04	4.99	15.34	8.41	0.98	32.15	0.72	28.32	10.99	4.99	1.70	2.32
Time FE	X	X	X	X	X	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X	X	X	X	X	X
YouTuber Time trend	X	X	X	X	X	X	X	X	X	X	X	X	X	X
YouTubers	2,302	1,543	1,382	838	1,776	1,764	827	4,724	769	4,207	1,462	906	868	430
Videos	93,616	87,945	25,512	25,963	96,645	62,041	82,650	200097	15,831	224,739	77,593	43,446	13,951	11,622

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a popular keyword, and 0 otherwise. Each column displays the results of a 2SLS estimation including only the observations from one particular video category. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Section 2.4.2 for details): the 1^{st} to 10^{th}, the 10^{th} to 25^{th}, the 25^{th} to 50^{th}, the 50^{th} to 75^{th}, and the 75^{th} to 100^{th} percentile. Then, I replace the dependent variable $Mainstream_{vit}$ in equation (2.2) with each of these dummies and estimate equations (2.2) and (2.3) by 2SLS.

The results in Table 2.10 illustrate the pattern. The estimate for β in column 1 is similar to its counterpart in Table 2.3: an increase in the feasible number of ad breaks per video leads to a 20% percentage point reduction in the probability to upload a video that is given a keyword from the 1^{st} to 10^{th} percentile of the distribution of most-viewed keywords. The effect size corresponds to nearly 50% of a standard deviation in the dependent variable. The estimate, however, decreases by half in columns 2 and 3, and by about two-thirds in column 4. Finally, in column 5, the estimate switches its sign and becomes positive. The effect size, however, is small: it corresponds to 15% of a standard deviation in the dependent variable. All estimates for β are statistically significant.

To interpret these results, note that a video is given around eleven keywords on average and that this number is constant over time. Hence, a video can be given keywords from several parts of the distribution of most-viewed keywords. Bearing this mind, the estimates in Table 2.10 demonstrate that the YouTubers who could increase the feasible number of ad breaks per video do not move from exclusively uploading mainstream only to uploading non-mainstream content only. Rather, they change the "mixture" of topics in a video: they abandon covering mainstream topics – the more mainstream, the stronger the effect – and cover more of the less mainstream topics instead. Thereby, the probability to upload very mainstream content decreases, while the probability to upload not so mainstream content remains unchanged or increases only slightly. Indeed, when I count each video's number of "affiliations" to the categories displayed in Table 2.10 and use this count as dependent variable in equation (2.2), a 2SLS estimation shows that videos from YouTubers who could increase the feasible number of ad breaks per video are given keywords from fewer different categories after Oct 2015 than before (column 6 in Table 2.10).

2.7 Mechanism

This section studies the economic mechanism that drives the results from Section 2.6. In particular, I show that YouTubers who increase the feasible number of ad breaks per video avoid competition in the ad "price": since mainstream content is also supplied by many YouTubers, viewers could easily switch to a different channel if a YouTuber increased her ad "price." Switching becomes less likely, however, when the YouTuber uploads content that is less mainstream and thereby supplied by fewer competitors. I define a measure for "competitive content", i.e., a measure for the most-supplied content on YouTube, and show that it is highly correlated to mainstream content. Then, I demonstrate that an increase in the feasible number of ad breaks per video reduces the

Table 2.10: Differentiation along the tail

	1^{st} to 10^{th} percentile (1)	10^{th} to 25^{th} percentile (2)	25^{th} to 50^{th} percentile (3)	50^{th} to 75^{th} percentile (4)	75^{th} to 100^{th} percentile (5)	Sum of affiliations (6)
$D_i * post_t$	-0.200***	-0.102***	-0.104***	-0.068*	0.081**	-0.376***
	(0.0420)	(0.0380)	(0.0400)	(0.0356)	(0.0315)	(0.0959)
First stage	0.0292***	0.0306***	0.0306***	0.0306***	0.0306***	0.031***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	152.86	166.70	166.70	168.04	169.44	169.44
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
YouTuber Time Trend	X	X	X	X	X	X
YouTubers	10,599	10,591	10,591	10,590	10,589	10,589
Videos	1,064,248	1,033,666	1,033,666	1,031,051	1,028,446	1,028,446

Notes: Robust standard errors in parentheses. Each column displays the results of a 2SLS estimation. In column 1, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 1^{st} to 10^{th} percentile of the distribution of most-viewed keywords. In column 2, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 10^{th} to 25^{th} percentile of the distribution of most-viewed keywords. In column 3, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 25^{th} to 50^{th} percentile of the distribution of most-viewed keywords. In column 4, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 50^{th} to 75^{th} percentile of the distribution of most-viewed keywords. In column 5, the dependent variable is an indicator equal to one if video v of YouTuber i in month t is given a keyword from the 75^{th} to 100^{th} percentile of the distribution of most-viewed keywords. In column 6, the dependent variable is the sum of a video's percentile indicators. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

YouTubers' probability to upload competitive content. Since competitive content is typically also mainstream – i.e., content in high demanded is also supplied by many YouTubers – competition in the ad "price" ultimately leads to a reduction in the probability to upload mainstream content. Finally, I support this result by demonstrating that the audience of YouTubers who could increase the feasible number of ad breaks per video becomes more stable, i.e., the viewers become less likely to switch to competitors. In contrast to that, I find no evidence for a YouTuber learning effect (see Appendix A.2.3).

2.7.1 Definition of competitive content

First, I construct a measure of "competitive content", i.e., a measure for the most-supplied content on YouTube. The procedure is analogous to the definition of "mainstream content" (see Section 2.4.2). For each month, for each video category, I compute how many times a certain keyword has been *used* and rank them in descending order; the upper one percent of this distribution is classified as "competitive." Then, I assign a dummy variable that is equal to one to all videos equipped with a competitive keyword. Note that a competitive keyword is not necessarily a mainstream keyword, too. A keyword may attract many views although it is not used by many YouTubers; similarly, a keyword may be used by many YouTubers, but does not attract many views. In my sample, the correlation between mainstream and competitive content is equal to 0.57.

Take the category "Science & Technology" in April 2015 as an example again. The three most-used keywords are "deutsch", "test", and "review" (note that they are different from the three most-viewed keywords, see Section 2.4.2). Figure 2.18 shows that the distribution of usages over keywords is heavily skewed. For instance, the upper one percent of keywords accounts for 17.4%, while the lowest ten percent account for 4.4% of all keyword usages.[27] The numbers are similar for other categories and other points in time.

[27] Many keywords are used rarely – e.g., once or twice – which is responsible for the steps in the plot.

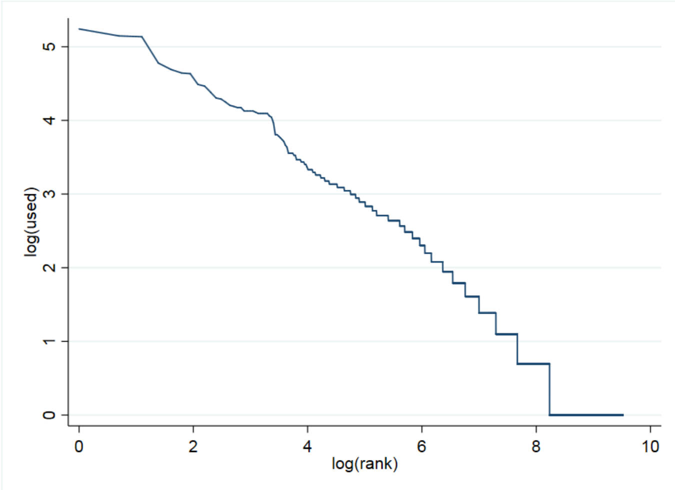

Figure 2.18: Log-log plot of the number of usages of a keyword and their associated rank in the category "Science & Technology" in March 2015.

2.7.2 IV regression

Analogous to Section 2.5, the baseline regression equation is given by

$$Competitive_{vit} = \beta' D_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \epsilon_{vit}, \tag{2.8}$$

where the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a competitive keyword as defined above. Thus, I estimate a Linear Probability Model, where the parameter β' measures the average percentage point change in the probability to upload competitive content for YouTubers in the treatment relative to the control group.

As for equation (2.2), an OLS estimation of equation (2.8) is unlikely to yield the causal effect of advertising on the probability to upload competitive content for three interrelated reasons (see Section 2.5.2 for a detailed discussion of these concerns). First, YouTubers can self-select into the treatment group. Second, omitted YouTuber specific time-varying factors might drive $Competitive_{vit}$ and D_i at the same time. Third, reverse causality may be an issue. To account for the endogeneity in a YouTuber's treatment status, I use equation (2.3) as a first stage again and estimate equations (2.3) and (2.8) by 2SLS.

Finally, the reduced form of equations (2.3) and (2.8) is given by

$$Competitive_{vit} = \gamma' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_{it} + \upsilon_{vit}, \tag{2.9}$$

where γ' measures the effect of an additional unit of $close_i$ on the probability to upload competitive content.

2.7.3 Results

YouTubers avoid competition

The estimates for β' in Table 2.11 are similar to the estimates for β in Table 2.3. Columns 1 to 3 in Table 2.11 show the results from a potentially biased OLS estimation of equation (2.8). The estimates are close to zero and not statistically significant. In contrast to that, the estimates obtained by a 2SLS estimation of equations (2.3) and (2.8), displayed in columns 4 to 6, are negative and statistically significant at the 1%-level. According to these estimates, an increase in the feasible number of ad breaks per video decreases the probability to upload competitive content by about twenty percentage points; the effect size corresponds to 42% of a standard deviation in the dependent variable $Competitive_{vit}$ and to around 30% of its baseline value 0.65. As in Section 2.6.1, the large difference between the OLS and the 2SLS estimates confirms the endogeneity concerns about a YouTuber's treatment status D_i. Finally, columns 7 to 9 show the reduced form estimates of equation (2.9). Consistent with the results from the 2SLS estimation, the estimates are negative: a one unit increase in $close_i$ leads to a 0.6 percentage point reduction in the probability to upload competitive content. Though small, the estimates are significant at the 1%-level.

The results in Table 2.11 show that an increase in the feasible number of ad breaks per video reduces the YouTubers' probability to upload competitive content. Given that the dependent variables $Mainstream_{vit}$ and $Competitive_{vit}$ are highly correlated, this is no surprise. Thus, the estimates confirm that YouTubers who increase their ad "price" avoid competition over competitive content, which is a plausible economic mechanism that drives the results from Section 2.6.

Validity checks

Although the measures are highly correlated, "competitive content" is conceptually different from "mainstream content." This section conducts three validity checks to show that the empirical strategy from Section 2.5 is also valid when I use $Competitive_{vit}$ as dependent variable in equation (2.8). First, I confirm the plausibility of the exclusion restriction. Second, I show that the effect of an increase in the feasible number of ad breaks per video on the probability to upload competitive content is driven by videos that are ten minutes or longer. Finally, I rule out that video duration as such has a direct effect on the probability to upload competitive content. See Appendix A.1.7 for further robustness checks.

Table 2.11: Mechanism: Competitive content

	OLS	OLS	OLS	2SLS	2SLS	2SLS	Red. Form	Red. Form	Red. Form
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
$D_i * post_t$	-0.003	-0.004	-0.005	-0.212***	-0.210***	-0.179***			
	(0.0075)	(0.0074)	(0.0073)	(0.0481)	(0.0477)	(0.0456)			
$close_i * post_t$							-0.006***	-0.006***	-0.005***
							(0.0013)	(0.0013)	(0.0012)
First stage				0.0286***	0.0286***	0.0290***			
				(0.0023)	(0.0023)	(0.0024)			
F-test of excluded instruments				144.13	143.85	151.32			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Competitive_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a competitive keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Exclusion restriction To show that the instrument $close_i$ has no direct effect on the dependent variable $Competitive_{vit}$, I conduct an event study as outlined in Section 2.5.2. Based on the reduced form regression equation (2.9), I interact $close_i$ with each monthly dummy, using Oct 2015 $(t = 34)$ as the baseline.

Figure 2.19 shows the results. The estimates for $\gamma'_t,\ t \in [1, 33]$, fluctuate around zero without a visible trend; the lion's share of the estimates is not statistically significant at the 5%-level. In contrast to that, the estimates for $\gamma'_t,\ t \in [35, 49]$, are negative with a downwards trend, and most of them are statistically significant. Thus, while $close_i$ has no clear and statistically significant effect on $Competitive_{vit}$ before Oct 2015, its impact is negative and increasingly strong after Oct 2015, when it could operate through the treatment status D_i.

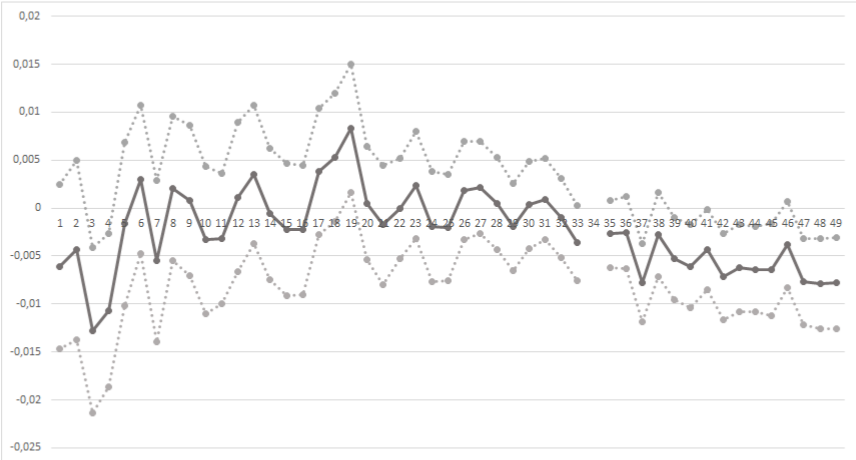

Figure 2.19: Event study competitive content (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}'_t$ from equation (2.9). The dashed line depicts a 95%-confidence interval.

As a further plausibility check, I conduct the event study on the subsample of non-advertising YouTubers. If $close_i$ affects $Competitive_{vit}$ only through an increase of the feasible number of ad breaks per video, then *all* estimates for γ'_t should be close to zero and statistically insignificant when considering the non-advertising YouTubers only. Figure 2.20 shows that this is the case. Although there is a downward trend in the estimates after Oct 2015, about a third of them has a positive sign, all of them are small, and no estimate is statistically significant at the 5%-level.

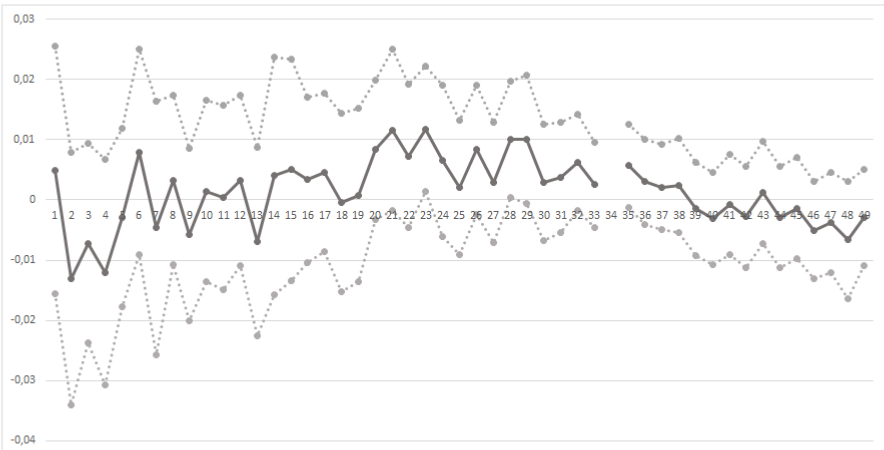

Figure 2.20: Event study competitive content (non-advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}'_t$ from equation (2.9). The dashed line depicts a 95%-confidence interval.

Evidence from the video level Similar to β, the parameter β' in equation (2.8) aggregates the effect of an increase in the feasible number of ad breaks on the probability to upload competitive content on the YouTuber level. Analogous to the approach from Section 2.5.2, I estimate

$$I(\geq 10)_{vit} = \alpha' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \upsilon_{vit} | Competitive_{vit}, \qquad (2.10)$$

by OLS for competitive and non-competitive content separately. If the aggregate effect from Table 2.11 is driven by the videos that are ten minutes or longer, the OLS estimate for α' should be positive and statistically significant when I condition on non-competitive content, but close to zero for competitive content (see Section 2.5.2 for a discussion). The results in Table 2.12 show that this is the case: the estimate for α' is close to zero and not statistically significant for competitive (columns 1 to 3), but several times as large and significant at the 5%-level for non-competitive content (columns 4 to 6).

Table 2.12: Competitive content – Evidence from the video level

	Competitive (1)	Competitive (2)	Competitive (3)	Non-comp. (4)	Non-comp. (5)	Non-comp. (6)
$close_i * post_t$	0.0011	0.0011	0.0010	0.003**	0.003**	0.003**
	(0.0010)	(0.0010)	(0.0010)	(0.0014)	(0.0014)	(0.0013)
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	10,332	10,332	10,332	9,550	9,550	9,550
Videos	693,449	693,449	693,449	373,444	373,444	373,444

Notes: Robust standard errors in parentheses. The dependent variable is $I(\geq 10)_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is ten minutes or longer, and 0 otherwise. All estimates are obtained by OLS and based on using the advertising YouTubers only. In addition, the estimates in Columns 1 to 3 are based on videos classified as competitive. The estimates in Columns 4 to 6 are based on videos classified as non-competitive. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Video duration and competitive content To check if video duration as such has no effect on the probability to upload competitive content, I estimate

$$Competitive_{vit} = \delta' duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \qquad (2.11)$$

by OLS, including only observations from before Oct 2015. Table 2.13 shows that the estimate for δ' is close to zero and not statistically significant (see Section 2.5.2 for a discussion). Thus, I find no evidence in my data that video duration as such directly affects $Competitive_{vit}$.

Table 2.13: Video duration and competitive content

	OLS (1)	OLS (2)	OLS (3)
video duration	-0.0000298 (0.000116)	-0.0000666 (0.000114)	-0.0000695 (0.0001139)
Time FE	X	X	X
YouTuber FE	X	X	X
Controls		X	X
YouTuber Time Trend			X
YouTubers	10,113	10,113	10,113
Videos	566,079	566,079	566,079

Notes: Robust standard errors in parentheses. The dependent variable is $Competitive_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is classified as competitive, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, the estimates are based on a regression that excludes all months $t \geq 34$. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.7.4 Commentator fluctuation

Finally, I provide evidence of a decrease in the viewer fluctuation of YouTubers who could increase the feasible number of ad breaks per video. If a YouTuber uploads less mainstream content, she decreases the probability that viewers switch to competitors, because the video supply is lower. Thus, the YouTuber's viewer fluctuation should go down, which means that a given number of views should be generated by a smaller number of different viewers than before.

Since data on a YouTuber's viewers is not available, I use her videos' commentators as a proxy variable and define YouTuber i's commentator fluctuation as

$$fluctuation_i = \frac{commentators_i}{comments_i}, \tag{2.12}$$

where the numerator refers to the number of unique commentators of YouTuber i and the denominator refers to the total number of comments she received. If each comment on i's videos is left by a different commentator, $fluctuation_i$ is equal to 1. If several comments are written by the same commentator, $fluctuation_i$ is smaller than 1. Finally, if YouTuber i never receives any comment, $fluctuation_i$ is not defined.

Next, I compute each YouTuber's *change* in $fluctuation_i$ before and after Oct 2015,

$$\Delta fluctuation_i = fluctuation_{i,post} - fluctuation_{i,pre}, \tag{2.13}$$

where $fluctuation_{i,post}$ is based on the fifteen months after, and $fluctuation_{i,pre}$ is based on the fifteen months before and including Oct 2015.[28] A decrease in YouTuber i's commentator fluctuation after Oct 2015 implies that $\Delta fluctuation_i < 0$; an increase implies that $\Delta fluctuation_i > 0$.

To check if an increase in the feasible number of ad breaks leads to a decrease in the YouTubers' commentator fluctuation, I use $\Delta fluctuation_i$ as dependent variable in the regression equation

$$\Delta fluctuation_i = \rho_0 + \rho_1 D_i + \epsilon_i, \tag{2.14}$$

where ρ_1 measures how the YouTubers' treatment status D_i affects their average change in commentator fluctuation. To account for endogeneity in D_i, I use

$$D_i = \psi_0 + \psi_1 close_i + e_i, \tag{2.15}$$

as a first stage and estimate equations (2.14) and (2.15) by 2SLS (see Section 2.5.2 for a detailed discussion of the endogeneity concerns). Since $fluctuation_i$ is sensitive to additional commentators when the total number of comments is small – for instance, if a YouTuber has only received three comments, it makes a big difference if they are written by two or three different commentators – I restrict the analysis to YouTubers who received at least 25 comments before and after Oct 2015 (see Appendix A.1.7 for alternative thresholds).

Table 2.14 shows the results. Column 1 presents the potentially biased OLS estimate for ρ_1. The estimate is negative: an increase in the feasible number of ad breaks per video leads to a decrease in $\Delta fluctuation_i$. The 2SLS estimate in column 2 is negative, too, but more than three times larger in absolute value than the OLS estimate. The effect size corresponds to 18% of a standard deviation in the dependent variable $\Delta fluctuation_i$. The reduced form estimate in column 3 is consistent with the 2SLS estimate in column 2. All estimates are statistically significant. In contrast to that, the 2SLS and reduced form estimates are not statistically significant when I conduct the same analysis on the subsample of non-advertising YouTubers (Table 2.15). Thus, I find that an increase in the feasible number of ad breaks leads to a decrease in the YouTubers' commentator fluctuation, which supports the plausibility of the results from Section 2.7.3 along with the argument that YouTubers upload less mainstream content to avoid competition in the ad "price."

[28] Since I have 34 observation periods before and including Oct 2015, but only fifteen observation periods afterwards, I restrict the computation of $fluctuation_{i,pre}$ to the fifteen most recent ones to increase the comparability to $fluctuation_{i,post}$.

Table 2.14: Commentator analysis

	OLS (1)	2SLS (2)	Red. Form (3)
D_i	-0.0142***	-0.0475**	
	(0.0038)	(0.0232)	
$close_i$			-0.0014**
			(0.0006)
First stage		0.0289***	
		(0.0023)	
F-test of excluded instruments		159.78	
YouTubers	5,907	5,907	5,907

Notes: Robust standard errors in parentheses. The dependent variable is $\Delta fluctuation_i$, which is the difference in the commentator fluctuation before and after Oct 2015 for YouTuber i. The estimates are based on the advertising YouTubers only. Only YouTubers who received more than 25 comments before and after Oct 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table 2.15: Commentator analysis – Non-advertising YouTubers

	OLS (1)	2SLS (2)	Red. Form (3)
D_i	-0.022***	-0.039	
	(0.008)	(0.055)	
$close_i$			-0.0009
			(0.0013)
First stage		0.024***	
		(0.004)	
F-test of excluded instruments		34.33	
YouTubers	1,462	1,462	1,462

Notes: Robust standard errors in parentheses. The dependent variable is $\Delta fluctuation_i$, which is the difference in the commentator fluctuation before and after Oct 2015 for YouTuber i. The estimates are based on the non-advertising YouTubers only. Only YouTubers who received more than 25 comments before and after Oct 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.8 Differentiation in the aggregate

Up to this point, I have studied how advertising affects *individual* YouTubers' content choice. In contrast to that, this section shows how content differentiation develops *in the aggregate*. In particular, I study if the tail of keywords becomes "longer" (i.e., if the total number of keywords increases), and if the tail becomes "fatter" (i.e., if the concentration of videos on keywords decreases). I do not make causal claims here; rather, I pursue a descriptive before-after comparison to put the results from Sections 2.6 and 2.7 into a broader context.

I have two options to analyze content differentiation in the aggregate: I could continue to focus on the subsample of YouTubers whom I selected for the main analysis in Section 2.4.3 or I could examine the entire population of German YouTubers. Proceeding with the subsample has the advantage of computing aggregate measures that are solely based on YouTubers who have the option to increase their feasible number of ad breaks per video, but would not reveal how the *entire* video supply on YouTube develops after Oct 2015. Even YouTubers who are not directly affected by the launch of the new ad break tool may adapt their video content as a reaction to their competitors' change in content; thus, studying the population of YouTubers might be more informative about aggregate developments. On the other hand, the content choices of YouTubers whom I did not select for the main analysis could be driven by motives that are orthogonal to the launch of the new ad break tool and its consequences; such effects might superimpose the treatment's aggregate effect on content differentiation and complicate the interpretation of the net effect. Since no approach clearly excels the other, I pursue both options and interpret the results adequately.

2.8.1 The tail becomes longer

To show that the tail of keywords becomes longer both within the subsample and the entire population of YouTubers, I compute the absolute number of unique keywords before and after Oct 2015. As I observe 34 months before (and including) Oct 2015, but only 15 months afterwards, I limit the analysis to the 15 most recent months before (and including) Oct 2015.

In the subsample, there exist $607,358$ unique keywords before, and $875,503$ unique keywords after Oct 2015, which corresponds to an absolute increase of $268,145$ unique keywords and to a relative increase of 44.15%. Considering the population of YouTubers, I find that there exist $1,090,355$ unique keywords before, and $2,096,373$ unique keywords after Oct 2015, which corresponds to an absolute increase of $1,006,018$ keywords and to a relative increase of 92.27%. The results match the findings from Sections 2.6 and 2.7: it is plausible that the total number of unique keywords increases when the YouTubers reduce the probability to upload mainstream or competitive content. The difference in the results could stem from entry: by construction, the population includes all YouTubers who entered the platform after Oct 2015, which may further increase the number of unique keywords that exist after Oct 2015.

2.8.2 The tail does not become fatter

To study if the tail becomes "fatter", I compute a Gini coefficient for the concentration of videos on keywords before and after Oct 2015.[29] Again, I restrict the analysis to the 15 most recent months before (and including) Oct 2015. Note that the Gini coefficient for the subsample measures the concentration of videos on keywords that occur *within the subsample*, while the Gini for the population measures the concentration of *all* videos on *all* keywords.

The Gini coefficient for the subsample is high and remains nearly unchanged: it is equal to 0.800 before, and equal to 0.806 after Oct 2015, which corresponds to an increase of 0.75%. Thus, the YouTubers whom I initially selected for the main analysis do not differentiate from each other after Oct 2015. The result does not contradict the findings from Section 2.7, though. My measures for mainstream and competitive content are based on *all* active German YouTubers. It is therefore possible that the YouTubers in the subsample decrease their probability to upload competitive content, where competitive content takes *the population of* YouTubers into account, but that the concentration of videos on keywords *within* the subsample remains nearly unchanged. In addition to that, the tail of keywords becomes longer after Oct 2015 (see Section 2.8.1). If many of those additional keywords are used by a small number of videos, the Gini coefficient as a relative measure of concentration remains unchanged even if the concentration of videos on the remaining keywords decreases.

The Gini coefficient for the population of German YouTubers increases from 0.848 before to 0.862 after Oct 2015, which corresponds to an increase of 1.65%. Here, too, the increase in the relative concentration measure could be due to the large amount of additional keywords. It is also possible that further developments – orthogonal to the launch of the new ad break tool – superimpose the effect of an increase in the feasible number of ad breaks on content differentiation in the aggregate. For instance, the growing popularity of the platform may have led to a large number of entrants who copy from the most popular YouTubers and thereby increase the concentration of videos on keywords.

[29] I.e., the keywords replace the households, and the number of videos that use a certain keyword replaces the income in a conventional Gini computation. Note, also, that I do not use absolute measures of concentration such as the Herfindahl index, because the number of keywords before and after Oct 2015 is different.

2.9 Quality

As an extension of the main analysis, this section studies the effect of an increase in the feasible number of ad breaks on video quality. Two predictions compete. On the one hand, a higher number of ad breaks per video implies that each viewer is c.p. more valuable than before; hence, the incentive to provide high quality goes up. In addition, the YouTubers may want to counterbalance their viewers' increased ad nuisance costs. On the other hand, YouTubers could not only avoid competition in the ad "price", but also competition in terms of video quality when they reduce their probability to upload competitive content (see, e.g., Bourreau, 2003; Armstrong and Weeds, 2007; Weeds, 2013); as a result, the incentive to provide high quality diminishes. Moreover, YouTubers deviate from the content they were providing before and which they might have been more intrinsically motivated to cover. A lack of passion could have a negative effect on their videos' quality (see Sun and Zhu, 2013, for a similar argument). The results from two different measurement approaches are ambiguous: while an increase in the feasible number of ad breaks per video leads to a decrease in the fraction of likes, it leads to an increase in the number of views.

2.9.1 Likes and dislikes

First, I use a video's number of likes and dislikes to measure its quality. To this end, I normalize the number of likes of video v by YouTuber i in month t by its sum of likes and dislikes: $\frac{Likes}{Likes+Dislikes}_{vit}$. Though straightforward to interpret, this measure reflects the viewers' general satisfaction with a video, which is determined by its quality *and* the viewers' ad aversion. Thus, even if an increase in the feasible number of ad breaks led to an increase in video quality, a video's fraction of likes could decrease if the viewers' additional ad nuisance costs prevail.

I replace the dependent variable $Mainstream_{vit}$ in equation (2.2) with $\frac{Likes}{Likes+Dislikes}_{vit}$ and estimate equations (2.2) and (2.3) by 2SLS. Table 2.16 shows the results. Again, the potentially biased OLS estimates of equation (2.2) in columns 1 to 3 are close to zero and not statistically significant. In contrast to that, the 2SLS estimates in columns 4 to 6 are negative and statistically significant at the 1% level: an increase in the feasible number of ad breaks leads to a 4 percentage point reduction in the fraction of likes. The effect size corresponds to around 25% of a standard deviation in the dependent variable $\frac{Likes}{Likes+Dislikes}_{vit}$ and to 4.4% of its baseline value 0.91. The reduced form estimates in columns 7 to 9 are in line with these results. Note that I lose $77,066$ videos that have not received any likes or dislikes.

The results in Table 2.16 illustrate that viewer satisfaction has gone down. It is, however, unclear if the effect is driven by a decrease in video quality or by the viewers' irritation from additional ad breaks. See Appendix A.1.8 for validity checks.

Table 2.16: Quality - Likes / (Likes + Dislikes)

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	-0.0001	0.0001	0.0001	-0.041***	-0.040***	-0.039***			
	(0.0015)	(0.0015)	(0.0015)	(0.0108)	(0.0108)	(0.0107)			
$close_i * post_t$							-0.001***	-0.001***	-0.001***
							(0.0003)	(0.0003)	(0.0003)
First stage				0.028***	0.028***	0.028***			
				(0.0025)	(0.0024)	(0.0024)			
F-test of excluded instruments				130.86	130.60	137.57			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,594	10,594	10,594	10,594	10,594	10,594	10,594	10,594	10,594
Videos	990,476	990,476	990,476	990,476	990,476	990,476	990,476	990,476	990,476

Notes: Robust standard errors in parentheses. The dependent variable is $\frac{Likes}{Likes+Dislikes}_{vit}$, i.e., the share of positive ratings for video v of YouTuber i in month t. The estimates are based on using the advertising YouTubers only. Videos that received no likes nor dislikes are excluded from the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.9.2 Views

Second, viewers "vote with their feet." Hence, I use a video's number of views as a further measure of quality. YouTube counts a view if the video is watched for at least thirty seconds; if the video is shorter than that, it must be watched entirely.[30] If an increase in the feasible number of ad breaks led to an increase in video quality, more viewers may watch the video for more than thirty seconds. In addition, more viewers may watch the video repeatedly.

Analogous to Section 2.9.1, I replace the dependent variable $Mainstream_{vit}$ in equation (2.2) with the logarithm of the number of views of video v by YouTuber i in month t: $log(Views)_{vit}$. Then, I estimate equations (2.2) and (2.3) by 2SLS. Table 2.17 shows the results. The potentially biased OLS estimates in columns 1 to 3 are positive and statistically significant at the 1%-level. According to these estimates, an increase in the feasible number of ad breaks leads to a 20% increase in views. The 2SLS estimates in columns 4 and 5 are more than twice as large and statistically significant at the 1%-level, too. The 2SLS estimate is, however, sensitive to including a YouTuber specific linear time trend: in column 6, it diminishes by about a third relative to columns 4 and 5. Moreover, the estimate is only weakly statistically significant at the 10%-level. The

[30] See www.tubics.com/blog/what-counts-as-a-view-on-youtube/ (May 2019).

reduced form estimates match the pattern. They are positive and statistically significant at the 1%-level in columns 7 and 8, but only at the 5%-level when I add a YouTuber specific linear time trend in column 9.

There are two potential explanations for the differences to Section 2.9.1. First, video quality may enhance, whereby more (repeated) viewers are attracted. At the same time, however, viewers express their dissatisfaction with the additional breaks by a disliking the video. Second, there could be algorithmic confounding of the data (Salganik, 2017, Ch. 3). YouTube, too, earns a fraction of the YouTubers' ad revenue. Thus, the platform has an incentive to treat videos with many ad breaks favorably, for instance, through its ranking algorithm. In this case, the number of views was not informative about a video's quality, but only about an algorithmic advantage. See Appendix A.1.8 for validity checks.

Table 2.17: Quality - log(Views)

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. Form (7)	Red. Form (8)	Red. Form (9)
$D_i * post_t$	0.207***	0.210***	0.216***	0.436***	0.469***	0.297*			
	(0.0201)	(0.0261)	(0.0260)	(0.1586)	(0.1592)	(0.1518)			
$close_i * post_t$							0.0125***	0.0134***	0.0086**
							(0.0044)	(0.0044)	(0.0044)
First stage				0.0285***	0.0286***	0.0290***			
				(0.0023)	(0.0023)	(0.0024)			
F-test of excluded instruments				144.26	143.98	151.44			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081	1,067,081

Notes: Robust standard errors in parentheses. The dependent variable is $log(Views)_{vit}$, which is the logarithm of the views video v of YouTuber i uploaded in month t has received. The estimates are based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

2.10 Conclusion

This paper demonstrates that an increase in the feasible number of ad breaks per video leads to an increase in content differentiation between several thousand YouTubers. In particular, I find that an increase in the feasible number of ad breaks per video reduces the YouTubers' probability to duplicate mainstream content by about twenty percentage points, because YouTubers avoid competition in the ad "price." The results provide empirical evidence for predictions from economic theory: models that acknowledge the conceptual equivalence between direct prices and consumers' nuisance costs from advertising find that media outlets prefer to differentiate from each other to avoid ruinous competition in the ad "price."

The paper advances debates on the effect of advertising on content differentiation. In particular, showing that advertising does *not* lead to the duplication of mainstream content entails two implications for present policies. First, advertising quantities are often restricted in an attempt to protect consumers.[31] The Audiovisual Media Services Directive, for instance, requires that the proportion of television advertising and teleshopping spots within a given clock hour shall not exceed 20% (Article 23 §1). My paper demonstrates that consumers may *benefit* from advertising, because it increases content differentiation; policy makers need to take this additional effect into account when they determine advertising quantity restrictions. Similarly, public interventions in television markets – i.e., public service broadcasters – grow from the claim that advertising funded broadcasting fails to serve all viewers' preferences over content (Armstrong and Weeds, 2007). My results controvert this argument: advertising leads to *more* content differentiation. Thus, while valuable contributions to culture, education, and the public discourse certainly justify public service broadcasting, concerns about content duplication by advertising funded broadcasters do not.

My paper is limited in at least four respects. First, although I present competition in the ad "price" as a plausible mechanism for my main results and rule out a YouTuber learning effect, I cannot exclude the possibility that there are other potential mechanisms. For instance, YouTubers might not only avoid competition to other YouTubers and acquire a more stable audience when they upload less mainstream content, but the characteristics of their viewers may change, too. Viewers of less mainstream content could be generally less ad averse or have a higher valuation of the video content such that they are willing to endure more ads.

Second, I cannot evaluate the effect of advertising on welfare, because I lack measures for consumer and producer surplus. Although I demonstrate that advertising leads to more content differentiation – which is likely to raise consumer surplus (Brynjolfsson et al., 2003) – the viewers must also pay an increased ad "price", which works into the opposite direction. Since I obtain no estimates for the viewers' ad aversion, my setup does not answer which effect overweights. On the producer side, I remain agnostic about the effect of advertising on the surplus of YouTube itself, the YouTubers, and the advertisers. YouTube as a platform is likely to benefit from advertising,

[31] See `www.ofcom.org.uk/__data/assets/pdf_file/0021/19083/advertising_minutage.pdf` (Dec 2018).

though. Advertising leads to more content differentiation, which attracts more viewers; more viewers, in turn, generate more ad revenue. Likewise, the YouTubers' surplus benefits from an increase in ad revenue; it is, however, unclear how their utility from covering different topics than before is affected. Finally, the advertisers' surplus may go up or down. On the one hand, a higher ad quantity makes it more likely that potential customers click on their ads and buy their products. On the other hand, the advertisers cannot influence where exactly their ads appear, whereby it is unclear how well the audience is targeted. Hence, it is possible that the additional costs of advertising surmount the additional revenues.

The third limitation of my paper is that the YouTubers' per-view-revenue from advertising is unaffected by the degree of targeting. Media outlets' revenue per ad usually increases in the degree of targeting, because the advertisers' willingness to pay is higher. On YouTube, in contrast, the price per ad is constant, whereby my results cannot be extrapolated to an environment where the per-ad-revenue increases if a narrow and specific audience is attracted. It is likely, however, that the effect of an increase in the feasible number of ad breaks was higher, because the YouTubers had an additional incentive to differentiate their content.

Finally, I do not discuss any concerns related to commercial media bias, i.e., advertisers exerting pressure on the media outlets' content decisions. As argued, however, there is no direct relationship between YouTubers and advertisers whose ads appear as breaks during the videos, so the issue is of small importance in my application. Yet, it is possible that commercial media bias arises from product placement contracts between advertisers and YouTubers, for instance, if the advertisers want their products to appear within friendly and uncontroversial videos; studying the relationship between product placement and commercial media bias on YouTube would be an interesting question for further research.

3 Incumbency Dominance in Letters to the Editor: Field Experimental Evidence

With Markus Dertwinkel-Kalt and Johannes Münster. Published in: *Political Communication (2019), Vol. 36:3, 337-356.*

3.1 Introduction

Free and unbiased media are important prerequisites for democracy, as media inform, set the agenda and influence both voters and politicians. The neutrality of the media can be compromised, however, by biases of the media themselves, as well as by the outside interference of actors such as advertisers, lobbyists, domestic or even foreign governments (Shoemaker and Reese, 1996).

While the theoretical literature has singled out many different reasons for media bias (for an overview see Lichter, 2017), its manifestations fall into one of two categories: the way bits of information are *culled* and *crafted* by gatekeepers into media messages (Shoemaker and Vos, 2009). In other words, media content can become distorted through newsmakers choosing "which events or information to cover, and how to cover them" (Groeling, 2013). Similarly, Gentzkow et al. (2016) differentiate between *filtering* and *outright distortion.* The latter refers to a distorted representation of facts whereas the former refers to a partisan omission of facts in news coverage. According to Gentzkow et al. (2016) and Puglisi and Snyder (2016), filtering—the strategic selection of facts to be reported—is more pervasive than an outright distortion of given facts.[1]

In order to test for the filtering of media content, we conducted a randomized field experiment in Germany three weeks before the federal election in 2017. In the experiment, we wrote four different versions of a letter to the editor. The versions differed in the subject the letter was about, the chancellor Angela Merkel versus the main challenger Martin Schulz, and in the evaluation of this subject, positive versus negative. In all other respects, the letters were identical. We sent one randomly drawn version of the letter to each of over 200 German daily newspapers and observed

[1]Filtering has been documented, for instance, by Larcinese et al. (2011). They unveil a significant correlation between the endorsement policy of newspapers and the differential coverage of bad/good economic news as a function of the president's political affiliation.

whether the letter was published or rejected.

The goal of the experiment is to test whether different theories about filtering media content apply to the selection of letters to the editor: *political bias*, *negativity bias*, and *incumbency dominance*. We compare the acceptance rates of politically left-leaning versus right-leaning letters, of negatively framed versus positively framed letters, and of letters about the incumbent versus the main challenger, defining biases as unequal acceptance rates of different types of letters.[2] The experiment uses a between-subjects design: every newspaper received exactly one of our letters. Therefore, our study focuses on investigating whether the German newspaper system as a whole has a systemic or structural bias, while we do not test for biases of individual outlets. In an additional analysis, however, we investigate whether newspapers are more or less likely to accept a letter with a political message that is close to their own political position.

Gatekeeping of media messages has long been studied in the communications literature (Shoemaker and Vos, 2009). Moreover, a large literature has studied the news factors (which are intrinsic properties of potential news items) and the news values (their importance for gatekeepers) shaping media content (Galtung and Ruge, 1965; O'Neill and Harcup, 2009; Harcup and O'neill, 2017). Our study contributes to this literature in two ways. First, letters to the editor are "an important but poorly understood form of voluntary political participation" (Cooper et al., 2009, p.131). We study how central topics in political communication – gatekeeping, political biases, and news values – play out in the newspapers' selection of letters. Second, the field experimental approach allows for clean tests of several types of media bias by leveraging the methodological benefits of randomized controlled trials.

Not much is known about biases in letters to the editor even though they belong to the most read sections of editorial pages (Hynds, 1994). They may sometimes even directly influence politicians' behavior; for example, it has been argued that Barry Goldwater overestimated his chances of winning with a conservative platform partly because of the conservative tone of many letters to the editor (Converse et al., 1965). Moreover, Richardson and Franklin (2004) present evidence that, in election races, political parties orchestrate letter campaigns. Research on letters to the editor has mostly studied their content, the demographic characteristics of letter writers, and whether the published letters adequately gauge public opinion (Cooper et al., 2009). The process of selecting letters for publication as well as the news values that guide editorial decisions have been studied by qualitative methods such as in-depth interviews with letter editors (Wahl-Jorgensen, 2007; Raeymaeckers, 2005). Closest related to our study is the field experiment by Butler and Schofield (2010) who compare whether a letter supporting McCain or Obama was more likely

[2] This concept of bias is related to "gatekeeping bias" as defined in D'Alessio and Allen (2000) and "selection bias" as defined in Groeling (2013).

to be published during the 2008 US presidential election. Our experiment adds an additional dimension of comparison: the positive versus negative evaluation of the respective candidate. Hence, unlike Butler and Schofield (2010), our study allows to disentangle political bias from incumbency dominance as well as to test for negativity bias.

Germany is an ideal country for our experiment because of its comparatively high number of independent newspapers (Noam, 2016). It is also an interesting case from a media systems perspective. Germany is a multi-party democracy with a democratic-corporatist media system (Hallin and Mancini, 2004). In contrast, previous field experimental work on letters to the editor has studied newspapers in the USA (Butler and Schofield, 2010), and the comparative political communication literature has demonstrated that results concerning the liberal media system of the USA do not necessarily carry over to other media systems (see de Vreese, 2017, for a survey).

Empirical studies of media bias face the methodological challenge that the researchers cannot observe the population of all possible news items from which the media select what they publish. Therefore, it is hard to establish any systematic tendencies or biases in the mapping from all possible news items to actual media content. Groeling (2013) calls this the *Problem of the Unobserved Population.* One research strategy to overcome this problem consists of narrowing down the set of possible news items to a specific subset, such as press releases (e.g., Grimmer, 2013; Haselmayer et al., 2017), war fatalities (Aday, 2010), or official economic statistics (Larcinese et al., 2011; Soroka, 2012), where all items in the subset are known to the researchers (see Groeling, 2013, p. 144 for a review of further studies of this type). Alternatively, researchers can create the population of news items themselves. As Groeling (2013, p.145) puts it, "Perhaps the ultimate way to observe the unobserved population is to actually create it." For instance, field experiments in economics that use the well-established *correspondence method* follow this approach. Here, fictitious CVs are sent in order to study discrimination in the labor market (Bertrand and Mullainathan, 2004; Bartoš et al., 2016; Bertrand and Duflo, 2017; Riach and Rich, 2002; Guryan and Charles, 2013). Relatedly, King et al. (2014) generated social media posts in order to study censorship in China in a field experiment. In order to study filtering, our study also follows the approach of generating the otherwise unobserved population of news items the newspaper has to select from, which are in our case letters to the editor.[3]

Most of the research on letters to the editor has focused on the printed letters, without observing the population of all letters sent to the media. A few studies (Foster and Friedrich, 1937; Renfro, 1979) compared the population of all letters received by specific newspapers with all letters they

[3]In particular, our paper therefore adds to the literature on field experiments using the media (Panagopoulos and Green, 2008; Gerber et al., 2009; Butler and Schofield, 2010; Green et al., 2014, 2017), and to a small experimental literature on the content selection of newspapers (Butler and Schofield, 2010; Helfer and Aelst, 2016).

printed.[4] Unfortunately, data on the population of all letters received by all newspapers in a country are not available. It is difficult, moreover, to entirely eliminate confounding factors in such observational studies of the relation between the content of potential items for publication and their actual coverage in the media (see Grimmer, 2013, p.129). The experimental research design of Butler and Schofield (2010) offers a clever way to address both the unobserved population problem and confounding effects by experimentally creating the population. In our randomized controlled trial, the population is the set of letters that we have sent, and the different versions of the letters are randomly assigned, alleviating concerns about confounding.

3.2 Design and hypothesis

The aim of the experiment is to test for three different potential manifestations of media bias, that is, *political bias*, *negativity bias*, and *incumbency dominance*. We therefore designed four versions of a letter to the editor that could be classified as (1) pro chancellor Merkel, (2) contra challenger Schulz, (3) pro challenger Schulz, and (4) contra chancellor Merkel. For brevity, Appendix 3 contains only the translations of the letters (2) and (3) into English.

First, we would like to know whether newspapers are biased, on aggregate, toward the right or left, when comparing Merkel and Schulz. More precisely, we ask whether letters praising Merkel or criticizing Schulz are more or less likely to be accepted than letters criticizing Merkel or praising Schulz.

There are several competing theories about *political bias* or *partisan bias* in the media (see Gentzkow et al., 2016; Lichter, 2017, for surveys). Since many newspapers are profit-maximizing firms, one might conjecture that their political position is more in line with the center-right and the comparatively business-friendly Merkel. If this is the case, and newspapers tend to select letters to the editor that are in line with their own political position, the right-leaning versions of our letter (i.e., positive letters about Merkel and negative letters about Schulz) should have a better chance of publication than the left-leaning versions (i.e., positive letters about Schulz and negative letters about Merkel). On the other hand, many journalists in Germany are themselves more politically left leaning, and therefore the left-leaning letter might have a better chance of being published (Kepplinger, 2011). Moreover, newspapers might counterbalance their own political position by preferring letters expressing different political opinions (Butler and Schofield, 2010), or they might adjust their content to the political opinions of their readers (Gentzkow et al., 2016; Haselmayer et al., 2017). We therefore have no strong prior expectation whether left-leaning or

[4] Perrin and Vaisey (2008) also observe all letters received by one newspaper within a three-month period, and focus on the content of these letters.

right-leaning letters are published more often.

Hypothesis I (Political Bias). *The acceptance rate of letters "pro Merkel" and "contra Schulz" is different from the acceptance rate of letters "pro Schulz" and "contra Merkel."*

Next we ask whether newspapers are more likely to publish positive letters (stating that one candidate is clearly better, with an optimistic outlook in case this candidate wins), or negative letters (stating that one candidate is clearly worse, with a pessimistic view of the future in case this candidate wins).[5]

Galtung and Ruge (1965) argued that negativity is one of the important news factors in the selection of news (see also O'Neill and Harcup, 2009). Psychological studies document a *negativity-bias* (Ito et al., 1998; Rozin and Royzman, 2001; Soroka and McAdams, 2015) whereby people pay more attention and react more strongly to bad news than to good news. Similarly, Trussler and Soroka (2014) show that politically interested news consumers have a preference for negative content. In line with this, studies on media content have found that newspapers are more likely to cover negative news. For example, Heinz and Swinnen (2015) show that German newspapers report 20 times as much about downsizing firms than about firms creating an equal number of new jobs. Relatedly, Niven (2001) and Garz (2014) find a dominance of negative reports on unemployment in the United States and Germany, respectively. Baumgartner and Chaqués Bonafont (2015) document a strong negativity bias in general political coverage so that partisan media focus on the opponent's failures instead of the own party's virtues. We test for a negativity bias among letters to the editor by investigating whether a negative, critical letter stating a worried outlook for the future is more likely to be published than a positive and optimistic letter.

Hypothesis II (Negativity Bias). *Letters "contra Schulz" and "contra Merkel" are more often printed than letters "pro Schulz" and "pro Merkel."*

The third type of media bias that we test for is *incumbency dominance* whereby incumbents obtain more media coverage than their challengers. Incumbency dominance describes the phenomenon that politicians in the government obtain more media coverage than those in the opposition (see Vos, 2014, for a survey of the correlates of the media coverage of individual politicians, and Vos and Van Aelst, 2018, for a recent contribution). Schoenbach et al. (2001) documented incumbency dominance in TV news in Germany in the 1990s. More recently, Holtz-Bacha et al. (2014) found that the press coverage of the 2009 general election in Germany concentrated on the two main candidates, with more coverage of the chancellor than of the challenger.

[5] By changing as little as possible only between the positively and the negatively connotated letter, we can be confident that differences in print probabilities can be traced back to the use of the positive and the negative connotation, and not to differences in the content of the letter, for instance.

Different explanations for incumbency dominance have been proposed. On the one hand, it can be explained by referring to the "universal news value of political power" (Van Dalen, 2012) according to which incumbents have a higher news value than their competitors due to the political power they wield. In particular, in the German political system, the chancellor has a powerful position, which makes him or her comparatively newsworthy (Hopmann et al., 2011). On the other hand, incumbency dominance can be explained by the "watchdog role of the media" (Green-Pedersen et al., 2017) whereby media make societal problems a subject of discussion and therefore put an emphasis on the responsibility of those that design policy—the incumbents.

Not all of this coverage is positive, however; incumbency dominance often also means more critical coverage (Green-Pedersen et al., 2017).[6] In contrast to a political bias, incumbency dominance would predict that a letter about Merkel is more likely to be published than an otherwise identical letter about Schulz, irrespective of whether these letters denounce or applaud their subject.

Hypothesis III (Incumbency Dominance). *Letters "pro Merkel" and "contra Merkel" are more often printed than letters "pro Schulz" and "contra Schulz".*

We test Hypotheses I to III using the non-parametric Fisher's exact test. In a robustness check, we control for circulation as Butler and Schofield (2010) have found that larger newspapers were less likely to publish their letters. One possible explanation is that bigger newspapers receive more letters and can thus be more selective. Controlling for circulation might therefore improve the accuracy of our estimates. Similarly, national newspapers may handle letters in a different way than regional newspapers. We therefore also perform a regression analysis that controls for national (as opposed to regional) newspapers. In a further robustness check, we control for the state in which a newspaper is published. Finally, we run a regression where we weight each newspaper by its circulation.

Ethical issues. The experimental design may raise ethical concerns since the methodology does not allow us to obtain informed consent by the experimental subjects. These concerns apply similarly to audit and correspondence studies, which represent established and widely accepted methodologies for studying discrimination (see Riach and Rich, 2002; Guryan and Charles, 2013; Bertrand and Duflo, 2017, for surveys). By one count, there are 117 studies from 17 different countries using this approach (Salganik, 2017). In this literature, four conditions are pointed out that jointly justify forgoing informed consent (Riach and Rich, 2004; Pager, 2007; Salganik, 2017): (i) any potential harm to subjects is minimal, (ii) the study generates socially valuable insights that (iii) cannot be achieved with other empirical methods, and (iv) the experiment takes place

[6] In the literature, the term *incumbency bonus* is widespread. We instead speak of *incumbency dominance* as this highlights that coverage does not have to be a bonus for the incumbent, but can be positive or negative.

in a context where some forms of deceptions are not unheard of, so that it does not "pollute an already pristine ethical landscape" (Salganik, 2017, p. 304).

We believe our experiment fulfills these requirements.[7] We consciously designed the experiment to minimize any potential harm for the newspapers' readers and the wider public. The number of letters that we sent is small in comparison to the overall number of letters in German newspapers.[8] Sending equal numbers of letters supporting and criticizing Merkel and Schulz ensures that our experimental intervention is politically balanced, although a biased selection of the letters by the newspapers might result in an unbalanced effect on voters. The letters do not contain any wrong or misleading statement of facts. They contain an expression of a personal opinion about which candidate is better or worse, but do not contain an argument supporting that opinion—after all, the text had to fit all four versions of the letter equally well. All letters call for fairness in reporting and high participation in the election; these are widely shared democratic values to which we fully subscribe ourselves.

Potentially adverse effects of our study for newspapers could comprise the time required to read and process our letter, or the opportunity cost of newspaper space when printing our letter. We believe these costs to be minimal. The decision to print the letter is, after all, the newspaper's decision. Dealing with letters to the editor is a typical everyday activity both for newspapers and readers. Our experiment does not expose anyone to a harm or discomfort greater than those ordinarily encountered in daily life, and thus meets the appropriate standard for minimal risk (Morton and Williams, 2010, pp. 479-483).

One might also be concerned that a newspaper might suffer a loss of reputation when it becomes known as biased. Our study is not, however, designed to test for discrimination by individual newspapers. Indeed our method does not allow this type of inquiry, which would require sending different letters to the same newspaper. We test for the prevalence of biases across the newspaper landscape, and no newspaper can be identified from our research as biased, as this would require sending different letters to the same newspaper.

To summarize, we designed and executed the experiment to ensure minimal harm (condition i). Media bias is a hotly debated topic both in academia and the public more generally. Our study contributes scientifically to this debate, generating socially valuable insights (condition ii). As mentioned in the introduction, any study of media bias faces the unobserved population problem, and creating the population in a field experiment is a unique way to overcome this challenge

[7] At the time of writing, the faculty of economics and social sciences at our university is establishing an ethics commission. At the time of our experiment, however, no such internal review board had been established, so we could not have asked it for approval.

[8] To give a rough estimate, 1604 letters have been published in our study period (September 5 to 24, 2017) in newspapers that make the letters to the editor available on Nexis; 12 of them (or about 0.7%) stemmed from our experiment. Note that these are not all letters published (see Section 3.3 for details on our data collection).

(condition iii). Finally, given the recent debates about media bias, the media are rarely perceived as a "pristine landscape" that our experiment might pollute (condition iv).

3.3 Implementation and data collection

Between September 5 and 8 in 2017—in the third week before the general elections—we sent out letters to the editor to over 200 daily German newspapers.[9] Our letters were highly topical as we referred to the reporting of the TV debate between chancellor Merkel and her challenger Schulz that took place shortly before, on September 3. According to election forecasts, the study period was quite representative of the last few months before the election, with no major swings for the biggest parties.[10]

For the selection of newspapers in the study, we relied on a compilation of German daily newspapers published by the Federation of German Newspaper Publishers (Bundesverband Deutscher Zeitungsverleger).[11] We visited the websites of all these newspapers to find out whether there is an online form or an e-mail address for submitting letters to the editor. If no form and no specific e-mail address for letters to the editors was presented, we used the e-mail address of the editors. Many newspapers in Germany have common owners, common publishing houses, or share one and the same section on federal politics. To prevent interference between experimental units, we submitted only one letter to different newspapers that handle letters to the editor through one and the same online form or e-mail address. On the other hand, when newspapers that have common owners or different local editions handle letters to the editor by independent editorial departments, we treated them as independent experimental units.[12]

One of our four different letters was sent to each newspaper, either via the contact form of the homepage or via e-mail. In addition, we provided the contact details of a fictitious sender, "Annamarie Richter." We wanted to use a common name that would not raise any suspicions, while at the same time making sure we would not accidentally write a letter with the name and address of an actual person. We chose a common German family name, "Richter", and combined it with a comparatively rare first name, "Annamarie", which is a version of the more common first name "Annemarie." On online telephone books, such as `dastelefonbuch.de`, no person with the name *Annamarie Richter* can be found. Her address was always the same, except for the city she lived in. For each newspaper we chose the address *Hauptstr. 14*, in or near the city

[9] Our randomization ensures that the version of the letter is independent of the time the letter was sent.

[10] See, for example, `http://www.wahlrecht.de/umfragen/allensbach.htm` (accessed on Dec. 9, 2017).

[11] See `http://www.bdzv.de/maerkte-und-daten/zeitungslandschaft` (accessed on Aug. 9, 2017).

[12] This leaves us with about two thirds of the 333 newspapers that are registered in Germany according to BDZV.

where the newspaper has its headquarters, and the respective postal code.[13][14] *Hauptstrasse* is by far the most common street name in Germany ("main street") and can be found in most cities and communities.[15] If requested on the online form, we also provided the mobile number of one of the co-authors, but we did not answer any phone calls. Mobile numbers in Germany cannot be ascribed to particular locations or cities. We answered e-mails that asked for further contact details (such as the telephone number) with one standard e-mail providing the requested details. All letters to the editor and all e-mails included the statement that we would appreciate receiving a notification of whether the letter would be printed as we would not be reading the newspaper in the following weeks due to a vacation.[16]

To find out which newspapers did publish our letter, we collected notifications by e-mail and telephone from the newspapers as to whether our letter had been printed or rejected. In addition, we searched newspapers' websites and the Internet using the general search engines *Google* and *Bing*, for the name "Annamarie Richter." [17] We also made use of the Nexis and the Genios newspaper databases, which contain about one half of the newspapers in our study. The coverage of letters to the editor in these databases, however, is not 100% complete: some newspapers that are available in Nexis or Genios had published our letter, but it was impossible to find our letter in these databases. Therefore, we contacted the remaining newspapers—those that had not informed us directly and where our search had not shown positive proof that the letter had been published—by e-mail and telephone after the federal election. For two newspapers, the status of our letter remained unclear, and we had to leaf through their print issues in order to learn whether our letter had been printed or not.[18]

[13]If the contact form did not indicate that the address was obligatory we did not provide it.

[14]As most daily newspapers are regional newspapers we had to vary the postal code and the city of the address across newspapers to avoid suspicions.

[15]See `http://www.strassen-in-deutschland.de/die-haeufigsten-strassennamen-in-deutschland.html` (accessed on Aug. 9, 2017).

[16]Vote shares of the two main parties differ significantly across the German Bundeslaender (states). We therefore stratified on the state level from BDZV (2017), which reports the newspapers' state by merging the city-states of Berlin, Bremen, and Hamburg with the surrounding territorial states, so that there are 13 different states.

[17]German newspapers typically publish letters to the editor with the full name. As "Annamarie" is very similar to the more common name "Annemarie", we also searched for "Annemarie Richter." We also searched for "A. Richter", and for snippets from the text of the letter, but this did not result in additional information.

[18]We cannot rule out that some newspapers realized that the letter had already been published elsewhere and therefore rejected the letter. This would have reduced the number of printed letters. Importantly, the probability of a newspaper discovering the letter elsewhere should be independent of the treatments. Moreover, it seems plausible to assume that a newspaper that discovers any version of our letter published in another newspaper will reject our letter, no matter what version itself has received. Under this assumption (which is unfortunately not testable with our data), such cases of interference would bias our results towards zero and thus render our tests conservative as the treatment effects might have been stronger absent the interference.

3.4 Results

Out of 214 letters in our data set, 89 (i.e., 41.6%) were printed.[19] Table 3.1 gives an overview of the data.

Table 3.1: Overview of the data

	Sent	Printed	Printed (%)
Version 1 (*pro Merkel*)	51	24	47%
Version 2 (*contra Schulz*)	53	21	40%
Version 3 (*pro Schulz*)	54	16	30%
Version 4 (*contra Merkel*)	56	28	50%

Notes: Column one gives the number of observations in each treatment. Column two states for each treatment how many letters were printed. Column three gives the share of printed letters in percentages.

When submitting letters we supplied all the information asked for by the newspaper. Conditional on this information, 28 newspapers (13.1%) asked for further information by e-mail or by calling the phone number we provided. Out of these 28 newspapers, seven eventually printed the letter. Moreover, four newspapers informed us that they would print the letter if they could reach us by telephone, but did not print it finally since we never answered their phone calls.[20]

3.4.1 Nonparametric tests

In order to test our three hypotheses we use the Fisher's exact test. The subject pool in our experiment is not a sample from some bigger population, but rather the population of all German daily newspapers that handle letters to the editor independently. Therefore, a test which relies on the randomization distribution for inference is appropriate.

First, as we are agnostic about the direction of a potential political bias, we used a two-sided test to test Hypothesis I. We do not find a political bias: out of 104 pro-Merkel and contra-Schulz letters

[19]Among the newspapers we initially selected for our study, one turned out to be out of business and one was strikebound. We used defunct e-mail addresses for two further newspapers. In addition, some newspapers informed us that they had forwarded our letter to a central editorial board dealing with letters to several newspapers, potentially inducing interference between the corresponding experimental units. This eliminated 24 further newspapers from our sample. In a robustness check, we included these 24 newspapers as not having published the letter. Our results from the non-parametric tests on Hypotheses I to III remain largely unaffected in the sense that we do not find evidence of a political bias or a statistically significant negativity bias, but for incumbency dominance ($p = 0.047$, one-sided).

[20]As a randomization check, we regressed newspaper circulation and the left-right score by Garz et al. (2019) on the three main indicator variables. None of the coefficients is significant, so our randomization seems to have worked. Moreover, there is no significant difference in the acceptance rate of letters submitted via online form or via e-mail. The former were accepted in 40.8%, the latter in 42.9% of the cases.

45 (43.3%) were printed, while out of the 110 contra-Merkel or pro-Schulz letters 44 (40.0%) were printed ($p = 0.678$). Thus, publications in our data set do not seem to be biased toward the political left- or right-wing.[21]

Second, we test for negativity bias. Since we had a clear prediction that negative letters were more likely to be printed, we use a one-sided test here. Out of 109 negatively connotated letters 49 (45.0%) were printed; out of 105 positively connotated letters only 40 (38.1%) were printed. Thus, negatively connotated letters are more likely to be printed, but this effect is not statistically significant ($p = 0.190$).[22]

Third, we find a significant effect of incumbency dominance ($p = 0.026$, one-sided) whereby letters about Merkel were more likely to be printed than letters about the challenger. Out of 107 letters about Merkel, 52 (48.6%) were printed, while only 37 out of 107 letters (34.6%) about the challenger were printed. Hence, a letter about Merkel had a $48.6/34.6 - 1 \approx 40.5\%$ higher chance of publication. Thus, newspapers were more likely to publish letters about Merkel than about Schulz. This indicates a higher newsworthiness of the Merkel letters.

Notably, testing three hypotheses on the same sample may increase the false discovery rate. To take this into account, we conduct a Bonferroni correction by multiplying the *p*-values by the number of hypotheses, that is, $n = 3$. Although this is a very conservative correction, incumbency dominance stays (weakly) significant ($p = 0.078$).

3.4.2 Regression Analysis

As argued in Section 3.2, larger newspapers or national newspapers might have a different approach to handling letters to the editor than smaller newspapers or regional newspapers, for example, because they may receive more letters and can thus be more selective. Controlling for these covariates might thus improve the precision of our estimation. For ease of interpretation, we estimate the linear probability model[23]

$$Print_i = \beta_0 + \beta_1 Left_i + \beta_2 Negative_i + \beta_3 Incumbent_i + \beta_4 X_i + \varepsilon_i. \qquad (3.1)$$

[21] Using Butler and Schofield (2010) result of a 26% acceptance rate as a baseline, a power calculation shows that given our sample size and a desired power of 80%, the minimal effect size for a political bias that we would have been able to detect on a significance level of 5% would have been 0.19 (0.17 on the 10%-level). Our estimate, however, is close to zero (0.033).

[22] Using, as before, a 26% acceptance rate as a baseline, the minimal effect size for negativity bias (i.e., the difference in acceptance rates of negative and positive letters) that we would have been able to detect on a significance level of 5%, with a desired power of 80% and our given sample size, would have been 0.17 (0.15 on the 10%-level). Our estimate, however, is considerably smaller (0.069). To obtain a statistical significance of 5% for that effect size, we would have needed a sample of more than 1000 newspapers.

[23] We also ran a logistic regression and the results are similar.

The dependent variable $Print_i$ is a dummy variable that equals 1 if newspaper i has printed the letter. $Left_i$ is an indicator variable for the left-leaning letters (the versions pro Schulz and contra Merkel), $Negative_i$ is an indicator variable for the negative letters (contra Merkel and contra Schulz), and $Incumbent_i$ is an indicator for letters about the incumbent chancellor Merkel (the versions pro Merkel and contra Merkel).[24] X_i contains the control variables. We use quarterly circulation (measured in 1,000 units, as reported in BDZV (2017), a dummy for national newspapers, and dummies for the state (Bundesland) where the newspaper is published. The ε_i is mean zero noise.

It is straightforward to see that β_1 gives the expected difference between the average acceptance rates of a left-leaning and a right-leaning letter. Similarly, β_2 is the expected difference between the average acceptance rates of a negative and a positive letter. Finally, β_3 is the expected difference between average acceptance rates of a letter about Merkel and Schulz.

Table 3.2 reports the results. The dependent variable is an indicator of whether the newspaper printed the letter. Column 1 does not control for any covariates and basically reproduces our results above: only the incumbent dummy is statistically significant (at the 5% level), and indicates that the probability of a letter on Merkel being accepted is 13.9 percentage points higher than for a letter on Schulz.[25] Next, we control for circulation and for national newspapers. We do not enter these control variables simultaneously, since they are highly correlated: national newspapers have a higher circulation. As expected, larger newspapers were less likely to publish our letter. The dummy for national newspapers is statistically significant, while circulation is not. Finally, we add state dummies to these regressions. None of the state dummies are statistically significant. The estimation results for the parameters of interest remain basically unchanged. In particular, across all specifications, the coefficient of the incumbent dummy is estimated between 13.6 and 15.8 percentage points.

We conclude that our results are robust when controlling for the covariates. Our results are also robust to different choices of standard errors, that is, classical standard errors and bootstrapping. Furthermore, our results are robust to running separate regressions for the three hypotheses. One further robustness check concerns the publishing houses. Newspapers within the same publishing house may not decide independently whether to print the letter. As argued in Section 3.3, we only

[24]Note that the three dummy variables $Left_i$, $Negative_i$, and $Incumbent_i$ are not perfectly collinear.

[25]Apart from being based on a different statistical method, this test differs in a minor detail from the nonparametric tests discussed above. The OLS estimator of, for example, β_1 can be shown to be equal to the difference between (i) the unweighted average of the acceptance rates of versions pro Schulz and contra Merkel, and (ii) the unweighted average of the acceptance rates of versions pro Merkel and contra Schulz. In contrast, the test statistic for Hypothesis I used above considers the differences between the respected weighted averages, with weights equal to their share of observations. This also explains why the coefficients in Table 3.2, column 1, are slightly different from the raw differences used in the non-parametric tests. Since we have an almost equal number of versions, this difference is minor.

Table 3.2: Regression results

	Print	Print	Print	Print	Print	Print
Left Letter	-0.0353	-0.0373	-0.0406	-0.0511	-0.0540	-0.0607
	(0.0672)	(0.0674)	(0.0671)	(0.0670)	(0.0671)	(0.0666)
Negative Letter	0.0647	0.0702	0.0644	0.0561	0.0622	0.0548
	(0.0672)	(0.0677)	(0.0671)	(0.0676)	(0.0679)	(0.0672)
Incumbent Letter	0.139**	0.136**	0.139**	0.156**	0.153**	0.158**
	(0.0672)	(0.0675)	(0.0671)	(0.0673)	(0.0676)	(0.0668)
Circulation		-0.000209			-0.000282*	
		(0.000132)			(0.000146)	
National Newspaper			-0.263*			-0.425**
			(0.151)			(0.164)
State dummies				X	X	X
Constant	0.332***	0.345***	0.342***	0.391	0.432	0.397
	(0.0672)	(0.0692)	(0.0676)	(0.339)	(0.361)	(0.342)
N	214	214	214	214	214	214
R^2	0.026	0.029	0.034	0.112	0.117	0.129

Notes: Robust standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variable is a dummy variable equal to 1 if newspaper i has printed the letter. Left Letter is an indicator variable for left-leaning letters, Negative Letter an indicator for negative letters, and Incumbent Letter an indicator for letters on Merkel. Control variables include the quarterly circulation of a newspaper in thousands, a dummy variable equal to 1 if newspaper i is a national newspaper, and state dummies.

sent one letter per publishing house if the associated newspapers' editorial departments are not independent. To account for potential non-independent decisions in the few cases where we sent more than one letter per publishing house, we clustered standard errors at the publishing house level and our results are unaffected.

Weight by newspaper circulation. So far, all observations are given equal weight in the analysis, but typically only a handful of newspapers shape the whole market. For an equal number of right- and left-leaning letters being printed, our preceding analysis would not indicate a political bias, even if the newspapers that print the left-leaning letter have a much higher readership than those printing the right-leaning letter. As a consequence, it might be more natural to use $\frac{c_i}{\sum_j c_j} \cdot Print_i$ as the dependent variable, where c_j gives the circulation of newspaper j. Whereas our above regression answers whether the average newspaper has certain biases, the weighted regression answers whether the newspaper of the average reader has these biases.[26] Table 3.3 shows our results. While the signs of the coefficients are analogous to our findings from Table 3.2, their magnitudes are not directly comparable. We can, however, compare the effect sizes in terms of standard deviations of the dependent variables. For instance, according to our main regression (Table 3.3, Column 1) a letter on Merkel leads to a change that corresponds to 0.28 standard deviations, while the change corresponds to 0.32 standard deviations if we weight by circulation (Table 3.3, Column 1). Hence, our results are robust to weighting the newspapers by their share of circulation.

Exploring the role of previously published letters. To put our findings into perspective, we investigated how many letters to the editor about Merkel, Schulz, or both of them had been published in the three months before our experiment. We relied on the coverage of letters to the editor in all German newspapers available on Nexis.[27] The search resulted in 270 letters about Merkel and 158 about Schulz, so there were 70.9% more published letters about Merkel than Schulz. Another 62 letters contained references to both candidates. Our finding that letters about Merkel have a 40% higher chance getting published can partially explain the higher number of published letters on Merkel.[28]

[26] Note that this analysis differs from weighting the squared residual for each observation with its respective circulation. The latter leads to similar results as our main regression.

[27] We searched Nexis for publications containing the key words "Merkel" and "Schulz" in combination with the German word for letter to the editor ("Leserbrief"), in the time period June 1 to August 31, 2017. A research assistant read through all the results in order to count only those letters that were really about the respective candidates, as opposed to, e.g., letters about (or written by) some other person named Merkel or Schulz.

[28] Our estimation results are conditional on the given supply of other letters to the editor reaching the newspapers. To explore whether the acceptance decisions depend on the number of letters concerning Merkel or Schulz that have already been published in a newspaper, we had a closer look at all letters published in the three weeks prior

Table 3.3: Circulation

	Print(W)	Print(W)	Print(W)	Print(W)	Print(W)	Print(W)
Left Letter	0.000398	0.000452	0.000376	0.000304	0.000342	0.000249
	(0.000456)	(0.000443)	(0.000456)	(0.000435)	(0.000429)	(0.000435)
Negative Letter	0.000798*	0.000652	0.000797*	0.000536	0.000457	0.000528
	(0.000456)	(0.000425)	(0.000456)	(0.000417)	(0.000405)	(0.000414)
Incumbent Letter	0.00109**	0.00116**	0.00109**	0.00121***	0.00125***	0.00122***
	(0.000456)	(0.000445)	(0.000456)	(0.000427)	(0.000425)	(0.000425)
Circulation		0.00000553			0.00000367	
		(0.00000468)			(0.00000385)	
National			-0.00108*			-0.00242***
			(0.000625)			(0.000832)
State Dummies				X	X	X
Constant	0.000548	0.000184	0.000590	0.00711	0.00658	0.00714
	(0.000456)	(0.000495)	(0.000459)	(0.00566)	(0.00543)	(0.00569)
N	214	214	214	214	214	214
R^2	0.044	0.089	0.046	0.219	0.237	0.231

Notes: Robust standard errors in parantheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variable is a dummy variable equal to 1 if newspaper i has printed the letter weighted by i's share of overall newspaper circulation. Left Letter is an indicator variable for left-leaning letters, Negative Letter an indicator for negative letters, and Incumbent Letter an indicator for letters on Merkel. Control variables include the quarterly circulation of a newspaper in thousands, a dummy variable equal to 1 if newspaper i is a national newspaper, and state dummies.

3.4.3 Additional findings: Do newspapers prefer letters that oppose their political position?

Our study also allows us to test whether newspapers are more likely to print a letter to the editor that reflects a political view which is opposite to the political orientation of the newspaper. The German Press Code, a voluntary ethical agreement among German publishers, explicitly encourages the publication of such letters. It states under Guideline 2.6: "The Press Code must be observed when publishing readers' letters. It is in the interest of informing the public to allow opinions not shared by the editorial team to be expressed in the Readers' Letters section."[29] Moreover, Butler and Schofield (2010) found evidence of such a balancing behavior in the context of the 2008 US presidential election.

German newspapers do not publish election endorsements, however, which makes it harder to determine their political positions. This problem is amplified by an internal plurality of German newspapers whereby the political orientation changes over time or between different parts of the newspapers (e.g., politics vs. feuilleton). That a newspaper's political position is less clear for German than for US newspapers might also be driven by the larger number of parties in the parliament—Germany has a multi-party system—and by a greater similarity of the biggest parties, that is, CDU/CSU and SPD in Germany are more similar than the Democrats and Republicans in the US. Nevertheless, recent studies have attempted to classify a number of German newspapers according to their political position. Dewenter et al. (2016) rely on Media Tenor's human coding of the tonality of news about the biggest parties, CDU/CSU and SPD. Their data cover only seven newspapers used in our study. Due to the small number of observations, we just classify these newspapers as left-leaning (*taz*, *Frankfurter Rundschau*, and *Süddeutsche Zeitung)* versus right-leaning (*Frankfurter Allgemeine Zeitung, Welt*, *BILD* and *Berliner Zeitung)*, based on the Index of Political Coverage from Dewenter et al. (2016).[30] Among those seven newspapers, three received a letter that opposed their political position, and they printed it; the remaining four received a letter that matched their position, and none of them printed it. This fits nicely with the idea that newspapers prefer to publish letters that oppose their own position.

Garz et al. (2019) have ordered a number of German newspapers according to their relative

to our study in a convenience sample of 21 newspapers in our data, which make the letters available on Nexis. The letters were coded as *pro Merkel*, *contra Merkel*, *pro Schulz*, *contra Schulz*, and*other* by two student research assistants (intercoder reliability as measured by percent agreement of 96%). Dissenting codings have been decided by discussion between the students. In this sample, we found evidence for incumbency dominance ($p < 0.05$), but no effects of previously published letters on the printing decision.

[29] See `presserat.de/fileadmin/user_upload/Downloads_Dateien/Pressekodex13english_web.pdf` (accessed on Dec. 8, 2017).

[30] In an earlier study, Eilders (2002) investigated the positioning for five of these newspapers (taz, FR, FAZ, SZ and Welt) and came to a similar conclusion on their positioning.

political position. They rely on an automated text analysis comparing word frequencies in media outlets and party programs. Using their left-right scores, we obtain the political position of 46 of the newspapers used in our study. Since Garz et al. (2019) selected the media outlets for their study by audience reach, this subsample represents newspapers with a larger audience.[31]

Table 3.4.3 reports regression results with the left-right score of the respective newspapers and an interaction term as regressors. The estimation equation is[32]

$$Print_i = \beta_0 + \beta_1 Left_i + \beta_5 Score_i + \beta_6 Score_i \cdot Left_i + \varepsilon_i. \tag{3.2}$$

A higher score indicates a more conservative newspaper. The coefficient of interest is that of the interaction term. It shows that conservative newspapers are more likely to publish a left-leaning letter. The effect size is substantial. Consider two newspapers that differ by one standard deviation in their left-right score, which is about 0.014. Then the estimated difference across the two newspapers, between the differences in acceptance rates of left- and right-leaning letters, is about $23.91 \cdot 0.014 \cdot 100 \approx 33$ percentage points.[33]

In summary, our data support the hypothesis that newspapers tend to publish letters that oppose their own political position.[34]

3.5 Discussion and Concluding Remarks

This study reports results from a field experiment on letters to the editor in order to test for different implications of media bias. In particular, it allows us to test for three different manifestations of filtering bias with respect to media coverage of political content, that is, political bias, negativity bias, and incumbency dominance. Our between subjects design allows for a test of systemic biases of the German newspapers as opposed to individual outlet bias. We found no political bias among German newspapers with respect to the publication of letters to the editor. Moreover, the newspapers seem to follow the recommendation in the German press code to print letters that oppose their own political position. Indeed, we find that letters that oppose a newspaper's political

[31] The average quarterly circulation of a newspaper in the subsample is 177,566 copies as opposed to 68,071 copies for the whole dataset.

[32] Alternatively, we ran a logistic regression and the results are similar.

[33] This result is partly driven by one observation with an extreme score, that is, the self-proclaimed "Socialist" newspaper Neues Deutschland. When removing this observation from the sample, the estimated coefficient still has the same sign, but it is about 30% smaller in absolute value and loses statistical significance.

[34] The results in this section also show that the regression in the main results section do not capture all relevant explanatory variables. Given random assignment, however, the estimates of (3.1) should nevertheless be unbiased. Moreover, controlling for newspapers' political position in a test of political bias of the newspapers' landscape as a whole would control for too much: the object of interest is the average reactions of newspapers, not the average reactions conditional on their political leanings.

Table 3.4: Additional results

	Print	Print	Print	Print
Left Letter	-0.00469	-0.00586	-0.0152	-0.0752
	(0.139)	(0.149)	(0.153)	(0.155)
Score	-22.18***	-22.38***	-21.45***	-27.25***
	(2.982)	(4.451)	(4.520)	(4.624)
Left Letter x Score	22.85***	23.76***	22.97***	23.91***
	(5.840)	(6.909)	(7.043)	(6.514)
Negative Letter		0.0602	0.0680	0.0282
		(0.160)	(0.162)	(0.159)
Incumbent Letter		0.0446	0.0348	0.0388
		(0.160)	(0.163)	(0.155)
Circulation			-0.000155	
			(0.000118)	
National Newspaper				-0.425***
				(0.138)
Constant	0.390***	0.336**	0.368**	0.452**
	(0.0958)	(0.151)	(0.163)	(0.174)
N	46	46	46	46
R^2	0.103	0.110	0.115	0.176

Notes: Robust standard errors in parentheses. $^{*}p < 0.1$, $^{**}p < 0.05$, $^{***}p < 0.01$. The dependent variable is a dummy variable equal to 1 if newspaper i has printed the letter. Left Letter is an indicator variable for left-leaning letters, Negative Letter an indicator for negative letters, and Incumbent Letter an indicator for letters on Merkel. Score represents the political position of newspaper i on a left-right score and ranges from -1 (extremely left) to +1 (extremely right). Control variables include the quarterly circulation of a newspaper in thousands and a dummy variable equal to 1 if newspaper i is a national newspaper.

position have a higher chance for publication in this newspaper. In line with the literature, we observed that our negatively connotated letters were printed more often than positively connotated ones. They were not statistically significantly more likely to be printed, however. Finally, we observed a strong effect of incumbency dominance, which gives a relation to the literature on incumbency advantages in elections. Incumbents are more likely to win elections, and the effect that the incumbent gets more media attention may contribute to this fact. Some care needs to be taken when interpreting the incumbency dominance we found, however, since we cannot control for other differences between the candidates.

We stay agnostic with respect to the reasons for media bias. When testing for political bias, negativity bias, and incumbency dominance, we did not impose any assumptions on whether these effects should be driven by the intrinsic preferences of the journalists or by readers' demand for biased media coverage, for instance. While we document significant support for incumbency dominance, the reasons for the occurrence of this effect need to be addressed in future research.

Our study was inspired by Butler and Schofield (2010) who conducted a similar experiment in the US in 2008 with letters to the editor that were either supportive of McCain or Obama. Our 2×2 design has various advantages over the design of the original study. If we had sent only supporting versions for either of the candidates (Versions 1 and 3), we might have observed a political bias as the pro-Merkel-letter would have been printed more often. By including the criticizing versions, however, we see that in fact we do not have a political bias, but an effect of incumbency dominance: Not only letters supportive of Merkel, but also those that criticize Merkel are more likely to be printed. The original study could also neither elicit this effect nor test for negativity bias as only supportive letters were sent. Finally, it could not test for incumbency dominance as none of the candidates—neither Obama nor McCain—were incumbents. Our data also have advantages over those by Butler and Schofield (2010). While we count publications, they treated publications and contacts made by the newspaper equivalently. Just making a contact by the newspaper, however, does not necessarily indicate that the newspaper wants to print the letter. Furthermore, our sample is much larger as we included all the German daily newspaper that handle letters to the editor independently. Altogether, our study has various advantages over the original study with respect to the experimental design and the data set.

Finally, our study contributes to the debate on the manipulation of user-generated content (see, e.g., Mayzlin et al., 2014). We have observed that it is relatively easy to place a fictitious letter in a newspaper. While our experiment was balanced with respect to the political message (as we sent roughly equal numbers of all versions of the letter), our findings raise the question whether it is possible to affect the press and therefore also public opinion through fake letters that are less balanced in the aggregate.

4 Coverage Bias on Wikipedia? Evidence from Biographies of German Members of Parliament

with Johannes Münster

4.1 Introduction

User-generated content is becoming more and more important. Consumers turn to Yelp to find a restaurant, to TripAdvisor to plan a vacation, and to Wikipedia to search for information (Luca, 2016b). Three out of the top five websites by Internet traffic – YouTube, Facebook, and Wikipedia – are based on user-generated content.[1] These websites rely on millions of voluntary contributions, each being a public good in its own, with no price mechanism steering the contributors' choices, and little if any centralized coordination of efforts. Research on the private supply of such *multiple* public goods shows that the equilibrium *mix* of public goods is typically inefficient; this inefficiency comes in addition to the well known result that overall contribution levels fall below the Samuelson optimum (e.g., Bilodeau, 1994; Ghosh et al., 2007; Cornes and Itaya, 2010). Whether the individual contributions to user-generated content websites lead to an overall balanced mix of content, or to systematic coverage biases, is therefore unclear.

This paper studies if Wikipedia, the world's largest online encyclopedia, has a coverage bias in its biographies of German members of Parliament (MPs). That is, we ask if the individual contributions of Wikipedia users lead to the unbalanced coverage of otherwise comparable MPs from different political parties. Our main proxy for coverage is biography length. Biography length is easy to measure objectively; moreover, the MPs' biographies contain little if any criticism. In addition, biography length is a meaningful measure in itself: the length of a biography may influence opinion formation if voters interpret it as an indication of an MP's valence characteristics.[2] Hence,

[1]See `https://www.alexa.com/topsites` (Dec 2018).

[2]To check the plausibility of this claim, we conducted a classroom survey, finding that a longer biography signals knowledge, strength as a public servant, and the ability to inspire people (see Appendix C.1).

a coverage bias on Wikipedia – in other words, an unbalanced coverage of comparable MPs from different political parties in terms of their biography length – may affect the workings of democracy.

A major challenge is to disentangle the effect of party affiliation on biography length from the effect of an MP's characteristics. Partisan contributors could amplify the biographies of MPs from a specific party, leading to a coverage bias as defined above. If, on the other hand, MPs about whom there is more to write self-select into a particular party, differences in Wikipedia coverage between MPs from different political parties would – according to our definition – not be classified as bias. Since many MP characteristics are unobserved (e.g., ability, wittiness, or looks), we cannot easily distinguish their impact on biography length from the impact of the MPs' party affiliation.

We address this issue in two steps. First, we study a sample of relatively homogeneous observations. We consider the 18th German Bundestag (2013 to 2017) and focus on MPs from Germany's two biggest political parties, the center-right CDU/CSU and the center-left SPD, who jointly comprised more than three quarters of all MPs and formed a coalition government. In particular, when we compare the biographies of MPs from the CDU/CSU and the SPD, differences in length cannot originate from differences in government versus opposition parties, centrist versus more extreme parties, or big versus fringe parties.[3] To further increase comparability, we exclude MPs in distinguished offices such as Chancellor Angela Merkel, ministers, and party heads from the analysis.

As a second step, we compare the length of German and English Wikipedia biographies in a difference-in-differences framework. Partisan contributors are less likely to amplify the English biographies, since German voters are unlikely to read them. Assuming that unobserved MP characteristics affect the German and the English biography length equivalently, a difference-in-differences estimation using language as a first, and party affiliation as a second difference, yields unconfounded estimates of the effects of party affiliation on biography length. Since English biographies are only available for about a quarter of our observations, we also take potential selection effects into account.

We find that biographies of MPs from the SPD are, on average, about half a page shorter than biographies of MPs from the CDU/CSU. These differences remain after controlling for gender, political experience, outside earnings, education, and MPs' constituency demographics. To put the effect size into perspective, note that the average biography length is about 2.33 pages, the median length is about 1.66 pages, and the standard deviation in biography length is about two pages. The difference-in-differences estimate for coverage bias, i.e., the effect of party affiliation on biography length, is about twice as large as our basic estimate, confirming that the unequal

[3]The literature on the news coverage of politicians has found a clear incumbency effect on coverage (Vos, 2014).

coverage between MPs from CDU/CSU and SPD is not driven by unobserved MP characteristics.

While biography length is our main proxy for coverage, biographies of MPs from the SPD also differ from those of the CDU/CSU in several additional dimensions of coverage: they exhibit fewer images, fewer adjectives, a lower adjective to word ratio, and a smaller number of links to external websites under the control of the MP or her party than biographies on MPs from the CDU/CSU. Images and adjectives brighten texts and contribute to a more positive coverage, and a high number of weblinks under party control indicates that Wikipedia is used more extensively for election campaigns. Biographies of MPs from the SPD are also assigned to a lower number of Wikipedia categories, which makes them harder to find.

As argued, differences in coverage between SPD and CDU/CSU may be driven by partisan contributions to Wikipedia. We check the plausibility of this explanation in three ways. First, we identify all anonymous edits conducted from the Bundestag network by tracking the users' IP addresses. Consistent with our explanation, we find that biographies of MPs from the CDU/CSU are edited nearly 50% more often from the Bundestag building than biographies of MPs from the SPD. Second, we document that there are fewer authors who repetitively contribute to SPD biographies only. Finally, if partisan contributions drive the differences in coverage, they should generate debates within the Wikipedia community. A testable implication is that Wikipedia's talk pages for MPs from the SPD should be shorter. Indeed, there are fewer talk pages for MPs from the SPD, and these pages are also shorter on average.

Our paper makes two contributions. First, our results provide empirical evidence that the individual contributions to user-generated content websites can lead to systematic coverage biases. Wikipedia is an interesting case in point, since it is a well known privately provided public good (e.g., Zhang and Zhu, 2011) that attracts millions of readers every month.[4] Second, we present a novel empirical strategy to detect coverage biases on user-generated content platforms. While existing approaches use variation in institutional features *across* or *within* platforms (e.g., Anderson and Magruder, 2012; Mayzlin et al., 2014), we exploit variation between different language versions of a platform.

The paper proceeds as follows. Section 4.2 discusses the related literature. Section 4.3 describes our dataset. The empirical strategy and the results with respect to biography length are presented in Section 4.4; Section 4.5 investigates the occurrence of adjectives, images, categories, and weblinks under party control as further measures of coverage and shows that the results are similar. In Section 4.6, we show that partisan contributors are a likely driver of coverage bias. Section 4.7 confirms that negative coverage of MPs does not confound our analysis. Section 4.8 discusses the external validity and limitations of our analysis.

[4] See `https://en.wikipedia.org/wiki/Wikipedia:Statistics` (Dec 2018).

4.2 Related literature

The paper relates to three overlapping strands of literature: user-generated content, media bias, and the private supply of public goods. The literature on user-generated content and social media has shown that user-generated content has large causal effects on economic and social outcomes (see Luca, 2016b, for a survey). Our paper is related to Mayzlin et al. (2014) who study promotional content in hotel reviews. Moreover, there are several studies showing that Wikipedia exhibits only few factual errors (Mesgari et al., 2015), but has frequent errors of omission (Brown, 2011), which motivates our focus on coverage. It is often stated that coverage in Wikipedia suffers from a systemic bias induced by its contributors' demographics: a majority is English-speaking, white, male, and Internet affine (Halavais and Lackaff, 2008), which leads, for instance, to the underrepresentation of women (Reagle and Rhue, 2011; Hinnosaar, 2019). Our analysis shows that while biographies of women are shorter on average, the effect is not statistically significant and even changes sign when adding controls for the demography of the MPs' constituencies.

Most of the research on media bias (see Gentzkow et al., 2016; Puglisi and Snyder, 2016, for surveys) has focused on traditional media – such as newspaper and television – while comparatively little is known about political biases in user-generated content. An exception is Greenstein and Zhu (2012) who apply automated text analysis to measure the slant in Wikipedia articles on political issues. In their paper, slant is an intrinsic property of an article, measuring whether its language is more typical for the Republican or Democratic party in the USA. In contrast to that, our measure of coverage bias is based on the overall coverage of German MPs from different political parties. In addition, research on media bias has mainly focused on media in the USA, while research on biases of media in other languages is rare. The political systems and the media systems of the USA and Germany differ along many dimensions (Persson and Tabellini, 2005; Hallin and Mancini, 2004), and it is a priori unclear whether results obtained in the US will hold in the German case. Indeed, a naive extrapolation of the result that the English Wikipedia has a pro-liberal bias (Greenstein and Zhu, 2012) might suggest that, when comparing the two biggest German parties, the German Wikipedia is biased in favor of the center-left SPD relative to the center-right CDU/CSU. Our results, however, show the exact opposite.

A large literature has studied public goods (see Batina and Ihori, 2005, for a book length survey). As argued, theory has shown that the equilibrium mix of multiple privately provided public goods may be inefficient (Bilodeau, 1994; Ghosh et al., 2007; Cornes and Itaya, 2010). Further related papers are Ihori et al. (2014), who apply the theory of multiple public goods to the burden sharing in the NATO, and Hellwig (2007) and Fang and Norman (2010) who study the provision of multiple excludable public goods. We provide empirical evidence on the mix of multiple privately provided public goods, a topic that has not received much attention in the empirical literature, except in the contexts of charities (Vesterlund, 2016). Our study also relates to the literature on the motivation and incentives of individual contributors (e.g., Zhang and Zhu, 2011) by pointing out that partisan political activity may be a motivation of some Wikipedia authors.

4.3 Data

Kürschner (2015) provides a list of all members of the 18th Bundestag, including information on their education, political experience, and party affiliation. Data on the MPs' offices during the 18th Bundestag stems from `bundestag.de`. Information on the MPs' ancillary incomes is collected from `abgeordnetenwatch.de`: German MPs are obliged to declare their ancillary income by means of ten different categories; following the literature, we use these categories' mean values in our analysis (e.g., Becker et al., 2009).[5] Data on constituency demographics stems from the electoral management body.[6]

From the initial list of MPs, we exclude Chancellor Angela Merkel, 35 MPs in distinguished offices (party heads and ministers from the 18th or a preceding Bundestag), and nine MPs who had already left the 18th Bundestag before we started our data collection.[7] Of 598 remaining MPs in our dataset, 294 MPs are from the CDU/CSU, 184 from the SPD, 62 from the Left, and 58 from the Greens; no MP switched parties during the observation period. Moreover, 266 of the 598 MPs were directly elected to the Bundestag.[8]

German Wikipedia biographies exist for all MPs in our dataset, English biographies are available for 138 MPs. The numbers of characters, words, adjectives, images, and categories per biography are obtained via Wikipedia's API. In addition, each biography links to a background site that provides a list of unique authors and the number of edits. Information on the number of weblinks under party control, translation indicators, MPs' English homepages, and criticism is hand coded. Table 4.1 provides summary statistics of all the variables used in our analysis.

[5]See `www.abgeordneten.watch.de/blog/nebeneinkuenfte2016` (June 2015).

[6]See `www.bundeswahlleiter.de/de/bundestagswahlen/BTW_BUND_13/strukturdaten/` (Dec 2017).

[7]See Appendix C.2 for robustness checks that include these observations.

[8]In Germany, each voter casts two votes in elections to the Bundestag. The first vote decides which local candidate from each of Germany's 299 constituencies will be sent to the Bundestag. The second vote is cast for a party list and determines the parties' relative strength in the Bundestag. The Bundestag has a minimum number of 598 seats. In its 18th election term, it was amplified by four additional "overhang" seats, since the CDU won more constituency seats than it would have been entitled to based on the second-vote share. A further 29 "balance seats" sustain the parties' relative strengths, leading to a total number of 631 seats.

Table 4.1: Summary statistics

Variable	Mean	Std. Dev.	Min.	Max.	N
Biography length German (characters)	5901.577	4890.87	1118	45316	598
Biography length English (characters)	4909.507	4119.552	838	28132	138
CDU/CSU	0.492	0.5	0	1	598
SPD	0.308	0.462	0	1	598
Greens	0.097	0.296	0	1	598
Left	0.104	0.305	0	1	598
Female	0.366	0.482	0	1	598
Former periods in Bundestag	1.577	1.748	0	9	598
Ancillary income (in 1000 Euros)	25.609	106.476	0	1411	596
Directly elected	0.453	0.498	0	1	598
Population density	910.613	1471.26	38	12842.9	271
Fraction population aged 18 to 35	20.114	2.7	15.6	28.9	271
Fraction population with Abitur	36.417	7.473	21.7	59.7	271
Number of international offices	0.209	0.496	0	3	598
Former periods in European Parliament	0.012	0.135	0	2	598
Number of adjectives	65.502	61.972	8	618	598
Number of images	2.475	1.198	0	11	598
Number of categories	9.206	2.446	5	24	598
Number of weblinks under party control	1.487	0.669	0	4	598
Number of unique authors	52.05	59.534	5	580	598
Edited from Bundestag network (German)	0.527	0.5	0	1	598
Number of edits from Bundestag network (German)	4.427	5.399	1	40	309
Characters added from Bundestag network (German)	533.176	919.954	0	7201	306
Characters deleted from Bundestag network (German)	198.373	648.166	0	6744	306
Net character change from Bundestag network (German)	171.321	801.527	-6743	7201	598
Edited from Bundestag network (English)	0.058	0.235	0	1	138
Length (talk pages)	5206.723	13030.45	7	130746	292
Criticizing sentences	0.281	1.041	0	10	598
Number of words	547.518	469.732	76	4326	598
English homepage	0.022	0.146	0	1	598
Translation template	0.017	0.128	0	1	598

Notes: This table presents the summary statistics of all variables used in the analysis. The variables *CDU/CSU, SPD, Greens, Left, Female, Directly elected, Edited from Bundestag network (German), Edited from Bundestag network (English), English homepage*, and *Translation template* are dummy variables. The variables *Population density, Fraction population aged 18 to 35*, and *Fraction population with Abitur* are available only for a subset of MPs who are directly elected. The variables *Biography length English (characters)* and *Edited from Bundestag network (English)* are available only for MPs for whom an English Wikipedia biography exists.

4.4 Empirical strategy and results

Our goal is to answer if there is a coverage bias on Wikipedia, i.e., if comparable MPs from different political parties are covered differently in terms of their biography length. We proceed in two steps. First, we focus on the biographies of a relatively homogeneous group of backbenchers from CDU/CSU and SPD. Any differences in biography length can therefore not originate from differences in government versus opposition parties, centrist versus more extreme parties, or big versus fringe parties. Moreover, the personalities of single prominent MPs cannot affect our results. Second, to disentangle the effect of party affiliation from unobserved MP characteristics, we compare the MPs' German and English biography length in a difference-in-differences framework.

4.4.1 Basic estimation

Figure 4.1 shows the average biography length per party in characters. Since $2,500$ characters roughly correspond to one DIN-A4 page of a biography's PDF print version, biographies of MPs from the CDU/CSU and the Greens are around two and a half pages, biographies of MPs from the SPD are around two pages, and biographies of MPs from the Left nearly three pages long. To put these numbers into perspective, note that the average biography length across all MPs is around 2.33 pages (5901.6 characters) and the median around 1.66 pages (4487.5 characters), with a standard deviation of nearly two pages (4890.87 characters). Hence, the difference in average biography length between the CDU/CSU and the SPD corresponds to roughly 25% of a standard deviation.

To control for observable MP characteristics, we further estimate the regression equation

$$length_i^G = \beta_0^G + \beta_1^G P_i + \beta_2^G X_i + u_i^G, \tag{4.1}$$

by OLS, where $length_i^G$ corresponds to the German biography length of MP i, P_i is a vector of party dummies with the SPD as omitted category, and X_i is a vector of control variables including MP i's gender, political experience, education, ancillary income, and – if MP i is directly elected – constituency demographics.[9] The parameter of interest is β_1^G, as it measures the effect of party affiliation on biography length relative to the omitted category SPD. In other words, β_1^G corresponds to the average coverage bias relative to the SPD.

Table 4.2 shows the results. Column 1 replicates the party averages from Figure 4.2 relative to the omitted category SPD, i.e., the CDU/CSU coefficient corresponds to around half a DIN-A4 page.

[9] Several large cities such as Cologne and Berlin are divided into several electoral districts. The data on those cities are aggregated and therefore not independent. We account for that by clustering the standard errors at the city level, and obtain 249 clusters.

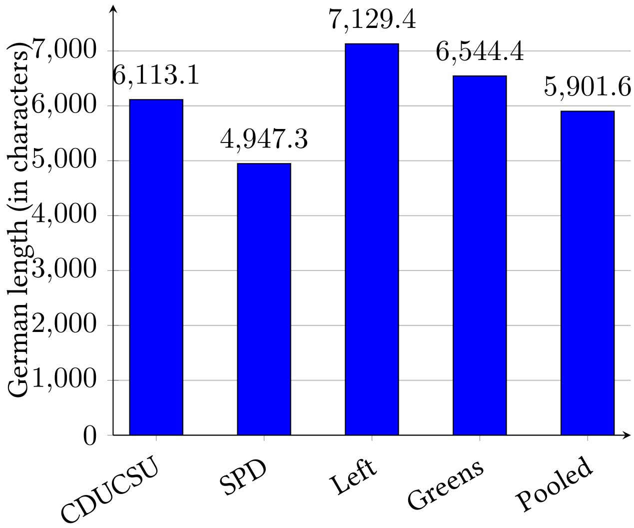

Figure 4.1: German biography length across parties in characters

When we control for gender (column 2), political experience (column 3), and doctoral degrees as well as ancillary incomes (column 4), the size of the CDU/CSU coefficient decreases to a third of a DIN-A4 page (which corresponds to 15% of a standard deviation in the dependent variable), but remains statistically significant. Plausibly, when we consider only the directly elected MPs and account for their constituency demographics in column 5, population density has a highly significant effect on biography length: broadband connections in urban areas are usually better than in rural areas, which facilitates the use of Wikipedia and increases the biographies' likelihood of being read. In contrast to that, education or a constituency's share of voters aged 18 to 35 – those who are most prone to use the Internet as a source of political information – have no statistically significant effect. The CDU/CSU coefficient is statistically significant despite the decreased sample size and corresponds to around 30% of a standard deviation in the dependent variable.[10]

In addition, we perform several two-sided t-tests to check whether differences in coverage between other parties are statistically significant. In columns (1) and (2), no other differences in biography length between any two parties are significant, while in columns (3) and (4), the difference between CDU/CSU and Left is significant at the 5%-level.

[10]Note that we do not consider observations from the Greens and the Left here, since only three Left MPs and one Green MP are directly elected.

Table 4.2: Basic results

	(1) Length	(2) Length	(3) Length	(4) Length	(5) Length
CDU/CSU	1165.8***	1057.0***	971.3***	870.4**	1614.8**
	(374.1)	(385.9)	(366.6)	(353.1)	(738.7)
Left	2182.2***	2297.7***	2388.3***	2396.4***	
	(728.7)	(730.6)	(694.6)	(656.1)	
Greens	1597.2	1692.7*	1736.4*	1733.5*	
	(972.0)	(985.2)	(906.8)	(917.7)	
Female		-689.1	-474.8	-216.0	299.0
		(435.6)	(403.9)	(402.0)	(749.8)
Former periods in BT			869.6***	875.5***	723.1***
			(161.5)	(160.2)	(197.6)
Ancillary Income				2.002	1.681
				(1.702)	(1.621)
PhD				1915.3***	886.2
				(608.9)	(868.8)
Population density					0.896***
					(0.278)
Fraction pop. 18–35					255.2
					(155.5)
Fraction pop. with Abitur					-1.997
					(48.21)
Constant	4947.3***	5231.9***	3810.6***	3327.6***	-2578.2
	(232.7)	(305.5)	(347.0)	(373.3)	(3436.9)
N	598	598	598	596	266
R^2	0.021	0.025	0.121	0.150	0.185

Notes: Robust standard errors in parentheses. The dependent variable $length_i^G$ measures Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i, $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenwatch.de`. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18-35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i's constituency demography. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

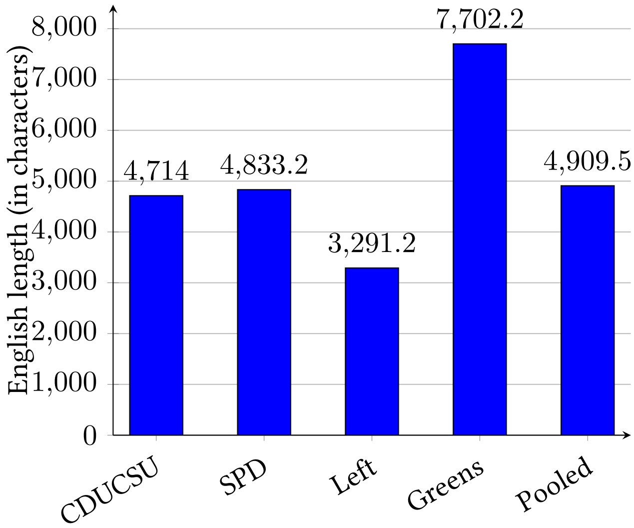

Figure 4.2: English biography length across parties in characters

4.4.2 Difference-in-differences estimation

If party affiliation is correlated to unobservable MP characteristics (e.g., if more salient politicians self-select into a particular party) an OLS estimation of equation (4.1) might suffer from an omitted variable bias. To disentangle the effect of party affiliation on biography length from the effect that unobserved MP characteristics could have, we exploit variation between the MPs' biographies in the German and in the English Wikipedia. Under the identifying assumption that unobserved MP characteristics affect the German and the English biography length equivalently, a difference-in-differences estimation using language as a first, and party affiliation as a second difference, yields unconfounded estimates of the effects of party affiliation on biography length.

English Wikipedia biographies exist for 138 MPs in our sample; Figure 4.2 shows their average length. English biographies of MPs both from the CDU/CSU and the SPD are nearly two pages, English biographies of MPs from the Greens are around three pages, and English biographies of MPs from the Left are around one and a third page long. To put these numbers into perspective, note that the average English biography length across all MPs is around two pages (4909.5 characters) and the median around one and a half pages (3817 characters), with a standard deviation of 1.66 pages (4119.6 characters).

Suppose that the MPs' English biography length is determined analogously to equation (4.1). Then,

$$length_i^E = \beta_0^E + \beta_1^E P_i + \beta_2^E X_i + u_i^E. \tag{4.2}$$

Differencing equations (4.1) and (4.2) yields

$$length_i^G - length_i^E = \beta_0 + \beta_1 P_i + \beta_2 X_i + u_i, \tag{4.3}$$

where $\beta_k = \beta_k^G - \beta_k^E$ and $u_i = u_i^G - u_i^E$. The parameter of interest in equation (4.3) is β_1, whose interpretation hinges on further assumptions on β_1^E. As argued, biography length may serve as a positive signal about MPs' valence characteristics (see Appendix C.1), so partisan contributors have an incentive to amplify the Wikipedia biographies of MPs from one specific party. While partisan contributions may lead to a coverage bias in the German Wikipedia, this is unlikely to occur in the English Wikipedia version, since German voters are unlikely to read their MPs' English biographies. Thus, we assume that $\beta_1^E = 0$; see Appendix C.3 for further discussion. If party affiliation has no effect on the MPs' English biography length, β_1 corresponds to β_1^G in equation (4.1). Hence, under the identifying assumption that unobserved MP characteristics affect the German and the English biography length equivalently and thereby cancel out, equation (4.3) yields an unbiased estimate for β_1^G, the average coverage bias relative to the SPD.

Naively regressing equation (4.3) on the subsample of MPs where we observe English biographies may lead to sample selection bias, though. Thus, we also consider a selection model consisting of equation (4.3) and the selection equation

$$d_i = 1[\delta Z_i + \epsilon_i] > 0, \tag{4.4}$$

where the dependent variable d_i indicates if there exists an English biography for MP i (see also Greene, 2003). For the selection model to work well, Z_i has to include additional variables that determine whether or not there exists an English Wikipedia biography for MP i, but do not affect the dependent variable in equation (4.3). Thus, $Z_i = (P_i, X_i, I_i)$, where I_i is a vector of variables that determine the international political relevance of MP i: the number of election terms in the European Parliament and the number of international offices during the 18th Bundestag (Commission of Foreign Affairs, Commission of European Affairs, and being head of an international parliamentary group). While international offices plausibly increase the probability that an English biography will be set up, they are not very salient, and hardly anything is written on them. We estimate the selection model by Heckman two-step and by maximum likelihood.[11]

Table 4.3 shows the results. Column 1 shows the results of an OLS estimation of equation (4.3). Although the size of the CDU/CSU coefficient is considerable – it corresponds to nearly a DIN-A4 page, which is nearly 60% of a standard deviation in the dependent variable and around twice as

[11] We do not believe that the probability of having an English Wikipedia biography is affected by MP preferences: only five of the 138 MPs with an English Wikipedia biography even provide their personal homepage in English.

much as in Table 4.2 – it is not statistically significant. Columns 2 and 4 show the results from a two-step and a maximum likelihood estimation of equations (4.3) and (4.4), respectively. Since the magnitude of the coefficients is not directly comparable to the OLS estimates in column 1 of Table 4.3, we also display their marginal effects at the mean (MEM) in columns 3 and 5, which are similar to the OLS estimates.[12] In contrast to the OLS regression, the CDU/CSU coefficient is statistically significant at the 10%-level in the two-step, and at the 1%-level in the maximum likelihood specification. Hence, the difference-in-differences estimation provides even stronger evidence of a coverage bias against MPs from the SPD than our results from section 4.4.1.

4.5 Further dimensions of coverage

This section shows that our results on coverage bias from Section 4.4.1 hold for alternative dependent variables in equation (4.1): adjectives, images, categories, and weblinks under party control. First, we study the occurrence of adjectives in the biographies. Adjectives can make a text more lively and colorful. Moreover, the literature on sentiment analysis shows a correlation between the presence of adjectives and the subjectivity of a sentence, i.e., the degree to which opinions are expressed (e.g., Bruce and Wiebe, 1999; Wiebe et al., 2004; Pang and Lee, 2008). We find that biographies on MPs from the CDU/CSU have around ten more adjectives and a higher adjective-to-word ratio than biographies on MPs from the SPD; the former difference corresponds to around 18% of a standard deviation in the dependent variable and is statistically significant at the 5%-level (Table 4.4, columns 1 and 2).

Similarly, images make biographies more lively and attractive. We find that biographies of MPs from the CDU/CSU contain on average 0.33 images (30% of a standard deviation) more than biographies of MPs from the SPD; the difference is statistically significant at the 1%-level (Table 4.4, column 3).

Next, Wikipedia articles are usually assigned to categories, that "help readers to find, and navigate around, a subject area, to see pages sorted by title, and to thus find article relationships".[13] Thus, assigning a biography to many different categories enhances its chances to be found by readers. Biographies on MPs from the CDU/CSU are assigned to 0.65 more categories than biographies on MPs from the SPD; this difference corresponds to 25% of a standard deviation in the dependent variable and is statistically significant at the 1%-level (Table 4.4, column 4).

[12] MEMs measure the effect of a change in one of the regressors on the conditional mean of the difference in biography length, evaluated at the mean values of all other covariates, and given that this difference is observed. MEMs for dummy variables show how the biography length changes as the dummy changes from 0 to 1, holding all other covariates at their mean values.

[13] See `https://en.wikipedia.org/wiki/Help:Category` (Dec 2018.)

Table 4.3: Difference-in-difference results

	(1) OLS	(2) Two-step	(3) MEM Two-step	(4) ML	(5) MEM ML
CDU/CSU	2230.8	3165.4^{*}	2275.5	6010.6^{***}	2963.0
	(1378.3)	(1749.1)		(1720.4)	
Left	6447.0^{***}	7892.4^{***}	6488.2	11842.6^{***}	6843.8
	(1627.8)	(2489.3)		(2365.9)	
Greens	2901.8	4365.4^{*}	3183.5	7941.9^{***}	3385.0
	(2805.8)	(2510.7)		(2432.1)	
Female	-1237.8	-1409.9	-1184.7	-1443.3	-356.8
	(1265.4)	(1284.4)		(1419.8)	
Former periods in BT	670.0^{**}	1030.4^{**}	577.0	1902.7^{***}	187.7
	(332.0)	(472.2)		(370.7)	
PhD	2509.2^{*}	3180.6^{**}	2536.5	3845.3^{**}	849.9
	(1477.9)	(1518.9)		(1599.4)	
Ancillary Income	10.18	9.837	11.32	1.398	5.105
	(9.853)	(8.039)		(7.474)	
Constant	-245.1	-5430.1		-18997.1^{***}	
	(1607.3)	(5347.4)		(2375.3)	
Mill's		2837.6		10558.6^{***}	
Lambda		(2792.9)		(906.1)	
N	136	596	596	596	596
R^2	0.124				

Notes: Robust standard errors in parentheses. The dependent variable in columns (1), (2), and (4) is $length_i^G - length_i^E$ that measures the differences in Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i, $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenwatch.de`. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18-35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i's constituency demography. The MEMs in columns (3) and (5) show the coefficients' marginal effect at the mean, i.e., the change in the difference in biography length, given that it is observed, and holding all other factors at their mean. MEMs for dummy variables show the effect in the dependent variable for a discrete change from 0 to 1. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Finally, many Wikipedia biographies provide weblinks to external websites. A link is "under party control" if it directs to a website that is under the obvious and substantial influence of an MP's party (e.g., MPs' personal or party homepages). Weblinks under party control facilitate the use of Wikipedia as a platform for political campaigns. We find that biographies on MPs from the CDU/CSU contain about half a weblink more than biographies on MPs from the SPD (60% of a standard deviation); the difference is statistically significant at the 1%-level (Table 4.4, column 5).[14]

4.6 Partisan contributions

Next, we demonstrate that differences in partisan contributions to Wikipedia are a likely driver of the unbalanced coverage of MPs from CDU/CSU and SPD. To this end, we study authorship patterns and the biographies' talk pages.

4.6.1 Authorship patterns

This section shows that there are more Wikipedia authors who repetitively contribute to the biographies of MPs from the CDU/CSU than to biographies of MPs from the SPD. Moreover, we show that biographies of MPs from the CDU/CSU are edited more often from the Bundestag building.

First, we check if there exist authors who repetitively amplify the biographies of MPs from one specific party. For each article, Wikipedia displays either the authors' user name or, in case of anonymous contributions, their IP address. We identify all authors who contribute to at least 10% of the biographies of MPs from the CDU/CSU or from the SPD and classify them as "repetitive contributors." Next, we check which of these repetitive contributors amplify the biographies of just one specific party and classify them as "party-specific repetitive contributors." We find that there exist 37 repetitive and three party-specific repetitive contributors for the SPD. Moreover, there exist 42 repetitive and five party-specific repetitive contributors for the CDU/CSU.

Next, we track all contributions of anonymous users whose IP addresses are displayed. Building on a study by Bayerischer Rundfunk (2017), we consider all IP adresses that can be linked to the Bundestag building.[15] We find that 50.3% of the biographies of MPs from the CDU/CSU, and 52.2% of the biographies of MPs from the SPD were edited at least once from the Bundestag. Moreover,

[14] The finding is supported by one particular incident. According to the Wikipedia talk pages, it is difficult to incorporate weblinks into articles, partly because of the German umlauts (ä, ö, ü). There exists, however, a user called "Cducsu" who has written a program to facilitate the procedure and has used it to install weblinks underneath biographies of MPs affiliated to the CDU/CSU that redirect to the homepage of the CDU/CSU parliamentary group.

[15] According to `bundesedit.de` (June 2018), the digits "193.17." at the beginning of an IP address indicate the Bundestag network.

Table 4.4: Further results

	(1) Adjectives	(2) Adj / Words	(3) Images	(4) Categories	(5) Weblinks
CDU/CSU	10.35**	0.00123	0.344***	0.648***	0.408***
	(4.376)	(0.00170)	(0.0987)	(0.195)	(0.0567)
Left	28.08***	-0.00184	0.454***	2.772***	0.643***
	(8.294)	(0.00244)	(0.171)	(0.325)	(0.103)
Greens	19.24*	-0.00563**	0.735***	0.510	0.399***
	(11.55)	(0.00237)	(0.257)	(0.360)	(0.0946)
Female	-4.064	-0.00237	-0.173*	-0.0350	-0.0573
	(4.992)	(0.00148)	(0.101)	(0.184)	(0.0555)
Former periods in BT	11.46***	0.000925**	0.131***	0.469***	0.0620***
	(2.039)	(0.000396)	(0.0422)	(0.0581)	(0.0139)
Ancillary Income	0.0198	-0.000011**	0.00009	0.0002	-0.0003
	(0.0211)	0.000005)	(0.00004)	(0.00123)	(0.0002)
PhD	22.35***	-0.00406**	0.284**	0.433*	0.00588
	(8.269)	(0.00192)	(0.124)	(0.254)	(0.0714)
Constant	34.12***	0.118***	1.983***	7.737***	1.111***
	(4.722)	(0.00175)	(0.109)	(0.188)	(0.0559)
N	596	596	596	596	596
R^2	0.149	0.039	0.087	0.212	0.133

Notes: Robust standard errors in parentheses. The dependent variable in column (1) is $Adjectives_i$, which measures the number of adjectives in the Wikipedia biography of MP i. The dependent variable in column (2) is $Adjectives/Words_i$, which measures the adjective-to-word ratio of MP i. The dependent variable in column (3) is $Images_i$, which measures the number of images in the biography of MP i. The dependent variable in column (4) is $Categories_i$, which measures the number of categories that the biography of MP i is assigned to. The dependent variable in column (5) is $Weblinks_i$, which measures the number of weblinks under party control underneath the biography of MP i. CDU/CSU_i, $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenwatch.de`. PhD_i is equal to 1 if MP i has a PhD. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

biographies of MPs from the CDU/CSU were edited on average 5.2 times, while biographies of MPs from the SPD were edited 3.8 times on average. There exist also differences in terms of the number of characters added or deleted: while on average 627.02 characters are added and 482.20 characters deleted from biographies of MPs from the CDU/CSU by Bundestag contributors, only 232.49 characters are added and 157.65 characters deleted from biographies of MPs from the SPD. Note that the difference in the net number of characters added between the CDU/CSU and the SPD – around thirty – is much smaller than the effect we find in Section 4.4. Hence, edits by anonymous users from the Bundestag alone cannot explain the coverage bias documented above.

4.6.2 Talk pages

Partisan contributions are likely to entail discussions if additional content should be included or not. The purpose of a Wikipedia talk page "is to provide space for editors to discuss changes to its associated article or project page."[16] Hence, existence and length of a talk page can indicate the occurrence of partisan contributions.

We find that talk pages exist for 132 MPs from the CDU/CSU and for 79 MPs from the SPD. Moreover, talk pages on MPs from the SPD are on average half a page shorter than talk pages on MPs from the CDU/CSU (3589.52 versus 5029.30 characters), which corresponds to 10% of a standard deviation in talk page length.

4.7 Negative coverage

Our analysis assumes that Wikipedia coverage is beneficial for MPs. Media coverage is generally beneficial for politicians, e.g., because it increases name recognition (Burden 2002); this applies in particular to less well-known politicians such as the MPs in our dataset. In addition, we have conducted a classroom survey that shows that a longer biography signals knowledge, strength as a public servant, and the ability to inspire people (see Appendix C.1). Wikipedia coverage could hurt MPs, however, if the biographies contained large amounts of criticism. To resolve such concerns, this section shows that negative coverage is a minor concern.

To explore the extent of negative coverage, we systematically identify negative sentences in the biographies. In a first step, we search each biography for sentences that contain the word stems of "Kritik" ("criticism"), "Diskussion" ("discussion"), "Rück- / Austritt" ("resignation"), "Skandal" ("scandal"), and "Affaire" ("affair"). Next, we determine if these sentences actually criticize the MP. We find that negative coverage is a minor issue: only 7% of the biographies contain more than one sentence of negative coverage, and 90% do not contain any negative coverage at all.

[16] See https://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines (Dec 2018).

To confirm that the results from Section 4.4 are not driven by different amounts of negative coverage, we estimate equations (4.1), (4.3) and (4.4) on the subsample of MPs whose biographies do not contain any criticism at all. Table 4.5 shows the results. Although the estimates are smaller and not as statistically significant as in Tables 4.2 and 4.3, they are qualitatively similar.[17]

4.8 Conclusion

This paper demonstrates that the individual contributions to the privately provided public good Wikipedia lead to a coverage bias against the SPD. To disentangle the effect of party affiliation on coverage from the effect that unobserved MP characteristics may have, we focus on a sample of relatively homogeneous backbenchers from the 18th German Bundestag. Moreover, we compare the length of German and English Wikipedia biographies in a difference-in-differences framework. We also present empirical evidence that supports partisan contributions as a likely driver of our results.

Our analysis is limited in at least two ways. First, to disentangle the effect of party affiliation on coverage from the effect that MP characteristics may have, we focus on MPs from just one Bundestag. The external validity of our study is, however, supported by similar patterns of Wikipedia coverage of members of the 16 German State Parliaments and of the European Parliament.[18] Biographies of CDU/CSU affiliates in a State Parliament are on average about a quarter page (or about 0.15 standard deviations) longer than biographies of SPD affiliates. Similarly, biographies of CDU/CSU affiliates in the European Parliament are on average about half a page (or about 0.25 standard deviations) longer than biographies of SPD affiliates. These numbers suggest that our main results reflect a general pattern, rather than being specific to the sample at hand. Interestingly, when we compare the Wikipedia coverage of judges in the German Constitutional Court – who are usually nominated by a particular party – we do not find differences in coverage. These judges are, however, elected for a lifetime and not by the public, and thus have no incentive to amplify their Wikipedia biographies. These observations therefore also fit our hypothesis about partisan contributions as the main driver of our results.

Second, our analysis cannot explain why there is a coverage bias against MPs from the SPD.[19] On the one hand, the SPD has fewer potential voters than the CDU/CSU. Hence, there may

[17] We also perceive pure vandalism as a minor issue: false statements are quickly detected by Wikipedia's control mechanisms and are thereupon erased. Moreover, if MPs are involved in scandals such as plagiarism or the consumption of illegal drugs, they usually resign, and these observations are excluded from our analysis.

[18] The data was collected on March 29, 2017.

[19] We contacted the parties' press offices to inquire whether there are coordinated party activities in Wikipedia. According to all replies, there exist no official guidelines for the handling of Wikipedia; every MP is responsible herself for her Wikipedia biography.

Table 4.5: Negative coverage

	(1) Length	(2) Length	(3) Length	(4) Length	(5) Length	(6) ML
CDU/CSU	510.0**	451.1*	439.5*	454.7*	704.2	2927.9***
	(257.5)	(255.0)	(247.3)	(238.3)	(493.0)	(1103.2)
Left	575.3	634.3*	728.9**	814.6**		3973.5**
	(356.9)	(362.6)	(349.0)	(338.5)		(1807.3)
Greens	131.9	185.5	285.4	323.8		1424.4
	(312.0)	(309.2)	(300.6)	(295.7)		(1718.0)
Female		-396.5*	-312.1	-179.7	-132.9	-1091.6
		(226.5)	(219.1)	(214.4)	(336.7)	(995.8)
Former periods in BT			370.2***	387.1***	354.1***	531.5*
			(79.43)	(78.72)	(118.7)	(276.0)
Ancillary Income				1.662	1.656	5.578
				(1.117)	(1.201)	(6.731)
PhD				1006.7***	446.8	2016.5*
				(344.0)	(564.5)	(1117.0)
Population density					0.719***	
					(0.200)	
Fraction pop. 18-35					39.03	
					(117.9)	
Abitur					-35.23	
					(30.28)	
Constant	4539.5***	4703.6***	4121.9***	3810.0***	3781.0*	-8857.4***
	(175.4)	(196.1)	(198.6)	(188.9)	(1951.8)	(1993.7)
Mills						1.662***
Lambda						(0.275)
N	533	533	533	532	236	532
R^2	0.009	0.014	0.069	0.095	0.132	

Notes: Robust standard errors in parentheses. The results are based on a subsample of observations whose biographies do not contain any criticizing sentences. The dependent variable in columns (1) to (5) is $length_i^G$ which measures Wikipedia coverage of MP i in terms of her biography length in characters. The dependent variable in column (6) is $length_i^G - length_i^E$, which measures the differences in Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i, $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenwatch.de`. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18-35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i's constituency demography. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

be relatively less demand for Wikipedia biographies of MPs from the SPD, such that partisan contributions have a comparatively smaller payoff. Although voters of CDU/CSU and SPD are equally Internet affine (Forschungsgruppe Wahlen, 2014), potential voters of the SPD may also perceive the new media as a less relevant information source. On the other hand, it is possible that the differences in partisan activity reflect the parties' perceptions of how important an extensive Internet presence is. For instance, Peter Tauber, who was secretary general during our observation period, provides a social media compendium that also points to the importance of Wikipedia (Tauber, 2013, p.12), while nothing comparable exists for the SPD. A comprehensive analysis of these channels promises insightful future research.

5 Selective Sharing of News Items and the Political Position of News Outlets

with Julian Freitag and Johannes Münster

5.1 Introduction

State-of-the-art research shows that the media have a causal effect on the economic and political choices of individuals (DellaVigna and La Ferrara, 2015). The media are, however, repeatedly accused of being biased towards the political left or right. For instance, six-in-ten US citizens see political bias in the news media[1] and four-in-ten German voters think that the government exerts pressure on the media.[2] Are the media really biased? A growing body of literature addresses these questions by developing methods to assess political biases of news outlets (Groeling, 2013; Puglisi and Snyder, 2016).

Measuring the political position of news outlets is challenging, though. In particular, researchers must find ways to overcome problems of subjectivity and the absence of suitable baselines against which to assess bias (e.g., Groeling, 2013). Existing approaches are based on in-depth content analyses – either by human or by automated coding – or on determining the political position of the news outlets' audience (Puglisi and Snyder, 2016). Many of these procedures are data demanding, computationally burdensome, and time consuming. Easy to implement methods to assess the political position of news outlets, in contrast, are rare.

We present a novel approach to measure the political position of online news outlets that is based on the selective sharing of news items by politicians on social media. Our central argument is that politicians predominantly share news items that are in line with their own political position, i.e., left-wing politicians prefer to share news items from left-wing news outlets, while right-wing

[1] See `https://news.gallup.com/poll/207794/six-partisan-bias-news-media.aspx`, viewed: Feb 2019.

[2] See `https://docplayer.org/43364962-Glaubwuerdigkeit-der-medien-eine-studie-im-auftrag-des-westdeutschen-rundfunks-dezember-2016.html`, viewed: Feb 2019.

politicians prefer to share news items from right-wing news outlets.[3] Consequently, we can utilize the politicians' revealed preferences over news items to infer the political position of the news outlets.

Formally, we compute a Spearman rank correlation coefficient for each news outlet under consideration. The Spearman rank correlation coefficient measures the correlation between the rank of the political position of the politicians' parties (from most left-wing to most right-wing on a one-dimensional scale) on the one hand, and the rank of the politicians' share of referrals to a particular news outlet – aggregated on the party level – on the other hand. A positive correlation indicates that the news outlet is positioned to the right, a negative correlation indicates that the news outlet is positioned to the left.

Our approach has a number of advantages. First, it is quick and easy to implement. We infer the political position of a news outlet from the selective sharing of news items by politicians, whose political position is clear. Moreover, since sharing news items via social media channels is nowadays part of the politicians' profession, we observe the politicians' choices over news items in a setting where they have an incentive to reveal their preferences consciously and truthfully. Thus, our approach does not require any elaborate content analysis, but circumvents problems of subjectivity and the absence of suitable baselines against which to assess bias nonetheless.

Second, the results from our procedure are straightforward to interpret. The Spearman rank correlation coefficient for a particular news outlet is either positive, negative, or equal to zero, whereby the news outlet can directly be classified as biased to the left, biased to the right, or unbiased.

Finally, our approach is applicable widely beyond this paper. In particular, while many existing procedures are limited to assessing political media bias in two-party democracies, our approach can also be applied to multi-party democracies, as long as the parties' political position can be measured on an ordinal, one-dimensional scale.[4] In addition to that, our approach is not data demanding and can thus be applied to small datasets, too.

We apply our procedure to twelve major online news outlets in Germany and consider the selective sharing of news items of German MPs via Twitter. The Spearman rank correlation coefficient is positive for five news outlets, but only statistically significant for two of them (*BILD* and *Welt*). The Spearman rank correlation coefficient is negative for seven further news outlets, and statistically significant for three of them (*Zeit*, *Spiegel*, and *Deutschlandfunk*). Following the above considerations, we conclude that *BILD* and *Welt* are positioned on the right, *Zeit*, *Spiegel*, and *Deutschlandfunk* are positioned on the left, and the remaining seven news outlets are positioned

[3]A robustness check to the application of this measure confirms that less than four percent of the politicians' referrals criticize the news item or the news source.

[4]Machine learning techniques, for instance, usually struggle when multiple parties are involved (Colleoni et al., 2014).

in the center of the political spectrum. Several robustness checks support our main results.

The remainder of the paper is organized as follows. Section 5.2 discusses the related literature. Section 5.3 illustrates the application of the Spearman rank correlation coefficient as a measure of the political position of news outlets more closely. In Section 5.4, we describe the data collection procedure and the data preparation process. Section 5.5 presents the results of our application, compares them to existing measures of the political position of German news outlets, and demonstrates their robustness. Section 5.6 concludes.

5.2 Related literature

Our paper is related to three strands of literature. First, our approach to measure the political position of news outlets contributes to the literature on political media bias (see Groeling, 2013; Gentzkow et al., 2016; Puglisi and Snyder, 2016, for surveys). It is especially close to papers that develop alternative approaches to measure political bias in the US (e.g., Groseclose and Milyo, 2005; Ho and Quinn, 2008; Gentzkow and Shapiro, 2010) and in Germany (e.g., Dallmann et al., 2015; Dewenter et al., 2016; Garz et al., 2019). We contribute to this literature by presenting a novel approach to determine the political position of news outlets that is easy to implement, straightforward to interpret, and applicable to multi-party democracies and small datasets.

Next, our paper is related to the growing literature on the selective sharing of information on social media. This literature is divided into two fields. One group of papers infers the political position of users *from* their selective sharing of information whose political position is clear (e.g., Barberá et al., 2015; Boutet et al., 2012; Colleoni et al., 2014). A second group of papers takes the reverse approach and provides *evidence of* the selective sharing of information by users whose political position is clear. Adamic and Glance (2005), for instance, demonstrate that political bloggers prefer to share hyperlinks that match their own political opinion. In addition, Shin and Thorson (2017) and Aruguete and Calvo (2018) show that Twitter users selectively share messages that are in line with their political position; An et al. (2014) provide analogous evidence for selective sharing on Facebook. Our approach builds on the findings from the latter group of papers, since our measure is based on politicians' selective sharing of news items that are in line with their own political position.

Finally, we contribute to studies that examine the role of social media in political processes (Luca, 2016b, provides a survey on the economic and political impact of social media and user-generated content). The selective exposure to social media content has attracted particularly much attention (e.g., Bakshy et al., 2015; Knobloch-Westerwick and Meng, 2009, 2011; Garrett, 2009a,b). Selective exposure to social media content is conceptually closely related to selective sharing of information;

the two are often called "complementary processes" (Shin and Thorson, 2017). Most closely related to our approach is An et al. (2012) who create a one-dimensional map of the political position of US news media based on Twitter users' subscription and interaction patterns.[5] On top of that, our paper adds to studies on politicians' usage of Twitter and other social media (Jungherr, 2016), coverage of politicians on Twitter (Jungherr, 2014), and the contribution of social media in political mobilization (Bond et al., 2012) and social movements (Hermida et al., 2014).

5.3 Method

As argued, the idea of the approach is to assess the correlation between the political position of a party and its politicians' number of referrals to a specific news outlet. To avoid over-fitting, we use the *Spearman rank correlation coefficient*, a well known non-parametric measure of the correlation between two ordinal variables.[6]

Regarding our application to the selective sharing of news items by German MPs on Twitter, let o, $o = 1, \ldots, 12$, denote the twelve news outlets under consideration (see Section 5.4 for details on the selection process). Moreover, let i, $i = 1, \ldots, 7$, denote the seven parties in the 19th Bundestag. The political position of party i is denoted by $x_i \in \mathbb{R}$.[7] Let n_{io} denote the absolute number of tweets from MPs of party i that contain a reference to outlet o. These raw counts will depend on the number of MPs belonging to party i and on how active they are on Twitter, two factors that are not informative about the political position of outlet o. Therefore, our main measure considers the *relative* number of Twitter referrals by party i to outlet o,

$$y_{io} = \frac{n_{io}}{\sum_{r=1}^{12} n_{ir}}. \tag{5.1}$$

We observe seven different values y_i – one for each party i – for each news outlet o.[8]

Next, the parties' political positions, x_i, are assigned to integer ranks $rg(x_i)$, where rank 1 is given to the most left-wing, and rank 7 is given to the most right-wing party. Moreover, fix a news outlet o and consider the relative numbers of Twitter referrals y_{io} by parties $i = 1, \ldots, 7$ to this news outlet o. Assign integer ranks $rg(y_{io})$ from rank 1 to rank 7, where the smallest

[5]The intuition is that the closer the political position of two media sources, the more their audiences overlap.

[6]See Siegel and Castellan (1988), for a detailed discussion on the Spearman rank correlation coefficient.

[7]By aggregating tweets on the party level, we abstract from political heterogeneity within parties. This simplification makes our analysis more transparent and less data demanding. Moreover, we cannot use the MPs' voting history to determine their individual political position as, e.g., the ADA in the US (see `https://adaction.org/ada-voting-records/`), because such data are not available for the AfD, who was never part of the Bundestag before.

[8]See Section 5.5.3 for a discussion of the advantages and disadvantages of using the relative number of Twitter referrals.

referral share to news outlet o is given the smallest rank. For outlet o, let ρ_o denote the correlation coefficient between $rg(x_i)$ and $rg(y_{io})$. It is then given by

$$\rho_o = \frac{\sum_{i=1}^{7}(rg(x_i) - \overline{rg(x)})(rg(y_{io}) - \overline{rg(y_o)})}{\sqrt{\sum_{i=1}^{7}(rg(x_i) - \overline{rg(x)})^2}\sqrt{\sum_{i=1}^{7}(rg(y_{io}) - \overline{rg(y_o)})^2}}, \tag{5.2}$$

where $\overline{rg(x)}$ and $\overline{rg(y_o)}$ denote the average ranks of x and y for outlet o. In other words, equation (5.2) gives the Spearman rank correlation coefficient between the political position of a party and the relative number of Twitter referrals from this party mentioning outlet o.

The values of ρ_o lie in the interval $[-1, 1]$. If $\rho_o > 0$, the parties' ranked political position and their respective ranked relative number of Twitter referrals to outlet o are *positively* correlated. Thus, news items from o are shared relatively more often by right-wing MPs, which indicates that news outlet o is biased to the right. If, on the other hand, $\rho_o < 0$, the parties' ranked political position and their respective ranked relative number of Twitter referrals to outlet o are *negatively* correlated. Thus, news items from o are shared relatively more often by left-wing MPs, which indicates that outlet o is biased to the left. Finally, if $\rho_o = 0$, the parties' ranked relative number of Twitter referrals to outlet o is independent from their political position; in this case, news outlet o is unbiased.

We test the statistical significance of ρ_o against the null hypothesis that $rg(x_i)$ and $rg(y_{io})$ are independent, i.e., we test

H_0: *There is no correlation between the parties' ranked political position and their ranked referral shares to outlet o.*

against

H_1: *There is a correlation between the parties' ranked political position and their ranked referral shares to outlet o.*

Given the small number of observations per news outlet o ($N = 7$), we consider the exact p-values of the Spearman rank correlation coefficients ρ_o, which we take from Owen (1962). Since we test H_0 for twelve news outlets, we also take multiple hypotheses testing into account with the Bonferroni correction.

5.4 Data

To carry out the analysis, we first determine which news outlets to consider. Our approach is based on the assumption that the selective sharing of news items reveals politicians' preferences over the news outlets' content. Hence, a major requirement on the news outlets is that all German MPs can potentially select from all outlets' news items. Local and specialized news outlets (i.e., those that focus on a particular topic such as sports, fashion or economics) are thus excluded from the analysis.[9] Moreover, we do not consider news aggregators such as Google news or mixed content providers such as e-mail providers. We retrieve the ten largest national online outlets (by number of visits) from `ivw.de`.[10] Nine out of these ten news outlets meet the requirements discussed above.[11] In addition, we include the online news sites of the two major German public TV broadcasters and the major German public radio news broadcaster into the analysis, such that we end up with twelve national online news outlets.[12]

Next, we collect tweets from all MPs of the seven parties in the 19^{th} Bundestag (2017–) who are active on Twitter via the Twitter API. In a first step, we retrieve every tweet by every MP between Oct 24, 2017 (first session of the newly elected Bundestag), and May 11, 2018. Next, we check which tweets share news items published by one of the twelve selected online news outlets and aggregate these tweets on the party level (Table 5.1).[13] The corresponding relative number of Twitter referrals to each outlet for each party is shown in Table 5.2.

Table 5.1: Absolute number of Twitter referrals by party

	BILD	Spiegel	Focus	n-tv	Welt	Zeit	SZ	Stern	F.A.Z.	ARD	ZDF	D.funk	Total
LINKE	16	551	67	79	200	275	264	65	113	314	92	125	2,161
Grüne	99	896	63	64	430	425	619	32	382	349	94	217	3,670
SPD	73	749	58	51	324	290	347	48	237	188	78	133	2,576
FDP	123	293	102	85	542	128	117	23	375	88	37	66	1,979
CDU	280	283	130	89	707	145	193	28	484	256	69	140	2,804
CSU	17	22	4	3	54	6	23	0	25	9	5	7	175
AfD	674	564	912	286	2,175	255	202	86	572	363	81	95	6,265
Total	1,282	3,358	1,336	657	4,432	1,524	1,765	282	2,188	1,567	456	783	19,630

Notes: Table 5.1 shows the absolute number of Twitter referrals by party to each news outlet under consideration. Newspapers and magazines: *BILD* refers to news items from `bild.de`. *Spiegel* refers to news items from `spiegel.de`. *Focus* refers to news items from `focus.de`. *Welt* refers to news items from `welt.de`. *Zeit* refers to news items from `zeit.de`. *SZ* refers to news items from `sueddeutsche.de`. *Stern* refers to news items from `stern.de`. *F.A.Z.* refers to news items from `faz.net`. Public service broadcasters (television): *ARD* refers to news items from `tagesschau.de`. *ZDF* refers to news items from `zdf.de/nachrichten`. Public service broadcasters (radio): *D.funk* refers to news items from `deutschlandfunk.de`. Other online news outlets: *n-tv* refers to news items from `n-tv.de`.

[9] See Section 5.6 for a discussion on how to apply the approach to local news outlets.

[10] The IVW ("Information Community for the Assessment of the Circulation of Media") certifies and audits the circulations of major publications, including newspapers and magazines, within Germany.

[11] We excluded *upday* from the analysis, which is a news aggregator pre-installed on all Samsung mobile devices.

[12] The top ten news outlets by number of visits include all major German national news outlets. Technically, our analysis could be extended to more online news outlets. The smaller the news outlet, however, the less likely it is to meet the requirements.

[13] This includes reactions and comments on re-tweets that originally shared news items. Some illustrative examples are displayed in Appendix D.

Table 5.2: Relative number of Twitter referrals by party

	BILD	Spiegel	Focus	n-tv	Welt	Zeit	SZ	Stern	F.A.Z.	ARD	ZDF	D.funk	Total
LINKE	.0074	.2550	.0310	.0366	.0925	.1273	.1222	.0301	.0523	.1453	.0426	.0578	1
Grüne	.0270	.2441	.0172	.0174	.1172	.1158	.1687	.0087	.1041	.0951	.0256	.0591	1
SPD	.0283	.2908	.0225	.0198	.1258	.1126	.1347	.0186	.0920	.0730	.0303	.0516	1
FDP	.0622	.1481	.0515	.0430	.2739	.0647	.0591	.0116	.1895	.0445	.0187	.0334	1
CDU	.0999	.1009	.0464	.0317	.2521	.0517	.0688	.01	.1726	.0913	.0246	.0499	1
CSU	.0971	.1257	.0229	.0171	.3086	.0343	.1314	.0	.1429	.0514	.0286	.04	1
AfD	.1076	.0900	.1456	.0457	.3472	.0407	.0322	.0137	.0913	.0579	.0129	.0152	1

Notes: Table 5.2 shows the relative number of Twitter referrals by party to each news outlet under consideration. The relative numbers are computed based on the absolute numbers in Table 5.1. Newspapers and magazines: *BILD* refers to news items from `bild.de`. *Spiegel* refers to news items from `spiegel.de`. *Focus* refers to news items from `focus.de`. *Welt* refers to news items from `welt.de`. *Zeit* refers to news items from `zeit.de`. *SZ* refers to news items from `sueddeutsche.de`. *Stern* refers to news items from `stern.de`. *F.A.Z.* refers to news items from `faz.net`. Public service broadcasters (television): *ARD* refers to news items from `tagesschau.de`. *ZDF* refers to news items from `zdf.de/nachrichten`. Public service broadcasters (radio): *D.funk* refers to news items from `deutschlandfunk.de`. Other online news outlets: *n-tv* refers to news items from `n-tv.de`.

Finally, for each outlet o, we assign the ranks 1 to 7 to the seven parties referral shares to o, where rank 1 is given to the smallest, and rank 7 to the largest referral share. For the ranking of the political parties we rely on Forschungsgruppe Wahlen (2017) who order the parties from left to right in the political spectrum. An overview of all ranks is given in Table 5.3.

Table 5.3: Overview of the ranks

	Party	BILD	Spiegel	Focus	n-tv	Welt	Zeit	SZ	Stern	F.A.Z.	ARD	ZDF	D.funk
LINKE	1	1	6	4	5	1	7	4	7	1	7	7	6
Grüne	2	2	5	1	2	2	6	7	2	4	6	4	7
SPD	3	3	7	2	3	3	5	6	6	3	4	6	5
FDP	4	4	4	6	6	5	4	2	4	7	1	2	2
CDU	5	6	2	5	4	4	3	3	3	6	5	3	4
CSU	6	5	3	3	1	6	1	5	1	5	2	5	3
AfD	7	7	1	7	7	7	2	1	5	2	3	1	1

Notes: Table 5.3 shows the ranks (i) for the parties' political position from most left-wing to most right-wing and (ii) for the parties' relative number of Twitter referrals to the twelve news outlets. The ranks of the referral shares are computed based on Table 5.2. Newspapers and magazines: *BILD* refers to news items from `bild.de`. *Spiegel* refers to news items from `spiegel.de`. *Focus* refers to news items from `focus.de`. *Welt* refers to news items from `welt.de`. *Zeit* refers to news items from `zeit.de`. *SZ* refers to news items from `sueddeutsche.de`. *Stern* refers to news items from `stern.de`. *F.A.Z.* refers to news items from `faz.net`. Public service broadcasters (television): *ARD* refers to news items from `tagesschau.de`. *ZDF* refers to news items from `zdf.de/nachrichten`. Public service broadcasters (radio): *D.funk* refers to news items from `deutschlandfunk.de`. Other online news outlets: *n-tv* refers to news items from `n-tv.de`.

5.5 Results

5.5.1 Main results

Table 5.4 and Figure 5.1 show the results from computing the Spearman rank correlation coefficient ρ_o for all twelve news outlets. We find that ρ_o is positive for five news outlets, but only statistically significant for two of them (*BILD* and *Welt*). Moreover, we find that ρ_o is negative for seven further news outlets, and statistically significant for three of them (*Zeit*, *Spiegel*, and *Deutschlandfunk*). Hence, following our considerations from Sections 5.1 and 5.3, we conclude that *BILD* and *Welt* are positioned on the right, *Zeit*, *Spiegel*, and *Deutschlandfunk* are positioned on the left, and the remaining seven news outlets are positioned in the center of the political spectrum.[14]

Table 5.4: Main results

	BILD	Spiegel	Focus	n-tv	Welt	Zeit	SZ	Stern	F.A.Z.	ARD	ZDF	D.funk
ρ_o	0.964***	-0.857**	0.571	0.179	0.964***	-0.964***	-0.571	-0.393	0.286	-0.679	-0.679	-0.857**
p-value	(0.0028)	(0.024)	(0.200)	(0.714)	(0.0028)	(0.0028)	(0.200)	(0.396)	(0.556)	(0.110)	(0.110)	(0.024)

Notes: Table 5.4 shows the *Spearman rank correlation coefficient* computed for each news outlet under consideration based on the ranks given in Table 5.3. Newspapers and magazines: *BILD* refers to news items from `bild.de`. *Spiegel* refers to news items from `spiegel.de`. *Focus* refers to news items from `focus.de`. *Welt* refers to news items from `welt.de`. *Zeit* refers to news items from `zeit.de`. *SZ* refers to news items from `sueddeutsche.de`. *Stern* refers to news items from `stern.de`. *F.A.Z.* refers to news items from `faz.net`. Public service broadcasters (television): *ARD* refers to news items from `tagesschau.de`. *ZDF* refers to news items from `zdf.de/nachrichten`. Public service broadcasters (radio): *D.funk* refers to news items from `deutschlandfunk.de`. Other online news outlets: *n-tv* refers to news items from `n-tv.de`. Exact p-values in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

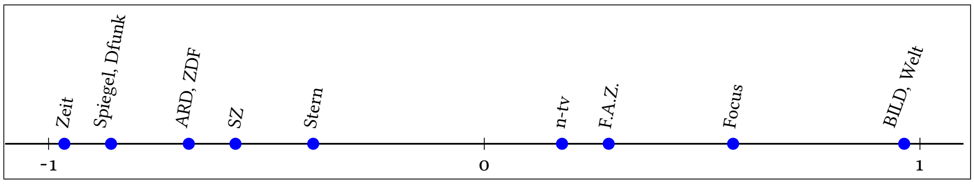

Figure 5.1: Main results: The Spearman rank correlation coefficient for each news outlet under consideration.

5.5.2 Comparison to existing measures

The political position of German news outlets has been measured before; recent approaches include Dallmann et al. (2015), Dewenter et al. (2016), and Garz et al. (2019). In this section, we compare the results from our novel approach with the findings from these papers.

Using automated text analysis, Dallmann et al. (2015) develop several distinct measures for the political media bias in four online news outlets, including three of whom we consider, too. The authors find that *F.A.Z.* tends to favor the more right-wing parties CDU, CSU, and FDP, while their results for *Spiegel* and *Zeit* are ambiguous. This matches or result that *F.A.Z.* is more right-wing than *Spiegel* and *Zeit*, and that *Spiegel* and *Zeit* are similar in their political position.

Dewenter et al. (2016) introduce a political coverage index (PCI) that is based on human coding of the tonality of media reports about Germany's two major parties, the center-right CDU/CSU and

[14]Under the Bonferroni correction, ρ_o is statistically significant (5% level) for three news outlets: *BILD*, *Welt*, and *Zeit*.

the center-left SPD.[15] As in our case, values of the PCI lie in the interval $[-1, 1]$, where negative values of the PCI indicate a bias to the left and positive values indicate a bias to the right. The analysis by Dewenter et al. (2016) includes nine news outlets that we cover, too.[16] Figure 5.2 shows that the values of the PCI are strongly correlated to our Spearman rank correlation coefficient (correlation of 0.77). Moreover, the measures agree on the direction of biases in eight out of nine cases.

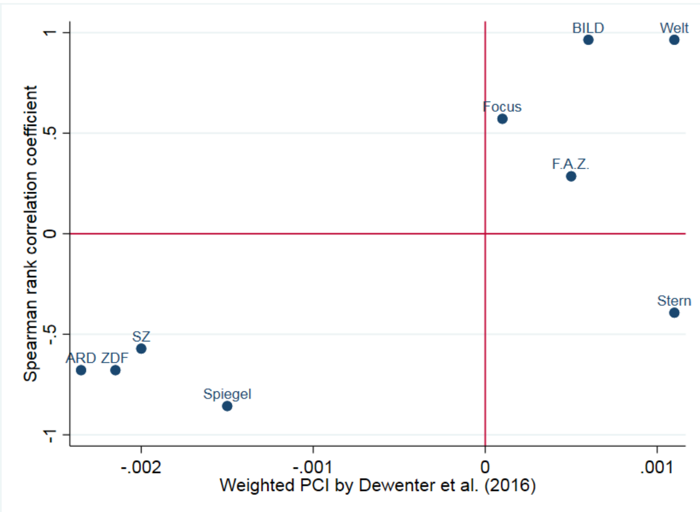

Figure 5.2: Comparison of the Spearman rank order coefficient to the weighted PCI by Dewenter et al. (2016).

Garz et al. (2019) construct an index of media outlets' political position that is based on comparing the language of the outlet with the language of the Facebook pages of Germany's main political parties. Here, too, the index lies in the interval $[-1, 1]$, where negative values indicate a bias to the left and positive values indicate a bias to the right. The analysis by Garz et al. (2019) includes eleven news outlets that we cover, too. Figure 5.3 shows that the values of their index are strongly correlated to our Spearman rank correlation coefficient (correlation of 0.74). Moreover, the measures agree on the direction of biases in nine out of eleven cases. In sum, these comparisons support the validity of our approach.

[15] The authors use tonality data from MediaTenor.

[16] Dewenter et al. (2016) analyzed two different news sources by ARD and ZDF, respectively. We used the mean values of the PCI for these news outlets to conduct the comparison.

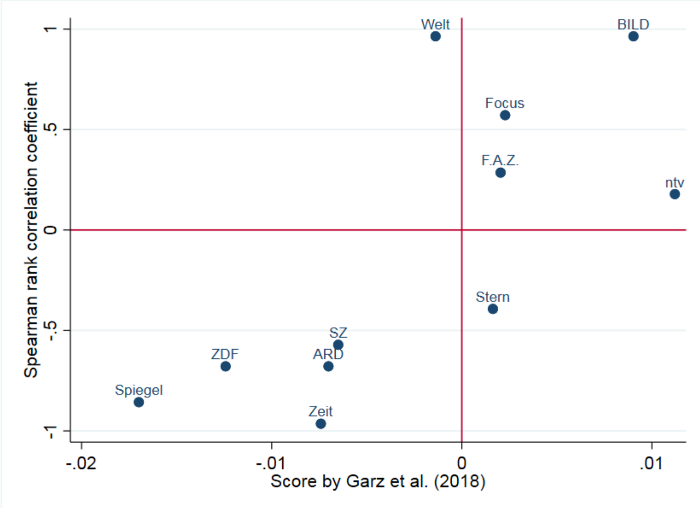

Figure 5.3: Comparison of the Spearman rank order coefficient to the score by Garz et al. (2019).

5.5.3 Robustness checks

Tonality check

In this section, we probe the robustness of our results. First, our analysis is based on the assumption that politicians share only news items that are in line with their own political position. This assumption would be violated if, for instance, politicians shared news items in order to criticize the item itself or the respective news source, or if they disagree with a re-tweet that originally shared the news item. To support the plausibility of our assumption, we let two Research Assistants read $2,998$ randomly drawn tweets from our dataset.[17] The Research Assistants were asked to determine if a tweet criticizes the shared news item or its outlet, if it criticizes the content of a re-tweet that shared a news item, if it criticizes the news item or its news outlet in a re-tweet, or if it does not contain any of these. Appendix D displays some illustrative examples of tweets that the Research Assistants classified as criticizing or non-criticizing. In sum, 113 tweets – i.e., 3.8% – were classified as criticizing the news outlet (inter coder reliability of 99%). This small fraction supports the plausibility of our basic assumption that the MPs share news items via Twitter that are in line with their own political position.

As a further robustness check, we excluded these 113 criticizing tweets from the randomly

[17] We initially decided that the Research Assistants could code $3,000$ tweets within a reasonable amount of time. The random tweets were drawn proportionally to the total amount of tweets. E.g., if the Twitter referrals of party i to news outlet o constituted 1% of all tweets, we would randomly draw $1\% * 3,000 = 30$ tweets by party i to news outlet o for the Research Assistants to check. Rounding of non-integer numbers of tweets resulted in $2,998$ instead of $3,000$ tweets.

drawn subsample of 2,998 tweets and computed the Spearman rank correlation coefficient on the basis of the remaining 2,885 tweets. Since the random subsample was drawn proportionally to the entire sample, the referral shares y_{io} – and thereby the Spearman rank correlation coefficient ρ_o – can only be affected if these criticizing tweets are unevenly distributed across parties *and* outlets; otherwise, our results would remain unchanged.[18] The first row of Table 5.5 shows that although the magnitude of the Spearman rank correlation coefficient underlies small changes compared to the results shown in Table 5.4, our main results are robust to taking out the criticizing tweets. In addition, ρ_o is weakly statistically significant (10% level) for two further outlets: *Focus* and *ARD*, where the former is positioned on the right, and the latter is positioned on the left.[19]

Table 5.5: Robustness Checks

	BILD	Spiegel	Focus	n-tv	Welt	Zeit	SZ	Stern	F.A.Z.	ARD	ZDF	D.funk	taz
No criticizing tweets	0.929^{***}	-0.857^{**}	0.750^{*}	0.179	0.964^{***}	-1.000^{***}	-0.464	-0.500	0.429	-0.750^{*}	-0.464	-0.857^{**}	
p-value	(0.0068)	(0.024)	(0.066)	(0.714)	(0.0028)	(0.0004)	(0.302)	(0.266)	(0.354)	(0.066)	(0.302)	(0.024)	
Exclude LINKE	0.943^{**}	-0.886^{**}	0.771	0.371	0.943^{**}	-0.943^{**}	-0.771	-0.029	-0.143	-0.486	-0.486	-0.829^{*}	
p-value	(0.0166)	(0.034)	(0.102)	(0.498)	(0.0166)	(0.0166)	(0.102)	(1.000)	(0.802)	(0.356)	(0.356)	(0.058)	
Exclude AfD	0.943^{**}	-0.771	0.314	-0.314	0.943^{**}	-1.000^{***}	-0.314	-0.657	0.714	-0.714	-0.486	-0.771	
p-value	(0.0166)	(0.102)	(0.564)	(0.564)	(0.0166)	(0.0028)	(0.564)	(0.176)	(0.136)	(0.136)	(0.356)	(0.102)	
Exclude LINKE and AfD	0.900^{*}	-0.800	0.600	-0.100	0.900^{*}	-1.000^{**}	-0.600	-0.400	0.500	-0.500	-0.100	-0.700	
p-value	(0.084)	(0.134)	(0.350)	(0.950)	(0.084)	(0.0166)	(0.350)	(0.516)	(0.450)	(0.450)	(0.950)	(0.234)	
Include *taz*	0.964^{***}	-0.857^{**}	0.571	0.321	0.964^{***}	-0.929^{***}	-0.464	-0.393	0.429	-0.571	-0.571	-0.857^{**}	-0.929^{***}
p-value	(0.0028)	(0.024)	(0.200)	(0.498)	(0.0028)	(0.0068)	(0.302)	(0.355)	(0.556)	(0.200)	(0.200)	(0.024)	(0.0068)
Absolute no. of tweets	0.607	-0.464	0.321	0.321	0.429	-0.643	-0.643	-0.179	0.357	-0.107	-0.571	-0.500	
p-value	(0.166)	(0.302)	(0.498)	(0.498)	(0.354)	(0.138)	(0.138)	(0.714)	(0.444)	(0.840)	(0.200)	(0.266)	

Notes: The first row of Table 5.5 shows the Spearman rank correlation coefficient computed based on a randomly drawn subsample of tweets, excluding all tweets that were classified as criticizing (Section 5.5.3). The second, third, and forth rows show the Spearman rank correlation coefficient computed when excluding the Tweets by LINKE, AfD, and both at the same time, respectively (Section 5.5.3). The fifth row shows the Spearman rank correlation coefficients when *taz* is considered (Section 5.5.3). The sixth row shows the Spearman rank correlation coefficient based on the *absolute* number of Twitter referrals as given in Table 5.1 (Section 5.5.3). Newspapers and magazines: *BILD* refers to news items from `bild.de`. *Spiegel* refers to news items from `spiegel.de`. *Focus* refers to news items from `focus.de`. *Welt* refers to news items from `welt.de`. *Zeit* refers to news items from `zeit.de`. *SZ* refers to news items from `sueddeutsche.de`. *Stern* refers to news items from `stern.de`. *F.A.Z.* refers to news items from `faz.net`. *taz* refers to news items from `taz.de`. Public service broadcasters (television): *ARD* refers to news items from `tagesschau.de`. *ZDF* refers to news items from `zdf.de/nachrichten`. Public service broadcasters (radio): *D.funk* refers to news items from `deutschlandfunk.de`. Other online news outlets: *n-tv* refers to news items from `n-tv.de`. Exact p-values in parentheses.
* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Exclude extreme parties

Next, we confirm that our approach does not hinge on the selective sharing of the politically extreme parties, LINKE and AfD, alone. It is, for instance, possible that only these parties follow distinct patterns in their sharing behavior, while the sharing behavior of the more centrist parties

[18] For instance, if only a particular party criticizes all news outlets, but does so proportionally across outlets such that for all outlets the same fraction of tweets is critical, its relative number of Twitter referrals is not affected when dropping those negative tweets. Similarly, if only a particular news outlet is being criticized, but proportionally so by all parties, the ranking within that outlet would not be affected, either.

[19] Under the Bonferroni correction, ρ_o is weakly statistically significant (10% level) for three news outlets: *BILD*, *Welt*, and *Zeit*.

is similar and thereby uninformative about the political position of the news outlets.[20] To this end, we exclude (i) LINKE, (ii) AfD, and (iii) LINKE and AfD at the same time from the analysis and compute the Spearman rank correlation coefficient based on the relative number of Twitter referrals by the remaining parties, respectively. Rows two, three, and four of Table 5.5 show the results. The magnitude of the Spearman rank correlation coefficient underlies small changes compared to the results shown in Table 5.4. Moreover, given the smaller number of observations, our results are less statistically significant. While ρ_o is statistically significant for *BILD*, *Welt*, and *Zeit* in all three analyses, it is not statistically significant for *Spiegel* and *Deutschlandfunk* when excluding AfD (row three) or both AfD and LINKE (row four).[21]

Relative number of Twitter referrals

Third, we use the parties' *relative* number of Twitter referrals to each of the twelve news outlets as a basis for their ranking (see Section 5.4). The major advantage over using the absolute number of Twitter referrals is that the parties who are most active on Twitter are not automatically given high referral ranks for each news outlet, which would undermine the idea of our measure. The main disadvantage of this approach is, however, that the Spearman rank correlation coefficient that we compute for each outlet is dependent on the other news outlets included into the analysis, because a party's referral share to news outlet o – and thereby its rank – depends on the referrals to all other news outlets that we consider.

We consider this to be a minor disadvantage. As argued, we are interested in the sign of the Spearman rank correlation coefficient and not in its precise magnitude. Thus, even if ρ_o's magnitude was affected by considering more or less news outlets, it would not be a threat to our approach as long as its sign does not switch. Three robustness checks support this view. First, we included Twitter referrals to *taz*, which is known to be a very left-wing news outlet, into the analysis (Table 5.5, row five). News items by *taz* are relatively often shared by left-wing, but not by right-wing parties; as a result, the referral shares to the original twelve outlets change for the left-wing, but not for the remaining parties. Accordingly, we find that ρ_o decreases for *Zeit*, but is still statistically significant at the 1% level. Moreover, ρ_o for *taz* itself is negative and also statistically significant at the 1%-level, hence, *taz* is positioned on the left as expected. The results

[20] On the other hand, it has recently been argued that the extremely left-wing and the extremely right-wing parties have become quite similar regarding certain topics such as immigration; see, e.g., `https://www.zeit.de/politik/deutschland/2017-07/afd-linke-rechts-links-waehler-gemeinsamkeiten`, viewed Feb 2019. If this was the case, our main results would be even too conservative.

[21] Under the Bonferroni correction, ρ_o is statistically significant for four outlets: *Zeit* and *taz* (10% level), *BILD* and *Welt* (5% level).

for the remaining news outlets are unaffected.[22]

Next, we successively exclude the Twitter referrals to one of the originally selected twelve news outlets and check how the results for the remaining eleven news outlets change.[23] In each case, the magnitude of the Spearman rank correlation coefficient changes slightly, but never switches sign. Moreover, with one exception, the news outlets that are classified as positioned on the political left or right in Section 5.5.1 remain to be classified as such unless it is their turn to be excluded (when we exclude *Welt*, ρ_o for *Deutschlandfunk* is no longer statistically significant).

Third, we compare our results from Section 5.5.1 with the results we would have obtained when using the absolute instead of the relative number of Twitter referrals.[24] The most right-wing party AfD – whose members are most active on Twitter – would then be given one of the highest ranks for each news outlet, while the second-most right-wing party CSU – whose members are least active on Twitter – would be given one of the lowest. As a result, the Spearman rank correlation coefficients would be very different from those presented in Table 5.4 (Table 5.5, row six). In particular, the magnitude of the coefficients computed based on the absolute number of Twitter referrals is smaller and none of them is statistically significant.

5.6 Conclusion

We present a novel and easy to implement measure for the political position of news outlets that is based on the selective sharing of news items by German MPs. Its application to twelve major German online news outlets shows that two news outlets, *BILD* and *Welt*, are biased to the right, three news outlet, *Zeit*, *Spiegel*, and *Deutschlandfunk*, are biased to the left, and the remaining outlets are unbiased. These results are in line with earlier findings on the political position of German news outlets.

Our approach is limited in at least two respects. First, while our approach can assess whether a news outlet is positioned on the left or on the right, it is agnostic about the type of bias, i.e., whether there is a selection or a distortion bias or both. Similarly, we cannot determine whether the bias is demand or supply driven.

Second, the approach is applicable only to online news outlets. Yet, since nowadays every major news outlet also operates online, we do not consider this as an important caveat. Relatedly, the measure cannot be applied to small news outlets whose news items are never shared by politicians.

[22] Under the Bonferroni correction, ρ_o is statistically significant (10% level) for three news outlets: *BILD*, *Welt*, and *Zeit*.

[23] These results are unreported, but available upon request.

[24] Using the relative number of referrals to a news outlet also distinguishes our approach from a recent study by the Pew Research Center that classifies the political position of a number of US news outlets based on the absolute number of Facebook shares by members of the 114th and 115th US Congresses. See `http://www.people-press.org/2017/12/18/sharing-the-news-in-a-polarized-congress/`, viewed: Feb 2019.

This does not, however, generally preclude the investigation of local news outlets. One could, for instance, study the sharing patterns of *local* politicians to determine the political position of *local* online news outlets, which would be an interesting direction for further research.

6 Quantity Restrictions on Advertising, Commercial Media Bias, and Welfare

with Johannes Münster. Published in: *Journal of Public Economics (2015), Vol. 131, 124-141.*

6.1 Introduction

It is widely agreed that a free and independent media is important for society and democracy. While the independence of media can be endangered from many directions, recent discussions both in academia and policy circles have shown that commercial media bias is an important concern. Commercial media bias arises out of a conflict of interest between advertisers and audiences over media content. Studies from marketing have shown that advertisers prefer lighter content and genres that put consumers in a more advertising receptive mood.[1] Moreover, advertisers may prefer the media not to report critically about their products.[2] There are indications that the topic of commercial media bias has become especially important in recent years. The FCC (2011) has reported worries about the "crumbling ad-edit wall" in broadcast television: due to their difficult financial situation, media enterprises may be particularly vulnerable to advertisers' pressures.[3]

This paper investigates the welfare effects of limits for the quantity of advertising in a commercial free-to-air television market in the presence of commercial media bias. Since free-to-air

[1]For example, Wilbur (2008, p.373) finds that "advertiser genre preferences are nearly opposite those of viewers": viewers prefer action and news, while advertisers prefer reality and comedy. A case in point is that Coca-Cola and General Foods have refused to advertise during news broadcasts, as "bad" news might affect consumers' perception of their products (Hawkins and Mothersbaugh, 2009).

[2]For example, tobacco advertisers have pressured media outlets to suppress information concerning the health risks of smoking. Blasco and Sobbrio (2012) provide a survey of the evidence.

[3]For example, the FCC (2011) describes the case of a local Fox channel, KBTC-TV, that featured a story on a new electronic rehabilitation system for injured kids. The reporter was introduced to the audience in a way that suggested an independent report by the channel. The reporter did not work for KTBC, however, but for the Cleveland Clinic. Lieberman (2007) reports that this is not an isolated case: "a hybrid of news and marketing (...) has spread to local TV newsrooms all across the country (...). Viewers who think they are getting news are really getting a form of advertising. And critical stories – hospital infection rates, for example, or medical mistakes or poor care – tend not to be covered." Recent academic contributions on commercial media bias include Reuter and Zitzewitz (2006), Ellman and Germano (2009) and Germano and Meier (2013); we review the literature in Section 6.2. The issue has also recently raised the interest of the FTC, which hosted a workshop on the "blurred lines" between advertising and content in December 2013.

broadcasters do not collect direct payments from their audiences, they may be especially susceptible to advertisers' influence. In our model, broadcasters choose the quality of their program for the viewers, and the quantity of advertising. The conflict of interest between viewers and advertisers gives rise to a trade-off: making the program more attractive for viewers increases the number of viewers, but lowers the willingness to pay of the advertisers. Ceteris paribus, a cap on advertising quantity will drive the per viewer price of advertising up, since the inverse ad demand is decreasing in quantity. A higher per viewer price of advertising makes it more profitable for broadcasters to attract additional viewers. Therefore, the non-advertising program content of the media will become more aligned with viewers' preferences.

This result may help to understand cross country differences in television content. News belong to the viewers' (but not advertisers') most preferred television genres (Wilbur, 2008). We would therefore expect that the supply of news is higher when advertising quantity is restricted, and indeed Aalberg et al. (2010) show that the supply of news and current affairs by the biggest commercial broadcasters during prime time is drastically higher in several European countries, where advertising quantity is restricted, than in the USA, where it is not.[4]

A quantity restriction on advertising increases consumer surplus, but decreases producer surplus. We study the conditions under which welfare (which we take to be the sum of consumer surplus and all profits) increases. In particular, due to its effect on media content, a cap may improve welfare even when consumers do not directly suffer from advertising, or can easily avoid ads by the use of ad avoidance technologies such as digital video recorders (DVRs).

Competition between many independently owned broadcasters helps overcoming commercial media bias. Surprisingly, it increases at the same time the likelihood that a cap on advertising improves welfare. Therefore, competition and regulation of advertising should not be seen as substitutes; rather, they complement each other. The key reason for the complementarity is as follows. A cap that marginally reduces advertising quantity crowds out the marginal advertisers. The associated loss in producer surplus depends on the willingness to pay of the marginal advertisers, which in equilibrium equals the price of an advertising spot. Competition on the media market decreases this price. Correspondingly, the marginal advertiser has a lower willingness to pay, and the loss in producer surplus from a cap is lower.

[4] Aalberg et al. (2010) compare the two biggest commercial broadcasters in each of five European countries with the two biggest commercial broadcasters in the USA. During peak hours, in 2007 the biggest two commercial broadcasters in the USA devoted an average of 6 minutes a day on news and current affairs. In comparison, the biggest two commercial broadcasters in Belgium provided 42, in the Netherlands 20, in Norway 19, in Sweden 27, and in the UK 37 minutes. Of course, there are many differences between these European countries and the USA. The authors stress the higher importance of public service broadcasting in Europe. Within countries, public service broadcasters have a higher news supply than commercial broadcasters; across countries, the biggest public service broadcasters in the European countries show more news during prime time than those in the USA.

The complementarity between regulation and competition is thus tightly linked to the effect of competition on advertising prices. Empirically, it seems that competition on the broadcasting market reduces advertising prices (see Brown and Alexander, 2005). As has been pointed out by Athey et al. (2013, p.6) and Anderson et al. (2012), this poses a puzzle in media economics, since standard models of free TV give the opposite prediction: competition between broadcasters for viewers decreases advertising quantities since viewers are ad averse, and thereby *increases* advertising prices. Our model provides a potential explanation for the empirical results. Strong competition among broadcasters leads to low advertising quantities, but also to viewer friendly programs. Other things being equal, the reduction of advertising quantity increases advertising prices, as in the standard models. A more viewer friendly program, however, lowers the advertisers' willingness to pay and thus equilibrium advertising prices. We show that the latter effect dominates the former one.

Our model also contributes to understanding the "crumbling ad-edit wall" diagnosed by some observers of today's media markets. In times of low ad demand, for example due to advertisers moving online or due to general economic conditions, the price of an ad per viewer is lower. Therefore, attracting viewers is less important for the broadcasters. As a consequence, in equilibrium media content will be more aligned with advertiser preferences.

A cap on advertising lowers broadcasters' profits, and may thus induce exit and a higher concentration on the media market. We show, however, that our main results are qualitatively similar when taking endogenous entry into account. In particular, a "local" cap (that slightly reduces advertising quantity) improves consumer surplus, and is more likely to be welfare enhancing when competition is fierce. In contrast, a proportional tax on advertising revenues has rather different implications than a cap. The reason is that a cap reduces advertising quantity, while a tax increases it in the long run. Marginal costs are zero in television markets. A tax on advertising revenue is therefore a tax on variable profits, and for a given number of broadcasters, equilibrium decisions are unchanged. The tax lowers broadcasters' profits, however, and thus induces exit, and the reduced competition leads to an increase of advertising quantity.

Our paper contributes to two classic topics in public finance, the private provision of public goods, and the comparison of price versus quantity instruments, in the specific setting of advertising financed media. Media markets are of general interest since the working of these markets affects not only their active participants, but also generates important externalities, for example by helping citizens to take well-informed political decisions. For an adequate analysis, the structure of media markets needs to be modelled in more detail than is customary in the theory of public goods. We thus build on modeling tools developed in the economic analysis of advertising and in the theory of two-sided markets, which is a comparatively new topic in public finance.

The paper is organized as follows. The next section provides the background by reviewing (i) the empirical literature on the influence of advertisers on media content, (ii) the conflict of interest between viewers and advertisers, (iii) the regulation of television advertising, and (iv) the related literature. Section 6.3 gives a simple and highly stylized example that illustrates the main effects in our model. Section 6.4 introduces the model, briefly mentions its microfoundations (which are presented in detail in Appendix F.2), and discusses the assumptions underlying our welfare analysis. Section 6.5 characterizes the equilibrium and its welfare properties, investigates the welfare effects of a cap, determines the welfare maximizing cap without and with endogenous entry in the broadcasting market, and discusses advertising taxes. Section 6.6 studies Pay TV and ad avoidance technologies. Section 6.7 summarizes our findings, discusses robustness issues and extensions (laid out in detail in Appendix F), and briefly mentions the testable predictions of the model. Proofs are relegated to Appendix E, and some lengthy technical proofs Appendix F.

6.2 Background

Ads influence editors. Many media platforms depend heavily on advertising revenues. For example, the 2014 Pew report on the state of the news media finds that advertising accounts for 69% of US news revenues. At the same time, media reports about firms and their products influence profits.[5] The media's dependence on advertising revenues, combined with its impact on firms' profits, implies that advertisers have economic incentives to influence editorial decisions. Several recent papers have shown econometrically that indeed advertisers systematically influence media content.[6] Advertisers' influence on media content can be expected to be especially strong in solely advertising funded media such as free TV or radio broadcasting. Moreover, the difficult financial situation of the news media today afflicts the quality of news coverage (Pew Research Center, 2013) and has led to a reconsideration of the traditional separation between media companies' news and business divisions (FCC, 2011).

Conflict of interest between viewers and advertisers. At the center of our model is a conflict of interest between viewers and advertisers over media content. Here we discuss the empirical literature that motivates this assumption. First, there is good evidence that viewers favor different genres than advertisers. Wilbur (2008) estimates a two-sided empirical model of viewer demand for programs and advertiser demand for audiences. In his data, viewers' two most preferred programs are action and news, accounting for 16% of program network hours, whereas advertisers' two most

[5] See Appendix F.2 for empirical references.

[6] The seminal contribution is Reuter and Zitzewitz (2006). See Table F.1 in Appendix F for an overview of econometric evidence. As we describe in Appendix F.1, interviews and surveys of key players in the market also confirm that advertisers influence media content.

preferred programs are reality and comedy, accounting for 47% of program network hours. His results suggest that advertisers' preferences have a bigger impact on the networks than the viewers' preferences. Similarly, Brown and Cavazos (2005, p.30) find that "broadcast television programs receive large and statistically significant premia or discounts based on their content, holding constant the number, income, age and gender of the viewers these programs attract. Sitcoms receive large premia, while news shows and police dramas receive large discounts." In their sample, adjusting for the length of these program types, sitcoms aired more than one-and-a-half time more often than news shows and police dramas combined.

A potential explanation for advertisers' genre preferences is provided by the experimental research concerning context effects on advertising effectiveness. Goldberg and Gorn (1987) show that happier program content puts viewers in a more advertising receptive mood. Relatedly, Mathur and Chattopadhyay (1991) find that it improves viewers' message recall as well as their cognitive responses towards the commercials. Advertisers take these issues seriously. Hawkins and Mothersbaugh (2009, p.298) report that "Coca-Cola and General Foods have refused to advertise some goods during news broadcasts because they believe that 'bad' news affect the interpretation of their products. According to a Coca-Cola spokesman: 'It's a Coca-Cola policy not to advertise on TV news because there is going to be some bad news in there, and Coke is an up-beat, fun product'."

A second issue is that viewers but not advertisers may favor accurate reporting of any defects, risks, or negative externalities of products (see Blasco and Sobbrio, 2012, for a review, and Appendix F.2). An important and well documented case in point is the media coverage of the health risks of smoking. Another important case is the media coverage of anthropogenic climate change, where the discourse in the news media has significantly diverged from the scientific consensus. As pointed out by Ellman and Germano (2009), one potential reason behind this biased media coverage is the influence of big advertisers such as car manufacturers or airlines.

Regulation of TV advertising. The regulation of the quantity of advertising in television differs markedly across countries. In the European Union, for example, the Audiovisual Media Services Directive requires that the "proportion of television advertising spots and teleshopping spots within a given clock hour shall not exceed 20%" (Article 23 §1.). In contrast, in the United States there are no such rules, except for children's programs. Economic theory has identified two countervailing considerations concerning the welfare effects of limits for the quantity of advertising (Anderson and Coate, 2005): On the one hand, broadcasters are often competitive bottlenecks and have market power over advertisers, suggesting that advertising quantity may be too low from a welfare perspective, for the usual reason why firms with market power restrict quantities below the efficient level. On the other hand, consumers may have a disutility from advertising, suggesting that there may be too much advertising in free TV, since the free TV broadcasters cannot perfectly internalize the effect of advertising on their viewers.[7] Indeed, regulation authorities describe protecting consumers as the most important function of the quantity restrictions (e.g., OFCOM, 2011).

Today consumers can, however, avoid contact with annoying advertisements by the use of ad avoidance technologies such as digital video recorders (DVRs). In the EU, about 30 percent of all households already use such technologies (IP Network, 2013). In the US, 47% of TV households have at least one digital video recorder (Leichtmann Research Group, 2013), and about 23% have DVRs on more than one TV set. The average US American watches 25 minutes of DVR playback a day (Nielsen Media Research, 2013). The traditional argument for quantity restrictions on advertising may become less compelling under these conditions. Our paper shows that, however, a cap on advertising makes the non-advertising content of the media more aligned with viewers' preferences. Therefore, a cap may increase welfare even if no consumer is directly affected by advertising.

Related literature. Our paper is related to four strands of the literature. First, broadcasting is a prime example of the private provision of a public good. For this reason, our paper contributes to the broad literature on public good provision (see Batina and Ihori, 2005, for a review). The provision of public goods via advertising is studied in (Luski and Wettstein, 1994) and Anderson and Coate (2005). Our paper goes beyond these papers by studying advertisers' impact on media content.

Second, our paper contributes to the literature on price versus quantity instruments (Weitzman, 1974), as we compare the welfare implications of a tax on advertising with the effects of a quantity regulation. Our contribution to this literature is to focus on a specific industry, namely advertising supported media.

Third, we contribute to the growing work on media bias (see Prat and Strömberg, 2013, for a survey). The economics' literature has mainly focused on political media bias. We focus on advertisers' influence and commercial media bias. Our analysis is closely linked to Ellman and Germano (2009). In their setting, consumers value accurate news, while advertisers value ad-receptive consumers. They show that a monopoly newspaper will underreport news that sufficiently reduces advertiser profits. Interestingly, in a newspaper duopoly, commercial media bias will be eliminated when advertising demand is sufficiently high, unless advertisers are able to commit to withdraw ads from newspapers if they report too critically. Germano and Meier (2013) study the case of many competing horizontally differentiated media outlets to investigate how

[7] Wilbur (2008) estimates that a 10% reduction of advertising quantity in television leads to a 25% increase in audience size.

media diversity and ownership concentration affect commercial media bias. Blasco et al. (2016) and Spiteri (2015) show that commercial media bias is a concern in particular if all advertisers share the same preferences over media content, as in the tobacco example, where the tobacco industry had a shared interest in eliminating coverage of the health risks of smoking. Otherwise, competition in the product market may help overcome commercial media bias. Blasco and Sobbrio (2012) provide a survey on competition and commercial media bias. However, the quantity of advertising chosen by free TV broadcasters, its interaction with program quality and commercial media bias, and the welfare effects of a cap on advertising, have not been formally studied yet. The present paper attempts to close this gap. We ask how a quantity restriction on advertising influences commercial media bias, analyze its welfare properties, and compare the effects of a quantity restriction with those of a tax on advertising revenues.

Our model also relates to the literature on political media bias and media capture. In some settings politicians or governments are in fact major advertisers. Politicians that aim to be (re)elected inform the voters on their manifestos via canvassing television ads; they prefer the broadcasters not to report on any scandals or former mistakes that could reduce their chances. Voters, on the other hand, wish to be properly informed about the candidates. Suppressed information on politicians can prevent them from making an appropriate choice and hence lead to distorted political outcomes. Empirical evidence on this mechanism is given by Di Tella and Franceschelli (2011) who show in a study of Argentinian newspapers that government advertising is associated with a reduced coverage of the government's corruption scandals. Moreover, as reported above, news are among the most preferred genres of viewers, but not of advertisers. News consumption may have positive externalities by improving citizens' political decisions, and consumers will not internalize the large social gains associated with an informed electorate. Therefore, there could be a demand driven media bias of too little informative news even without any interference from advertisers (Gentzkow and Shapiro, 2008). Commercial media bias against news aggravates this concern.

Fourth, in order to model advertising supported media adequately, we build on the literature on advertising (see Bagwell, 2007, for a survey) and two-sided markets (see Anderson and Gabszewicz, 2006, for a survey). There are three major views in the economic analysis of advertising. According to the informative view, advertising provides customers with information about the existence, price, or qualities of the products. The persuasive view holds that advertising changes consumers' tastes. The complementary view holds that advertising raises the true utility of the advertised goods. The literature has ambiguous results on whether there is too much or too little advertising from a welfare perspective. Moreover, the empirical literature indicates that no single view captures all the relevant aspects (Bagwell, 2007). In this paper, we aim to show that a cap on advertising improves welfare under some conditions. To make our case strong, we take a rather benign view of advertising and model advertising as informative.[8]

In the theory of two sided markets, our paper is closely related to the seminal work of Anderson and Coate (2005), who argue that from a welfare perspective equilibrium advertising quantities in a two-sided media market may be too high or too low, mainly depending on consumers' ad aversion. Their model has been extended to a more detailed analysis of horizontal product differentiation by Peitz and Valletti (2008). Our model of entry in two-sided markets TV is related to Choi (2006) and Crampes et al. (2009). The studies by Ellman and Germano (2009), Germano and Meier (2013), and Blasco et al. (2016) discussed above pioneered using models of two-sided markets for the analysis of commercial media bias. As we compare the welfare implications of a tax on advertising with the effects of quantity regulation, our work is furthermore related to Kind et al. (2008) who examine taxes in two-sided markets. They study the cases of a monopoly platform, and of perfect competition, assuming that the platforms' marginal costs are strictly positive, and show that taxes can help to accomplish the social optimum if the platform causes overprovision. Our paper, in contrast, focuses on endogenous program quality, and in particular on commercial media bias, in television markets, which are typically oligopolistic. Moreover, in television markets marginal costs are negligible. Thus revenue taxes are taxes on variable profits and affect entry but cannot be used to fine-tune economic decisions in the short run. Finally, our paper can also be linked to work on ad avoidance technologies (e.g., Anderson and Gans, 2011).

6.3 Example

In this section, we illustrate the main effects in our model with a simple and highly stylized example, deferring a more detailed discussion of our assumptions to the next section. In the example, a monopoly broadcaster chooses its program quality v and its advertising quantity a. Consumers are uniformly distributed on a Hotelling line $[0, 1]$, the broadcaster is located at 0. Consumers have linear travel costs: a viewer who is located at a distance $x \in [0, 1]$ from the broadcaster has utility $v - x$ from watching television. Consumers are ad neutral: their utility from watching television is independent of the advertising quantity. A consumer watches television whenever his utility exceeds his outside option of zero. The total number of consumers is normalized to one, thus the number of viewers is simply equal to the program quality v.

To capture the conflict of interest between advertisers and viewers, we assume that advertisers' willingness to pay for advertising spots decreases in program quality. To be specific, let r denote

[8] Section 6.4.2 discusses the assumptions underlying our welfare analysis in more detail. In Appendix F.3, we also explore the case of misleading advertising.

the per viewer price of an advertising spot, and suppose that the inverse ad demand per viewer is $r = 1 - v - a$.

The broadcaster is financed by advertising, has zero variable costs, and fixed costs $F > 0$. To ensure viability of the market, let $F < 1/27$. The broadcaster's revenue is equal to the number of viewers, times the prices of an ad per viewer, times the number of ads; its profit is $\pi = v(1 - v - a)a - F$.

For a given advertising quantity $a > 0$, the profit maximizing program quality v is determined by the first order condition

$$1 - v - a = v. \tag{6.1}$$

Equation (6.1) illustrates the fundamental trade-off in our model. The left hand side of (6.1) describes the marginal gain of the broadcaster from higher quality, on a per advertising spot basis: higher quality increases the number of viewers, and on each viewer the broadcaster earns the price of an ad per viewer. The right hand side of (6.1) describes the marginal costs of the broadcaster from higher quality, per advertising spot: higher quality decreases the price of an ad per viewer, and the loss of revenue is equal to the number of viewers, which is equal to v in the example.

Solving equation (6.1) for the profit maximizing program quality gives $v = v^*(a) := (1 - a)/2$. Substituting $v^*(a)$ into the broadcaster's profit function leads to $\pi = (1 - a)^2 a/4 - F$. Without a cap on advertising quantity, the profit maximizing choices of the broadcaster are $a = v = 1/3$, resulting in a profit $1/27 - F > 0$. If there is a cap $\bar{a} < 1/3$, the broadcaster's profit maximizing choices are $a = \bar{a}$ and $v = v^*(\bar{a})$ as long as the resulting profit is positive; otherwise, the broadcaster shuts down.

Note that the profit maximizing quality *increases* when a binding cap is introduced. The reason is straightforward to see from the first order condition (6.1): if the advertising quantity is lower due to a cap, ceteris paribus the price of an ad per viewer is higher, and this gives the broadcaster an incentive to increase its quality in order to attract additional viewers.

Now consider the welfare effects of a cap on advertising quantity. We measure consumer surplus by the consumers' aggregate utility from watching television, $CS = \int_0^v (v - x)dx$. Inserting $v^*(a)$ shows that $CS = (1 - \bar{a}^2)/8$ is decreasing in $\bar{a}$. Because a cap increases program quality, consumers are better off, even though they are ad neutral in our example. On the other hand, the cap decreases producer surplus, as measured by the area under the per-viewer inverse advertising demand curve multiplied by the number of viewers. To see this, insert $a = \bar{a}$ and $v = v^*(\bar{a})$ into $PS = v \int_0^a (1 - v - x)dx$ to get $PS = \bar{a}(1 - \bar{a})/4$, which is increasing in $\bar{a}$ in the relevant range $\bar{a} \leq 1/3$. Welfare (the sum of consumer surplus and produces surplus minus fixed costs) is $(1 - \bar{a}^2)/8 - F$ and thus decreasing in $\bar{a}$ in the relevant range. The benefits of the consumers from a cap outweigh the losses of producers. Of course, if the cap is too tight it will drive the broadcaster out of business, to the detriment of both consumer surplus and producer surplus. The welfare maximizing cap is as tight as possible, subject to the broadcaster breaking even.

6.4 The model

6.4.1 Economic agents

There are $N \geq 2$ advertising funded media outlets. Our prime application is to free-to-air television broadcasters, but the model is also applicable to other advertising funded media, such as radio broadcasting. Broadcaster i chooses its program quality $v_i \in \mathbb{R}$ and its quantity of advertising $a_i \in \mathbb{R}_+$.[9] We study a model of a circular town in the spirit of Salop (1979), this is perhaps the best known textbook model that allows for horizontal product differentiation and a flexible number of firms.[10] Broadcasters are evenly spaced on a circle with unit circumference. A mass n of viewers is uniformly distributed on the circle. Viewers single home: each viewer watches only one broadcaster.[11] The utility of a viewer located at a distance x from broadcaster i is

$$w + v_i - \delta a_i - \tau x. \tag{6.2}$$

Here, $w > 0$ is an exogenous parameter sufficiently big to ensure the market is covered in equilibrium; it represents a viewer's utility from a program located at his ideal point with zero advertising and program quality. The viewers' utility increases in program quality v_i. The parameter $\delta \geq 0$ captures ad aversion; consumers are ad averse when the parameter δ is strictly positive, and ad neutral when $\delta = 0$.[12] Transportation costs are linear with a transportation cost parameter $\tau > 0$ which can be regarded as a measure of the broadcasters' substitutability; the lower τ, the easier it is to substitute for broadcaster i's program.

There is a mass m of producers. Each of them produces and advertises one good at constant marginal costs normalized to zero. We refer to the producers also as the advertisers. Advertising

[9] We take v_i from the real numbers to avoid corner solutions which are less interesting from an economic point of view.

[10] In Appendix F, we show that the main results do not hinge on specific features of the Salop circle model. We introduce a more general model of television viewing behavior that nests the Salop model and several other textbook models of discrete choice. We find that the conditions under which a local cap improves welfare are qualitatively similar. The precise quantitative implications, however, depend on the assumed model of television viewing.

[11] Note that the assumption of single homing viewers makes the case for advertising restrictions stronger. Single homing implies that broadcasters have market power on the advertising market and will restrict advertising quantities in order to drive up the price per ad per viewer. Therefore, as argued by Anderson and Coate (2005), equilibrium advertising levels may be too low in equilibrium. If we had competition among broadcasters for advertisers, we would rule out by assumption an important argument why equilibrium advertising quantities may be too low.

[12] In the main model, we assume all viewers dislike ads to the same degree. In order to investigate the impact of ad avoidance technologies, an extension where consumers differ in ad aversion is studied in Section 6.6.2.

is informative: consumers are initially unaware of the existence of a good, but become informed when watching a channel that is airing an ad for the good. Producers are characterized by the quality of their goods, denoted by $\tilde{\sigma}$. We assume that $\tilde{\sigma}$ is uniformly distributed on $[0, \sigma]$. The parameter $\sigma > 0$ corresponds to the highest possible quality of a consumption good.

To model the conflict of interest over media content between viewers and advertisers, we assume that a consumer watching a channel with quality v_i is willing to pay up to $\tilde{\sigma} - \beta v_i$ for a product of quality $\tilde{\sigma}$, where $\beta > 0$. High quality television program reduces the perceived benefits of the products, and thus the consumers' willingness to pay for them. Following Anderson and Coate (2005), we assume that producers capture the willingness to pay of the consumer on the product market. Therefore, the willingness to pay of an advertiser of type $\tilde{\sigma}$ for informing a viewer who watches broadcaster i is $\tilde{\sigma} - \beta v_i$, as well. Thus, viewers' utility increases in v_i, while advertisers' willingness to pay decreases in v_i. In this way, our model captures the conflict of interests over media content between viewers and advertisers.[13] The model combines elements from the classic study of welfare in broadcasting markets by Anderson and Coate (2005) with ideas from the literature on commercial media bias. If program quality is exogenous[14], our model is close to Anderson and Coate (2005).[15] We take from Ellman and Germano (2009) and Germano and Meier (2013) the assumption that program quality decreases the willingness to pay of advertisers.[16]

Advertisers multi-home. Denote the per-viewer price of an ad on broadcaster i by r_i. Assuming $\sigma > \beta v_i + r_i \geq 0$,[17]. Advertising demand is

$$a_i = m \Pr(\tilde{\sigma} - \beta v_i > r_i) = m(1 - \frac{\beta v_i + r_i}{\sigma}).$$

Solving for r_i gives inverse ad demand per viewer, which is

$$r_i = \sigma - \beta v_i - \frac{a_i \sigma}{m}, \tag{6.3}$$

[13]Microfoundations are mentioned in Section 6.4.3 and discussed in detail in Appendix F.2.

[14]If there is an upper bound $\bar{v}$ on program quality, one can also think of the standard model as the case where $\beta = 0$. Then broadcasters will choose the program quality as high as possible. In the main part of the paper, we assume that the upper bound on quality is not binding. We come back to this issue in Section 6.6.1.

[15]One remaining difference is that we study a Salop model with a circular town, whereas Anderson and Coate (2005) consider a linear Hotelling specification.

[16]In our main model, we assume that a higher program quality reduces the willingness to pay of all advertisers by the same amount; that is, advertisers have a shared interest in reducing program quality. We study the robustness of our results in two extensions. In Appendix F.3 we assume that only a subset of advertisers has an interest in reducing program quality; this also allows to study sector specific regulation. In Appendix F.4, it depends on the quality $\tilde{\sigma}$ of the advertised good how much the willingness to pay changes with television quality. We find that if one plausibly assumes that the willingness to pay of producers of high quality is less affected by program quality, the quality enhancing effect of a cap is reinforced.

[17]The second inequality ensures we can safely ignore corner solutions where every advertiser advertises; this will be the case in equilibrium if $N \geq \beta\tau\left(\sigma + m\beta\delta\right)$

whenever $\sigma - \beta v_i \geq a_i\sigma/m$; otherwise inverse ad demand is zero. The broadcaster's revenue per viewer is $r_i a_i$.

Suppose all broadcasters $j \neq i$ behave symmetrically, and let $u := v_j - \delta a_j$. Moreover, suppose that there is an indifferent viewer located between broadcaster i and its closest competitors.[18] Denote the distance between the indifferent viewer and broadcaster i by $\hat{x}$. Then

$$v_i - \delta a_i - \tau\hat{x} = u - \tau\left(\frac{1}{N} - \hat{x}\right).$$

Any viewer with distance less than $\hat{x}$ watches broadcaster i. Therefore, the fraction of viewers watching broadcaster i is

$$2\hat{x} = \frac{1}{N} + \frac{v_i - \delta a_i - u}{\tau}.$$

The profit of broadcaster i is

$$\pi_i = n\left(\frac{1}{N} + \frac{v_i - \delta a_i - u}{\tau}\right)\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right)a_i - F, \tag{6.4}$$

where $F > 0$ are the fixed costs of operation. In Sections 6.5.1 to 6.5.3 we consider the case where the number of broadcasters is exogenous, and assume that the fixed costs are sufficiently small such that broadcasters make positive profits in equilibrium. Section 6.5.4 shows that our main results are robust if we study a model with free entry and endogenize the number of broadcasters with a zero profit condition.

6.4.2 Welfare

In the main part of the paper, our welfare analysis assumes that the willingness to pay of advertisers correctly captures the social benefit of advertisements. As argued by Anderson and Coate (2005), this is a neutral benchmark case, abstracting away from several countervailing considerations (see also Bagwell, 2007, for a review of the economics of advertising). On the one hand, consumers benefit from informative advertising, and when these gains cannot fully be appropriated by the producers, their willingness to pay underestimates the welfare gains of advertising (Shapiro, 1980). On the other hand, if there is competition between producers of products, the advertisers'

[18] This is the case if $u - \tau/N < v_i - \delta a_i < u + \tau/N$. If $v_i - \delta a_i < u - \tau/N$, broadcaster i has no viewers. If $v_i - \delta a_i > u + \tau/N$, broadcaster i is said to undercut its rivals, which will not happen in equilibrium. Due to the linear travel costs, the profit of broadcaster i is discontinuous when broadcaster i just undercuts its rivals: if broadcaster i increases its quality and/or reduces its advertising so much that a viewer whose location is at the location of broadcaster $i + 1$ prefers broadcaster i, then broadcaster i gains all the viewers of broadcaster $i + 1$, including those located between $i + 1$ and $i + 2$. This is a standard property of the Salop (1979) model with linear transportation costs. We carefully spell out profits from undercutting in the proofs.

willingness to pay overestimates the true welfare gains from advertising due to the business stealing effect (Grossman and Shapiro, 1984).

Moreover, persuasive or misleading advertising may make consumers buy products at a price higher than their "true" utility gains from them, which is another reason why advertisers' willingness to pay may overestimate the welfare gains from advertising (Dixit and Norman, 1978). Incorporating these effects would give additional reasons why a cap on advertising can increase welfare. The literature on commercial media bias reviewed in Section 6.2 has argued that there are empirically large and important externalities from advertising due to such effects, for example when the advertised products involve health risks and the media do not disseminate this information. We abstract from these considerations in our main model, but take them into account in an extension in Appendix F.3, where we show that they make the case for advertising restrictions stronger.[19]

Our analysis will focus on symmetric equilibria where all broadcasters choose the same quantity a and quality v. Then consumer surplus CS is given by

$$CS = n(w + v - \delta a) - \frac{n\tau}{4N}. \tag{6.5}$$

Producer surplus PS is the surplus of the broadcasters and the advertisers; in other words, PS equals the sum of advertisers' profits, broadcasters' profits, and fixed costs. In our setting, PS is equal to the total revenue of the advertisers. PS can be calculated as the area under the per-viewer inverse demand curve for advertising spots, multiplied by the number of viewers:[20]

$$PS = n \int_0^a \left(\sigma - \beta v - \frac{\sigma x}{m} \right) dx. \tag{6.6}$$

Total revenues of the broadcasters equal $n(\sigma - \beta v - \sigma a/m)a$, that is, the number of viewers, times the price of an ad per viewer, times the number of ads per broadcaster. The profits of the

[19] Another interesting benchmark is to be agnostic about the value of advertising, and therefore to give it no positive or negative weight in the welfare analysis at all (see Peitz and Valletti, 2008, p.16). Then welfare is a function of program quality alone. A cap on advertising increases welfare according to this standard if it improves the equilibrium program quality. We show this is the case when the number of broadcasters is exogenous, but need not be the case with free entry.

[20] One way to see this is to calculate the revenues of the advertisers. Recall that the mass m of advertisers is uniformly distributed on $[0, \sigma]$. If a is the number of advertising spots, then the marginal advertiser z is given by $(\sigma - z)m/\sigma = a$, i.e. $z = \sigma - a\sigma/m$. Advertisers with $\tilde{\sigma} > z$ advertise, those with $\tilde{\sigma} < z$ do not. The per viewer revenue of an advertiser of a type $\tilde{\sigma} > z$ is equal to $\tilde{\sigma} - \beta v$. Thus advertisers' total revenue is

$$n \int_{\sigma - \frac{a}{m}\sigma}^{\sigma} (\tilde{\sigma} - \beta v) \frac{m}{\sigma} d\tilde{\sigma} = n \int_0^a \left(\sigma - \beta v - \frac{\sigma x}{m} \right) dx.$$

advertisers are the difference between their revenues and the payments to the broadcasters,

$$n \int_0^a \left(\sigma - \beta v - \frac{\sigma x}{m}\right) dx - n \left(\sigma - \beta v - \frac{\sigma a}{m}\right) a = \frac{1}{2}\frac{a^2}{m} n\sigma. \tag{6.7}$$

For a given advertising quantity, advertisers' total profits (6.7) do not depend on program quality. To understand why, note that an increase in program quality implies a parallel downward shift of the inverse ad demand function; for advertising quantity to stay constant, the price of an advertising spot must decrease by the same amount. Moreover, given advertising quantity, advertisers' total profits (6.7) is independent of the number of broadcasters N.

With (6.7), equation (6.6) can also be written as

$$PS = \underbrace{n\left(\sigma - \beta v - \frac{\sigma a}{m}\right) a}_{\text{Revenues of broadcasters}} + \underbrace{\frac{1}{2}\frac{a^2}{m} n\sigma}_{\text{Profits of advertisers}}. \tag{6.8}$$

This formulation is helpful in the analysis in Section 6.5.4, where broadcasters' profits are driven down to zero by free entry, i.e., the revenue of the broadcasters equals their fixed costs NF.

Finally, welfare W is the sum of consumer surplus and total profits of broadcasters and advertisers: $W = CS + PS - NF$.

6.4.3 Microfoundations

A central assumption of our paper is that the willingness to pay of an advertiser for reaching a consumer decreases in the quality of the program the viewer watches. Four different microfoundations for this assumption can be provided. First, for viewers, high quality television may be a substitute for consumption, and thus lower their willingness to pay on the product markets. Second, viewers' recall of an ad may depend on the program it is embedded in. Third, the television program may impact the moods of boundedly rational consumers and thereby, in turn, their purchase behavior. Fourth, high quality television may contain useful information that counteracts deceptive advertising and thereby lowers consumer demand on the product market. We discuss these microfoundations in more detail in Appendix F.2, where we also provide references to the underlying empirical literature.

These microfoundations are not mutually exclusive. Great television programs may at the same time be substitutes for consumption goods, generate lower attention to and recall of advertisements, influence boundedly rational moods, and inform and counteract deceptive advertising. All these microfoundations imply that there is a conflict of interest between advertisers and viewers over television content, and lead to the same positive predictions of the model.[21] For normative questions, our main model builds on the first two microfoundations, where consumers' willingness to pay for a product accurately captures their true benefits from the product. If boundedly rational moods have an impact on purchase behavior, or advertising is deceptive, consumers may have losses on the product market since their perceived gains from the products are not equal to their true gains. The magnitude of these losses may depend both on advertising quantity and on television program quality. In Appendix F.4, we study an extension of our main model that takes these considerations into account.

6.5 Results

6.5.1 Equilibrium

For a given advertising quantity $a_i > 0$, the profit maximizing quality is determined by the first order condition

$$\frac{n}{\tau}\underbrace{\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right)}_{\equiv r_i} = \beta n \left(\frac{1}{N} + \frac{v_i - \delta a_i - u}{\tau}\right). \tag{6.9}$$

The left hand side of equation (6.9) describes the marginal gain of broadcaster i from increasing its program quality, on a per advertising spot basis. Higher quality increases the number of viewers by n/τ, and on each viewer the broadcaster earns the price of an ad per viewer r_i. The right hand side of equation (6.9) describes the broadcaster's marginal costs from increasing its quality, per advertising spot. Higher quality decreases the price of an ad per viewer by β, and the loss of revenue is equal to β times the number of viewers.

The first order condition for the profit maximizing advertising quantity is

$$\frac{\partial \pi_i}{\partial a_i} = -\frac{n\delta}{\tau} r_i a_i - n\left(\frac{1}{N} + \frac{v_i - \delta a_i - u}{\tau}\right)\frac{\sigma}{m} a_i + n\left(\frac{1}{N} + \frac{v_i - \delta a_i - u}{\tau}\right) r_i = 0. \tag{6.10}$$

If broadcaster i shows more ads, he loses viewers because of ad aversion (the first term), achieves a lower per-viewer price of ads (the second term), but generates additional revenue on the additional advertising quantity (the third term). The profit maximizing quantity balances the marginal benefits and costs. Our first result characterizes the symmetric equilibrium.

Proposition 1. *Suppose there is no quantity restriction on advertising. There is a symmetric*

[21] While all the microfoundations are consistent with our assumption that a consumer's willingness to pay for a product of type $\tilde{\sigma}$ is equal to $\tilde{\sigma} - \beta v$, the willingness to pay may also be a nonlinear function. We further discuss this in Appendix F.2.

equilibrium where for $i = 1, ..., N$:

$$a_i = \frac{m\beta\tau}{N(\sigma + m\beta\delta)}, \tag{6.11}$$

$$v_i = \frac{\sigma}{\beta} - \frac{\tau(2\sigma + m\beta\delta)}{N(\sigma + m\beta\delta)}. \tag{6.12}$$

Inverse ad demand per viewer is $r_i = \beta\tau/N$. *The equilibrium profit of broadcaster* i *is*

$$\pi_i = \frac{nm\beta^2\tau^2}{N^3(\sigma + m\beta\delta)} - F.$$

Equations (6.11) and (6.12) can easily be derived from the first order conditions (6.9) and (6.10), assuming that all broadcasters behave symmetrically.[22] Proving equilibrium existence is, however, somewhat more challenging for several interrelated reasons (see Appendix F.5). The profit functions are third order polynomials in advertising quantity and thus not everywhere concave; global optimality needs to be established. Moreover, in the classic Salop (1979) model, undercutting the rivals leads to a nonpositive profit. In contrast, in our model undercutting rivals can lead to a positive profit; we thus need to establish that the profit from undercutting is smaller than the equilibrium profit.

Proposition 1 implies that, when N increases or τ decreases, there will be fewer ads and higher program quality. More competition between broadcasters, be it through lower distances between two adjacent broadcasters or due to better substitutability of their programs, makes viewers better off.[23] This is in line with the results in Ellman and Germano (2009) and Germano and Meier (2013).

Higher competition has two countervailing effects on the equilibrium per viewer price of advertising. On the one hand, it lowers advertising quantity and thereby increases the per viewer price. On the other hand, it increases program quality and thereby decreases the per viewer price. Equation (6.9) reveals that in any symmetric equilibrium,

$$\frac{nr_i}{\tau} \equiv \frac{n}{\tau}\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) = \frac{n\beta}{N}.$$

Recall that the right hand side can be interpreted as broadcaster i's marginal costs of program quality, and equals β times the number of viewers of the broadcaster. In any symmetric equilibrium,

[22] This argument also shows that the equilibrium is unique in the class of all symmetric equilibria whenever $F > 0$ (see Appendix E).

[23] In the classic Salop model, more competition leads to lower prices. In a free TV regime, prices are zero anyhow. However, broadcasters compete in program quality and in advertising time. One can interpret advertising as an implicit price for the program; the result that there is lower advertising in our model is similar to the result that prices are lower in the Salop model.

each broadcaster has $1/N$ viewers; thus the marginal costs of quality decrease in N. At the profit maximizing quality, the marginal benefit of higher quality, which is proportional to r_i, must therefore also be lower. The model thus predicts that more competition on the media market leads to a lower price of an advertising spot. This prediction is in line with the empirical results of Brown and Alexander (2005), who show that a higher concentration on the media market goes along with higher advertising prices.[24]

In contrast, in models where program quality is exogenous as in Anderson and Coate (2005) or Choi (2006), more competition only leads to a lower advertising quantity, and since inverse demand is falling in quantity, advertising prices increase. As noted, this prediction seems at odds with the empirical evidence, and it is a puzzle in media economics to explain the discrepancy. Anderson et al. (1992) and Athey et al. (2013) propose explanations based on multi-homing viewers. Our model offers a complementary explanation in a model where viewers single home. Broadcasters compete for viewers not only in advertising quantity, but also in program quality. More competition on the broadcasting market leads to higher quality, and because of the conflict of interest between viewers and advertisers, higher quality decreases advertising prices.

Proposition 1 also shows that program quality increases in the mass of advertisers m. To understand this result, note that for any given advertising quantity, the profit maximizing program quality is determined by the trade-off described in (6.9): a higher program quality attracts more viewers, but leads to a lower price of advertising spots per viewer. When m increases, ceteris paribus the price of an advertising spot per viewer becomes higher, therefore, it pays to attract additional viewers. This fits nicely with the claim of some observers of today's media markets that, as earning money through advertising is more difficult, be it because advertisers move online or simply because of the general economic conditions, advertisers' interests have a bigger impact on media content. The model may thus contribute to explaining what is sometimes called the "crumbling ad-edit wall" (see FCC, 2011, for a discussion).[25]

To start the welfare analysis, we investigate whether equilibrium program quality and advertising are too high or too low from a welfare perspective. That is, we consider exogenous changes of either program quality or advertising quantity, marginally changing one while holding the other

[24]Brown and Alexander (2005) estimate that a 20% increase in concentration local broadcast television markets would lead to a 9% increase in the per-viewer price of ads (p. 336).

[25]The result is related to the paradox noted by Ellman and Germano (2009) in their model of newspaper competition that increasing the mass of advertisers eventually eliminates commercial media bias. Indeed, it is often argued that advertising revenues help to have independent media (see, e.g., FCC, 2011). As argued above, this is partly reflected in our model, since a higher number of advertisers m implies a higher equilibrium program quality. Moreover, it can be shown that in our model a cap makes it easier to bribe the broadcasters to suppress information by bribes that are independent of the advertising quantity. The risk of such political media capture must be traded off against the commercial media bias we focus on.

constant.

Proposition 2. *A small exogenous increase of program quality of all broadcasters, holding the advertising quantities constant, increases consumer surplus and decreases producer surplus. Moreover, welfare increases if and only if*

$$N > \hat{N}_v := \frac{m\beta^2\tau}{\sigma + m\beta\delta}.$$

An increase of program quality means that the program content is more in line with viewers' preferences, which is the reason why consumer surplus increases. Producer surplus, on the other hand, decreases. A higher program quality of broadcaster i has a business stealing effect on the competing broadcasters: it induces viewers to switch from the competitors to broadcaster i. The profit maximizing quality balances this increase in viewers with the decrease in the prices of advertising (see the discussion of equation (6.9) above). If the quality of all broadcasters is increased simultaneously, as envisioned in Proposition 2, however, viewers do not switch; therefore, the higher quality has only costs but no benefits for the broadcasters. Consequently, broadcasters' profits decrease. In other words, from the point of view of the broadcasters, program quality has a negative externality due to the business stealing effect, and the equilibrium quality is inefficiently high when compared to the quality that maximizes the joint profits of the broadcasters.

The effect of an increase of program quality on consumer surplus does not depend on N. In contrast, its effect on producer surplus is $\partial PS/\partial v = -n\beta a$. Since the equilibrium value of advertising quantity a is decreasing in the number of broadcasters, the effect on producer surplus is less important (smaller in absolute value) when there are many competing broadcasters. This observation explains the result concerning welfare. When there are many independent broadcasters (and similarly when the program substitutability is high), competition for viewers is fierce. Thus in equilibrium there are relatively few ads, and an increase of program quality does not reduce producer surplus much. Hence the positive effect on consumer surplus dominates. Similarly, if consumers are very ad averse, there are few ads in equilibrium and an increase in quality does not reduce producer surplus much, hence a small exogenous increase of program quality increases welfare.

Towards an understanding of the effect of a cap on advertising quantity, we now consider the effect of a small exogenous decrease of advertising quantity.

Proposition 3. *A small exogenous decrease of advertising quantity of all broadcasters, holding program qualities constant, increases consumer surplus and decreases producer surplus. Moreover, welfare increases if and only if*

$$N > \hat{N}_a := \frac{\beta\tau}{\delta}.$$

Consumer surplus decreases in advertising quantity since consumers are ad averse. Producer surplus is increasing in advertising quantity for the usual reason that a monopolist reduces quantities below the efficient level. Here, since viewers single-home, each broadcaster is in a monopoly position with respect to the attention of his viewers. When consumers are not very ad averse (δ sufficiently small), reducing advertising while keeping program quality v constant reduces welfare. This is in line with the results by Anderson and Coate (2005): when the quality of the broadcasters' content is not at stake, and consumers are not very ad averse, the equilibrium quantity of advertising is too low. Conversely, if ad aversion is severe, there is too much advertising in equilibrium.

Again, the effect on consumer surplus is independent of N, while the effect on producer surplus is less important when competition is high.[26] To understand why, note that the effect of a small exogenous decrease of advertising quantity on producer surplus is that the marginal advertiser is crowded out. The corresponding loss of producer surplus equals the willingness to pay of the marginal advertiser, which in turn equals the equilibrium per-viewer price r of an advertising spot times the number of viewers n. As discussed above, r is decreasing in N and increasing in τ. Therefore, the effect on producer surplus is small in absolute value when competition is high.[27]

6.5.2 Effects of a cap on advertising

While the program quality may be hard to regulate,[28] advertising can be restricted. As reported in Section 6.4.3, many countries impose a cap on the time devoted to ads on free TV. To analyze the effects of such a cap, we need to take into consideration its effect on program quality chosen by the broadcasters.[29] We now consider the effect of a quantity restriction on advertising $\bar{a}$ that constraints all broadcasters to choose $a_i \leq \bar{a}$. The following lemma studies the effect of a binding cap.

Lemma 1. *Suppose that there is a cap*

$$\bar{a} \in \left(0, \frac{m\beta\tau}{N\left(\sigma + m\beta\delta\right)}\right)$$

[26]Formally, $\frac{\partial PS}{\partial a} = n\left(\sigma - \beta v - \frac{\sigma a}{m}\right)$. Evaluating the derivative at the equilibrium values of a and v gives $\frac{\partial PS}{\partial a} = \frac{n\beta\tau}{N}$.

[27]We point out that this result does not hold in models where program quality is exogenous, such as Anderson and Coate (2005). In these models an increase in N increases the equilibrium price of advertising. Correspondingly, the loss in producer surplus due to a decrease of advertising quantity is higher when there is fierce competition on the media market.

[28]Moreover, a direct regulation of program quality may raise issues of free speech and media capture by state authorities.

[29]We focus on the effects of a cap on a free TV market. Of course, a cap (and similarly advertising taxes analyzed in Section 6.5.5) will also change the relative profitability of free TV and pay TV. See Section 6.6.1 for an analysis of pay TV.

on advertising. Then there is an equilibrium where broadcaster $i = 1, ..., N$ *chooses* $a_i = \bar{a}$ *and*

$$v_i = \frac{\sigma}{\beta} - \frac{1}{N}\tau - \frac{1}{m}\frac{\sigma}{\beta}\bar{a}. \tag{6.13}$$

Profit equals

$$\pi_i = \frac{n\bar{a}\beta\tau}{N^2} - F.$$

As in the absence of a cap, equilibrium program quality is high when competition is high. Moreover, the equilibrium per-viewer price of an ad is decreasing in N and increasing in τ.

Equilibrium quality is decreasing in $\bar{a}$: the more stringent the cap (i.e., the lower $\bar{a}$), the higher program quality. The main reason is that a cap reduces advertising quantity and thus, since inverse ad demand is decreasing in ad quantity, ceteris paribus increases the price of an ad per viewer. Therefore, attracting additional viewers is more profitable for the broadcasters, and thus the equilibrium program content is more in line with viewers' preferences. To understand the logic in more detail, consider Figure 6.1, which plots the marginal benefits and costs of quality from equation (6.9) as a function of v_i. A cap shifts the inverse ad demand function upward, since the advertising quantity on broadcaster i decreases; ceteris paribus, the price of an ad increases. Simultaneously, the competing broadcasters increase their quality, as predicted by (6.13). When broadcaster i leaves its quality unchanged, i has less viewers than before. Therefore, the marginal cost curve shifts downwards. These two reasons imply that broadcaster i has an incentive to increase its program quality. For future reference, note that the effect of the cap $\bar{a}$ on equilibrium program quality is independent of N.

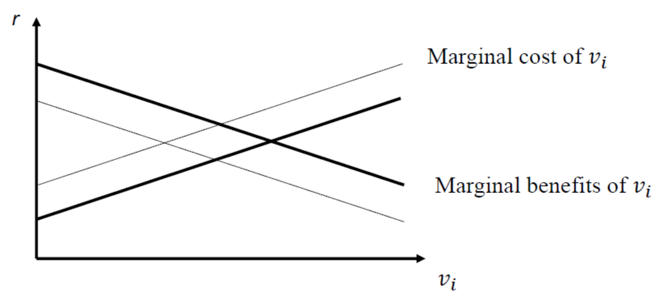

Figure 6.1: Left hand side: Parallel downwards shift of the inverse ad demand function. Right hand side: Turn of the inverse ad demand function.

As noted in the introduction, the supply of news and current affairs during peak hours by the biggest commercial broadcasters in several European countries is drastically higher than that of the biggest commercial broadcasters in the USA (Aalberg et al., 2010). This is in line with our result that a cap increases program quality. Advertising quantity is restricted in the European countries, but not in the USA. Other things being equal, our model predicts that television content is more viewer friendly in Europe, and indeed news belong to the viewers' (but not advertisers') most preferred genres (Wilbur, 2008); our model may thus contribute to an explanation of the cross-country differences.

We now analyze the welfare effects of a "local" cap that reduces advertising quantity slightly below the equilibrium level.

Proposition 4. *A local cap on advertising increases consumer surplus but decreases producer surplus. Welfare increases if and only if*

$$N > \hat{N}_{cap} := \frac{(2\sigma + m\beta\delta)\, m\beta^2\tau}{(\sigma + m\beta\delta)^2}. \tag{6.14}$$

The critical values satisfy $\hat{N}_v < \hat{N}_{cap} < \hat{N}_a$.

The cap reduces advertising and increases program quality; both effects increase consumer surplus and reduce producer surplus. The effect on welfare hinges on the relative importance of these effects. Note that, when $N > \hat{N}_a$, reduced advertising quantity increases welfare, and (since $\hat{N}_a > \hat{N}_v$) an increased program quality does as well; in this case, a cap will surely increase welfare. When this sufficient condition is not satisfied, the direct effect of an advertising cap on welfare is negative; the total effect, however, may nevertheless still be positive.

Proposition 4 implies that, as should be expected, ad aversion makes it more likely that a cap increases welfare. Moreover, whenever (6.14) holds, the size of the welfare gain through a local cap is increasing in δ. However, even if consumers are not ad averse, a cap on advertising can improve welfare.

Figure 6.2 plots the cutoffs as a function of δ. A decrease in advertising quantity, holding program quality constant, increases welfare above $\hat{N}_a$ (the dotted line). Clearly, this can only happen if $\delta > 0$, and the higher δ, the more likely it is. Above $\hat{N}_v$ (the thin line), an increase in v holding a constant increases welfare. Interestingly, the cutoffs differ most dramatically when δ is small. Figure 6.3 plots the cutoffs as a function of m. $\hat{N}_a$ and $\hat{N}_{cap}$ differ most when m is small. In times where advertisers move online and m decreases, the quality enhancing effect of a cap becomes more relevant.

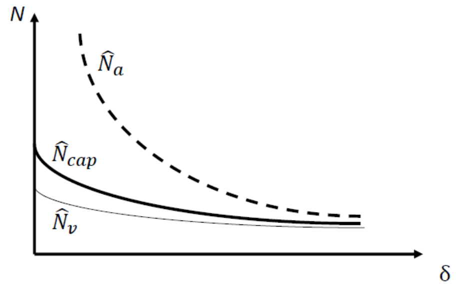

Figure 6.2: Left hand side: Parallel downwards shift of the inverse ad demand function. Right hand side: Turn of the inverse ad demand function.

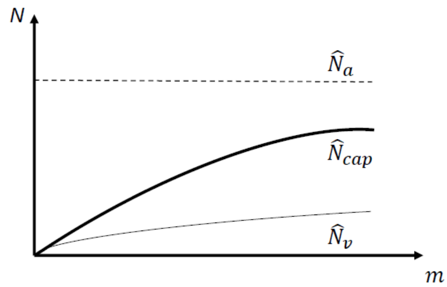

Figure 6.3: Left hand side: Parallel downwards shift of the inverse ad demand function. Right hand side: Turn of the inverse ad demand function.

The most surprising insight from Proposition 4 is that more competition, be it through a higher number of broadcasters (high N) or better substitutability (low τ), makes it *more* likely that a local cap improves welfare. Moreover, the quantitative importance of the welfare gains is greater when there is more competition. This is surprising since, as pointed out above, more competition *increases* equilibrium program quality. Therefore, while competition is helpful to increase program quality, it is not a substitute for regulating the market. Indeed, the marginal welfare gains from a local cap are increasing in N and decreasing in τ; in this sense, there is a complementarity between regulation and competition. Especially in a market with many independent broadcasters, a cap on advertising may improve welfare. A policy implication is that successful competition policy does not automatically make regulation of the advertising quantities dispensable.

To understand the result, recall three observations pointed out above: (i) the effects of a and v on consumer surplus does not depend on N, (ii) the effects of a and v on producer surplus gets smaller (in absolute value) when N increases, and (iii) the effect of a cap on equilibrium program quality is independent of N. These observations imply that the negative impact of a cap on producer surplus gets less important when N increases, while the positive impact on consumer surplus is not affected; hence it is more likely that welfare increases.[30]

6.5.3 The optimal cap

This section studies the problem to maximize welfare by choosing a cap $\bar{a}$ subject to not changing the number of broadcasters,

$$\max_{\bar{a}} W \text{ s.t. } \pi_i = n\bar{a}\beta\tau/N^2 \geq F.$$

For a given number of broadcasters, welfare is a convex function of $\bar{a}$: consumer surplus is linear in $\bar{a}$, while producer surplus is quadratic in $\bar{a}$ (see (6.8) in combination with (6.13)). Therefore, it is either optimal to have no cap on advertising, or a cap that brings profits down to zero, i.e., $\bar{a} = N^2F/\left(n\beta\tau\right)$. In particular, inequality (6.14) is a sufficient but not a necessary condition for the optimal cap to be binding.

Figure 6.4 illustrates. Broadcasters' profits are positive below the zero profit line (in bold); the area above it is ruled out by the assumption that equilibrium profits (absent a cap) are positive. A cap that drives broadcasters' profits to zero is welfare maximizing below the thin line; above it, laissez-faire is optimal. The two lines intersect only once, at $\hat{N}_{cap}$: on the zero profit line a local cap is a zero profit cap. Figure 6.4 also illustrates that a cap can be optimal even when $N < \hat{N}_{cap}$.

[30] We point out that this result hinges on endogenous program quality and a conflict of interest between viewers and advertisers. In models where program quality is exogenous, such as Anderson and Coate (2005) or Choi (2006), more competition implies a higher price of advertising, and correspondingly a larger negative impact of a cap on producer surplus; thus the marginal welfare gains of a local cap are smaller when competition is intense.

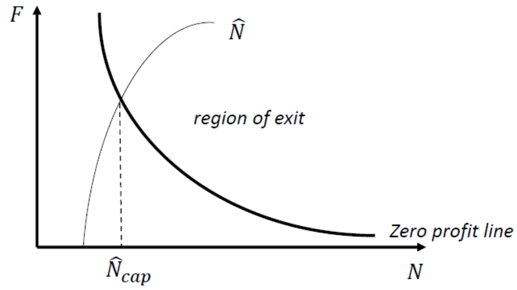

Figure 6.4: Left hand side: Parallel downwards shift of the inverse ad demand function. Right hand side: Turn of the inverse ad demand function.

Proposition 5. *There exists a critical value $\hat{N} \leq \hat{N}_{cap}$ such that the welfare maximizing cap is $\bar{a} = N^2 F/(n\beta\tau)$ if $N > \hat{N}$, and laissez-faire is optimal if $N \leq \hat{N}$. Moreover, $\hat{N}$ increases in F, m, β, and τ ;$\hat{N}$ decreases in n, δ, and σ.*

We now compare the results on the optimal cap with our results from Section 6.5.2 on a local cap. While the welfare gains of a local cap are increasing in N, the same is not everywhere true for the optimal cap. The reason is that, with higher N, the nonnegativity constraint on profits is more stringent.[31] On the other hand, the conditions under which a cap raises welfare are qualitatively similar for a local cap and the optimal cap. In particular, a more competitive broadcasting market, or higher ad aversion, increases the attractiveness of a cap. There are just two differences: the impact of the number of viewers n, and the fixed costs F. For a local cap, these do not matter. For the optimal cap, the higher n, and the lower F, the more stringent a cap can be before inducing exit; therefore, it is more likely that a zero profit cap raises welfare.

The optimal cap is not continuous in the parameters of the model. In Figure 6.4, when we cross the thin line from the left, the optimal policy jumps from laissez-faire to a cap that drives profits down to zero. This is somewhat disconcerting since the optimal policy is not robust with respect to small perturbations. The discontinuity disappears, however, once we consider endogenous entry.

[31]To see this, suppose F is below the point where the two lines cross in Figure 6.4. If N is small (to the left of the thin line), a cap lowers welfare. For intermediate values of N (between the thin and the bold line), a cap increases welfare. On the bold line, the zero profit cap is equivalent to no cap at all, and the associated welfare gains are zero. Therefore, the welfare gains from a cap are not monotone in N.

6.5.4 Endogenous number of broadcasters

This section endogenizes the number N of broadcasters by assuming free entry into the broadcasting market. We follow the standard approach to model entry in a two stage game. In stage 1, a large number of potential broadcasters decide whether or not to enter. Upon entry, a broadcaster has to invest the fixed costs F. A broadcaster who stays out has a profit of zero. In stage 2, broadcasters that have entered choose their advertising quantity and program quality.

The number of broadcasters is then determined by the condition that the broadcasters' profits (given in Proposition 1 and Lemma 1) equal zero.[32] As shown above, for any fixed number of broadcasters, a cap on advertising lowers the profits of the broadcasters. Therefore, under free entry, a cap will reduce competition on the broadcasting market.

Consider the welfare effects of a cap in the model with free entry. Since the broadcasters' profits equal zero by free entry, total profits equal the profits of the advertisers. A cap reduces advertising quantity, and by (6.7), advertisers' profits decreases. Hence a cap decreases total profits, as in the model with an exogenous number of broadcasters. Concerning consumer surplus, however, there are additional effects that can reverse our findings above. A cap induces broadcasters to exit, and exit has two negative consequences for consumers. First, ceteris paribus, exit leads to a lower program quality. This counteracts the quality enhancing effect of a cap studied in Section 6.5.2. The net effect of a cap on program quality depends on the relative strength of these effects. Second, when consumers have fewer broadcasters to choose from, the match between consumers and programs becomes worse (consumers have higher transportation costs). Indeed, a cap that is too stringent decreases consumer surplus. Nevertheless, as our next result shows, a local cap that slightly decreases the advertising quantity below its laissez-faire equilibrium level increases consumer surplus.

Proposition 6. *Consider the model with free entry on the broadcasting market. (i) A local cap increases consumer surplus. (ii) Suppose that*

$$F < \hat{F}_{capwithexit} := \frac{27n\left(\sigma + m\beta\delta\right)^5}{512m^2\sigma^3\beta^4\tau}.$$

Then a local cap increases welfare, and there is a (uniquely defined) optimal cap a^*, which is decreasing in δ and n, and increasing in τ and in F. (iii) If $F \geq \hat{F}_{capwithexit}$, laissez-faire is optimal.

A comparison of Proposition 6 with Proposition 5 shows that our results that better program

[32] We follow Salop (1979) and assume that, after entry or exit, broadcasters automatically relocate such that they are equidistant. We ignore the integer constraint on N for convenience. When fixed costs are high or the cap on advertising is very stringent, only a monopolist broadcaster may be active, or even all broadcasters may exit. We focus on the case where some competition prevails.

substitutability, higher ad aversion, and a larger viewer market increase the attractiveness of a cap, are robust to endogenous entry. Moreover, with endogenous entry, the number of broadcasters depends on the fixed costs: the lower F, the more competition on the broadcasting market. Therefore our result that, with entry, a cap improves welfare if F is sufficiently small, is similar to our result in Section 6.5.3 that a cap improves welfare if N is sufficiently large.

The effects of a cap can be decomposed into the effects for a fixed number of broadcasters, and the effects from the changing number of broadcasters. Holding the number of broadcasters constant, a cap lowers advertising quantity, which directly affects welfare, and induces an increase in program quality that affects welfare, too. In addition to that, a cap on advertising leads to a lower number of broadcasters. A lower N, in turn, affects welfare directly by changing total transportation costs and total fixed costs, and induces a decrease in program quality that affects welfare, too.

Depending on the parameters, endogenous entry can make it more or less likely that a local cap increases welfare.[33] When $\delta > 2\sigma/(3m\beta)$, the exit induced by a cap makes it more likely that a local cap increases welfare.[34] On the other hand, when $\delta < 2\sigma/\left(3m\beta\right)$, the exit makes it less likely that a local cap raises welfare. Therefore, with an endogenous number of broadcasters, the case for a cap is stronger when ad aversion is severe, and weaker when viewers are not very ad averse.[35]

It is not surprising that endogenous entry can tilt the desirability of a cap in both ways. While in the classic Salop model, entry is excessive, Choi (2006) has shown that both excessive and insufficient entry are possible in a Salop model of free TV (see also Crampes et al., 2009). Our results indicate a related ambiguity in the present context. The possibility of excess entry on media markets should not be dismissed as purely theoretical, however. Berry and Waldfogel (1999) show empirically that in the U.S. radio market, entry is excessive when evaluated from the point of view of the radio stations and the advertisers. While they cannot give a complete welfare analysis (due to lack of data on the listeners' value of programming), their results indicate that the business stealing effect of entry, which is one reason why entry may be excessive, is quantitatively important.

[33] A related but different concern is that content of higher quality may have higher costs. As argued by Anderson (2007), a cap can for this reason reduce program quality.

[34] In the laissez-faire equilibrium with free entry, the effect of a cap for given N can be signed as follows. From Proposition 4, we know that, for given N, a local cap raises welfare if and only if $N > \hat{N}_{cap}$. Setting N equal to the equilibrium number of broadcasters under free entry, and solving the inequality for F, reveals that the effects of a local cap for a given number of broadcasters increase welfare if and only if $F < \hat{F}_{cap} := \frac{n(\sigma+m\beta\delta)^5}{m^2\beta^4\tau(2\sigma+m\beta\delta)^3}$. Taking entry into consideration, a cap raises welfare if $F < \hat{F}_{capwithexit}$. Straightforward calculations show that, $\hat{F}_{capwithexit} > \hat{F}_{cap}$ if and only if $\delta > 2\sigma/\left(3m\beta\right)$.

[35] The additional effects due to endogenous entry also determine how m affects the probability that a cap raises welfare: $\hat{F}_{capwithexit}$ decreases in m if and only if $\delta < 2\sigma/\left(3m\beta\right)$.

6.5.5 The effects of an advertising tax

A tax on advertising seems to be a recurrent policy idea (Rauch, 2013). For example, the states of Iowa and Florida taxed advertising in the late 1980s, and advertising taxes have recently been discussed in Minnesota and Ohio.[36] While many countries impose a cap on advertising quantities, however, Austria (with a tax rate of 10%) is currently the only OECD country that taxes advertising revenues. In this section, we point out that a proportional tax on advertising revenues has quite different implications than a cap in our model.[37] We assume that the tax revenue is redistributed lump sum to the consumers, and call *net consumer surplus* the consumers' surplus before redistribution of tax revenues, given in (6.5). Welfare is the sum of net consumer surplus, tax revenue, and all profits.

Consider first the case of an exogenously given number of broadcasters. Since the marginal costs of broadcasters are equal to zero, a tax on advertising revenue is a tax on variable profits, and does not change the equilibrium advertising quantity or program quality. Advertisers' profits and net consumer surplus are unaffected. The broadcasters bear the burden of the tax, since they are monopolists on the advertising markets: due to single homing of consumers, each broadcaster is the only one that can sell access to his viewers. The tax just redistributes from the broadcasters to the government budget, and welfare is constant. In contrast, under the conditions of Proposition 5, a cap raises welfare. Quantity restrictions are a superior instrument to taxes on this market.

With free entry, a tax on advertising revenues leads to exit, and thereby to a higher advertising quantity and lower program quality. Moreover, consumers have fewer broadcasters to choose from, and thus higher transportation costs. These effects decrease net consumer surplus.[38] Interestingly, a tax on advertising increases advertisers' profits (and hence the sum of all profits, too). At first sight, this might be a surprising result; it stems from the two-sidedness of the market. The tax on advertising lowers the number of broadcasters and thus softens the competition for audiences. Therefore, equilibrium advertising quantities are higher and equilibrium program quality is lower. By (6.7), advertisers' profits increase. In contrast, a cap decreases advertising quantities and

[36] Relatedly, in the discussion on tax reform, U.S. House and Senate Committees introduced proposals to change the tax deductibility of advertising. See AdvertisingAge (2013).

[37] An excise tax based on the quantity of advertising, on the other hand, has similar effects as a cap. For given N, an excise tax leads to a lower advertising quantity, and higher program quality; for any cap $\bar{a}$, an equivalent tax rate can be found that leads to the same equilibrium advertising quantities and program qualities. Profits with the tax are lower by the tax revenue than with the cap (unless tax revenues are redistributed lump sum to the broadcasters, in which case the effect of the tax is exactly equal to that of the cap). Therefore, an excise tax on advertising that is, in the short run (for given N), equivalent to a cap, leads in the long run to higher concentration on the broadcasting market.

[38] Since the tax revenue is redistributed to consumers, these negative effects have to be balanced against the additional income from the redistribution of tax revenue. It can be shown that a small tax on advertising increases the sum of net consumer surplus and tax revenues if and only if $F > 27n\left(\sigma + m\beta\delta\right)^2 / \left(64m^2\beta^4\tau\right)$.

therefore advertisers' profits, as well.

Proposition 7. *With free entry on the broadcasting market, a small tax on advertising decreases net consumer surplus and increases profits. It increases welfare if, and only if,*

$$F > \hat{F}_{taxwithexit} := \frac{729n\left(\sigma + m\beta\delta\right)^5}{64m^2\beta^4\tau\left(4\sigma + 3m\beta\delta\right)^3}.$$

Moreover, $\hat{F}_{taxwithexit} < \hat{F}_{capwithexit}$ *if and only if* $\delta > 2\sigma/\left(3m\beta\right)$.

While a cap reduces advertising quantity, a tax on ad revenue increases it. Moreover, the cap increases program quality, while the tax reduces it. This explains why a tax decreases net consumer surplus and increases profits, while the effects of a cap are just the other way round. Moreover, the conditions under which these instruments raise welfare are qualitatively quite different. In particular, fixed costs F, program substitutability τ, the viewer market n, and ad aversion δ have the opposite effect on the probability that a tax, or a cap, raise welfare. To understand why, consider for example fixed costs F. As explained in Section 6.5.4 above, an advertising cap raises welfare if and only if F is sufficiently low – just as in the model with exogenous N a local cap raises welfare when N is sufficiently high. The intuition is that the cap raises v and lowers a, and both increases welfare when there is a lot of competition on the broadcasting market (compare Propositions 2 and 3), i.e., when F is low. Proposition 7, in contrast, shows that a tax on advertising revenue raises welfare if and only if F is sufficiently high. The tax increases a and lowers v, which increases welfare when there is not much competition on the broadcasting market, i.e., when F is high. The tax and the cap have in common, however, that they reduce the equilibrium number of broadcasters. As reported above, if $\delta > 2\sigma/\left(3m\beta\right)$, exit makes it more likely that a cap raises welfare. In this case $\hat{F}_{taxwithexit}$ is smaller than $\hat{F}_{capwithexit}$; thus there is a range of parameters where F is between these critical values, and both a cap and a tax raise welfare. Conversely, when $\delta < 2\sigma/\left(3m\beta\right)$, exit makes it less likely that welfare increases; then there is a range of parameters where neither the cap nor the tax raises welfare.

6.6 Extensions

This section explores two extensions of our model: pay TV, and consumers that differ in ad aversion and use ad avoidance technologies. (Extensions on producers that differ in how far they are affected by television program quality and sector specific regulation, and deceptive advertising are provided in Appendix F.) To keep the discussion short, we assume N to be exogenous and focus on the conditions under which a local cap raises welfare, as in Section 6.5.2 above.

6.6.1 Pay TV

Our model gives additional support to results by Anderson and Coate (2005) and Peitz and Valletti (2008) that a cap on advertising does *not* improve welfare in a pay TV market. Indeed, in a pay TV market, program quality will not be too low from a welfare perspective. To see this, suppose broadcaster i charges a price p_i. A viewer located at distance x from broadcaster i has utility $w + v_i - \tau x - \delta a_i - p_i$ from watching the broadcaster. The profit of the broadcaster is[39]

$$\pi_i = n\left(\frac{1}{N} + \frac{v_i - \delta a_i - p_i - u}{\tau}\right)\left(p_i + \left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right)a_i\right) - F,$$

when all other broadcasters $j \neq i$ offer the viewers the same gross of transportation costs utility $u = v_j - \delta a_j - p_j$. Consider the profit maximizing choice of v_i and p_i for given a_i. Increasing both v_i and p_i by the same amount increases profits if $a_i\beta < 1$. It is natural to assume that there is some upper bound $\bar{v}$ on program quality above which it cannot be improved. Whenever $p_i > 0$ in equilibrium, the broadcaster will increase its program quality as much as possible. This result fits the claim by Brown and Cavazos (2005) that the business strategy of the pay TV broadcaster HBO was to air explicitly darker, advertiser unfriendly material.[40]

6.6.2 Ad avoidance technologies

As argued in Section 6.2, viewers can today easily avoid contact with advertisements by using ad avoidance technologies such as ad blockers or digital video recorders. The traditional argument for a cap on advertising thus may seem less compelling: any viewer who is exposed to ads reveals by his behavior that he is not very ad averse. The point made in this paper, that a cap may improve welfare even if viewers do not directly suffer from exposure to ads, however, gets reinforced when there are ad averse viewers who use ad avoidance technologies.

To illustrate this, we consider an extension where there are two types of consumers: a mass n_1 of consumers who are intrinsically ad neutral ($\delta = 0$), and a mass $n_2 = n - n_1$ who are intrinsically ad averse and have a $\delta > 0$. Suppose that ad aversion is independent of the location of a consumer; both ad averse and ad neutral consumers are distributed uniformly on the circle. Moreover, suppose that ad avoidance technologies are freely available. Then viewers with $\delta > 0$ use ad avoidance technologies, and thus effectively *no* consumer is directly negatively affected by ads.

[39] As above, this implicitly assumes that the market share of broadcaster i is between zero and one, inverse ad demand is positive, and broadcaster i does not undercut its rivals.

[40] Note, however, that the result is driven by the assumption that all viewers have the same marginal rate of substitution between money and program quality. If viewers differ in these respects, the commercial media bias may reappear in equilibrium even in a pay TV regime, as in Ellman and Germano (2009).

Only those viewers who are intrinsically ad neutral are reached by ads, and only those play a role in the calculations of the media outlets and the advertisers. The other ones are affected, however, by the program quality chosen by the broadcasters. We can model this situation as above by setting $\delta = 0$, replacing n by n_1 in the formulas for profits and producer surplus, and adding a term $n_2 (w + v) - n_2\tau / (4N)$ to the consumer surplus to account for the consumers who use ad avoidance technologies. Thus, as compared to a situation where everyone is intrinsically ad neutral, there is an additional welfare benefit from higher program quality: the consumers using ad avoidance technologies do not figure in the broadcasters' or advertisers' decisions, but enjoy a higher program quality as well. For the welfare comparison in Proposition 4, this implies that the condition for when a local cap improves welfare (6.14) becomes less strict than when every consumer is intrinsically ad neutral.[41]

6.7 Conclusion

This paper has argued that a cap on advertising in free-to-air television (or other advertising funded media) drives up the per viewer price of advertising spots and thus induces the media to choose more viewer friendly program content. Due to this effect on non-advertising content an advertising cap can increase welfare, even when viewers are not directly ad averse or can use ad avoidance technologies. Competition between broadcasters helps overcoming commercial media bias. There is, however, a complementarity between competition and regulation: on a more competitive broadcasting market, the marginal welfare gains from a cap are higher. The paper also shows that endogenous entry into the broadcasting market can tilt the desirability of advertising caps in either way, but does not overturn the main insights from the model. Moreover, the paper compared advertising caps with taxes on advertising revenue, arguing that these two policy instruments are quite different in the present context.

We used the Salop model with linear transportation costs as our model of television viewing behavior. As we show in Appendix F.4, however, our results concerning an exogenous number of broadcasters extend to a far more general setting, which comprises other well known discrete choice models such as the Logit model. In particular, the conditions under which a local cap raises welfare are qualitatively similar, and the optimal cap has the same qualitative properties, in these alternative models of television viewing. An interesting question for further research is to generalize the analysis of entry beyond the Salop model.

Our model assumed that higher program quality reduces the willingness to pay of all advertisers by the same amount. We discuss two extensions in Appendix F that relax this assumption. First,

[41] Note that the condition does not depend on n, so replacing n by n_1 does not affect it.

in Appendix F.3, we study an extension where only some advertisers prefer low quality programs, while others are indifferent. If advertising demand from the latter type of advertisers is sufficiently high, the market solves the problem of commercial media bias. Otherwise, however, a cap may raise welfare in a larger set of circumstances, and the welfare gains from a cap may be higher, than in our main model. The reason is that, in the extension, the higher program quality induced by the cap does not decrease the profits of those advertisers who are indifferent over program quality, thus the negative effect of a cap on producer surplus is less important. In addition to that, the extended setting allows to study sector specific regulations such as, for example, a ban on tobacco advertising, and shows they can be even more beneficial.

Second, we discuss an extension where the effect of program quality on advertising demand depends on the quality of the advertised goods. Plausibly, producers of high quality goods have less to lose from high program quality. An advertising cap implies that the marginal advertiser sells a product of higher quality, and thus is less affected by an increase in program quality. We show in Appendix F.4 that this reinforces the effect of the cap on program quality: with a cap, inverse advertising demand is less sensitive to program quality, therefore broadcasters will increase quality further. We also discuss other nonlinearities in consumers' utility from watching television, and in the inverse demand for advertising spots.

Our welfare analysis is based on the view that advertising is informative, and on a rational choice model of consumer behavior. Of course, these assumptions are doubtful when purchase decisions are boundedly rational, or when advertising is suggestive or deceptive. In Appendix F.3, we study an extension of our model that takes these issues into account, and show that deceptive advertising makes the case for an advertising cap stronger.

The size and relative importance of the effects we identify is ultimately an empirical question. The model has several testable empirical implications, such as the comparative static of equilibrium advertising quantity and program quality with respect to competition on the television market, and with respect to the mass of advertisers. A particularly interesting exercise for future research would be to empirically study the effect of an advertising cap on program content. Moreover, our model considered a commercial television market. Public service broadcasters may be less susceptible to commercial media biased insofar as their funding is secured largely independent from advertising revenues. Since public service broadcasters also compete for viewers' attention, their presence may impact the program content of commercial broadcasters as well. Studying these interdependencies is an interesting topic for future research.

7 Conclusion

This thesis contributes to the understanding of media markets by studying the determinants of media outlets' content choice, by developing novel techniques to assess media bias, and by analyzing the welfare effects of potential regulations. Chapter 2 demonstrates that advertising has a causal positive effect on content differentiation on YouTube, in particular, an exogenous increase in the technically feasible advertising quantity reduces the YouTubers' probability to duplicate mainstream content. Chapters 3, 4, and 5 develop novel techniques and applications to measure political media bias. Finally, Chapter 6 shows that commercial media bias can be mitigated by a cap on advertising quantity. The results contribute to the literature on content differentiation in (digital) media markets (e.g., Anderson, 2006; Waldfogel, 2017, 2018; Goldfarb and Tucker, 2019), political media bias (e.g., Groeling, 2013; Gentzkow et al., 2016; Puglisi and Snyder, 2016), commercial media bias (e.g., Ellman and Germano, 2009; Blasco and Sobbrio, 2012; Germano and Meier, 2013), two-sided (media) markets (e.g., Anderson and Jullien, 2016), media content as a public good (e.g., Anderson and Coate, 2005; Batina and Ihori, 2005), and user-generated content (e.g., Luca, 2016b).

Moreover, the thesis points into several directions of future research. First and foremost, media consumption is moving online. As a consequence, the fixed costs of generating media content decrease, whereby entry in existing online media markets becomes easier; moreover, platforms with entirely new business models such as search engines, news aggregators, and blogs emerge. Although Chapters 2, 4, and 5 study phenomena of such newly emerged platforms, more research on different types of online platforms is necessary to draw broader conclusions on their effect on social and economic outcomes.

In addition to that, user-generated content is becoming more and more important and opens up novel research opportunities. Today, users can (anonymously) comment on products, news, and more or less random discussions on the Internet. While Chapters 2, 4, and 5 study contributions to YouTube, Wikipedia, and Twitter, the thesis does not cover anonymous comments within online discussion forums such as Reddit or online newspapers' forums, which would be an interesting direction for future research.

Relatedly, Chapters 3, 4, and 5 introduce novel techniques to asses political media bias, but the recent debate about fake news requires researchers to take one step further and study how to mitigate the emergence and spreading of fake news.

Finally, online media markets allow for new opportunities on behalf of advertisers. In particular, targeted advertising could be beneficial for consumers and advertisers, since it may reduce the search cost of consumers and increase the effectiveness of an ad. At the same time, however, researchers must better understand the role of privacy concerns. In sum, the recent developments in media markets and their importance for democracy and society, require further research on the digitization of media markets in general, and on online media platforms, user-generated content, and advertising in particular.

A Appendix to Chapter 2

A.1 Robustness checks

This section probes the robustness of my results. In particular, I show that the main results from Section 2.6.1 are robust to using an alternative observation period, to an alternative selection of YouTubers, to an alternative classification of the treatment group, and to alternative definitions of the instrument $close_i$ and of the dependent variable $Mainstream_{vit}$. In addition to that, I report the results of placebo regressions that support the plausibility of the exclusion restriction, I conduct several robustness checks on the results from Section 2.7, and I probe the validity of the empirical strategy when studying video quality.

A.1.1 Alternative observation period

First, I show that the results from Section 2.6.1 are robust to using an alternative observation period. As argued in Section 2.4.3, I cannot extend the analysis to earlier or later points in time; I can, however, select a shorter observation period. Table A.1 shows the results from estimating equations (2.2) and (2.3) on observations from Jan 2014 to July 2016 only (hence, I exclude twelve months before, and six months after Oct 2015). While the potentially biased OLS estimates in columns 1 to 3 are close to their counterparts in Table 2.3, the 2SLS estimates in columns 4 to 6 and the reduced form estimates in columns 7 to 9 are smaller by a third. This is no surprise: the event study in Section 2.6.2 illustrates that the effect of an increase in the feasible number of ad breaks on the probability to upload mainstream content becomes stronger over time. Thus, excluding the last six months from the analysis results in smaller estimates.

A.1.2 Alternative selections of YouTubers

Next, I demonstrate that the results from Section 2.6.1 are robust to alternative selections of YouTubers. As argued in Section 2.4.3, the final dataset includes only YouTubers whose median video duration before Oct 2015 is smaller than 10, because I want to focus on YouTubers who were ignorant of the ten minutes trick before the launch of the new ad break tool. One could argue, however, that the selection is too loose. For instance, if a YouTuber's median video duration

Table A.1: Alternative observation period

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.008	0.005	0.005	-0.131***	-0.130***	-0.124***			
	(0.008)	(0.008)	(0.008)	(0.048)	(0.047)	(0.046)			
$close_i * post_t$							-0.004***	-0.004***	-0.004***
							(0.001)	(0.001)	(0.001)
First stage				0.029***	0.029***	0.029***			
				(0.002)	(0.002)	(0.002)			
F-test of excluded instruments				147.47	147.48	153.11			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,513	10,513	10,513	10,513	10,513	10,513	10,513	10,513	10,513
Videos	745,219	745,219	745,219	745,219	745,219	745,219	745,219	745,219	745,219

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only the periods $t \in [13, 43]$ are included into the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

before Oct 2015 is equal to 9, a large share of her videos may already be ten minutes or longer, so she could have come across the ten minutes trick before the new ad break tool was launched. To rule out concerns about the selection of YouTubers, I estimate regression equations (2.2) and (2.3) on two subsamples: first, a subsample of YouTubers whose median video duration before Oct 2015 is smaller than 7.5, second, a subsample of YouTubers whose 90^{th} percentile of the distribution of video durations (not the median) is smaller than 10.

Tables A.2 and A.3 show the results. The potentially biased OLS estimates in columns 1 to 3 resemble their counterparts in Table 2.3. Similarly, the 2SLS estimates in columns 4 to 6 are close to the estimates based on the entire dataset. The first stage as well as the reduced form estimates (columns 7 to 9), however, are nearly twice as large as their counterparts in Table 2.3. A potential explanation is that the average YouTuber whom I consider in this section has more scope to react to the launch of the new ad break tool than the average YouTuber from the main analysis, which matches the considerations from Section 2.5.2. In sum, I do not find evidence of my main results being sensitive to alternative selections of YouTubers.

A.1.3 Alternative classifications of the treatment group

This section shows that the results from Section 2.6.1 are robust to alternative classifications of the treatment group. In particular, I show that neither the five percentage point cutoff nor considering

Table A.2: Alternative selection of YouTubers – median video duration < 7.5

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.007	0.005	0.005	-0.246***	-0.240***	-0.234***			
	(0.008)	(0.008)	(0.008)	(0.050)	(0.049)	(0.048)			
$close_i * post_t$							-0.010***	-0.010***	-0.009***
							(0.002)	(0.002)	(0.002)
First stage				0.040***	0.040***	0.040***			
				(0.003)	(0.003)	(0.003)			
F-test of excluded instruments				161.21	160.26	166.53			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	9,519	9,519	9,519	9,519	9,519	9,519	9,519	9,519	9,519
Videos	923,189	923,189	923,189	923,189	923,189	923,189	923,189	923,189	923,189

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers whose median video duration before Oct 2015 is smaller than 7.5 are included into the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.3: Alternative selection of YouTubers – 90^{th} percentile < 10 min

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.009	0.005	0.004	-0.213***	-0.234***	-0.234***			
	(0.011)	(0.010)	(0.011)	(0.057)	(0.056)	(0.057)			
$close_i * post_t$							-0.011***	-0.013***	-0.013***
							(0.003)	(0.003)	(0.003)
First stage				0.054***	0.053***	0.053***			
				(0.005)	(0.005)	(0.005)			
F-test of excluded instruments				135.13	134.03	134.40			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	6,891	6,891	6,891	6,891	6,891	6,891	6,891	6,891	6,891
Videos	610,496	610,496	610,496	610,496	610,496	610,496	610,496	610,496	610,496

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, only YouTubers whose 90^{th} percentile of the distribution of video durations before Oct 2015 is smaller than 10 are included into the analysis. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

only a YouTuber's videos between ten and fourteen minutes drive my results.

Table A.4 shows the 2SLS estimates from using two alternative cutoffs; YouTubers are classified as treated if their share of videos between ten and fourteen minutes has increased by at least one (columns 1 to 3) or by at least ten percentage points (columns 4 to 6). While the estimates in columns 1 to 3 are close to their counterparts in Table 2.3, the estimates in columns 4 to 6 are larger by a third: the probability to upload mainstream content decreases by around 35 percentage points for YouTubers in the treatment relative to the control group. The result is plausible: the average effect of an increase in the feasible number of ad breaks on the probability to upload mainstream content is stronger for YouTubers who increase their share of videos that are ten minutes or longer to a higher extent.

Next, I classify a YouTuber as treated if she increased her share of videos that are ten minutes or longer (instead of ten to fourteen minutes) by at least five percentage points. Table A.5 shows the results. The potentially biased OLS estimates in columns 1 to 3 are close to zero and not statistically significant. The 2SLS estimates in columns 4 to 6 are negative and statistically significant at the 1%-level, but smaller in absolute value than their counterparts in Table 2.3. A potential explanation is that considering *all* videos that are ten minutes or longer leads to more noise in the estimation, for instance, because videos that are more than "just" longer than ten minutes are less likely to indicate that a YouTuber exploits the ten minutes trick. Finally, the reduced form estimates in columns 7 to 9 are similar to the results from Section 2.6.1.

A.1.4 Alternative definitions of the instrument

Next, I confirm that the results from Section 2.6.1 are robust to alternative definitions of the instrument $close_i$: While it is equal to a YouTuber's *median* video duration before Oct 2015 in the main analysis, $close_i$ corresponds to the 75^{th} and to the 90^{th} percentile of the distribution of her video durations here. These two alternative definitions of $close_i$ may better capture a YouTuber's "closeness" to the ten minutes threshold before Oct 2015.

Table A.6 shows the results from a 2SLS estimation of equations (2.2) and (2.3) using a YouTuber's 75^{th} percentile (columns 1 to 3), and using a YouTuber's 90^{th} percentile of the distribution of video durations before Oct 2015 (columns 4 to 6) as an instrument for D_i. The estimates in columns 1 to 3 are negative, but smaller in absolute value than their counterparts in Table 2.3; they are also less statistically significant. The estimates in columns 4 to 6, in contrast, are larger than their counterparts in Table 2.3. The first stage estimates and the first F-statistics, however, are in both cases much smaller than in Table 2.3. Hence, the 75^{th} and the 90^{th} percentiles of the distribution of a YouTuber's video durations before Oct 2015 have less power to predict a YouTuber's treatment status D_i than the median.

Table A.4: Alternative classifications of the treatment group – cutoffs

	1%	1%	1%	10%	10%	10%
	(1)	(2)	(3)	(4)	(5)	(6)
$D_i * post_t$	-0.245***	-0.234***	-0.232***	-0.363***	-0.347***	-0.329***
	(0.062)	(0.060)	(0.060)	(0.090)	(0.087)	(0.084)
First stage	0.024 ***	0.024***	0.024***	0.016***	0.016***	0.017***
	(0.003)	(0.002)	(0.003)	(0.001)	(0.002)	(0.002)
F-test of excluded instruments	73.31	73.33	71.54	92.66	92.89	100.79
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. All estimates are 2SLS estimates. In columns 1 to 3, YouTubers who have increased their share of videos between ten and fourteen minutes by at least 1 percentage point after Oct 2015 are classified as treated. Analogously, in columns 4 to 6, YouTubers who have increased their share of videos between ten and fourteen minutes by at least 10 percentage point after Oct 2015 are classified as treated. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Table A.5: Alternative classifications of the treatment group – all videos $\geq$ 10 minutes

	OLS (1)	OLS (2)	OLS (3)	2SLS (4)	2SLS (5)	2SLS (6)	Red. F. (7)	Red. F. (8)	Red. F. (9)
$D_i * post_t$	0.005	-0.001	-0.0008	-0.151***	-0.145***	-0.141***			
	(0.007)	(0.007)	(0.007)	(0.036)	(0.035)	(0.035)			
$close_i * post_t$							-0.006***	-0.006***	-0.006***
							(0.001)	(0.001)	(0.001)
First stage				0.039***	0.039***	0.039***			
				(0.003)	(0.003)	(0.003)			
F-test of excluded instruments				207.91	208.62	210.86			
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. Moreover, YouTubers who have increased their share of videos that are ten minutes or longer by at least five percentage points are classified as treated. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

A.1.5 Alternative definitions of mainstream content

Here, I show that the results from Section 2.6.1 are robust to alternative definitions of mainstream content. To this end, I generate four alternative measures. First, I assign a dummy equal to one to all videos that are given a keyword from the upper half percent, second, I assign a dummy equal to one to all videos that are given a keyword from the upper two percent of the distribution of most-viewed keywords (see Section 2.4.2). Third, instead of using a share, I classify a fixed number of keywords per month per category as mainstream – 250 keywords for the categories "Entertainment", "People & Blogs", and "Let's Play", where I have the most observations, and 100 keywords for the remaining categories – and assign a dummy equal to one to all videos given mainstream a keyword such defined. Finally, instead of considering the views, for each month, for each category, I compute how many *Likes* a certain keyword has attracted and rank them in descending order; the upper one percent of this distribution is then classified as mainstream and all videos given such a keyword are assigned a dummy equal to one. Table A.7 provides an overview of how these measures are correlated.

Table A.8 shows the results from a 2SLS estimation of equations (2.2) and (2.3) using the four alternative definitions of $Mainstream_{vit}$. In columns 1 to 3, the estimates for β are negative, but much smaller in absolute value than their counterparts in Table 2.3 and not statistically significant. The estimates in columns 4 to 6, in contrast, are larger by a third than the estimates in Table 2.3 and

Table A.6: Alternative definitions of the instrument

	75^{th} perc. (1)	75^{th} perc. (2)	75^{th} perc. (3)	90^{th} perc. (4)	90^{th} perc. (5)	90^{th} perc. (6)
$D_i * post_t$	-0.131*	-0.148**	-0.140*	-0.377**	-0.425***	-0.413***
	(0.074)	(0.074)	(0.073)	(0.148)	(0.161)	(0.157)
First stage	0.007***	0.007***	0.007***	0.002***	0.002***	0.002***
	(0.001)	(0.001)	(0.001)	(0.0006)	(0.0006)	(0.0006)
F-test of excluded instruments	24.71	24.58	24.83	14.68	14.34	14.71
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls		X	X		X	X
YouTuber Time Trend			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. The dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. The estimates are based on using the advertising YouTubers only. All estimates are 2SLS estimates. In columns 1 to 3, the instrument $close_i$ is defined as the 75^{th} percentile in the distribution of video durations of YouTuber i before Oct 2015. In columns 4 to 6, the instrument $close_i$ is defined as the 90^{th} percentile in the distribution of video durations of YouTuber i before Oct 2015. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

statistically significant at the 1%-level. When I use a fixed number of keywords per category per month to define mainstream content (columns 7 to 9), the estimates are negative and statistically significant at the 1%-level, but around a fourth smaller than in Table 2.3. Finally, when I consider the keywords' number of Likes instead of their views in columns 10 to 12, the estimates are close to their counterparts in Table 2.3.

Table A.7: Correlation measures mainstream content

	1%	0.5%	2%	Fixed	Likes
1%	1.0000				
0.5%	0.8237	1.0000			
2%	0.8207	0.6761	1.0000		
Fixed	0.8495	0.8479	0.7528	1.0000	
Likes	0.8298	0.7688	0.7802	0.7598	1.0000

Notes: Correlation matrix for the different measures of mainstream content.

A.1.6 Placebo regressions

In this section, I conduct a series of placebo regressions to support the plausibility of the exclusion restriction as discussed in Section 2.5.2. To this end, I augment the reduced form equation (2.4) to

$$Mainstream_{vit} = \gamma^{Placebo} close_i * fakepost_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit} |\ t \leq 33, \quad \text{(A.1)}$$

where in the first placebo regression, $fakepost_t$ is equal to one if $t \geq 3$, in the second placebo regression $fakepost_t$ is equal to one if $t \geq 4$, and so on; I run 29 placebo regressions in sum. If $close_i$ has no direct effect on $Mainstream_{vit}$, all estimates for $\gamma^{Placebo}$ should be close to zero and not statistically significant. The idea is similar to the event study in Section 2.5.2: YouTubers with different values of $close_i$ must not have been on different trends in terms of $Mainstream_{vit}$ before Oct 2015.

Of 29 placebo regressions, the estimate for $\gamma^{Placebo}$ is in three cases statistically significant at the 5%-level; these estimates are, however, positive. Thus, the results provide additional support for the plausibility of the exclusion restriction.

A.1.7 Robustness checks: mechanism

Next, I conduct several robustness checks on the results from Section 2.7. Since I know from Section 2.7.1 that the dependent variables $Mainstream_{vit}$ and $Competitive_{vit}$ are highly correlated, I do not repeat all the analyses from above, though. Instead, I focus on the robustness checks on the

Table A.8: Alternative definitions of mainstream content

	0.5% (1)	0.5% (2)	0.5% (3)	2% (4)	2% (5)	2% (6)	Fixed (7)	Fixed (8)	Fixed (9)	Likes (10)	Likes (11)	Likes (12)
$D_i * post_t$	-0.029	-0.0198	-0.0198	-0.361 ***	-0.355***	-0.343***	-0.142***	-0.143***	-0.133***	-0.252***	-0.231***	-0.232***
	(0.0464)	(0.0451)	(0.0442)	(0.0555)	(0.0545)	(0.0533)	(0.0477)	(0.0469)	(0.0459)	(0.0519)	(0.0503)	(0.0496)
First stage	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	144.13	143.85	151.32	144.13	143.85	151.32	144.13	143.85	151.32	144.13	143.85	151.32
Time FE	X	X	X	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X		X	X
YouTuber Time Trend			X			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. In columns 1 to 3, the dependent variable $Mainstream_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper half percent of the distribution of most-viewed keywords. In columns 4 to 6, the dependent variable $Mainstream_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper two percent of the distribution of most-viewed keywords. In columns 7 to 9, the dependent variable $Mainstream_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from a fixed number of the distribution of most-viewed keywords. In columns 10 to 12, the dependent variable $Mainstream_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper one percent of the distribution of most-liked keywords. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

dependent variable as such, i.e., I study alternative definitions of competitive content and I run a series of placebo regressions to support the plausibility of the exclusion restriction from Section 2.7.2. In addition, I provide robustness checks on the commentator analysis in Section 2.7.4.

Alternative definitions of competitive content

Analogous to Appendix A.1.5, I show that the results from Section 2.7.3 are robust to alternative definitions of competitive content. Here, I generate three alternative measures. First, I assign a dummy equal to one to all videos that are given a keyword from the upper half percent, second, I assign a dummy equal to one to all videos that are given a keyword from the upper two percent of the distribution of most-*used* keywords (see Section 2.7.1). Third, instead of using a share, I classify a fixed number of keywords per month per category as competitive – 250 keywords for the categories "Entertainment", "People & Blogs", and "Let's Play", where I have the most observations, and 100 keywords for the remaining categories – and assign a dummy equal to one to all videos given competitive a keyword such defined. Table A.9 provides an overview of how these measures are correlated.

Table A.10 shows the results from a 2SLS regression of equations (2.2) and (2.3) using the three alternative definitions of $Competitive_{vit}$. All estimates are negative and statistically significant at the 1%-level. In columns 1 to 3, the estimates for β are similar to their counterparts in Table 2.3; the estimates in columns 4 to 9, in contrast, are a fourth to a fifth smaller in absolute value.

Table A.9: Correlation measures competitive content

	1%	0.5%	2%	Fixed
1%	1.0000			
0.5%	0.8424	1.0000		
2%	0.8415	0.7093	1.0000	
Fixed	0.8602	0.8552	0.7819	1.0000

Notes: Correlation matrix for the different measures of competitive content.

Placebo regressions

Analogous to Appendix A.1.6, I conduct a series of placebo regressions to support the plausibility of the exclusion restriction as discussed in Section 2.7.3. To this end, I augment the reduced form equation (2.9) to

$$Competitive_{vit} = \gamma^{Placebo'} close_i * fakepost_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit} | \, t \leq 33, \quad \text{(A.2)}$$

Table A.10: Alternative definitions of competitive content

	0.5% (1)	0.5% (2)	0.5% (3)	2% (4)	2% (5)	2% (6)	Fixed (7)	Fixed (8)	Fixed (9)
$D_i * post_t$	-0.232***	-0.231***	-0.208***	-0.178***	-0.176***	-0.143***	-0.176***	-0.183***	-0.154***
	(0.0519)	(0.0514)	(0.0494)	(0.0436)	(0.0434)	(0.0416)	(0.0483)	(0.0483)	(0.0465)
First stage	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***	0.0286***	0.0286***	0.0290***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	144.13	143.85	151.32	144.13	143.85	151.32	144.13	143.85	151.32
Time FE	X	X	X	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X	X	X	X
Controls		X	X		X	X		X	X
YouTuber Time Trend			X			X			X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. In columns 1 to 3, the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper half percent of the distribution of most-used keywords. In columns 4 to 6, the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from the upper two percent of the distribution of most-used keywords. In columns 7 to 9, the dependent variable $Competitive_{vit}$ is equal to one if video v of YouTuber i in month t is given a keyword from a fixed number of the distribution of most-used keywords. All estimates are based on 2SLS regressions. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

where in the first placebo regression, $fakepost_t$ is equal to one if $t \geq 3$, in the second placebo regression $fakepost_t$ is equal to one if $t \geq 4$, and so on; as above, I run 29 placebo regressions in sum. Of 29 placebo regressions, the estimate for $\gamma^{Placebo'}$ is in four cases statistically significant at the 5%-level; these estimates are, however, positive. Thus, the results provide additional support for the plausibility of the exclusion restriction.

Alternative cutoffs in the commentator analysis

Here, I show that the results from Section 2.7.4 are robust to alternative comment cutoffs. To this end, I restrict the analysis to all advertising YouTubers who received (i) more than one hundred, (ii) more than fifty, (iii) more than ten, and (iv) at least one comment before and after Oct 2015.

The results in Table A.11 confirm that the measure $fluctuation_i$ may lead to unreasonable results when the total number of comments is small. When I restrict the analysis to YouTubers with at least one hundred comments (columns 1 to 3) or to YouTubers with at least 50 comments before and after Oct 2015 columns (4 to 6), the potentially biased OLS estimate is smaller, while the 2SLS and the reduced form estimates are larger than their counterparts in Table 2.14. In contrast to that, when I restrict the analysis to YouTubers with at least ten comments before and after Oct 2015 (columns 7 to 9), the OLS is estimate larger, and the 2SLS and the reduced form estimates are smaller than in the main part and not statistically significant. Finally, when I consider all YouTubers who have received at least one comment (columns 10 to 12), the 2SLS and the reduced form estimate even switch their sign and become positive, but are not statistically significant.

A.1.8 Validity checks: Quality

Finally, I check if the empirical strategy from Section 2.5 is valid when I use $\frac{Likes}{Likes+Dislikes}_{vit}$ and $log(Views)_{vit}$ as dependent variables.

Exclusion restriction

To confirm the plausibility of the exclusion restriction, I conduct two further event studies. In particular, I estimate the augmented reduced form regression equations

$$\frac{Likes}{Likes + Dislikes}_{vit} = \sum_{t=1}^{33} \gamma_t'' close_i * pre_t + \sum_{t=35}^{49} \gamma_t'' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit}. \tag{A.3}$$

and

$$log(Views)_{vit} = \sum_{t=1}^{33} \gamma_t''' close_i * pre_t + \sum_{t=35}^{49} \gamma_t''' close_i * post_t + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + v_{vit}. \tag{A.4}$$

by OLS. If $close_i$ has no impact on the dependent variables, then all estimates for γ_t'' and γ_t''', $t \in [1, 33]$, should be close to zero without being statistically significant.

Figures A.1 and A.2 show the results. The estimates for γ_t'' and γ_t''', $t \in [1, 33]$ are not statistically significant and fluctuate around zero, which supports the plausibility of the exclusion restriction. Yet, the lion's share of the estimates is not statistically significant at the 5%-level after Oct 2015, either.

Table A.11: Alternative cutoffs for the commentator analysis

	>100 OLS (1)	>100 2SLS (2)	>100 Red. Form (3)	>50 OLS (4)	>50 2SLS (5)	>50 Red. Form (6)	>10 OLS (7)	>10 2SLS (8)	>10 Red. Form (9)	>0 OLS (10)	>0 2SLS (11)	>0 Red. Form (12)
D_i	-0.009**	-0.066***		-0.01**	-0.050**		-0.017***	-0.035		-0.035***	0.038	
	(0.004)	(0.024)		(0.004)	(0.022)		(0.004)	(0.025)		(0.005)	(0.031)	
$close_i$			-0.002***			-0.002**			-0.0010			0.001
			(0.0007)			(0.0007)			(0.0459)			(0.0008)
First stage		0.031***			0.031***			0.028***			0.028***	
		(0.0028)			(0.0025)			(0.0021)			(0.0019)	
F-test of excluded instruments		120.47			150.96			176.90			221.06	
YouTubers	3,989	3,989	3,989	4,924	4,924	4,924	7,098	7,098	7,098	8,916	8,916	8,916

Notes: Robust standard errors in parentheses. The dependent variable is $\Delta fluctuation_i$, which is the difference in the commentator fluctuation before and after Oct 2015 for YouTuber i. The estimates are based on the advertising YouTubers only. In columns 1 to 3, only YouTubers who received more than 100 comments before and after Oct 2015 are included in the analysis. In columns 4 to 6, only YouTubers who received more than 50 comments before and after Oct 2015 are included in the analysis. In columns 7 to 9, only YouTubers who received more than 10 comments before or after Oct 2015 are included in the analysis. In columns 10 to 12, all YouTubers who received at least one comment before and after Oct 2015 are included in the analysis. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$ * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

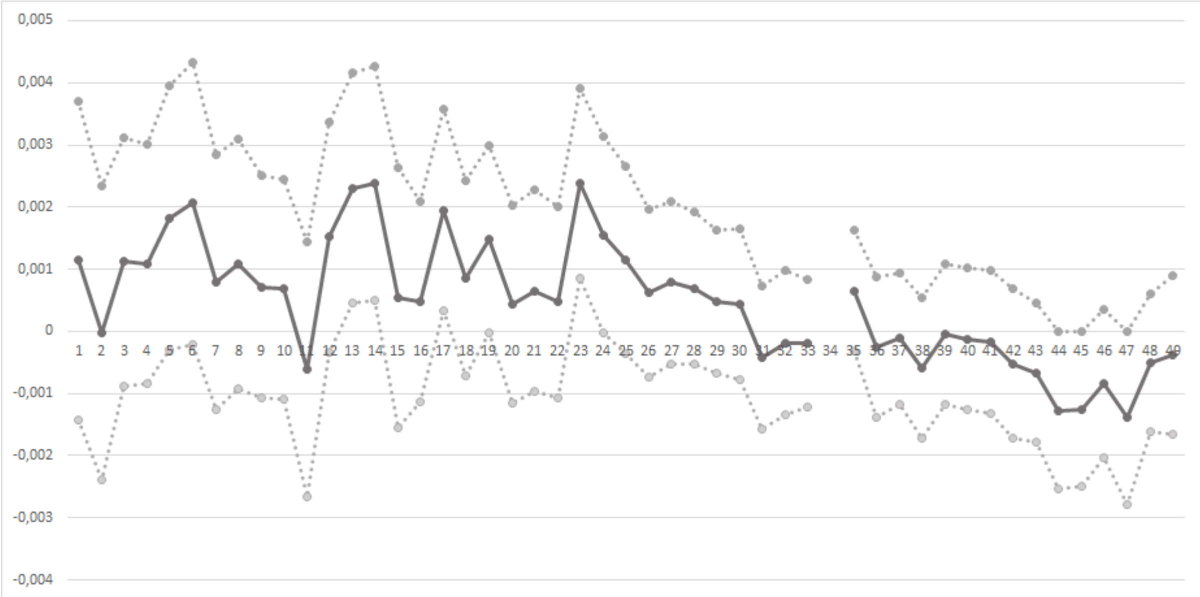

Figure A.1: Event study likes/(likes+dislikes) (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t''$ from equation (A.3). The dashed line depicts a 95%-confidence interval.

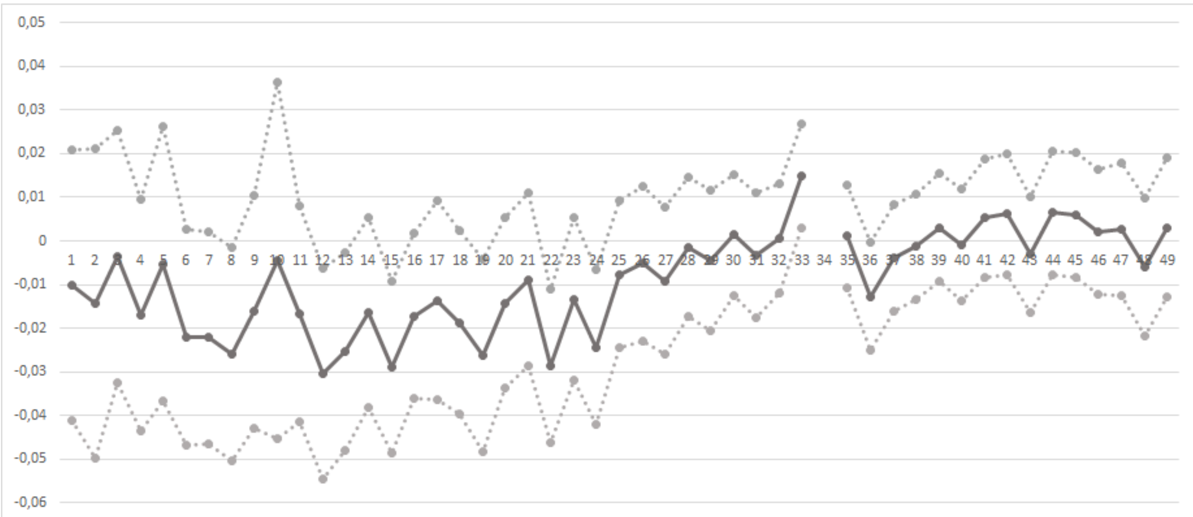

Figure A.2: Event study log(views) (advertising YouTubers). The dark grey dots and the solid line represent the estimates $\hat{\gamma}_t'''$ from equation (A.4). The dashed line depicts a 95%-confidence interval.

Video duration, likes, and views

Next, I check if video duration as such affects the dependent variables $\frac{Likes}{Likes+Dislikes}_{vit}$ and $log(Views)_{vit}$ (see Section 2.5.2). To this end, I estimate

$$\frac{Likes}{Likes + Dislikes}_{vit} = \delta'' duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \tag{A.5}$$

and

$$log(Views)_{vit} = \delta''' duration_{vit} + \theta X_{vit} + \phi_i + \phi_t + \tau t_i + \epsilon_{vit} \mid t \leq 33 \tag{A.6}$$

by OLS.

Table 2.5 shows the results. The size of the estimates for δ'' (columns 1 to 3), though statistically significant at the 1%-level, is negligible: a one second increase in video duration corresponds to a 0.0001 percentage point increase in the fraction of likes. The estimates for δ''' in columns 4 to 6, though, are relatively large and statistically significant at the 1%-level, too. According to these estimates, one further second in video duration leads on average to about 1.5 percent more views. These estimates may reflect the algorithmic drift discussed in Section 2.9.2. YouTube wants to keep its viewers as long as possible on the platform to show as many ads as possible to them. As a result, longer videos get higher rankings and are watched more often.

A.2 Further discussions

This section revisits a number of topics that could not be covered in the main part of the paper. In particular, I discuss the consequences of misclassifying advertising and non-advertising YouTubers, I show that no YouTube platform event beyond the launch of the new ad break tool affects my results, and I discard a YouTuber learning effect as a potential economic mechanism behind content differentiation.

A.2.1 Misclassification of advertising and non-advertising YouTubers

As explained in Section 2.4.1, I cannot retrieve data on the YouTubers' monetization settings on the video level. Instead, I pick twenty randomly drawn videos per YouTuber, and classify her as advertising YouTuber if I detect at least one ad break. In this section, I amplify how measurement errors during this procedure could affect my results. In addition, I discuss the consequences of sample migration between advertising and non-advertising YouTubers.

Potential consequences of measurement error

In this section, I illustrate that a potential measurement error would have only minor consequences. First, note that I could erroneously classify an advertising YouTuber as non-advertising, but not vice versa: if a YouTuber never permits for ad breaks, my algorithm cannot classify her as "advertising" by definition. Second, note that I do not use the classification dummy in a regression framework; hence, the regression results do not suffer from an errors-in-variables bias (e.g., Durbin, 1954). Yet, I *split* my sample into advertising and non-advertising YouTubers. Thus, misclassifying some advertising as non-advertising YouTubers might lead to selection bias in the subsamples.

If I misclassified some advertising as non-advertising YouTubers, the estimates in Table 2.3 may be too large. YouTubers who fall through the grid of the algorithm seldom permit for ad breaks and do not follow strict commercial incentives. Thus, they are on average more reluctant to adapt their content after Oct 2015 than the average YouTuber whom the algorithm detects. On the other hand, the YouTubers whom I missed might not even increase their share of videos between ten minutes and fourteen minutes. Thus, they are not affected by the instrument $close_i$ and their first stage is equal to zero. In this case, the LATE (see Section 2.5.2) was the same whether or not I classified some advertising as non-advertising YouTubers.

If some advertising YouTubers were included into the subsample of non-advertising YouTubers, the estimates in Table 2.6 may be too large, too. This would, however, strengthen my results: Section 2.6.2 demonstrates that there is no effect of an increase in the feasible number of ad breaks on the non-advertising YouTubers' content choice; if the estimates were even closer to zero, the validity check would be even more convincing.

Potential consequences of sample migration

An advertising YouTuber may have been non-advertising in the past and vice versa. Potential sample migration between advertising and non-advertising YouTubers, however, is unproblematic for three reasons. First, I do not directly compare advertising to non-advertising YouTubers. Second, many advertising YouTubers may have started as non-advertising YouTubers in the beginning of their career. If they became advertising YouTubers as a result of the treatment, they may have adapted their content with a delay, which may lead to an underestimation of the effect of advertising on content differentiation. Finally, if former advertising YouTubers have migrated to the subsample of non-advertising YouTubers, I might overestimate the main effect , which would – as argued in the previous subsection – make the validity check more convincing.

A.2.2 Platform events during the observation period

Next, I provide a systematic review of all platform "events" during my observation period, i.e., technical novelties or changes in YouTube's monetization policy beyond the launch of the new ad break tool. Note that an event can only affect my results if it is correlated to a YouTuber's probability to upload mainstream content *and* to her value of $close_i$ – no such event exists during the observation period. Since YouTube has no serious competitors, I remain agnostic about events at competing video sharing platforms.

Data collection

I collect information on all events from the YouTube Creators Blog, which announces YouTube news, introduces technical features, and gives general advice to YouTubers.[1] In a first step, I retrieve all blog posts from Jan 2013 to Jan 2017. Next, I manually exclude any post that does not deal with a platform event, such as YouTube promotion for academies, awards, (real world) events, and YouTuber portraits. The remaining 42 posts are listed in Table A.12. In a last step, I review all posts from Table A.12 and indicate if a YouTuber's monetization options or her probability to upload mainstream content could be affected. Thirteen events require further investigation; I discuss them chronologically.

Platform events in 2013

First, in March, YouTubers' access to their financial data changed. This event applies to all YouTubers equivalently, has no effect on their content choice, and is therefore unproblematic.

In May, selected YouTubers from the U.S., and in October, selected YouTubers worldwide were given the option to raise a subscription fee of 0.99$ per month. The pilot was, however, extremely limited: not even 100 YouTubers worldwide participated.[2] Thus, my results are unlikely to be affected by these events.

Next, YouTube launched its "Fan Finder": a YouTuber could let the platform turn one of her videos into an "ad" and show it to viewers of a different channel in place of a conventional ad; this was supposed to enlarge a YouTuber's fan base. Since YouTubers were asked to produce special videos that advertise their channel, the event may have affected their content choice. Yet, all YouTubers with at least $1,000$ subscribers could participate and there were no restrictions

[1] See `youtube-creators.googleblog.com/` (May 2019).

[2] E.g., `www.fastcompany.com/3020553/the-most-popular-youtube-channels-might-start-charging-you-to-watch`, `www.bbc.com/news/business-22474715`, or `searchenginewatch.com/sew/news/2267170/youtube-launches-paid-channels-subscription-fees-start-at-usd099-per-month` (May 2019).

Table A.12: YouTube platform events

	Date		Summary of the event	Moneti-zation	Content choice
1	2013	Jan	The channel view count only includes views from publicly available videos from now.		
2	2013	Feb	It is now technically feasible to update several video updates at the same time.		
3	2013	Mar	YouTube changes the interaction with AdSense: a YouTuber's financial overview is now available at YouTube Analytics	X	
4	2013	Mar	The new channel design "YouTube One" is available for all YouTubers.		
5	2013	Apr	Users see more videos in their homepage feed.		
6	2013	May	YouTubers receive an e-mail once a video upload has finished.		
7	2013	May	The new channel design "YouTube One" is mandatory for all YouTubers.		
8	2013	May	Selected YouTubers from the US may raise a subscription fee of 0.99$ per month.	X	X
9	2013	June	Mobile users (Android and iOS) may follow links embedded into videos from now.		
10	2013	July	YouTubers may now connect multiple channels via a Google+ page.		
11	2013	Aug	Improved mobile features for users.		
12	2013	Sept	Launch of the YouTube Audio Library (150 royalty-free tracks).		
13	2013	Sept	Improved tools for moderating comments.		
14	2013	Sept	New tools to identify and interact with one's top viewers.		
15	2013	Sept	YouTubers may now feature playlists from other channels.		
16	2013	Oct	Selected YouTubers from outside the US may also raise a subscription fee of 0.99$ per month.	X	X
17	2013	Nov	A YouTuber may let the platform turn her video into an ad that is then shown to viewers from different channels.		X
18	2013	Dec	Live streams are now technically feasible.		X
19	2014	Feb	YouTube validates a video's view count repeatedly from now on.		
20	2014	Feb	Users can create their own playlists.		
21	2014	Apr	Enhanced playlist tools in YouTube Analytics are launched.		
22	2014	June	New messaging and commenting features for YouTubers.		
23	2014	June	YouTube removes blocked users from a channel's subscriber count.		
24	2014	Nov	New YouTube homepage for music videos.		
25	2015	Mar	360 degree videos are now technically feasible.		X
26	2015	May	60fps for live streams is now technically feasible.		
27	2015	June	New data tool Music Insights is available: shows the cities where an artist is most popular, top tracks by artist, and views from both artists' official music videos and fan uploads claimed using Content ID.		
28	2015	July	A new design for YouTube mobile app is launched.		
29	2015	Oct	YouTube Red is launched in the US.	X	X
30	2015	Nov	New language and translation tools are available.		
31	2015	Nov	New virtual reality tools are available.		X
32	2016	Jan	Users can donate to the YouTuber after watching a video.	X	X
33	2016	Feb	A new blurring tool (to blur faces etc.) is available.		
34	2016	Apr	YouTube withholds any ad revenue generated during content ID disputes from now.	X	
35	2016	June	Mobile live streams are now technically feasible.		X
36	2016	Sept	YouTube Analytics becomes easier to understand for YouTubers.		
37	2016	Sept	New tools for YouTubers to engage with their community.		
38	2016	Oct	An optional feature for paid promotion disclosure is available.	X	X
39	2016	Oct	Special video end screens are available.		X
40	2016	Nov	New comment features are available for users.		
41	2016	Dec	Launch of a new URL system that is independent from Google+.		
42	2017	Jan	User messages in a chat stream may be highlighted.		

Notes: Summary of YouTube platform events.

on the advertising video's duration. Hence, the event is not correlated to $close_i$ and thereby unproblematic.

Finally, live streams became technically feasible in December and may have influenced YouTuber's content choice. The feature is open to all YouTubers, though. Hence, the event is not correlated to $close_i$ and cannot affect my results.

Platform events in 2015

In March, 360 degree videos became technically feasible. Similar to the live streams, the event may have influenced YouTubers' content choice, but since it is open to all YouTubers, there is no correlation to $close_i$.

YouTube Red, a paid subscription service that provides advertising-free streaming of all videos and exclusive original content was launched in October. The availability of YouTube Red is, however, limited to the US. Since my dataset includes only German YouTube channels, the event cannot affect my results.

In November, several virtual reality tools became available. Again, YouTubers' content choice may have been affected, but since the features are open to all YouTubers, there is no correlation to $close_i$.

Platform events in 2016

In January, YouTube launched a "Donate Button": users who click on the button can donate to a YouTuber after watching her video. As with the technical novelties from above, this may have influenced YouTubers' content choice. In addition, their monetization options were affected. Still, the feature is open to all YouTubers and thereby not correlated to $close_i$.

Next, in April, YouTube announced that it would withhold (not block) all ad revenue generated during copyright disputes. This event applies to all YouTubers equivalently, has no effect on their content choice, and is therefore unproblematic.

Mobile live streams became technically feasible in June, i.e., YouTubers could stream from their mobile devices. Similar to the "stationary" live streams from 2013, the event is not correlated to $close_i$ and cannot affect my results.

In October, YouTube launched an optional feature for paid promotion disclosure: by checking the "video contains paid promotion" box in their settings, YouTubers can inform their audience about paid product placement and endorsements by third parties. This may influence their videos' content, but is unrelated to $close_i$.

Finally, in October, video end screens, that allow YouTubers to promote up to four different videos or playlists, became technically available. Although the event may have affected the

YouTubers' content choice, the feature is open to all YouTubers, thereby not correlated to $close_i$, and hence unproblematic.

A.2.3 YouTuber learning effect

Here, I discuss a YouTuber learning effect as an alternative explanation for the results from Section 2.6: YouTubers copy the most mainstream content in the beginning of their career, but deviate from the mainstream when they become more experienced and start to develop a personal style. If such a learning effect was positively correlated with $close_i$, it could be the driving force behind the decrease in the probability to upload mainstream content after Oct 2015 rather than an increase in the feasible number of ad breaks per video.

Three arguments, however, speak against a YouTuber learning effect. First, there exists no plausible reason why YouTubers with a high value of $close_i$ would experience a stronger learning effect than YouTubers whose value of $close_i$ is low. See Section 2.5.2 for a detailed discussion on the independence of $close_i$.

Second, t_i controls for a YouTuber's average change in the probability to upload mainstream (or competitive) content over time. Columns 1 and 4 in Table A.13 replicate the 2SLS results from Tables 2.3 and 2.11 and illustrate that a linear YouTuber learning effect is of minor importance. On the one hand, the estimates for β and β' are nearly unaffected when I control for t_i. On the other hand, the estimates for t_i, though negative, are extremely small. A YouTuber's probability to upload mainstream content decreases by 0.00008 percentage points for each additional video; similarly, her probability to upload competitive content decreases by 0.0003 percentage points for each additional video.

Third, allowing for a more flexible YouTuber specific time trend by adding t_i^2 and t_i^3 does not affect the estimates for β and β', either (columns 2, 3, 5, and 6 of Table A.13). It becomes, however, obvious that the YouTuber specific time trend is not linear. For instance, columns 2 and 5 illustrate that a YouTuber's probability to upload mainstream or competitive content increases in the beginning, but decreases from around her 160^{th} video, which is consistent with the story from above. Note that the average number of videos per YouTuber is 99.3 and the median number of videos is 64. Thus, many YouTubers in my sample do not reach the turning point of 160. In sum, even though I find some evidence for a YouTuber learning effect, it is not the driving force behind the main results from Section 2.6.

Table A.13: Learning

	Mainstream (1)	Mainstream (2)	Mainstream (3)	Competitive (4)	Competitive (5)	Competitive (6)
$D_i * post_t$	-0.192***	-0.201***	-0.196***	-0.179***	-0.188***	-0.185***
	(0.0480)	(0.0485)	(0.0482)	(0.0456)	(0.0460)	(0.0458)
t_i	-0.00008	0.0004***	0.0007***	-0.0003***	0.0002*	0.0005***
	(0.00005)	(0.0001)	(0.0001)	(0.00004)	(0.0001)	(0.0001)
t_i^2		-1.13e-06***	-3.62e-06***		-1.24e-06***	-3.36e-06***
		(2.54e-07)	(8.93e-07)		(2.52e-07)	(9.02e-07)
t_i^3			4.50e-09***			3.82e-09**
			(1.69e-09)			(1.71e-09)
First stage	0.0290***	0.0288***	0.0289***	0.0290***	0.0288***	0.0289***
	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)	(0.0024)
F-test of excluded instruments	151.32	148.55	150.05	151.32	148.55	150.05
Time FE	X	X	X	X	X	X
YouTuber FE	X	X	X	X	X	X
Controls	X	X	X	X	X	X
YouTubers	10,599	10,599	10,599	10,599	10,599	10,599
Videos	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542	1,067,542

Notes: Robust standard errors in parentheses. In columns 1 to 3, the dependent variable is $Mainstream_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a mainstream keyword, and 0 otherwise. In columns 4 to 6, the dependent variable is $Competitive_{vit}$ which is a dummy variable equal to 1 if video v of YouTuber i uploaded in month t is equipped with a competitive keyword, and 0 otherwise. All estimates are obtained by 2SLS and based on using the advertising YouTubers only. Standard errors are clustered on the YouTuber level. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

B Appendix to Chapter 3

In the following we provide translations of the letter that favors the challenger (top) and that gives a negative perspective on the challenger (bottom). The letters on the chancellor are analogous, with the only difference that "Schulz" is replaced by "Merkel."

> Dear Sir or Madam,
>
> I herewith refer to your news coverage of the local Bundestag election campaign and of Sunday's TV debate.
>
> Let me be forthcoming in saying that I think Schulz is definitely the better candidate for our country and for the region where we live. For precisely this reason I would urgently ask you to stop pretending between the lines as if the election result had already been determined. What if your editorials were to demobilize important voters!
>
> The region where we live is facing immense challenges: an ailing infrastructure, a lack of day-care centers, the integration of refugees. Under a Schulz administration, I as a young mother could be more confident of the future for my children and my home. Hence: Fairplay in news coverage for a high voter turnout!
>
> Yours sincerely,
>
> Annamarie Richter
>
> PS: As we are on vacation from Friday on, I would kindly ask you to inform me via e-mail whether and when you publish the letter to the editor (it is rather difficult to reach me via mobile phone). Thank you!

> Dear Sir or Madam,
>
> I herewith refer to your news coverage of the local Bundestag election campaign and of Sunday's TV debate.
>
> Let me be forthcoming in saying that I think Schulz is definitely the worse candidate for our country and for the region where we live. For precisely this reason I would urgently ask you to stop pretending between the lines as if the election result had already been determined. What if your editorials were to demobilize important voters!
>
> The region where we live is facing immense challenges: an ailing infrastructure, a lack of day-care centers, the integration of refugees. Under a Schulz administration, I as a young mother would have look with great worry into the future of my children and my home. Hence: Fairplay in news coverage for a high voter turnout!
>
> Yours sincerely,
>
> Annamarie Richter
>
> PS: As we are on vacation from Friday on, I would kindly ask you to inform me via e-mail whether and when you publish the letter to the editor (it is rather difficult to reach me via mobile phone). Thank you!

C Appendix to Chapter 4

C.1 Classroom survey

To support our argument that voters perceive an extensive Wikipedia biography as a positive signal, we conducted a classroom survey among sixty undergraduate students of economics.[1] The goal was to study if students rate unknown MPs with a longer Wikipedia biography better in terms of their valence characteristics, i.e., qualities of a politician on which all voters agree (Stokes, 1963). We used nine valence characteristics that are frequently discussed in the political science literature (e.g., Kinder et al., 1980; Funk, 1999; Stone and Simas, 2010).

We randomized the students into two groups. Students in group 1 received the instruction: "Consider a politician from your preferred party. The Wikipedia biography of this politician (politician A) is three pages long. Consider another politician from the same party. The Wikipedia biography of this politician is one page long. Please answer the following questions." Students in group 2 received the same instructions, only that the biography of politician A was one, and of politician B was three pages long. Next, we asked which politician would probably score better with respect to each of the nine valence characteristics. The students could either reply "Politician A", "Politician B", or "Don't know." We considered the students' replies if they answered all nine questions.

Table C.1 shows the results. Column 1 displays the valence characteristics that we consider. Columns 2 to 4 show the shares of students who opted for the politician with the three-page biography (s_3), the politician with the one-page biography (s_1), and "Don't know", respectively. Column 5 displays the difference between s_3 and s_1, which is positive for all valence characteristics except for intelligence and honesty. We test the statistical significance of this difference against the null hypothesis that $s_3 = s_1$ and find that the difference between s_3 and s_1 is statistically significant for knowledge, strength as a public servant, and inspiring ($p < 0.01$).

[1] The survey was carried out via classEx, a free software for interactive classroom experiments (Giamattei and Lambsdorff, 2019).

Table C.1: Survey

	Three-page biography	One-page biography	Don't know	Difference (1)-(2)
Knowledge	0.417	0.1	0.483	0.316*** (0.0833)
Intelligence	0.117	0.183	0.7	-0.07 (0.070)
Provides strong leadership	0.283	0.2	0.517	0.083 (0.089)
Sets a good moral example	0.217	0.167	0.617	0.05 (0.080)
Strength as public servant	0.367	0.133	0.5	0.233*** (0.086)
Empathy	0.25	0.117	0.633	0.133* (0.076)
Inspiring	0.4	0.117	0.483	0.283*** (0.085)
Decency	0.167	0.15	0.683	0.017 (0.072)
Honesty	0.083	0.183	0.73	-0.1 (0.065)

Notes: This table summarizes the results of our survey. Column (1) shows the nine valence characteristics under consideration. Column (2) shows the share of participants that opted for the politician with the three-page biography. Column (3) shows the share of participants that opted for the politician with the one-page biography. Column (4) shows the share of participants that opted for "Don't know." Column (5) gives the difference between columns (2) and (3) along with its standard error. We test against the null hypothesis that (2) = (3). * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

C.2 Robustness checks

In our main analysis, we excluded 35 MPs in distinguished positions (ministers and party heads) and nine further MPs who had already left the Bundestag. Table C.2 shows that including these observations does not affect our results. Moreover, the estimates that were added – dummies for early resignations and distinguished offices – are very large and statistically significant, legitimizing the presumption that they are not comparable to other MPs in the sample.

In addition to that, we prove the robustness of the results in Table 4.3 by taking into account that English texts are about a fourth to a fifth shorter than German texts.[2] To this end, we scale the MPs' German biography length in equation (2.2) with the factors 0.6, 0.75, 0.8, and 0.9, respectively. Moreover, we use the difference in logs of the biography lengths as a dependent variable. Table C.2 shows that our results are qualitatively unaffected.

C.3 Party affiliation and English biography length

In Section 4.4, we argue that it is plausible to assume that there are no effects of party affiliation on the English biography length, because partisan contributors have no incentive to contribute to the English Wikipedia. In this section, we perform four additional plausibility checks. First, only thirteen of 598 MPs in our dataset provide more than a short CV in English on their personal homepage, suggesting that they do not consider an English web presence as important. Second, only eight of 138 English biographies (5.8%) were edited from the Bundestag building; with one exception, only small changes were undertaken. Third, the lion's share of the English biographies is not translated from their German counterparts. Translated articles have to be marked by a translation template on the article's talk page and by a link to the source article; only ten out of 138 English biographies are marked like this, and no biography is translated from a foreign language into German. In addition, Wikipedia advises against one-to-one translation.[3] Finally, while the assumption of no party effects may fail for foreign languages that are spoken in countries adjacent to Germany or by large minorities, Germany has no direct border with any English speaking country, and a low number of immigrants whose native language is English (Statistisches Bundesamt, 2017).

[2] See, e.g., `www.orbis-uebersetzungen.de` (Feb 2016).
[3] See `https://en.wikipedia.org/wiki/Wikipedia:Translation` (Dec 2018).

Table C.2: Robustness checks I

	(1) Length	(2) Length	(3) Length	(4) Length
CDU/CSU	1106.7**	1903.7**	1047.9***	2167.2***
	(529.0)	(850.0)	(381.4)	(780.8)
Left	2554.5***		2171.8***	
	(758.7)		(662.4)	
Greens	1933.0*		1599.9*	
	(1000.1)		(911.8)	
Female	-694.2	330.5	-421.7	135.5
	(501.3)	(866.7)	(408.1)	(740.5)
Former periods in BT	679.5***	411.5*	905.5***	747.0***
	(173.4)	(217.6)	(159.4)	(197.5)
PhD	2085.3***	1562.8	1917.1***	778.6
	(691.5)	(1021.7)	(613.3)	(883.1)
Party head	13689.0***	7010.2		
	(4381.3)	(4710.3)		
Minister during 18th BT	9316.7**	11781.8***		
	(3732.3)	(4530.0)		
Former periods as minister	10701.3***	11764.1***		
	(2517.3)	(2765.1)		
Early resign			7629.5**	13072.9***
			(3397.5)	(4265.3)
Population density		0.732**		1.515***
		(0.306)		(0.201)
Fraction pop. 18 - 35		305.4		92.70
		(190.5)		(156.3)
Fraction pop. with Abitur		-36.26		5.823
		(55.49)		(47.63)
Constant	3805.8***	-1841.2	3402.3***	-433.3
	(594.6)	(3575.3)	(372.1)	(3410.8)
N	633	289	607	275
R^2	0.437	0.542	0.173	0.344

Notes: Robust standard errors in parentheses. The dependent variable $length_i^G$ measures Wikipedia coverage of MP i in terms of her biography length in characters. CDU/CSU_i, $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $FormerperiodsinBT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenwatch.de`. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18-35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i's constituency demography. $Party\ head_i$ is equal to 1 if MP i is chairman of her party or its parliamentary group. $Minister\ during\ 18th\ BT_i$ is equal to 1 if i was a minister in the 18th election term. $Former\ periods\ as\ minister_i$ counts the election terms during which MP i was minister before the 18th election term. $Early\ resign_i$ is equal to 1 if MP i left the Bundestag early. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Table C.3: Robustness checks II

	(1) DiD	(2) DiD	(3) DiD	(4) DiD	(5) DiD
CDU/CSU	3836.8***	4767.6***	5032.0***	5533.5***	0.314
	(1212.8)	(1405.9)	(1466.2)	(1590.2)	(0.233)
Left	8622.4***	9979.4***	10373.3***	11128.3***	1.035***
	(1714.2)	(1944.0)	(2023.7)	(2189.9)	(0.345)
Greens	4359.3**	5818.2***	6255.7***	7111.0***	0.00663
	(1745.9)	(1993.2)	(2076.7)	(2249.0)	(0.347)
Female	-1570.5	-1506.5	-1499.7	-1483.2	-0.232
	(999.5)	(1148.4)	(1200.0)	(1306.8)	(0.142)
Former periods in BT	987.7***	1348.2***	1458.8***	1678.5***	0.0645
	(265.4)	(303.4)	(316.2)	(342.9)	(0.0737)
PhD	2565.7**	3007.0**	3167.8**	3501.7**	0.138
	(1123.5)	(1293.2)	(1351.8)	(1473.3)	(0.193)
Ancillary Income	0.802	0.832	0.938	1.180	- 0.0000
	(5.490)	(6.145)	(6.392)	(6.916)	(0.0009)
Constant	-13683.6***	-15861.5***	-16490.7***	-17738.0***	-0.0713
	(1909.9)	(2043.0)	(2096.4)	(2220.4)	(0.952)
Mills	6793.5***	8244.2***	8699.7***	9621.6***	0.284
Lambda	(812.0)	(807.0)	(820.1)	(854.2)	(0.510)
N	596	596	596	596	596

Notes: Robust standard errors in parentheses. The results are based on an ML estimation of the DiD selection model. The dependent variable is $length_i^G - length_i^E$, which measures the differences in Wikipedia coverage of MP i in terms of her biography length in characters. In columns (1) to (4) the German length $length_i^G$ was scaled with the factors 0.6, 0.75, 0.8, and 0.9, respectively, before taking the difference. CDU/CSU_i, $Left_i$, and $Greens_i$ are dummy variables equal to 1 if MP i is affiliated to that party; SPD is the omitted category. $Female_i$ is equal to 1 if MP i is a woman. $Former\ periods\ in\ BT_i$ counts the election terms that MP i has been in parliament. $Ancillary\ income_i$ is the mean ancillary income of MP i during the 18th election term in 1,000 Euros based on the estimation of `abgeordnetenwatch.de`. PhD_i is equal to 1 if MP i has a PhD. $Population\ density_i$, $Fraction\ pop.\ 18-35_i$, and $Fraction\ pop.\ with\ Abitur_i$ refer to i's constituency demography. * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

D Appendix to Chapter 5

D.1 Tweet examples (translated from German into English)

D.1.1 Examples of non-criticizing tweets

Grüne tweeted:

> 12.000 plastic particles in one (!) litre of arctic ice. We do not only poison the fish in the sea, but everything will end up in our bodies. Time to act. Stop #plasticpollution #plastictax.

SPD tweeted:

> Civil insurance: Well explained on Spiegel Online.

LINKE tweeted:

> It is good that @Simone_Lange opposes #Hartz4 so clearly. Otherwise very sad. @dieLinke is the social alternative and will continue to exert pressure for a fundamentally different policy.

D.1.2 Examples of criticizing tweets

LINKE tweeted:

> I also accuse the SPD of playing a waiting game! But it is not correct that the LINKE supports the proposal by the FDP in its current form! #219a must be deleted. Induced abortion has no place in the penal code.

AfD tweeted:

> While @PoggenburgAndre is politically “classified”, @DLF of course abstains from doing so for the former secret police collaborator #Kahane. And you really wonder why fewer and fewer citizens trust your reporting?

E Appendix to Chapter 6

E.1 Proof of Proposition 1

Here we show that, for any $F > 0$, if a symmetric equilibrium exists, it is given by (6.11) and (6.12). In Section F.5, we establish that (6.11) and (6.12) indeed constitute an equilibrium.

Suppose that $F > 0$. In any symmetric equilibrium, $a_i > 0$ for otherwise broadcasters make losses $-F$. Therefore, the first order conditions (6.9) and (6.10) have to hold. By symmetry ($a_i = a_j$ and $v_i = v_j$) these conditions simplify to

$$\frac{1}{\tau}\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) - \frac{\beta}{N} = 0,$$

$$-\frac{\delta}{\tau}\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) a_i - \frac{1}{N}\frac{\sigma}{m}a_i + \frac{1}{N}\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) = 0.$$

It is easily verified that the unique solution to these equations is given by equations (6.11) and (6.12).

E.2 Proof of Proposition 2

From (6.5),

$$\frac{\partial CS}{\partial v} = n > 0. \tag{E.1}$$

Moreover, from (6.6),

$$\frac{\partial PS}{\partial v} = -n\beta a \tag{E.2}$$

which is strictly smaller than zero since $a > 0$ in equilibrium.

Finally, consider the marginal effect of v on welfare W,

$$\frac{\partial W}{\partial v} = \frac{\partial CS}{\partial v} + \frac{\partial PS}{\partial v} = n - n\beta a.$$

Inserting the equilibrium value of a gives

$$\frac{\partial W}{\partial v} = n - \frac{nm\beta^2\tau}{N\left(\sigma + m\beta\delta\right)},$$

which is strictly positive iff $N > \hat{N}_v$.

E.3 Proof of Proposition 3

From (6.5),

$$\frac{\partial CS}{\partial a} = -\delta n < 0. \tag{E.3}$$

Moreover, from (6.6),

$$\frac{\partial PS}{\partial a} = n\left(\sigma - \beta v - \frac{\sigma a}{m}\right) > 0, \tag{E.4}$$

which is strictly positive in equilibrium. The marginal effect of a on welfare is

$$\frac{\partial W}{\partial a} = \frac{\partial CS}{\partial a} + \frac{\partial PS}{\partial a} = -\delta n + n\left(\sigma - \beta v - \frac{\sigma a}{m}\right).$$

Inserting the equilibrium values from (6.11) and (6.12) gives

$$\frac{\partial W}{\partial a} = -\delta n + \frac{1}{N} n \beta \tau,$$

which is strictly negative iff $N > \hat{N}_a$.

E.4 Proof of Lemma 1

The proof is similar to the proof of Proposition 1, and hence omitted.

E.5 Proof of Proposition 4

The marginal effect of $\bar{a}$ on CS is given by

$$\frac{dCS}{d\bar{a}} = \frac{\partial CS}{\partial a} + \frac{\partial CS}{\partial v}\frac{dv}{d\bar{a}}.$$

From Lemma 1 it follows that

$$\frac{dv}{d\bar{a}} = -\frac{1}{m}\frac{\sigma}{\beta}, \tag{E.5}$$

so from (E.3), (E.1) and (E.5),

$$\frac{dCS}{d\bar{a}} = -\delta n - \frac{\sigma n}{m\beta} < 0.$$

The marginal effect of $\bar{a}$ on the producer surplus PS is given by

$$\frac{dPS}{d\bar{a}} = \frac{\partial PS}{\partial a} + \frac{\partial PS}{\partial v}\frac{dv}{d\bar{a}}.$$

From (E.2), (E.4), and (E.5) it follows that

$$\frac{dPS}{d\bar{a}} = n\left(\sigma - \beta v\right),$$

which is strictly positive since in equilibrium both inverse ad demand and advertising quantity are strictly positive, i.e., $\sigma - \beta v > \sigma a/m > 0$.

Finally, consider the effect of $\bar{a}$ on welfare W,

$$\frac{dW}{d\bar{a}} = \frac{dCS}{d\bar{a}} + \frac{dPS}{d\bar{a}} = -\delta n - \frac{\sigma n}{m\beta} + n\left(\sigma - \beta v\right).$$

From inserting the equilibrium value of v from Proposition 1 it follows that the total effect of $\bar{a}$ on W is

$$\frac{dW}{d\bar{a}} = -\delta n - \frac{\sigma n}{m\beta} + n\beta\frac{\tau\left(2\sigma + m\beta\delta\right)}{N\left(\sigma + m\beta\delta\right)}$$

which is strictly negative if and only if $N > \hat{N}_{cap}$.

E.6 Proof of Proposition 5

By inserting the equilibrium value of a and v into the welfare function, it follows that the laissez-faire welfare W^{LF}, that is achieved when there is no cap, equals

$$\begin{aligned}W^{LF} = n&\left(w + \frac{\sigma}{\beta} - \frac{\tau\left(2\sigma + m\beta\delta\right)}{N\left(\sigma + m\beta\delta\right)} - \delta\frac{m\beta\tau}{N\left(\sigma + m\beta\delta\right)}\right) - \frac{n\tau}{4N}\\ &+ n\int_0^{\frac{m\beta\tau}{N(\sigma+m\beta\delta)}}\left(\sigma - \beta\left(\frac{\sigma}{\beta} - \frac{\tau\left(2\sigma + m\beta\delta\right)}{N\left(\sigma + m\beta\delta\right)}\right) - \frac{\sigma x}{m}\right)dx - NF.\end{aligned}$$

With a cap $\bar{a} = N^2F/\left(n\beta\tau\right)$, welfare equals

$$\begin{aligned}W^{cap} = n&\left(w + \frac{\sigma}{\beta} - \frac{1}{N}\tau - \frac{1}{m}\frac{\sigma}{\beta}\frac{N^2F}{n\beta\tau} - \delta\frac{N^2F}{n\beta\tau}\right) - \frac{n\tau}{4N}\\ &+ n\int_0^{\frac{N^2F}{n\beta\tau}}\left(\sigma - \beta\left(\frac{\sigma}{\beta} - \frac{1}{N}\tau - \frac{1}{m}\frac{\sigma}{\beta}\frac{N^2F}{n\beta\tau}\right) - \frac{\sigma x}{m}\right)dx - NF.\end{aligned}$$

The difference is

$$\begin{aligned}&W^{cap} - W^{LF}\\ &= \frac{\left(FN^3\left(\sigma + m\beta\delta\right) - mn\beta^2\tau^2\right)\left(mn\beta^2\tau^2\left(3\sigma + 2m\beta\delta\right) + N\left(\sigma + m\beta\delta\right)\left(FN^2\sigma - 2n\tau\left(\sigma + m\beta\delta\right)\right)\right)}{2N^2mn\beta^2\tau^2\left(\sigma + m\beta\delta\right)^2}.\end{aligned}$$

Since by assumption laissez–faire equilibrium profits (given in Proposition 1) are positive,

$$FN^3(\sigma + m\beta\delta) < mn\beta^2\tau^2.$$

Therefore, $W^{cap} > W^{LF}$ if and only if

$$mn\beta^2\tau^2(3\sigma + 2m\beta\delta) + N(\sigma + m\beta\delta)\left(FN^2\sigma - 2n\tau(\sigma + m\beta\delta)\right) < 0$$

or, equivalently,

$$F < \hat{F}(N) := \frac{2Nn\tau(\sigma + m\beta\delta)^2 - mn\beta^2\tau^2(3\sigma + 2m\beta\delta)}{N^3\sigma(\sigma + m\beta\delta)}.$$

For any $N > \hat{N}_{cap}$, Proposition 4 has already established that a local cap raises welfare and thus clearly $W^{cap} > W^{LF}$. For the rest of the proof, consider the case where $N \leq \hat{N}_{cap}$.

By differentiating $\hat{F}(N)$ with respect to N, it can be shown that for all $N \leq \hat{N}_{cap}$, $\hat{F}(N)$ is strictly increasing in N. Therefore, for all $N \leq \hat{N}_{cap}$, one can invert $\hat{F}(N)$ to find a strictly increasing function $\hat{N}(F)$ such that $F < \hat{F}(N)$ if and only if $N > \hat{N}(F)$. The remaining properties of $\hat{N}(F)$ can be shown by the implicit function rule, taking into account that, in the relevant range, $\hat{N}(0) \leq N \leq \hat{N}_{cap}$.

E.7 Proof of Proposition 6

Without a cap on advertising, the number of firms and quantities of advertising are, in equilibrium,

$$N = N^{LF} := \left(\frac{nm\beta^2\tau^2}{F(\sigma + m\beta\delta)}\right)^{\frac{1}{3}},$$

$$a = a^{LF} := \frac{m\beta\tau}{\left(\frac{nm\beta^2\tau^2}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}(\sigma + m\beta\delta)} = \left(\frac{\beta\tau F m^2}{n(\sigma + m\beta\delta)^2}\right)^{\frac{1}{3}}.$$

With a binding cap $\bar{a}$, the number of firms equals $N = \sqrt{(n\bar{a}\beta\tau)/F}$.

Substituting v from Lemma 1 in equation (6.5), and inserting the equilibrium number of firms $N = \sqrt{(n\bar{a}\beta\tau)/F}$, shows that consumer surplus is

$$CS(\bar{a}) = n\left(w + \frac{\sigma}{\beta} - \frac{\sigma}{m\beta}\bar{a} - \delta\bar{a}\right) - \frac{5}{4}\frac{\sqrt{n\tau F}}{\sqrt{\bar{a}\beta}}.$$

As noted in the main text, in contrast to the case with a constant number of broadcasters, with free entry a cap on advertising does not necessarily increase CS. Indeed,

$$CS'\left(\bar{a}\right) = -\frac{n\sigma}{m\beta} - n\delta + \frac{5}{8}\frac{1}{\bar{a}^{\frac{3}{2}}}\sqrt{\frac{Fn\tau}{\beta}} \tag{E.6}$$

is positive when $\bar{a}$ is sufficiently small; thus when a very stringent cap is relaxed, viewers become better off.

To prove part (i) of Proposition 6, evaluate (E.6) at the laissez-faire equilibrium value of advertising. After straightforward calculations,

$$CS'\left(a^{LF}\right) = -\frac{n\sigma}{m\beta} - n\delta + \frac{5}{8}\frac{n\left(\sigma + m\beta\delta\right)}{m\beta} = -\frac{3n\left(\sigma + m\beta\delta\right)}{8m\beta} < 0.$$

Therefore, a local cap improves consumer surplus.

It remains prove parts (ii) and (iii). Summing profits and consumer surplus shows that, given a binding cap $\bar{a}$, welfare equals

$$W\left(\bar{a}\right) = n\left(w + \frac{\sigma}{\beta} - \frac{\sigma}{m\beta}\bar{a} - \delta\bar{a}\right) - \frac{5}{4}\frac{\sqrt{n\tau F}}{\sqrt{\bar{a}\beta}} + \frac{1}{2}\frac{\bar{a}^2}{m}n\sigma. \tag{E.7}$$

A welfare maximizing planner maximizes $W\left(\bar{a}\right)$ by choosing a cap $\bar{a} \leq a^{LF}$.[1] Here, choosing $\bar{a} = a^{LF}$ is equivalent to laissez-faire, i.e., imposing no cap at all.

Differentiating (E.7),

$$W'\left(\bar{a}\right) = -\frac{n\sigma}{m\beta} - n\delta + \frac{5}{8}\frac{1}{\bar{a}^{\frac{3}{2}}}\sqrt{\frac{Fn\tau}{\beta}} + \frac{\bar{a}}{m}n\sigma.$$

Evaluating $W'\left(\bar{a}\right)$ at $\bar{a} = a^{LF}$ and rearranging gives

$$W'\left(a^{LF}\right) = -\frac{n\sigma}{m\beta} - n\delta + \frac{5}{8}\frac{n\left(\sigma + m\beta\delta\right)}{m\beta} + F^{\frac{1}{3}}\frac{\left(\beta\tau\right)^{\frac{1}{3}} n^{\frac{2}{3}}\sigma}{\left(\sigma + m\beta\delta\right)^{\frac{2}{3}} m^{\frac{1}{3}}},$$

which is strictly negative if and only if

$$F < \left(\frac{\left(\frac{n\sigma}{m\beta} + n\delta - \frac{5}{8}\frac{n(\sigma+m\beta\delta)}{m\beta}\right)\left(\sigma + m\beta\delta\right)^{\frac{2}{3}} m^{\frac{1}{3}}}{\left(\beta\tau\right)^{\frac{1}{3}} n^{\frac{2}{3}}\sigma}\right)^3 = \hat{F}_{capwithexit}.$$

This shows that a local cap improves welfare if and only if $F < \hat{F}_{capwithexit}$.

Note W fulfills the Inada condition $\lim_{\bar{a}\to 0} W'\left(\bar{a}\right) = \infty$, hence $W'\left(\bar{a}\right) > 0$ for sufficiently small

[1] Equation (E.7) presupposes that $N \geq 2$. Of course, a cap that is too stringent will eliminate competition on the broadcasting market. As stated above, we focus on the case where some competition prevails.

$\bar{a}$. If a binding cap $\bar{a} < a^{LF}$ is optimal, it must fulfill the first order condition $W'(\bar{a}) = 0$. Although W is not necessarily concave on $\left(0, a^{LF}\right]$, nevertheless a sufficient second order condition (called pseudo-concavity) holds:

Lemma 2. *Suppose that $W'(a_0) = 0$ for some $a_0 \in \left(0, a^{LF}\right)$. Then (i) $W''(a_0) < 0$. Moreover, (ii) $W'(a) > 0$ for all $a < a_0$ and $W'(a) < 0$ whenever $a_0 < a < a^{LF}$.*

Proof. Differentiating $W'(a)$,

$$W''(a) = -\frac{15}{16a^{\frac{5}{2}}}\sqrt{F\frac{n}{\beta}\tau + \frac{n\sigma}{m}}.$$

Suppose that $W'(a_0) = 0$ and $0 < a_0 < a^{LF}$. Then $W''(a_0)$ has the same sign as $g(a_0)$, where

$$g(a) := aW''(a) - W'(a) = -\frac{25}{16a^{\frac{3}{2}}}\sqrt{\frac{Fn\tau}{\beta} + \frac{n\sigma}{m\beta}} + n\delta.$$

Note that $g'(a) > 0$ and

$$g\left(a^{LF}\right) = -\frac{25}{16\left(\frac{m\beta\tau}{\left(\frac{nm\beta^2\tau^2}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}(\sigma+m\beta\delta)}\right)^{\frac{3}{2}}}\sqrt{\frac{Fn\tau}{\beta} + \frac{n\sigma}{m\beta}} + n\delta = -\frac{9}{16m}\frac{n}{\beta}(\sigma + m\beta\delta) < 0.$$

It follows that $g(a_0) < 0$ and hence $W''(a_0) < 0$. This establishes (i). Part (ii) is obvious from (i). $\square$

To complete the proof of part (ii) of Proposition 6, suppose that $F < \hat{F}_{capwithexit}$. By the intermediate value theorem, since $W'(\bar{a}) > 0$ for sufficiently small $\bar{a}$ and $W'\left(a^{LF}\right) < 0$, there exists some $a^* \in \left(0, a^{LF}\right)$ such that $W'(a^*) = 0$. By Lemma 2, W has a strict global maximum at a^*. For the comparative statics of the optimal cap, recall that $W'(a^*) = 0$ and $W''(a^*) < 0$. By the implicit function rule, the sign of $\frac{da^*}{d\delta}$ is equal to the sign of $\frac{\partial W'(a^*)}{\partial\delta} = -n < 0$. Similarly, $\frac{\partial}{\partial\tau}W'(\bar{a}) > 0$ and $\frac{\partial}{\partial F}W'(\bar{a}) > 0$, thus a^* is increasing in τ and in F. Moreover,

$$\begin{aligned}\frac{\partial}{\partial n}W'(\bar{a}) &= \frac{\partial}{\partial n}\left(n\left(-\frac{\sigma}{m\beta} - \delta + \frac{5}{8}\frac{1}{\bar{a}^{\frac{3}{2}}}\sqrt{\frac{F\tau}{\beta n} + \frac{\bar{a}}{m}\sigma}\right)\right) \\ &= \left(-\frac{\sigma}{m\beta} - \delta + \frac{5}{8}\frac{1}{\bar{a}^{\frac{3}{2}}}\sqrt{\frac{F\tau}{\beta n} + \frac{\bar{a}}{m}\sigma}\right) - \frac{5}{16}\frac{1}{\bar{a}^{\frac{3}{2}}}\sqrt{\frac{F\tau}{\beta n}}.\end{aligned}$$

Note the bracket is $W'(\bar{a})/n$ and thus zero when evaluated at a^*. It follows that $\frac{\partial}{\partial n}W'(a^*) < 0$ and a^* is decreasing in n.

To prove part (iii) of Proposition 6, suppose that $F \geq \hat{F}_{capwithexit}$. Then $W'(a_{LF}) \geq 0$. Lemma 2 implies that $W'(\bar{a}) > 0$ for all $\bar{a} < a^{LF}$. Thus the optimal policy is laissez-faire.

E.8 Proof of Proposition 7

Suppose there is a tax t on advertising revenue. The profit of broadcaster i equals

$$\pi_i = n\left(\frac{1}{N} + \frac{v_i - \delta a_i - u}{\tau}\right)\left(\sigma - \beta v_i - \frac{\sigma}{m} a_i\right) a_i (1-t) - F.$$

For given N, the equilibrium advertising quantity a, program quality v, are as given in Proposition 1 above. Inverse ad demand per viewer equals $r = \beta\tau/N$, and net of taxes $(1-t)\beta\tau/N$. The equilibrium profit of a broadcaster is is

$$\pi_i = \frac{nm\beta^2\tau^2}{N^3(\sigma + m\beta\delta)}(1-t) - F.$$

With endogenous entry, the equilibrium number of broadcasters is

$$N = \left(\frac{nm\beta^2\tau^2(1-t)}{F(\sigma + m\beta\delta)}\right)^{\frac{1}{3}}.$$

Therefore, in equilibrium

$$a = \frac{m\beta\tau}{\left(\frac{nm\beta^2\tau^2(1-t)}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}(\sigma + m\beta\delta)}, v = \frac{\sigma}{\beta} - \frac{\tau(2\sigma + m\beta\delta)}{\left(\frac{nm\beta^2\tau^2(1-t)}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}(\sigma + m\beta\delta)}.$$

Moreover, net consumer surplus (i.e., before redistribution of tax revenues) is

$$\begin{aligned} CS_{net} &= n(w + v - \delta a) - \frac{n\tau}{4N} \\ &= nw + \frac{n\sigma}{\beta} - \frac{2\tau n}{\left(\frac{nm\beta^2\tau^2(1-t)}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}} - \frac{n\tau}{4\left(\frac{nm\beta^2\tau^2(1-t)}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}}. \end{aligned} \tag{E.8}$$

Thus, CS_{net} is decreasing in t. Advertiser profits equals $a^2 n\sigma/(2m)$ (see equation (6.8)). Inserting the equilibrium value of a gives

$$\frac{1}{2}\frac{a^2}{m}n\sigma = \frac{1}{2}\frac{n\sigma}{m}\left(\frac{m\beta\tau}{\left(\frac{nm\beta^2\tau^2(1-t)}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}(\sigma + m\beta\delta)}\right)^2. \tag{E.9}$$

Thus, advertiser profits are increasing in t. Tax revenue T is given by $T = nrat$. In equilibrium,

$$T = \frac{nm\left(\beta\tau\right)^2 t}{\left(\frac{nm\beta^2\tau^2(1-t)}{F(\sigma+m\beta\delta)}\right)^{\frac{2}{3}}\left(\sigma+m\beta\delta\right)}. \tag{E.10}$$

Welfare is $W = CS_{net} + PS - NF + T$. Note that, because of free entry, $PS - NF$ equals advertisers' profits (E.9).

Inserting (E.8), (E.9), and (E.10), into W, differentiating with respect to t, and evaluating at $t = 0$, shows that

$$\frac{\partial W}{\partial t}|_{t=0} = \frac{F\left(\frac{mn\beta^2\tau^2}{F(\sigma+m\beta\delta)}\right)^{\frac{1}{3}}}{12m\beta^2\tau\left(\sigma+m\beta\delta\right)}\left(4m\beta^2\tau\left(4\sigma+3m\beta\delta\right)-\left(\frac{mn\beta^2\tau^2}{F\left(\sigma+m\beta\delta\right)}\right)^{\frac{1}{3}}9\left(\sigma+m\beta\delta\right)^2\right).$$

Therefore, W is increasing in t if and only if $F > \hat{F}_{taxwithexit}$.

To complete the proof, straightforward calculation of $\hat{F}_{capwithexit} - \hat{F}_{taxwithexit}$ shows that $\hat{F}_{taxwithexit} < \hat{F}_{capwithexit}$ if and only if $\delta > 2\sigma/\left(3m\beta\right)$.

F Online-Appendix to Chapter 6

This chapter was originally published as Online-Appendix to "Quantity Restrictions on Advertising, Commercial Media Bias, and Welfare." Available at URL: https://ars.els-cdn.com/content/image/1-s2.0-S0047272715001528-mmc1.pdf

F.1 Additional background material

Media reports impact firm profits. We start by pointing to evidence that media reports impact firm profitability. For example, when the New York Times reported about a potential breakthrough in cancer research in its Sunday edition, it induced a permanent rise in share prices of EntreMed, a biotechnology company with licensing rights to the breakthrough - even though the information had already been published in Nature and various newspapers several month earlier (Huberman and Regev, 2001). Similarly, Engelberg and Parsons (2011) show that media reporting has a causal effect on investor behavior, and Liu et al. (2014) show that media coverage of IPOs has long-run effects on the stock's value. Media reporting may also bring public scrutiny to sensitive issues and lead to regulatory threats to firms. For example, Erfle and McMillan (1990) show that during the 1979 oil crisis, television reports on the oil crises influenced home heating oil price ratios, but not residual fuel oil price ratios, and argue that the different reactions are explained by the threat of government intervention.

Moreover, several studies have shown that critical media reports impact consumer behavior. Niederdeppe and Frosch (2009) provide evidence that news coverage on trans fat reduced sales of trans-fat-products. Schlenker and Villas-Boas (2009) find a significant decrease in beef sales following reports on mad cow disease. Wakefield et al. (2003) show that antismoking messages can lower youth smoking rates. Laugesen and Meads (1991) find that doubling the coverage of smoking issues in newspapers lowers cigarette consumption as much as a 10% price increase.

Advertisers influence editors. Several recent papers have shown econometrically that advertisers systematically influence media content. The papers listed in Table F.1 compare advertiser spending with media coverage or slant. An empirical challenge is to identify whether there is a causal effect of advertising on media content. The recent literature has used state-of-the-art instrumental variable techniques and natural experiments to overcome this challenge (e.g., Gurun and Butler, 2012).

Interviews and surveys of key players in the market also confirm that advertisers influence media content. An early survey by Soley and Craig (1992) on newspaper editors found that almost 9 out of 10 of their correspondents claim that advertisers have attempted to take influence on editorial decisions. In a survey of journalists by Center (2000), a third of the journalists stated an “intrusion of commercial interests” into editorial decisions (p. 3). Similarly, the survey by An and Bergen (2007) on 219 advertising directors at US daily newspapers reports frequent conflicts between the journalism side and the business side of newspapers. Further evidence of advertisers' influence comes from a content analysis of ostensibly noncommercial newscasts (Upshaw et al., 2007): 90% of the stations studied contained at least one instance per newscast of commercial messages outside regular commercial blocks.

Table F.1: Evidence on advertisers' influence on media content

Study	Media	Advertiser	Country	Main result
Warner et al. (1992)	Magazines	Tobacco	US	“strong statistical evidence that cigarette advertising (...) diminished the coverage of the hazards of smoking” (p. 305)
Reuter and Zitzewitz (2006)	Finance publications	mutual funds	US	“mutual fund recommendations are correlated with past advertising” (p.197)
Reuter (2009)	Wine publications	Wineries	US	“Wine Spectator appears largely to insulate reviewers from the influence of advertisers” (p. 125)
Rinallo and Basuroy (2009)	Newspapers, magazines	Fashion	I, F, G, UK, US	“there is evidence of a strong positive influence of advertising on coverage (...) exist in both Europe and the US” (p. 33)
Gambaro and Puglisi (2015)	Newspapers	All	I	“coverage of a given company is positively related with the amount of ads purchased on that newspaper by that company” (p. 1)
Di Tella and Franceschelli (2011)	Newspapers	Government	Ar	“One standard deviation increase in monthly government advertising is associated with a reduction in the coverage of the government's corruption scandals of (...) 18% of a standard deviation in coverage.” (p. 119)
De Smet and Vanormelingen (2012)	Newspapers	Big advertisers	B	“advertisers in Belgian Dutch-language newspapers receive a significantly higher coverage” (p. 1)
Gurun and Butler (2012)	Local newspapers	Local companies	US	“positive slant about local companies is strongly positively related to the local advertising budget of these companies” (p. 563)
Dewenter and Heimeshoff (2014)	Car magazines	Car producers	G	“evidence for media bias in test scores” (p. 17)

Conflict of interest between viewers and advertisers. At the center of our model is a conflict of interest between viewers and advertisers over media content. Here we provide additional empirical evidence that motivates this assumption, and in particular deepen the issue that viewers but not advertisers may favor accurate reporting of defects, risks, or negative externalities of products.

As we mention in the paper, two important cases of commercial media bias concern the health risks of tobacco and anthropogenic climate change. Reporting about the health risks of smoking is an important and well-documented case of commercial media bias. The tobacco industry is a major advertiser. According to the WHO (2013, p.22), "the tobacco industry spends tens of billions of US dollars worldwide each year on tobacco advertising, promotion and sponsorship (TAPS). In the United States alone, the tobacco industry spends more than US$ 10 billion annually on TAPS activities." The tobacco industry has suppressed reports on health risks of smoking, and induced media platforms to merely reprint statements claiming that there was no proven evidence for smoking inflicting health (Ben, 2004). For example, when the magazine *Mother Jones* published an article on smoking and health, the tobacco companies withdrew their ads and "made clear that *Mother Jones* would never get cigarette advertising again" (Whelan et al., 1981, p.34). Warner and Goldenhar (1989) and Warner et al. (1992) provide strong statistical evidence that cigarette advertising in magazines relates to less coverage of the health risks of smoking.

News coverage of anthropogenic climate change is an issue of global importance. Global warming imperatively requires an accurately informed public. The broad scientific consensus is that human activities affect the climate (Oreskes, 2004). The discourse in the news media, however, has significantly diverged from the scientific consensus, particularly in the US. Boykoff and Boykoff (2004) study the US press coverage of global warming between 1988 and 2002.[1] They find that 53% of the investigated articles gives equal weight to the scientific consensus opinion and the view that human activities are a negligible factor in overall changes in the climate. The difference between US television news coverage and the scientific consensus is even more severe: from 1996 to 2004, 70% of the television broadcasts on climate change provided a balanced view on its causes (Boykoff, 2008, p.6). As pointed out by Ellman and Germano (2009), one potential reason behind this biased media coverage is the influence of big advertisers such as car manufacturers or airlines.[2]

Critical media reporting also has an important role to counteract misleading advertising.[3]

[1] Their empirical evidence stems from randomly selected articles from four major US newspapers: New York Times, Los Angeles Times, Washington Post, and Wall Street Journal.

[2] See also Blasco and Sobbrio (2012) and Germano and Meier (2013).

[3] Sharp evidence on deceptive advertising is provided by Zinman and Zitzewitz (2016), who show that ski resorts report 23% more fresh snow during weekends, when potential skiers are more flexible to react on snow conditions. Glaeser and Ujhelyi (2010) survey the available evidence. Nagler (1993) and Glaeser and Ujhelyi (2010) explore

Again, tobacco is a case in point: cigarette advertising has often downplayed the associated health risks (see Glaeser and Ujhelyi, 2010). Arguably, deceptive advertising, combined with tobacco advertisers' influence on media content, explains why public awareness of the health risks lagged decades behind their scientific discovery. For example, in Gallup polls from 1980 every second woman did not know that smoking during pregnancy increases the risk of stillbirth and miscarriage (Myers et al., 1981). Similarly, the WHO (2011) report on the global tobacco epidemic points out that many smokers do not fully understand the health risks of smoking. The WHO (2011) also mentions that the news media are a key source of health information, and emphasizes the importance of media reporting on tobacco control.

Another well documented example is advertising for medical drugs. For example, Faerber and Kreling (2014) classify more than one half of all major claims in prescription and non-prescription drug ads on US television during 2008-2010 as potentially misleading. Furthermore, Faerber and Kreling (2014) report that consumers may see up to 30 hours of television drug advertising each year; in contrast, they spend 15 to 20 minutes at an average visit with their primary care physician. Misleading advertising may thus seriously impair consumers' ability to take well informed decisions (Brody and Light, 2011).

F.2 Microfoundations

A central assumption of our paper is that the willingness to pay of an advertiser for reaching a consumer *decreases* in the quality v of the program the viewer watches. This section discusses the microfoundations mentioned in Section 6.4.3 of our paper.

One possible interpretation of the variable v is that it corresponds to a genre preferred by viewers. The empirical results by Wilbur (2008) and Brown and Cavazos (2005) indicate that advertisers' preferences over genres differ from viewers' preferences: advertisers prefer lighter content. Similarly, experimental evidence from Goldberg and Gorn (1987) shows that happier program types put viewers in a more advertising receptive mood.

The implications for consumer welfare depend on the mechanism how program content impacts on advertising effectiveness. One potential reason is that television genres preferred by consumers are substitutes for consumption goods; therefore, the better the quality of the television program, the lower the willingness to pay for goods. In other words, viewing advertiser friendly genres is a complement for consumption, i.e., it raises the utility of the viewer from consuming goods. We call this the *complementary microfoundation*.[4] Consumers have stable preferences defined over

theoretical implications of misleading advertising, and study optimal policy responses. We show in an extension of our model in Appendix F.3 that misleading advertising strengthens the case for a cap on advertising.

[4]This microfoundation is similar to the complementary view of advertising (Stigler and Becker, 1977; Becker and

consumption and advertising. Television program quality enters their utility function directly, and it moreover affects the utility they gain from consumption goods. In particular, we assume that the gross utility gain of a consumer from buying a product of quality $\tilde{\sigma}$ is equal to $\tilde{\sigma} - \beta v$. If in addition producers capture all the gains on the product market, the willingness to pay of a producer of type $\tilde{\sigma}$ for an advertising slot is also $\tilde{\sigma} - \beta v$. In this microfoundation, consumers are rational, and the willingness to pay of consumers for goods indicates their true welfare gains from consumption.

Another reason why television content may have an impact on advertising effectiveness is that consumers' recall of an ad depends on the program it is embedded in. Mathur and Chattopadhyay (1991) show in an experiment that viewers recall an ad better if it is shown in the context of a program that puts the viewers in a happy mood. This finding inspires our a second microfoundation, the *recall microfoundation.*

Suppose that some consumers recall an ad after seeing it, while others forget it. The probability that a consumer forgets an ad for a product of type $\tilde{\sigma}$ placed in a program of quality v is $p(v, \tilde{\sigma})$. Plausibly, p is increasing in v, and decreasing in $\tilde{\sigma}$: the better the television program, and the lower the product's quality, the more likely the consumer is to forget the product. Moreover, recall of better products (i.e., those with a high $\tilde{\sigma}$) might be less affected by television genre. A functional form consistent with these properties is $p(v, \tilde{\sigma}) = \beta v / \tilde{\sigma}$. A consumer who saw an ad for a product, but doesn't recall it, does not buy the product; just as consumers who do not know the product exists. Consumers make rational decisions given their information.[5] As in the complementary microfoundation, the willingness to pay of an informed consumer captures the true welfare gains from consumption. The willingness to pay of an advertiser of type $\tilde{\sigma}$ for showing an ad to a consumer is $(1 - p(v, \tilde{\sigma}))\tilde{\sigma}$. Note this is decreasing in program quality v. Moreover, if $p(v, \tilde{\sigma}) = \beta v / \tilde{\sigma}$, the willingness to pay of the advertiser is $\tilde{\sigma} - \beta v$, as in our model.

Television genre may also impact advertising effectiveness since it influences the moods of boundedly rational consumers. For an example, recall the case of Coca-Cola refusing to advertise during news broadcasts out of a concern that "bad" news might counteract its positioning of Coke as an "up-beat, fun product." There is good empirical evidence that consumers' moods impact their economic decisions. Harlé and Sanfey (2007) experimentally induce different moods by showing short movie clips to their subjects prior to an ultimatum game experiment. Incidental sad moods result in lower acceptance rates of unfair offers. Harlé et al. (2012) confirm this finding and study

Murphy, 1993; Bagwell, 2007). While this view claims that advertising is a complement for the advertised consumption good, we claim here that television content may be both a complement and a substitute to consumption goods. We point out that our model of advertising itself is also consistent with the complementary view of advertising.

[5] Moreover, if producers capture all the rents on the product market as in our model, consumers have no incentive to remember the ads; forgetting is a form of rational ignorance.

the underlying neural mechanisms in an fMRI study. Consumers have also been found in field data to be more likely to engage in impulse buying when they are in a positive mood (Beatty and Ferrell, 1998; Flight et al., 2012; Faber and Vohs, 2011). Television induced moods may thus affect advertising effectiveness and purchase behavior, even when the "true" utility from consumption is not affected by television genre.[6] We call this the *moods microfoundation*. In such a situation, consumers' willingness to pay cannot simply be equated with their true utility gains from the products. If some genres put consumers in a spending happy mood such that they overestimate the true utility gains of the products, the welfare analysis has to take this into account; we do this in Section F.3.1 of this chapter.

A second possible interpretation of the variable v is that a high quality corresponds to more accurate and critical reporting over products, for example over any risks involved in the consumption. A program of higher quality can then be interpreted as containing more information that helps consumers making well-informed decisions. Good television programs may also contain information about the advertiser or producer, and any externalities that the products may have. An example is the case of government advertising and reporting of corruption scandals investigated by Di Tella and Franceschelli (2011).

Following Nagler (1993) and Glaeser and Ujhelyi (2010), this interpretation can be used to develop another microfoundation, the *deceptive advertising microfoundation*. They argue that advertising sometimes is misleading and makes consumers underestimate costs involved in the consumption of their products. Key cases are advertising for medicines, cigarettes, and fast food; see also the discussion in Section F.3.1 of this chapter. Following Glaeser and Ujhelyi (2010), suppose that consumption of a product has health costs c, so the true gain of consumers from consuming a product of type $\tilde{\sigma}$ is $\tilde{\sigma} - c$. Consumers' perception of these costs may differ from the true costs. We assume that the perceived costs depend on how accurate reporting on television is; the better the program quality, the higher the perceived costs.[7] Suppose that perceived costs are equal βv_i, and that, in the relevant range, consumers underestimate the costs, i.e., $c > \beta v_i$. Thus, the better the program, the smaller consumers' errors.

Under these assumptions, a consumer is willing to pay up to $\tilde{\sigma} - \beta v_i$ for a product of quality $\tilde{\sigma}$. As above, we assume that the producer can completely capture these perceived benefits. Therefore, the willingness to pay of the producer for informing the consumer is equal to $\tilde{\sigma} - \beta v_i$, as well.

[6] See also DellaVigna (2009) and Lerner et al. (2015) who survey the growing literature on the role of emotions in economic decisions.

[7] Another difference between our setup and Glaeser and Ujhelyi (2010) is that, in our model, all consumers watching the same broadcaster are identical and have single unit demand. Therefore, even though each producer is a monopolist, consumption is efficient if and only if consumers perceive the health cost correctly, and there is no efficiency enhancing role for misinformation.

Consumers are aware that a better program quality helps them making better decisions, and thus perceive a benefit v_i from watching the program. They are not aware that the products involve any health costs beyond βv_i. Therefore, their perceived benefit from watching the broadcaster is

$$w + v_i - \delta a_i - \tau x \tag{F.1}$$

as in the paper. In the welfare analysis, however, we need to take into account that consumers do not correctly perceive the costs involved in the consumption decisions on the product market.

As discussed in the paper, the microfoundations described above are not mutually exclusive, and they all lead to the same positive predictions of the model. For normative questions, the main model in the paper builds on the complementarity microfoundation or the recall microfoundation, where consumers' willingness to pay for a product accurately captures their true benefits from the product. The moods microfoundation and the deceptive advertising microfoundation, on the other hand, show that consumers may have losses on the product market since their perceived gains from the products are not equal to their true gains. The magnitude of these losses may depend both on advertising quantity and on television program quality. Section F.3.1 of this chapter studies an extension of our main model that takes these considerations into account.

F.3 Extensions

F.3.1 Deceptive Advertising

In the paper we assume that a consumer's willingness to pay for a product accurately captures the consumer's benefits from the product. As argued above, this assumption is doubtful when purchase decisions are boundedly rational, or when advertising is suggestive or deceptive. Then, consumers may take suboptimal decisions (for themselves) on the product markets. Moreover, the corresponding losses of the consumers will depend on television program quality. This section investigates how taking these considerations into account modifies our main results.

As in the paper, we assume that a television program with quality v reduces the willingness to pay for a product of type $\tilde{\sigma}$ to $\tilde{\sigma} - \beta v$. However, here we assume that one part, $\gamma\beta v$, of the reduction comes from consumers making smaller errors, and the remaining part $(1-\gamma)\,\beta v$ comes from good television being a substitute for consumption, where $0 \leq \gamma \leq 1$. Following Glaeser and Ujhelyi (2010), suppose that consumption of a product has a health costs γc, so the true gain of consumers from buying a product of type $\tilde{\sigma}$ is $\tilde{\sigma} - (1-\gamma)\,\beta v - \gamma c$. The consumer perceives the costs to be $\gamma\beta v$; we assume that in the relevant range, consumers underestimate the costs, i.e., $c > \beta v$. We scale both the true costs and the perceived costs with the same parameter γ in order to have one single parameter that captures the importance of deceptive advertising. The case where $\gamma = 0$ corresponds to our main model. The case $\gamma = 1$ corresponds to the deceptive advertising microfoundation discussed in Section 6.4.3. When $0 < \gamma < 1$, higher television quality both reduces the true utility of consumption, and informs consumers so that they have a more accurate estimate of the costs.

As in the paper, we assume that the producers can capture all the perceived benefits from the products by charging the price $\tilde{\sigma} - \beta v$. The net utility gain of a consumer from consuming a good of type $\tilde{\sigma}$ is then $\tilde{\sigma} - (1-\gamma)\beta v - \gamma c - (\tilde{\sigma} - \beta v) = \gamma(\beta v - c)$. The consumer is informed about, and consumes, in total a such products, thus the consumer's loss on the product market equals $a\gamma(c - \beta v)$. Consumers are aware that a better program quality helps to make better decisions, and perceive a benefit v from watching the program. They are not aware that the products advertised on television involve any health costs beyond $\gamma\beta v$. Thus the consumers' perceived benefit from watching a broadcaster is given by (6.2), and hence as in the paper. Consumer surplus, however, also has to take into account consumers' losses on the product market:

$$CS = n(w + v - \delta a) - \frac{n\tau}{4N} - na\gamma(c - \beta v). \tag{F.2}$$

Producer surplus is, as in the paper, given by

$$PS = n\int_0^a \left(\sigma - \beta v - \frac{\sigma x}{m}\right) dx. \tag{F.3}$$

Note that consumption of a product of type $\tilde{\sigma}$ raises welfare if and only if $\tilde{\sigma} > \gamma c$. Thus consumption of high-quality goods is welfare enhancing in our setting.

For the positive analysis, this model generates the same predictions as the model in the paper. A cap on advertising, however, now has additional benefits for the consumers: it improves their decisions on the product market, both by reducing the number of ads and by improving the program quality. Thus, the welfare gains due to a cap are higher than in the paper. To see this formally, note that the effects of lower advertising quantity, and higher program quality, on consumer surplus are

$$-\frac{\partial CS}{\partial a} = \delta n + n\gamma(c - \beta v),$$
$$\frac{\partial CS}{\partial v} = n + na\gamma\beta.$$

Therefore, both the direct (less advertising) and the indirect (higher program quality) effect of a cap on consumer surplus are more important when $\gamma > 0$. It is therefore more likely that the cap's positive effects on consumer surplus outweigh the negative effects on producer surplus.

Deceptive advertising thus makes the case for a cap stronger. It modifies, however, our result on the complementarity between competition and regulation. When there are many independent broadcasters, consumers' errors are small since program quality is high and advertising quantities are low; thereby consumers' gains from a cap are smaller. This works against the complementarity between competition and regulation. Indeed, if γ is close to 1, competition and regulation are no longer local complements.

We now show, however, that the local complementarity between competition and regulation holds whenever $\gamma < (2\sigma + m\beta\delta) / (3\sigma + m\beta\delta)$; a sufficient condition is $\gamma < 2/3$. Competition and regulation are local complements if the marginal welfare gains from a cap, $-\frac{dW}{d\bar{a}}$, are increasing in N. Here, $W = CS + PS - NF$, where CS is given in (F.2), and PS in (F.3). Substituting v from Lemma 1 of the paper into these expressions, and differentiating, one obtains

$$\frac{\partial}{\partial N}\left(-\frac{dW}{d\bar{a}}\right) = n\beta\tau\frac{2\sigma - 3\sigma\gamma + m\beta\delta - m\beta\gamma\delta}{N^2(\sigma + m\beta\delta)},$$

which is strictly positive if and only if $\gamma < (2\sigma + m\beta\delta) / (3\sigma + m\beta\delta)$.

F.3.2 Sector Specific Regulation

In the paper we assumed that advertisers have a shared interest in low program quality. This may be appropriate for specific industries where the qualities of the products sold are highly correlated. For example, all producers in the tobacco industry may suffer from an increased awareness of the health risks of smoking. In other industries, however, broadcasters may be less hostile to accurate reporting. As argued by Ellman and Germano (2009) and Germano and Meier (2013), and modelled in detail by Blasco et al. (2016) and Spiteri (2015), competition on the product market can ameliorate commercial media bias when advertisers have opposing interests.

In this section, we study an extension where advertisers are interested in low program quality in some industries, but not in others. This allows us to probe the robustness of our results, and to shed light on the rationale of sector specific bans on television advertising. For example, in the United States, broadcasters are not allowed to send commercials on tobacco by the Public Health Cigarette Smoking Act of 1970 (Tobacco Control Legal Consortium, 2012). The rules in the European Union are similar. The Audiovisual Media Services Direction (European Union, 2012) bans commercials on cigarettes and other tobacco products, medicinal products and medicinal treatment available only on prescription. The advertisements for alcoholic beverages shall not be aimed specifically at minors and shall not encourage immoderate consumption (Article 9). The restrictions can be more stringent in the member states. In Germany, gambling must not be advertised and commercials for alcohol must not appeal to children or teenagers (Seufert and Gundlach, 2012).

We give a new rationale for sector specific regulations based on their impact on non-advertising media content below. To keep the discussion short, we focus on the case where $\delta = 0$. Suppose there is a mass m_1 of type 1 advertisers characterized by $\beta = 0$. These advertisers are not interested in dumbing down media content. Moreover, there is a mass $m_2 = m - m_1$ of type 2 advertisers with $\beta > 0$; these advertisers prefer lower program quality. Suppose that the quality $\tilde{\sigma}$ of the product of any advertiser is drawn from the uniform distribution on $[0, \sigma]$; thus type 1 and type 2 advertisers do not differ in this respect. (In Section F.4.2 we explore an extension where advertisers' preferences for media content depend on their product's quality). Advertising demand of broadcaster i is then

$$a_i = m_1 \Pr(\tilde{\sigma} > r_i) + m_2 \Pr(\tilde{\sigma} - \beta v_i > r_i) = \begin{cases} m_1 \left(1 - \frac{r_i}{\sigma}\right) + m_2 \left(1 - \frac{r_i + \beta v_i}{\sigma}\right), \text{if } 0 \le r_i < \sigma - \beta v_i, \\ m_1 \left(1 - \frac{r_i}{\sigma}\right), \text{if } \sigma - \beta v_i \le r_i < \sigma. \end{cases} \tag{F.4}$$

Inverse ad demand per viewer is

$$r_i = \begin{cases} \sigma - \sigma \frac{a_i}{m_1}, & \text{if } a_i < \frac{m_1 \beta v_i}{\sigma}, \\ \sigma - \frac{m_2 \beta}{m} v_i - \frac{\sigma a_i}{m}, & \text{if } \frac{m_1 \beta v_i}{\sigma} \le a_i < \frac{1}{\sigma}(m\sigma - \beta m_2 v_i). \end{cases}$$

If m_1 is sufficiently big, then only the type 1 advertisers are advertising in equilibrium, since the willingness to pay of type-2 advertisers is lower. In this case the market solves the problem of commercial media bias. If m_1 is sufficiently small, however, there exists a symmetric equilibrium where both type-1 and type-2 advertisers are served.[8] In this case, the profit of broadcaster i is

$$\pi_i = n\left(\frac{1}{N} + \frac{v_i - \delta a_i - u}{\tau}\right)\left(\sigma - \beta_2 v_i - \frac{\sigma a_i}{m}\right) a_i$$

where $\beta_2 := m_2 \beta / m < \beta$. The equilibrium values of program quality and advertising quantity can be found by replacing β by β_2 in the formulas in Proposition 1 and Lemma 1 in the paper.[9]

Consumer surplus can be calculated as in the paper (see equation (6.5) there). To calculate

[8] If the maximum quality $\bar{v}$ is sufficiently high, there also exist an equilibrium where all broadcasters choose very high quality and sell advertising spots only to type-1 producers. Any broadcaster trying to sell to type-2 producers would have to lower its quality so much that it would not have any viewers.

[9] The proof is similar to the proofs of these results, with one additional consideration: one has to take into consideration deviations to a high quality, whereby the deviating broadcaster captures all viewers, and serves only type-1 advertisers. The profits from this deviation are proportional to m_1; the deviation does not pay of m_1 is sufficiently small.

producer surplus, we need to take the two different types of advertisers into account:

$$PS = n\int_0^{\frac{m_1\beta v}{\sigma}} \left(\sigma - \sigma\frac{x}{m_1}\right)dx + n\int_{\frac{m_1\beta v}{\sigma}}^{a}\left(\sigma - \frac{m_2\beta}{m}v - \frac{\sigma x}{m}\right)dx.$$

We now consider the effect of a general cap that applies to the quantity of all advertising by type 1 and type 2 producers.

Proposition 8. *Consider the extension where there is a mass m_1 of advertisers with $\beta = 0$ and a mass $m_2 = m - m_1$ of advertisers with $\beta > 0$. Let $\delta = 0$. If both types of advertisers are served in equilibrium, a local cap on advertising increases welfare if and only if*

$$N > \hat{N}_{cap2} := \frac{2\beta^2\tau m_2}{\sigma\left(\beta\left(m - m_2\right) + 1\right)}.$$

Proof. Since by assumption $\delta = 0$,

$$W = nw + nv + n\int_0^{\frac{m_1\beta v}{\sigma}} \left(\sigma - \sigma\frac{x}{m_1}\right)dx + n\int_{\frac{m_1\beta v}{\sigma}}^{a}\left(\sigma - \frac{m_2\beta}{m}v - \frac{\sigma x}{m}\right)dx - \frac{n\tau}{4N}.$$

The effect of a cap on welfare is

$$\frac{dW}{d\bar{a}} = \frac{\partial W}{\partial a} + \frac{dv}{d\bar{a}}\frac{\partial W}{\partial v},$$

where

$$\frac{\partial W}{\partial a} = n\left(\sigma - \frac{m_2\beta}{m}v - \frac{\sigma a}{m}\right),$$

$$\frac{\partial W}{\partial v} = n - n\frac{m_2\beta}{m}\left(a - \frac{m_1\beta v}{\sigma}\right).$$

Moreover, from Lemma 1, replacing β by β_2, we have

$$\frac{dv}{d\bar{a}} = -\frac{1}{m}\frac{\sigma}{\beta_2} = -\frac{\sigma}{\beta m_2}.$$

Thus

$$\frac{dW}{d\bar{a}} = n\left(\sigma - \frac{m_2\beta}{m}v - \frac{\sigma a}{m}\right) - \frac{\sigma}{\beta m_2}\left(n - n\frac{m_2\beta}{m}\left(a - \frac{m_1\beta v}{\sigma}\right)\right). \tag{F.5}$$

The equilibrium values of a and v can be taken from Proposition 1 in the paper, replacing β by β_2 and setting $\delta = 0$:

$$a = \frac{m\beta_2\tau}{N\sigma} = \frac{\beta\tau m_2}{N\sigma},$$

$$v = \frac{\sigma}{\beta_2} - \frac{2\tau}{N} = \frac{m\sigma}{m_2\beta} - \frac{2\tau}{N}.$$

Inserting these into equation (F.5) shows that the effect of a local cap is

$$\frac{dW}{d\bar{a}} = \frac{n\beta\tau m_2}{Nm} - \frac{\sigma}{\beta m_2}\left(n - n\frac{m_2\beta}{m}\left(\frac{\beta\tau m_2}{N\sigma} - \frac{m_1\beta}{\sigma}\left(\frac{m\sigma}{m_2\beta} - \frac{2\tau}{N}\right)\right)\right)$$
$$= \frac{n\beta\tau m_2}{Nm} - \frac{\sigma}{m_2\beta}\left(n + n\beta\left(m_1 - \beta\tau\frac{m_2^2 + 2m_1m_2}{Nm\sigma}\right)\right). \quad \text{(F.6)}$$

This is strictly negative if and only if

$$\frac{\beta\tau m_2}{Nm} < \frac{\sigma}{m_2\beta}\left(1 + \beta\left(m_1 - \beta\tau\frac{m_2^2 + 2m_1m_2}{Nm\sigma}\right)\right).$$

or, equivalently (since $m = m_1 + m_2$), $N > \hat{N}_{cap2}$. □

To compare this with our main model, first note that when $m_2 \to m$, we get the same condition as in Proposition 1 in the paper for the case $\delta = 0$. Moreover, $\hat{N}_{cap2}$ is increasing in m_2. Thus the lower m_2, the more likely it is that a cap improves welfare. Therefore, in the extension considered in Proposition 8, it is *more* likely than in our main model that a cap improves welfare. The intuition is that, since only some broadcasters suffer from higher program quality, the loss of producer surplus due to a cap is not as important as in the main model in the paper.

As reported above, many countries impose bans on advertising for specific sectors or products, for example tobacco or alcohol. To see the implications in our model, consider a sector specific advertising ban that excludes all type 2 advertisers. Then for the broadcasters there is no drawback from choosing high program quality; thus in equilibrium program quality will be equal to its highest possible level $\bar{v}$. If $\bar{v}$ is sufficiently high, a sector specific advertising ban leads to a higher welfare than laissez-faire, or a local cap on all advertising. While most rationales for regulating the content of advertising are built on bounded consumer rationality, this argument identifies conditions such that regulating advertising content is justifiable for the reason it decreases commercial media bias, even when consumers are perfectly rational.

F.4 Robustness

The following section contains robustness checks with respect concerning television viewing behavior and advertising demand.

F.4.1 Television viewing behavior

This section probes the robustness of our results with respect to the model of television viewing behavior, focusing on the case where the number of broadcasters is exogenous. One limitation of our results comes from the assumption that everyone watches television. If the number of viewers is endogenous, and viewers are ad averse, a cap on advertising will ceteris paribus increase the total number of viewers. This means that increasing v has higher costs for a broadcaster, because he loses βa on every viewer he has, countervailing the quality improving effects of a cap.[10]

Our results do not hinge, however, on specific features of the Salop circle model. To show this, we introduce a more general model of television viewing behavior that nests the Salop model and several other textbook models of discrete choice. Suppose that, if all broadcasters $j \neq i$ behave symmetrically, the fraction of viewers who watch broadcaster i depends only on the difference between the utility $v_i - \delta a_i$ offered by broadcaster i, and the utility offered by the competitors, scaled by a factor $1/\tau$. The share is given by $s\left(\frac{v_i - \delta a_i - u}{\tau}\right)$, where $u := v_j - \delta a_j$, and s is a strictly increasing function with $s(0) = 1/N$. In general, the function s will depend on N; we assume it to be independent of the other exogenous parameters of the model. We assume that the function s is sufficiently well behaved such that a symmetric equilibrium in pure strategies exists and can be characterized by the relevant first order conditions. This model nests the Salop model with linear transportation costs studied in the paper (given undercutting is not an issue), the Salop model with any convex (e.g. quadratic) transportation costs, the Logit model (see, e.g., Anderson et al., 1992), and the covered Spokes model introduced by Chen and Riordan (2007) and used in a study of commercial media bias by Germano and Meier (2013).

Sections F.4.3 to F.4.3 below substantiate these claims, and Section F.4.3 characterizes the symmetric equilibrium with and without a cap. Here we summarize the main results. A cap increases program quality, and indeed $dv/(d\bar{a}) = -\sigma/\beta m$ exactly as in the paper, see equation (6.13) there. The comparative static of the equilibrium depends on the behavior of $Ns'(0)$. The advertising quantity a is decreasing in N, and quality v is increasing in N, if and only if $Ns'(0)$ is increasing in N. Similarly, the price of an advertising spot per viewer r is decreasing in N if and only if $Ns'(0)$ is increasing in N. As discussed in the paper, this seems to be the empirically plausible case. The Salop model with linear or strictly convex transportation costs, and the Logit model share this property that $Ns'(0)$ is increasing in N. The Spokes model is a limit case where $Ns'(0)$ is independent of N.

For the welfare analysis, we assume that, for a given number of broadcasters N, welfare is given by

$$W = n(w + v - \delta a) + n\int_0^a \left(\sigma - \beta v - \frac{\sigma}{m}x\right)dx - NF + C(N),$$

where $C(N)$ is independent of a and v. This is the case in all the discrete choice models mentioned

[10] Indeed, if ad aversion is strong, a cap may decrease equilibrium program quality. The simplest way to see this is to reconsider the example from Section 6.3 of the paper, and to introduce ad aversion $\delta > 0$. Then a cap will increase program quality whenever $\delta < 1$, but it will decrease quality when $\delta > 1$.

above (e.g. in the Salop model, $C(N)$ equals aggregate transportation costs). As in the paper, a cap increases consumer surplus, and decreases producer surplus. Moreover, the welfare analysis of a local cap is quite similar to our main model. A local cap improves welfare if and only if $Ns'(0) > \hat{N}_{cap}$. This is, in addition, a sufficient but not necessary condition for the optimal cap subject to the constraint that profits are nonnegative to be binding. The optimal policy is either to choose a cap that drives profits down to zero, i.e., $\bar{a} = FN^2 s'(0) / (n\beta\tau)$, or laissez-faire.

As seen above, many standard discrete choice models imply that $Ns'(0)$ is increasing in N. Moreover, if $Ns'(0)$ is increasing in N, then the comparative statics of equilibrium advertising quantity, program quality, and price per ad per viewer go in empirically plausible directions. It is exactly this property that also gives rise to the complementarity between competition and regulation: If $Ns'(0)$ is increasing in N, then an increase in N makes it more likely that a local cap raises welfare.

Since $s(\cdot)$ is independent of the remaining parameters of the model, their impact is exactly as in the linear Salop model considered in the paper.[11] Table F.2 lists the market share, the condition under which a local cap improves welfare, and the optimal cap (if binding) for several discrete choice models nested in our general model. It shows that the conditions under which a local cap improves welfare are qualitatively similar. Moreover, the optimal cap has the same qualitative properties under these models. Table F.2 also shows, however, that the precise quantitative implications depend on the assumed model of television viewing. The condition under which a local cap improves welfare is the most restrictive in the Logit model, followed in decreasing strength by the Spokes model, the Salop model with linear transportation costs and the Salop model with quadratic transportation costs.

Table F.2: Different assumptions on television viewing behavior.

Model	$s\left(\frac{u_i-u}{\tau}\right)$	$Ns'(0)$	local cap improves W iff	zero-profit cap
Salop linear	$\frac{1}{N}+\frac{u_i-u}{\tau}$	N	$N > \hat{N}_{cap}$	$\frac{FN^2}{n\beta\tau}$
Salop quadratic	$\frac{1}{N}+\frac{N(u_i-u)}{\tau}$	N^2	$N^2 > \hat{N}_{cap}$	$\frac{FN^3}{n\beta\tau}$
Spokes	$\frac{1}{N}+\frac{u_i-u}{N\tau}$	1	$1 > \hat{N}_{cap}$	$\frac{FN}{n\beta\tau}$
Logit	$\frac{e^{u_i/\tau}}{e^{u_i/\tau}+(N-1)e^{u/\tau}}$	$\frac{N-1}{N}$	$\frac{N-1}{N} > \hat{N}_{cap}$	$\frac{F(N-1)}{n\beta\tau}$

Notes: Salop linear (quadratic) refers to the Salop model with linear (quadratic) transportation costs.

[11] Note that even in the Spokes model a lower τ, i.e., a more competitive broadcasting market since programs are better substitutes, makes a cap more likely to be welfare enhancing.

F.4.2 Advertising demand

In our paper we assume that a higher program quality reduces the willingness to pay of all advertisers by the same amount. In a diagram with advertising quantity on the horizontal axis, and the price of an ad per viewer on the vertical axis, the inverse demand curve for advertising spots is linear, and a higher program quality leads to a parallel downward shift of the inverse demand curve; see the left hand side of Figure F.1. The reduction of the willingness to pay, however, may depend on the quality of the good. It seems plausible to assume that the willingness to pay of producers of high (rather than low) quality goods is less affected by program quality. Moreover, there might be nonlinearities in the inverse advertising demand curve. For example, advertisers may have increasing marginal costs from program quality, or similarly, viewers' marginal utility from program quality may be decreasing.

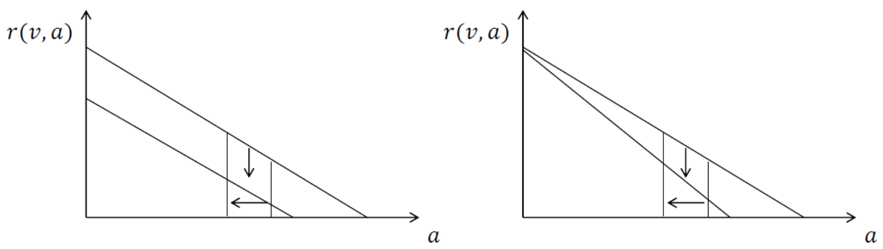

Figure F.1: Left hand side: Parallel downwards shift of the inverse ad demand function. Right hand side: Turn of the inverse ad demand function.

To study potential implications, suppose that the inverse demand for advertising is given by a function $r(a, v)$, which is decreasing in a and in v.[12] While a complete analysis is beyond the scope of this Appendix, we show that the plausible assumption that producers of high quality goods have little to lose from program quality strengthens the mechanism by which a cap increases program quality. Formally, the assumption means that the cross-partial derivative $r_{va} := \partial^2 r / (\partial v \partial a)$ is negative. In terms of the right hand side of Figure F.1, an increase in quality makes the inverse ad demand curve steeper.

[12] Note this can also capture decreasing marginal utility of the viewers. Suppose a viewer's utility is given by $w + f(v_i) - \delta a_i - \tau x$, where $f'(v) > 0 > f''(v)$, instead of equation (6.2), and r_i is as given by $r_i = \sigma - \beta v_i - \frac{a_i \sigma}{m}$. Then we can equivalently think of the broadcaster choosing $\tilde{v}_i := f(v_i)$; then viewers utility from watching is $w + \tilde{v}_i - \delta a_i - \tau x$ as in (6.2), but $r_i = \sigma - \beta f^{-1}(\tilde{v}_i) - \sigma a_i / m$, thus the advertisers have increasing marginal costs from program quality.

Consider the profit maximization problem of broadcaster i, given a binding cap $\bar{a}$:

$$\max_{v_i} ns\left(\frac{v_i - \delta\bar{a} - u}{\tau}\right) r\left(\bar{a}, v_i\right)\bar{a} - F$$

where $u = v_j - \delta\bar{a}$ for all $j \neq i$. Setting up the first order condition for the profit maximizing program quality, and then using symmetry, one obtains

$$\frac{s'\left(0\right)}{\tau} r\left(\bar{a}, v\right) + \frac{1}{N} r_v\left(\bar{a}, v\right) = 0$$

where r_v is the partial derivative $\partial r/\partial v$. Consider how the equilibrium value of v changes when the cap $\bar{a}$ changes. From the implicit function rule,

$$\frac{dv}{d\bar{a}} = -\frac{\frac{s'(0)}{\tau} r_a + \frac{1}{N} r_{va}}{\frac{s'(0)}{\tau} r_v + \frac{1}{N} r_{vv}}, \tag{F.7}$$

where as above subscripts indicate partial derivatives. The denominator is negative by the second order condition, and $r_a < 0 < s'\left(0\right)$ by assumption. Therefore, if $r_{va} = 0$, then $dv/da < 0$, and a cap improves program quality as in the paper; the sign of the second order partial derivatives r_{vv} and r_{aa} does not matter for this result.[13] As can be seen from (F.7), $r_{va} < 0$ reinforces the quality enhancing effect of the cap.

The economics behind this is as follows. A tighter cap implies that the marginal advertiser has a better product. When $r_{va} < 0$, it follows that the willingness to pay of the marginal advertiser decreases less in program quality. Therefore, broadcasters have an additional reason to increase their program quality. Similarly, any reason why the marginal advertiser has a lower stake in program quality works in the same direction. On the other hand, when the marginal advertiser has a higher stake in program quality (perhaps because of the advertisers' cost structure), this works against the quality enhancing effect of a cap.

[13]While the curvature of r in v does not matter for the sign of dv/da, it influences its absolute value. This can change the results on the desirability of a cap and on the local complementarity between competition and regulation. To see this, consider the case where there is some upper bound $\bar{v}$ on quality. Such an upper bound might endogenously arise as the result of viewers' utility being increasing in v only up to $\bar{v}$, or from advertisers' willingness to pay dropping rapidly when quality exceeds $\bar{v}$. Then, enough competition on the media market can be sufficient to ensure that equilibrium program quality is $\bar{v}$, and thus as high as it possibly can be. If, in addition, ad aversion is small, a cap on advertising will be detrimental to welfare.

F.4.3 Technical Details

Salop model with convex transportation costs

Suppose that the transportation costs of a viewer located at a distance x from the broadcaster equals $\tau l(x)$, where l is strictly increasing, strictly convex, and satisfies $l(0) = 0$. Let $u_i = v_i - \delta a_i$, and $u = v_j - \delta a_j$ for all $j \neq i$. Assuming that there is an indifferent viewer between broadcaster i and its closest competitors, the distance x between this viewers and broadcaster i solves

$$u_i - \tau l(x) = u - \tau l\left(\frac{1}{N} - x\right),$$

or equivalently

$$\frac{u_i - u}{\tau} = l(x) - l\left(\frac{1}{N} - x\right).$$

Let $\lambda(x, N) = l(x) - l\left(\frac{1}{N} - x\right)$. Since λ is strictly increasing in x, holding N fixed, an inverse function $\lambda^{-1}(\cdot, N)$ exists, and

$$\lambda^{-1}\left(\frac{u_i - u}{\tau}, N\right) = x.$$

Therefore, the market share of i is

$$s\left(\frac{u_i - u}{\tau}\right) = 2x = 2\lambda^{-1}\left(\frac{u_i - u}{\tau}, N\right).$$

Thus

$$s'\left(\frac{u_i - u}{\tau}\right) = \frac{2}{\lambda'\left(\lambda^{-1}\left(\frac{u_i - u}{\tau}, N\right)\right)} = \frac{2}{l'(x) + l'\left(\frac{1}{N} - x\right)}.$$

When $u_i - u = 0$, $x = 1/(2N)$. Thus

$$Ns'(0) = \frac{N}{l'\left(\frac{1}{2N}\right)}.$$

Therefore

$$\frac{d}{dN}\left(Ns'(0)\right) = \frac{l'\left(\frac{1}{2N}\right) + l''\left(\frac{1}{2N}\right)\frac{1}{2N}}{l'\left(\frac{1}{2N}\right)^2} > 0.$$

Thus $Ns'(0)$ is increasing in N.

It remains to establish that the first order conditions are sufficient for a maximum. Given the convex transportation costs, undercutting is not an issue. To show that any critical point is a global maximum, we now show that the variable profit, $\pi_i + F$, is log-concave in (a_i, v_i).

We begin by establishing that $\ln s\left(\frac{u_i-u}{\tau}\right)$ is strictly concave in u_i.

$$\frac{\partial^2}{\partial u_i^2}\left(\ln s\left(\frac{u_i-u}{\tau}\right)\right)=\frac{1}{\tau^2\left(s\left(\frac{u_i-u}{\tau}\right)\right)^2}\left(s\left(\frac{u_i-u}{\tau}\right)s''\left(\frac{u_i-u}{\tau}\right)-\left(s'\left(\frac{u_i-u}{\tau}\right)\right)^2\right)$$

is strictly smaller zero if, and only if,

$$s\left(\frac{u_i-u}{\tau}\right)s''\left(\frac{u_i-u}{\tau}\right)<\left(s'\left(\frac{u_i-u}{\tau}\right)\right)^2.$$

Here, this inequality is equivalent to

$$1+\lambda^{-1}\left(\frac{u_i-u}{\tau},N\right)\frac{\lambda''\left(\lambda^{-1}\left(\frac{u_i-u}{\tau},N\right)\right)}{\lambda'\left(\lambda^{-1}\left(\frac{u_i-u}{\tau},N\right)\right)}>0$$

which is true since $\lambda^{-1}>0$, $\lambda'>0$ and $\lambda''>0$.

It follows that

$$\ln\left(\pi_i+F\right)=\ln s\left(\frac{u_i-u}{\tau}\right)+\ln\left(\sigma-\beta v_i-\frac{\sigma a_i}{m}\right)+\ln a_i$$

is the sum of three functions, each of them being weakly concave in (a_i,v_i). Moreover, the first of these functions is strictly concave in (a_i,v_i) except along the line where $v_i-\delta a_i$ is constant; the second is strictly concave except along the line where $\beta v_i+\frac{\sigma a_i}{m}$ is constant. It follows that $\ln\left(\pi_i+F\right)$ is strictly concave in (a_i,v_i).

Logit

In the logit model, the market share of broadcaster i is

$$s\left(\frac{u_i-u}{\tau}\right)=\frac{e^{u_i/\tau}}{e^{u_i/\tau}+(N-1)\,e^{u/\tau}}$$

(see, e.g., Anderson et al., 1992). It is straightforward to calculate that

$$Ns'(0)=\frac{N-1}{N}.$$

Moreover, it can be shown that the variable profit (π_i+F) is log-concave in (a_i,v_i). Therefore, the first order conditions are sufficient for a maximum.

Spokes

In the covered Spokes model introduced by Chen and Riordan (2007) and used in Germano and Meier (2013),

$$s_i\left(\frac{u_i-u}{\tau}\right)=\frac{1}{N}+\frac{1}{N}\frac{u_i-u}{\tau}.$$

Then $Ns'(0)=1$ is independent of N. Again, the variable profit is log-concave in (a_i,v_i) and therefore the first order conditions are sufficient for a maximum.

Equilibrium characterization and welfare effects of a cap

The profit of broadcaster i is

$$\pi_i=ns\left(\frac{v_i-\delta a_i-u}{\tau}\right)\left(\sigma-\beta v_i-\frac{\sigma}{m}a_i\right)a_i-F.$$

The first order condition

$$\frac{\partial\pi_i}{\partial v_i}=s'\left(\frac{v_i-\delta a_i-u}{\tau}\right)\frac{1}{\tau}n\left(\sigma-\beta v_i-\frac{\sigma}{m}a_i\right)a_i-\beta ns\left(\frac{v_i-\delta a_i-u}{\tau}\right)a_i=0$$

simplifies to, assuming symmetry,

$$r\equiv\sigma-\beta v-\frac{\sigma}{m}a=\frac{\beta\tau}{Ns'(0)}.$$

Consider the case without a cap. The first order condition

$$\begin{aligned}\frac{\partial}{\partial a_i}\pi_i=ns'\left(\frac{v_i-\delta a_i-u}{\tau}\right)\left(-\frac{\delta}{\tau}\right)\left(\sigma-\beta v_i-\frac{\sigma}{m}a_i\right)a_i+ns\left(\frac{v_i-\delta a_i-u}{\tau}\right)\left(-\frac{\sigma}{m}\right)a_i\\+ns\left(\frac{v_i-\delta a_i-u}{\tau}\right)\left(\sigma-\beta v_i-\frac{\sigma}{m}a_i\right)=0\end{aligned}$$

simplifies, in any symmetric equilibrium, to

$$-\frac{\delta}{\tau}s'(0)\,ra-\frac{1}{N}\frac{\sigma}{m}a+\frac{r}{N}=0.$$

Inserting $r=\beta\tau/(Ns'(0))$ and solving for a gives

$$a=\frac{\beta}{Ns'(0)}\frac{m\tau}{\sigma+m\beta\delta}.$$

Moreover,

$$v = \frac{\sigma}{\beta} - \frac{\tau}{Ns'(0)} \frac{2\sigma + m\beta\delta}{\sigma + m\beta\delta}.$$

If there is a binding cap $\bar{a}$, then in any symmetric equilibrium

$$v = \frac{\sigma}{\beta} - \frac{\sigma}{\beta m}\bar{a} - \frac{\tau}{Ns'(0)}.$$

Inserting the equilibrium value of v into the welfare function, we get

$$\begin{aligned} W(\bar{a}) &= n\left(w + \frac{\sigma}{\beta} - \frac{\sigma}{\beta m}\bar{a} - \frac{\tau}{Ns'(0)} - \delta\bar{a}\right) + n\int_0^{\bar{a}} \left(\beta\left(\frac{\sigma}{\beta m}\bar{a} + \frac{\tau}{Ns'(0)}\right) - \frac{\sigma}{m}x\right) dx - NF + C \\ &= n\left(-\frac{\sigma}{\beta m}\bar{a} - \delta\bar{a}\right) + \frac{n\sigma}{2m}\bar{a}^2 + \frac{n\beta\bar{a}\tau}{Ns'(0)} + \text{terms independent of } \bar{a}. \end{aligned}$$

Consider the problem to maximize $W(\bar{a})$ by choosing a, subject to the constraint that profits are nonnegative:

$$\pi_i = \frac{n}{N}\frac{\beta\tau}{Ns'(0)}\bar{a} - F \geq 0.$$

Since $W(\bar{a})$ is convex in $\bar{a}$, either the optimal cap is driving profits to zero, or it is not binding. Moreover,

$$\frac{dW}{d\bar{a}} = n\left(-\frac{\sigma}{\beta m} - \delta\right) + n\frac{\bar{a}}{m}\sigma + \frac{n\beta\tau}{Ns'(0)}.$$

Evaluating this at the equilibrium level of a (absent a cap) gives

$$\frac{dW}{d\bar{a}} = n\left(-\frac{\sigma}{\beta m} - \delta\right) + \frac{n\frac{\beta}{Ns'(0)}\frac{m\tau}{\sigma + m\beta\delta}}{m}\sigma + \frac{n\beta\tau}{Ns'(0)} = n\left(-\frac{\sigma}{\beta m} - \delta\right) + \frac{n\beta\tau}{Ns'(0)}\frac{2\sigma + m\beta\delta}{\sigma + m\beta\delta}.$$

Rearranging shows that this is strictly negative if and only if

$$Ns'(0) > \frac{(2\sigma + m\beta\delta)\, m\beta^2\tau}{(\sigma + m\beta\delta)^2} = \hat{N}_{cap}.$$

Thus a local cap raises welfare if and only if $Ns'(0) > \hat{N}_{cap}$.

F.5 Omitted Steps of Proof of Proposition 1

In Appendix E, we have shown that if a symmetric equilibrium exists, it is as described in Proposition 1 in the paper. Here we prove existence of the equilibrium. We suppose all broadcasters $j \neq i$ behave as indicated in the Proposition and show that broadcaster i has no incentive to deviate. Since the proof is somewhat lengthy, we break it down into steps, for which we first briefly sketch the intuition. *Step 1* assumes that broadcaster i does not undercut and shows that, under this assumption, broadcaster i has no incentive to deviate. The remaining steps consider deviations that involve undercutting. *Step 2* prepares the ground by describing the range of a_i and v_i leading to undercutting. From here, it is straightforward to show that undercutting more than two rivals is not profitable: it leads to zero inverse ad demand and hence to a profit of zero (*Step 3*). The remaining steps consider undercutting two rivals. *Step 4* considers the case of $N = 3$. Here, by undercutting two rivals, broadcaster i captures all the market. It will choose v_i such that it just undercuts its two rivals since this is sufficient to make all viewers watch broadcaster i. We show that the resulting profit is smaller than the equilibrium profit. *Steps 5* and *6* consider undercutting two rivals in case of $N > 3$. Here, by undercutting two rivals, broadcaster i does not capture all viewers. *Step 5* shows that broadcaster i will not increase its program quality more than necessary to just undercut two rivals. The intuition is that at the equilibrium broadcaster i is already indifferent whether or not to increase its program quality a bit, thereby winning viewers but gaining a lower price for ads. At the considered deviation, broadcaster i already has more viewers than in equilibrium, and thus prefers not to increase its program quality any more. *Step 6* shows that the deviation profit from just undercutting two rivals is smaller than the equilibrium profit.

Step 1: Find the profit maximizing decisions of broadcaster i assuming that broadcaster i does not undercut any rival.

Suppose all broadcasters $j \neq i$ behave as indicated. Then

$$u = v_j - \delta a_j = \frac{\sigma}{\beta} - \frac{2\tau}{N}.$$

The profit of broadcaster i is

$$\pi_i = n\left(\frac{1}{N} + \frac{v_i - \delta a_i - \left(\frac{\sigma}{\beta} - \frac{2\tau}{N}\right)}{\tau}\right)\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) a_i$$

whenever

$$\left(1 - \frac{a_i}{m}\right)\frac{\sigma}{\beta} > v_i > \delta a_i + \left(\frac{\sigma}{\beta} - \frac{2\tau}{N}\right) - \frac{\tau}{N}.$$

Otherwise, profit is zero: if the first inequality does not hold, inverse ad demand is zero, if the second inequality does not hold, broadcaster i has no viewers. Therefore, broadcaster i will choose an $a_i > 0$ such that

$$\frac{3\tau}{N\left(\frac{\sigma}{\beta m}+\delta\right)} > a_i > 0. \tag{F.8}$$

We first consider the profit maximizing v_i for a given a_i satisfying (F.8). Note that π_i is strictly concave in v_i. Solving the first order condition

$$\frac{\partial \pi_i}{\partial v_i} = \frac{n}{\tau}\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) a_i - \beta n \left(\frac{1}{N} + \frac{v_i - \delta a_i - \left(\frac{\sigma}{\beta} - \frac{2\tau}{N}\right)}{\tau}\right) a_i = 0$$

for v_i shows that the profit maximizing program quality is

$$v_i^*\left(a_i\right) = \frac{1}{2}\left(\left(1 - \frac{a_i}{m}\right)\frac{\sigma}{\beta} + \delta a_i + \left(\frac{\sigma}{\beta} - \frac{2\tau}{N}\right) - \frac{\tau}{N}\right).$$

Substituting $v_i^*\left(a_i\right)$ into the profit of broadcaster i gives

$$\pi_i\left(a_i, v_i^*\left(a_i\right)\right) = \frac{n\beta}{4\tau}\left(\frac{3\tau}{N} - a_i\left(\frac{\sigma}{\beta m} + \delta\right)\right)^2 a_i.$$

The first order condition

$$\frac{d}{da_i}\pi_i\left(a_i, v_i^*\left(a_i\right)\right) = 0$$

has the solutions

$$a_{i1} = \frac{3\tau}{N\left(\frac{\sigma}{\beta m}+\delta\right)}$$

corresponding to the upper bound on a_i in (F.8), and

$$a_{i2} = \frac{a_{i1}}{3} = \frac{m\beta\tau}{N\left(\sigma + m\beta\delta\right)}$$

which is equation (6.11) in the paper. Moreover, it is straightforward to show that $\pi_i\left(a_i, v_i^*\left(a_i\right)\right)$ as a function of a_i is: zero at $a_i = 0$, strictly concave when $a_i < 2a_{i1}/3$, strictly convex when $2a_i/3 < a_i < a_{i1}$, and zero at $a_i = a_{i1}$. It follows that a_{i2} maximizes profit. Noting that $v_i^*\left(a_{i2}\right)$ is the value of v_i given in the Proposition completes step 1.

Step 2: Describe the range of a_i and v_i leading to undercutting.

Suppose broadcaster j is a distance k/N away from broadcaster i. A consumer with ideal point at the location of broadcaster j is indifferent between the products of broadcasters i and j if

$$v_i - \delta a_i - \frac{k\tau}{N} = v_j - \delta a_j.$$

Consider a unilateral deviation of broadcaster i, while all broadcasters except i stick to the equilibrium strategies, i.e., $v_j - \delta a_j = \frac{\sigma}{\beta} - \frac{2\tau}{N}$. Thus, the consumer is indifferent if

$$v_i - \delta a_i = \frac{\sigma}{\beta} - \frac{2\tau}{N} + \frac{k\tau}{N}.$$

For $k = 1, 2, \ldots$, the values of (a_i, v_i) satisfying this equation are the points of discontinuity of the demand of broadcaster i. Note that the discontinuity at $k = 1, 2, \ldots$ corresponds to just undercutting $2k$ rivals. For simplicity and w.l.o.g., we break all ties in favor of broadcaster i, i.e., we assume that if a broadcaster deviates then any consumer that is indifferent between the deviating and another broadcaster watches the deviating broadcaster.

To summarize, broadcaster i undercuts no rival if

$$v_i - \delta a_i < \frac{\sigma}{\beta} - \frac{2\tau}{N} + \frac{\tau}{N}.$$

Broadcaster i undercuts exactly $2k$ rivals, $k = 1, 2, \ldots$, if

$$\frac{\sigma}{\beta} - \frac{2\tau}{N} + \frac{k\tau}{N} \leq v_i - \delta a_i < \frac{\sigma}{\beta} - \frac{2\tau}{N} + \frac{(k+1)\,\tau}{N}.$$

Step 3: Broadcaster i cannot make a positive profit by undercutting more than 2 broadcasters.

By the symmetry of the behavior of broadcasters $j \neq i$, if broadcaster i undercuts more than 2 broadcasters, it undercuts 4 or more broadcasters. To undercut 4 or more broadcasters, the program quality of broadcaster i needs to be $v_i \geq \sigma/\beta + \delta a_i$. However, then $\sigma - \beta v_i - \sigma a_i/m < 0$, thus inverse ad demand is zero.

Step 4: Undercutting two rivals: the case $N = 3$.

Suppose that broadcaster i undercuts exactly 2 rivals. That is,

$$\frac{\sigma}{\beta} - \frac{\tau}{N} \leq v_i - \delta a_i < \frac{\sigma}{\beta}.$$

If $N = 3$, this means broadcaster i has a market share of 1. The profit of broadcaster i is then

$$n\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) a_i. \tag{F.9}$$

Note this is strictly decreasing in v_i, therefore, the optimal undercutting of two rivals satisfies

$$\frac{\sigma}{\beta} - \frac{\tau}{3} = v_i - \delta a_i.$$

Solving for v_i gives

$$v_i = \hat{v}_i(a_i) := \frac{\sigma}{\beta} - \frac{\tau}{3} + \delta a_i,$$

substituting in (F.9) shows that the profit is

$$\pi_i^{dev}(a_i, \hat{v}_i(a_i)) = n\left(\sigma - \beta\left(\frac{\sigma}{\beta} - \frac{\tau}{3} + \delta a_i\right) - \frac{\sigma a_i}{m}\right)a_i.$$

Note $\pi_i^{dev}(a_i, \hat{v}_i(a_i))$ is strictly concave in a_i. The first order condition

$$\frac{d}{da_i}\pi_i^{dev}(a_i, \hat{v}_i(a_i)) = 0$$

has the unique solution

$$a_i^{dev} = \frac{m\beta\tau}{6(\sigma + m\beta\delta)}.$$

Moreover,

$$\hat{v}_i\left(a_i^{dev}\right) = \frac{\sigma}{\beta} - \frac{\tau}{3} + \delta\frac{m\beta\tau}{6(\sigma + m\beta\delta)}.$$

The profit from the deviation is

$$\pi_i^{dev}\left(a_i^{dev}, \hat{v}_i\left(a_i^{dev}\right)\right) = \frac{3}{4}\frac{nm\beta^2\tau^2}{27(\sigma + m\beta\delta)},$$

which is $3/4$ of the equilibrium profit given in the Proposition.

It follows that in case $N = 3$, broadcaster i has no incentive to undercut two rivals.

Step 5: In case $N > 3$, from all strategies involving undercutting two rivals, the best strategy is at a point of discontinuity of demand.

Suppose broadcaster i undercuts exactly 2 rivals:

$$\frac{\sigma}{\beta} - \frac{\tau}{N} \leq v_i - \delta a_i < \frac{\sigma}{\beta}.$$

If $N > 3$, broadcaster i has a market share of

$$\frac{3}{N} + \frac{v_i - \delta a_i - \left(\frac{\sigma}{\beta} - \frac{\tau}{N}\right)}{\tau}.$$

Profit is

$$\pi_i^{Dev}(a_i, v_i) = n\left(\frac{3}{N} + \frac{v_i - \delta a_i - \left(\frac{\sigma}{\beta} - \frac{\tau}{N}\right)}{\tau}\right)\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right)a_i.$$

We will show that this profit is maximal whenever i just undercuts two rivals, i.e., when

$$\frac{\sigma}{\beta} - \frac{\tau}{N} = v_i - \delta a_i. \tag{F.10}$$

To see this, suppose that broadcaster i undercuts 2 rivals, and

$$\frac{\sigma}{\beta} - \frac{\tau}{N} < v_i - \delta a_i < \frac{\sigma}{\beta}.$$

For a fixed a_i,

$$\frac{\partial}{\partial v_i}\pi_i^{Dev} = \left(\frac{1}{\tau}\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) - \beta\left(\frac{3}{N} + \frac{v_i - \delta a_i - \left(\frac{\sigma}{\beta} - \frac{\tau}{N}\right)}{\tau}\right)\right) n a_i.$$

Note that π_i^{Dev} is strictly concave in v_i. Moreover, at $v_i = \frac{\sigma}{\beta} - \frac{\tau}{N} + \delta a_i$, we have

$$\begin{aligned}\left.\frac{\partial}{\partial v_i}\pi_i^{Dev}\right|_{v_i = \left(\frac{\sigma}{\beta} - \frac{\tau}{N} + \delta a_i\right)} &= \left(\frac{1}{\tau}\left(\frac{\beta\tau}{N} - \beta\delta a_i - \frac{\sigma a_i}{m}\right) - \beta\frac{3}{N}\right) n a_i \\ &= -\left(\frac{1}{\tau}\left(\beta\delta a_i + \frac{\sigma a_i}{m}\right) + \beta\frac{2}{N}\right) n a_i < 0.\end{aligned}$$

Therefore, in the relevant range π_i^{Dev} is strictly decreasing in v_i for fixed a_i. It follows that the best strategy involving undercutting two rivals must satisfy equation (F.10).

Step 6: In case $N > 3$, broadcaster i has no incentive to just undercut 2 rivals.

Suppose broadcaster i just undercuts 2 rivals, i.e., equation (F.10) holds. Then

$$\pi_i^{Dev} = n\frac{3}{N}\left(\sigma - \beta v_i - \frac{\sigma a_i}{m}\right) a_i.$$

Solve equation (F.10) for $v_i = \frac{\sigma}{\beta} - \frac{\tau}{N} + \delta a_i$ and substitute into π_i^{Dev} to get

$$\pi_i^{Dev} = n\frac{3}{N}\left(\sigma - \beta\left(\frac{\sigma}{\beta} - \frac{\tau}{N} + \delta a_i\right) - \frac{\sigma a_i}{m}\right) a_i.$$

Note that this is a strictly concave function of a_i. The first order condition

$$\frac{\partial}{\partial a_i}\pi_i^{Dev} = 0$$

has the unique solution

$$a_i = \frac{1}{2N} m\beta \frac{\tau}{\sigma + m\beta\delta}.$$

Inserting this into π_i^{Dev} gives

$$\pi_i^{dev} = \frac{3}{4}\frac{nm\beta^2\tau^2}{N^3\left(\sigma + m\beta\delta\right)}$$

which is $3/4$ of equilibrium profit. It follows that broadcaster i has no incentive to just undercut two rivals.

Bibliography

Aalberg, Toril, Peter Van Aelst, and James Curran (2010) "Media systems and the political information environment: A cross-national comparison," *The International Journal of Press/Politics*, **15** (3), 255–271.

Adamic, Lada A and Natalie Glance (2005) "The political blogosphere and the 2004 US election: divided they blog," in *Proceedings of the 3rd international workshop on Link discovery*, 36–43, ACM.

Aday, Sean (2010) "Chasing the bad news: An analysis of 2005 Iraq and Afghanistan war coverage on NBC and Fox News Channel," *Journal of communication*, **60** (1), 144–164.

AdvertisingAge, (2013) "Three Regulation Trends to Watch in 2013," *URL: https://adage.com/article/special-report-look-ahead-2013/regulation-trends-watch-2013/239090.*

An, Jisun, Meeyoung Cha, Krishna Gummadi, Jon Crowcroft, and Daniele Quercia (2012) "Visualizing media bias through Twitter," in *Sixth International AAAI Conference on Weblogs and Social Media*, 2–5, AAAI.

An, Jisun, Daniele Quercia, and Jon Crowcroft (2014) "Partisan sharing: Facebook evidence and societal consequences," in *Proceedings of the second ACM conference on Online social networks*, 13–24, ACM.

An, Soontae and Lori Bergen (2007) "Advertiser pressure on daily newspapers: A survey of advertising sales executives," *Journal of Advertising*, **36** (2), 111–121.

Anderson, Ashton, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec (2013) "Steering User Behavior with Badges," in *Proceedings of the 22Nd International Conference on World Wide Web*, WWW '13, 95–106, ACM.

Anderson, Chris (2006) *The long tail: Why the future of business is selling less of more*: Hachette Books.

Anderson, Michael and Jeremy Magruder (2012) "Learning from the Crowd: Regression Discontinuity Estimates of the Effects of an Online Review Database," *Economic Journal*, **122** (563), 957–989.

Anderson, Simon P (2007) "Regulation of television advertising," *The economic regulation of broadcasting markets*, **189**, 224.

Anderson, Simon P and Stephen Coate (2005) "Market provision of broadcasting: A welfare analysis," *The review of Economic studies*, **72** (4), 947–972.

Anderson, Simon P., Andre De Palma, and Jacques-Francois Thisse (1992) *Discrete choice theory of product differentiation*: MIT press.

Anderson, Simon P, Øystein Foros, Hans Jarle Kind, and Martin Peitz (2012) "Media market concentration, advertising levels, and ad prices," *International Journal of Industrial Organization*, **30** (3), 321–325.

Anderson, Simon P and Jean J Gabszewicz (2006) "The media and advertising: a tale of two-sided markets," *Handbook of the Economics of Art and Culture*, **1**, 567–614.

Anderson, Simon P and Joshua S Gans (2011) "Platform siphoning: Ad-avoidance and media content," *American Economic Journal: Microeconomics*, **3** (4), 1–34.

Anderson, Simon P. and Bruno Jullien (2016) "The advertising-financed business model in two-sided media markets," in *Handbook of Media Economics*, **1**: Elsevier, 41–90.

Angrist, Joshua D and Guido W Imbens (1995) "Identification and Estimation of Local Average Treatment Effects," Working Paper 118, National Bureau of Economic Research.

Aral, Sinan and Dylan Walker (2012) "Identifying Influential and Susceptible Members of Social Networks," *Science*, **337** (6092), 337–341.

Armstrong, Mark and Helen Weeds (2007) "Programme quality in subscription and advertising-funded television," Unpublished.

Aruguete, Natalia and Ernesto Calvo (2018) "Time to# protest: Selective exposure, cascading activation, and framing in social media," *Journal of Communication*, **68** (3), 480–502.

Athey, Susan, Emilio Calvano, and Joshua Gans (2013) "The impact of the internet on advertising markets for news media," working paper, National Bureau of Economic Research.

Bagwell, Kyle (2007) "The economic analysis of advertising," in *Handbook of industrial organization*, **3**: Elsevier, 1701–1844.

Bakshy, Eytan, Solomon Messing, and Lada A Adamic (2015) "Exposure to ideologically diverse news and opinion on Facebook," *Science*, **348** (6239), 1130–1132.

Barberá, Pablo, John T Jost, Jonathan Nagler, Joshua A Tucker, and Richard Bonneau (2015) "Tweeting from left to right: Is online political communication more than an echo chamber?" *Psychological science*, **26** (10), 1531–1542.

Bartoš, Vojtěch, Michal Bauer, Julie Chytilová, and Filip Matějka (2016) "Attention discrimination: Theory and field experiments with monitoring information acquisition," *American Economic Review*, **106** (6), 1437–75.

Batina, Raymond G and Toshihiro Ihori (2005) *Public goods: theories and evidence*: Springer Science & Business Media.

Baumgartner, Frank R and Laura Chaqués Bonafont (2015) "All news is bad news: Newspaper coverage of political parties in Spain," *Political Communication*, **32** (2), 268–291.

Bayerischer Rundfunk, (2017) "Wikipedia-Artikel über Abgeordnete vom Bundestag aus geschönt," *URL: https://www.br.de/bundestagswahl/wikipedia-artikel-abgeordnete-bundestag-100.html.*

BDZV, (2017) *Zeitungen 2017/18*: Bundesverband Deutscher Zeitungsverleger.

Beatty, Sharon E and M Elizabeth Ferrell (1998) "Impulse buying: Modeling its precursors," *Journal of retailing*, **74** (2), 169–191.

Becker, Gary S and Kevin M Murphy (1993) "A simple theory of advertising as a good or bad," *The Quarterly Journal of Economics*, **108** (4), 941–964.

Becker, Johannes, Andreas Peichl, and Johannes Rincke (2009) "Politicians' outside earnings and electoral competition," *Public Choice*, **140** (3-4), 379.

Beebe, Jack H. (1977) "Institutional structure and program choices in television markets," *The Quarterly Journal of Economics*, **91** (1), 15–37.

Behringer, Stefan and Lapo Filistrucchi (2015) "Hotelling competition and political differentiation with more than two newspapers," *Information Economics and Policy*, **30**, 36–49.

Ben, Bagdikian (2004) *The New Media Monopoly*: Beacon Press, Boston.

Berry, Steven T and Joel Waldfogel (1999) "Public radio in the United States: does it correct market failure or cannibalize commercial stations?" *Journal of Public Economics*, **71** (2), 189–211.

Bertrand, Marianne and Esther Duflo (2017) "Field experiments on discrimination," in *Handbook of economic field experiments*, **1**: Elsevier, 309–393.

Bertrand, Marianne and Sendhil Mullainathan (2004) "Are Emily and Greg more employable than Lakisha and Jamal? A field experiment on labor market discrimination," *American economic review*, **94** (4), 991–1013.

Bilodeau, Marc (1994) "Tax-earmarking and separate school financing," *Journal of Public Economics*, **54** (1), 51–63.

Blasco, Andrea, Paolo Pin, and Francesco Sobbrio (2016) "Paying positive to go negative: Advertisers' competition and media reports," *European Economic Review*, **83**, 243–261.

Blasco, Andrea and Francesco Sobbrio (2012) "Competition and commercial media bias," *Telecommunications Policy*, **36** (5), 434–447.

Bond, Robert, Christopher Fariss, Jason J. Jones, Adamd. I. Kramer, Cameron Marlow, Jaime E. Settle, and James H. Fowler (2012) "A 61-million-person experiment in social influence and political mobilization," *Nature*, **489** (7415), 295–298.

Bourreau, Marc (2003) "Mimicking vs. counter-programming strategies for television programs," *Information Economics and Policy*, **15** (1), 35–54.

Boutet, Antoine, Hyoungshick Kim, and Eiko Yoneki (2012) "What's in your tweets? I know who you supported in the UK 2010 general election," in *Sixth International AAAI Conference on Weblogs and Social Media*, AAAI.

Boykoff, Maxwell T (2008) "Lost in translation? United States television news coverage of anthropogenic climate change, 1995–2004," *Climatic Change*, **86** (1-2), 1–11.

Boykoff, Maxwell T and Jules M Boykoff (2004) "Balance as bias: global warming and the US prestige press," *Global environmental change*, **14** (2), 125–136.

Brody, Howard and Donald W Light (2011) "The inverse benefit law: how drug marketing undermines patient safety and public health," *American Journal of Public Health*, **101** (3), 399–404.

Brown, Adam R (2011) "Wikipedia as a data source for political scientists: Accuracy and completeness of coverage," *PS: Political Science & Politics*, **44** (2), 339–343.

Brown, Keith and Peter J Alexander (2005) "Market structure, viewer welfare, and advertising rates in local broadcast television markets," *Economics Letters*, **86** (3), 331–337.

Brown, Keith and Roberto Cavazos (2005) "Why is this show so dumb? Advertising revenue and program content of network television," *Review of Industrial Organization*, **27** (1), 17–34.

Bruce, Rebecca F and Janyce M Wiebe (1999) "Recognizing subjectivity: a case study in manual tagging," *Natural Language Engineering*, **5** (2), 187–205.

Brynjolfsson, Erik, Yu Hu, and Duncan Simester (2011) "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales," *Management Science*, **57** (8), 1373–1386.

Brynjolfsson, Erik, Yu Hu, and Michael D Smith (2003) "Consumer surplus in the digital economy: Estimating the value of increased product variety at online booksellers," *Management Science*, **49** (11), 1580–1596.

Butler, Daniel M and Emily Schofield (2010) "Were newspapers more interested in pro-Obama letters to the editor in 2008? Evidence from a field experiment," *American Politics Research*, **38** (2), 356–371.

Center, Pew Research (2000) *Self Censorship: How Often and Why. Journalists Avoiding The News*: Washington.

Chen, Yongmin and Michael H Riordan (2007) "Price and variety in the spokes model," *The Economic Journal*, **117** (522), 897–921.

Chevalier, Judith and Dina Mayzlin (2006) "The effect of word of mouth on sales: Online book reviews," *Journal of Marketing Research*, **43** (3), 345–354.

Choi, Jay Pil (2006) "Broadcast competition and advertising with free entry: Subscription vs. free-to-air," *Information Economics and Policy*, **18** (2), 181–196.

Coase, Ronald Harry (1974) "The market for goods and the market for ideas," *The American Economic Review*, **64** (2), 384–391.

Colleoni, Elanor, Alessandro Rozza, and Adam Arvidsson (2014) "Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data," *Journal of communication*, **64** (2), 317–332.

Converse, Philip E, Aage R Clausen, and Warren E Miller (1965) "Electoral myth and reality: the 1964 election," *American Political Science Review*, **59** (2), 321–336.

Cooper, Christopher, H Gibbs Knotts, and Moshe Haspel (2009) "The content of political participation: Letters to the editor and the people who write them," *Political Science & Politics*, **42** (1), 131–137.

Cornes, Richard and Junichi Itaya (2010) "On the private provision of two or more public goods," *Journal of Public Economic Theory*, **12** (2), 363–385.

Crampes, Claude, Carole Haritchabalet, and Bruno Jullien (2009) "Advertising, competition and entry in media industries," *The Journal of Industrial Economics*, **57** (1), 7–31.

D'Alessio, Dave and Mike Allen (2000) "Media bias in presidential elections: A meta-analysis," *Journal of communication*, **50** (4), 133–156.

Dallmann, Alexander, Florian Lemmerich, Daniel Zoller, and Andreas Hotho (2015) "Media bias in german online newspapers," in *Proceedings of the 26th ACM Conference on Hypertext & Social Media*, 133–137.

d'Aspremont, Claude, J Jaskold Gabszewicz, and J-F Thisse (1979) "On Hotelling's" Stability in competition"," *Econometrica*, **47**, 1145–1150.

De Palma, Andre, Victor Ginsburgh, Yorgo Y. Papageorgiou, and Jaques-Francois Thisse (1985) "The principle of minimum differentiation holds under sufficient heterogeneity," *Econometrica*, **53**, 767–781.

De Smet, Dries and Stijn Vanormelingen (2012) "The advertiser is mentioned twice. Media bias in Belgian newspapers," Unpublished.

DellaVigna, Stefano (2009) "Psychology and economics: Evidence from the field," *Journal of Economic literature*, **47** (2), 315–72.

DellaVigna, Stefano and Eliana La Ferrara (2015) "Economic and social impacts of the media," in *Handbook of media economics*, **1**: Elsevier, 723–768.

Dewenter, Ralf, Uwe Dulleck, and Tobias Thomas (2016) *Does the 4th estate deliver? Towards a more direct measure of political media bias*: DICE Discussion Paper.

Dewenter, Ralf and Ulrich Heimeshoff (2014) "Media Bias and advertising: Evidence from a German car magazine," *Review of Economics*, **65** (1), 77–94.

Di Tella, Rafael and Ignacio Franceschelli (2011) "Government advertising and media coverage of corruption scandals," *American Economic Journal: Applied Economics*, **3** (4), 119–51.

Dixit, Avinash and Victor Norman (1978) "Advertising and welfare," *The Bell Journal of Economics*, **9**, 1–17.

Downs, Anthony (1957) "An economic theory of political action in a democracy," *Journal of political economy*, **65** (2), 135–150.

Dukes, Anthony (2004) "The adverstising market in a product oligopoly," *The Journal of Industrial Economics*, **52** (3), 327–348.

Durbin, James (1954) "Errors in variables," *Revue de l'institut International de Statistique*, 23–32.

Easley, David and Arpita Ghosh (2013) "Incentives, gamification, and game theory: An economic approach to badge design," in *ACM Transactions on Economics and Computation (TEAC)*, ACM.

Economides, Nicholas (1986) "Minimal and maximal product differentiation in Hotelling's duopoly," *Economics Letters*, **21** (1), 67–71.

Eilders, Christiane (2002) "Conflict and consonance in media opinion: Political positions of five German quality newspapers," *European Journal of Communication*, **17** (1), 25–63.

Ellman, Matthew and Fabrizio Germano (2009) "What do the papers sell? A model of advertising and media bias," *The Economic Journal*, **119** (537), 680–704.

Engelberg, Joseph E and Christopher A Parsons (2011) "The causal impact of media in financial markets," *The Journal of Finance*, **66** (1), 67–97.

Erfle, Stephen and Henry McMillan (1990) "Media, political pressure, and the firm: The case of petroleum pricing in the late 1970s," *The Quarterly Journal of Economics*, **105** (1), 115–134.

European Union, (2012) *Directive 2010/13/EU of the European Parliament and of the Council, on the coordination of certain provisions laid down by law, regulation or administrative action in Member States concerning the provision of audiovisual media services (Audiovisual Media Services Directive)*: Saint Paul.

Faber, Ronald J and Kathleen D Vohs (2011) "Self-regulation and spending: Evidence from impulsive and compulsive buying," *Handbook of self-regulation: Research, theory, and applications*, 537–550.

Faerber, Adrienne E and David H Kreling (2014) "Content analysis of false and misleading claims in television advertising for prescription and nonprescription drugs," *Journal of general internal medicine*, **29** (1), 110–118.

Fang, Hanming and Peter Norman (2010) "Optimal provision of multiple excludable public goods," *American Economic Journal: Microeconomics*, **2** (4), 1–37.

FCC, (2011) "The information needs of communities. The changing media landscape in a broadband age."

Flight, Richard L, Melissa Markley Rountree, and Sharon E Beatty (2012) "Feeling the urge: Affect in impulsive and compulsive buying," *Journal of Marketing Theory and Practice*, **20** (4), 453–466.

Forschungsgruppe Wahlen, (2014) "Internet Strukturdaten 2014," *URL: http://www.forschungsgruppe.de/Umfragen/Internet-Strukturdaten/.*

Forschungsgruppe Wahlen, Mannheim (2017) "Politbarometer 2016 (Kumulierter Datensatz)," GESIS Datenarchiv, Köln. ZA6888 Datenfile Version 1.0.0 (2017), doi:10.4232/1.12824.

Foster, H Schuyler and Carl J Friedrich (1937) "Letters to the Editor as a Means of Measuring the Effectiveness of Propaganda," *American Political Science Review*, **31** (1), 71–79.

Funk, Carolyn L (1999) "Bringing the candidate into models of candidate evaluation," *The Journal of Politics*, **61** (3), 700–720.

Gabszewicz, Jean J., Didier Laussel, and Nathalie Sonnac (2001) "Press advertising and the ascent of the 'pensée unique'," *European Economic Review*, **45** (4-6), 641–651.

——— (2002) "Press advertising and the political differentiation of newspapers," *Journal of Public Economic Theory*, **4** (3), 317–334.

——— (2004) "Programming and advertising competition in the broadcasting industry," *Journal of Economics & Management Strategy*, **13** (4), 657–669.

Gabszewicz, Jean and Jaques-Francois Thisse (1986) "Spatial competition and the location of firms," in *Fundamentals of pure and applied economics*, **5**: Routledge, 1–71.

Gal-Or, Esther and Anthony Dukes (2003) "Minimum differentiation in commercial media markets," *Journal of Economics & Management Strategy*, **12** (3), 291–325.

Galtung, Johan and Mari Holmboe Ruge (1965) "The structure of foreign news: The presentation of the Congo, Cuba and Cyprus crises in four Norwegian newspapers," *Journal of peace research*, **2** (1), 64–90.

Gambaro, Marco and Riccardo Puglisi (2015) "What do ads buy? Daily coverage of listed companies on the Italian press," *European Journal of Political Economy*, **39**, 41–57.

Garcia Pires, Armando (2014) "Media diversity, advertising, and adaptation of news to readers' political preferences," *Information Economics and Policy*, **28**, 28–38.

Garrett, R Kelly (2009a) "Echo chambers online?: Politically motivated selective exposure among Internet news users," *Journal of Computer-Mediated Communication*, **14** (2), 265–285.

——— (2009b) "Politically motivated reinforcement seeking: Reframing the selective exposure debate," *Journal of Communication*, **59** (4), 676–699.

Garz, Marcel (2014) "Good news and bad news: evidence of media bias in unemployment reports," *Public Choice*, **161** (3-4), 499–515.

Garz, Marcel, Jil Sörensen, and Daniel F Stone (2019) "Partisan selective engagement: Evidence from Facebook."

Gentzkow, Matthew and Jesse M Shapiro (2008) "Competition and Truth in the Market for News," *Journal of Economic perspectives*, **22** (2), 133–154.

——— (2010) "What drives media slant? Evidence from US daily newspapers," *Econometrica*, **78** (1), 35–71.

Gentzkow, Matthew, Jesse M Shapiro, and Daniel F Stone (2016) "Media bias in the marketplace: Theory," in *Handbook of media economics*, **1**: Elsevier, 623–645.

Gerber, Alan S, Dean Karlan, and Daniel Bergan (2009) "Does the media matter? A field experiment measuring the effect of newspapers on voting behavior and political opinions," *American Economic Journal: Applied Economics*, **1** (2), 35–52.

Germano, Fabrizio and Martin Meier (2013) "Concentration and self-censorship in commercial media," *Journal of Public Economics*, **97**, 117–130.

Ghosh, Suman, Alexander Karaivanov, and Mandar Oak (2007) "A case for bundling public goods contributions," *Journal of Public Economic Theory*, **9** (3), 425–449.

Giamattei, Marcus and Johann Graf Lambsdorff (2019) "classEx—an online tool for lab-in-the-field experiments with smartphones," *Journal of Behavioral and Experimental Finance*, **22**, 223–231.

Glaeser, Edward L and Gergely Ujhelyi (2010) "Regulating misinformation," *Journal of Public Economics*, **94** (3-4), 247–257.

Goldberg, Marvin E and Gerald J Gorn (1987) "Happy and sad TV programs: How they affect reactions to commercials," *Journal of consumer research*, **14** (3), 387–403.

Goldfarb, Avi and Catherine Tucker (2019) "Digital economics," *Journal of Economic Literature*, **57** (1), 3–43.

Graitson, Dominique (1982) "Spatial competition a la Hotelling: A selective survey," *The Journal of Industrial Economics*, 11–25.

Green, Donald P, Brian R Calfano, and Peter M Aronow (2014) "Field experimental designs for the study of media effects," *Political Communication*, **31** (1), 168–180.

Green, Donald P, AJ Carnegie, Joel Middleton, K Kenski, and KH Jamieson (2017) "Political communication: insights from field experiments," *Oxford handbook of political communication*, 501–516.

Green-Pedersen, Christoffer, Peter B Mortensen, and Gunnar Thesen (2017) "The incumbency bonus revisited: Causes and consequences of media dominance," *British Journal of Political Science*, **47** (1), 131–148.

Greene, William H (2003) *Econometric analysis*: Pearson Education India.

Greenstein, Shane and Feng Zhu (2012) "Is Wikipedia Biased?" *American Economic Review*, **102** (3), 343–48.

Grimmer, Justin (2013) *Representational style in Congress: What legislators say and why it matters*: Cambridge University Press.

Groeling, Tim (2013) "Media bias by the numbers: Challenges and opportunities in the empirical study of partisan news," *Annual Review of Political Science*, **16**, 129–151.

Groseclose, Tim and Jeffrey Milyo (2005) "A measure of media bias," *The Quarterly Journal of Economics*, **120** (4), 1191–1237.

Grossman, Gene M and Carl Shapiro (1984) "Informative advertising with differentiated products," *The Review of Economic Studies*, **51** (1), 63–81.

Gurun, Umit G and Alexander W Butler (2012) "Don't believe the hype: Local media slant, local advertising, and firm value," *The Journal of Finance*, **67** (2), 561–598.

Guryan, Jonathan and Kerwin Kofi Charles (2013) "Taste-based or statistical discrimination: The economics of discrimination returns to its roots," *The Economic Journal*, **123** (572), F417–F432.

Halavais, Alexander and Derek Lackaff (2008) "An analysis of topical coverage of Wikipedia," *Journal of computer-mediated communication*, **13** (2), 429–440.

Hallin, Daniel C and Paolo Mancini (2004) *Comparing media systems: Three models of media and politics*: Cambridge university press.

Hamilton, James (2004) *All the news that's fit to sell: How the market transforms information into news*: Princeton University Press.

Harcup, Tony and Deirdre O'neill (2017) "What is news? News values revisited (again)," *Journalism studies*, **18** (12), 1470–1488.

Harlé, Katia M, Luke J Chang, Mascha van't Wout, and Alan G Sanfey (2012) "The neural mechanisms of affect infusion in social economic decision-making: a mediating role of the anterior insula," *Neuroimage*, **61** (1), 32–40.

Harlé, Katia M and Alan G Sanfey (2007) "Incidental sadness biases social economic decisions in the Ultimatum Game.," *Emotion*, **7** (4), 876.

Haselmayer, Martin, Markus Wagner, and Thomas M Meyer (2017) "Partisan bias in message selection: Media gatekeeping of party press releases," *Political communication*, **34** (3), 367–384.

Hawkins, D and D Mothersbaugh (2009) "Concept of Consumer Behavior," *Consumer behavior building Marketing Strategy, 12th edition (February 24, 2012)*.

Heinz, Matthias and Johan Swinnen (2015) "Media slant in economic news: A factor 20," *Economics Letters*, **132**, 18–20.

Helfer, Luzia and Peter Van Aelst (2016) "What makes party messages fit for reporting? An experimental study of journalistic news selection," *Political Communication*, **33** (1), 59–77.

Hellwig, Martin F (2007) "The provision and pricing of excludable public goods: Ramsey–boiteux pricing versus bundling," *Journal of Public Economics*, **91** (3-4), 511–540.

Herman, Edward S and Robert Waterman McChesney (1997) *The global media: The new missionaries of corporate capitalism*: Burns & Oates.

Hermida, Alfred, Seth C Lewis, and Rodrigo Zamith (2014) "Sourcing the Arab Spring: A case study of Andy Carvin's sources on Twitter during the Tunisian and Egyptian revolutions," *Journal of Computer-Mediated Communication*, **19** (3), 479–499.

Hinnosaar, Marit (2019) "Gender inequality in new media: Evidence from Wikipedia," *Journal of Economic Behavior & Organization*, **163**, 262–276.

Ho, Daniel and Kevin Quinn (2008) "Measuring explicit political positions of media," *Quarterly Journal of Political Science*, **3** (4), 353–377.

Holtz-Bacha, Christina, Ana Ines Langer, and Susanne Merkle (2014) "The personalization of politics in comparative perspective: Campaign coverage in Germany and the United Kingdom," *European Journal of Communication*, **29** (2), 153–170.

Hopmann, David Nicolas, Claes H de Vreese, and Erik Albæk (2011) "Incumbency bonus in election news coverage explained: The logics of political power and the media market," *Journal of Communication*, **61** (2), 264–282.

Hotelling, Harold (1929) "Stability in Competition," *Economic Journal*, **39**, 41–57.

Huang, Jason, David Reiley, and Nick Riabov (2018) "Measuring Consumer Sensitivity to Audio Advertising: A Field Experiment on Pandora Internet Radio."

Huberman, Gur and Tomer Regev (2001) "Contagious speculation and a cure for cancer: A nonevent that made stock prices soar," *The Journal of Finance*, **56** (1), 387–396.

Hynds, Ernest C (1994) "Editors at most US dailies see vital roles for editorial page," *Journalism Quarterly*, **71** (3), 573–582.

Ihori, Toshihiro, Martin McGuire, and Shintaro Nakagawa (2014) "International security, multiple public good provisions, and the exploitation hypothesis," *Defence and Peace Economics*, **25** (3), 213–229.

IP Network, (2013) "Television Key Facts 2013: Population and Audiovisual Equipment," *URL: http://www.international-television.org/tv-market-data/international-tv-key-facts.html*.

Ito, Tiffany A, Jeff T Larsen, N Kyle Smith, and John T Cacioppo (1998) "Negative information weighs more heavily on the brain: the negativity bias in evaluative categorizations.," *Journal of personality and social psychology*, **75** (4), 887.

Jungherr, Andreas (2014) "The logic of political coverage on Twitter: Temporal dynamics and content," *Journal of communication*, **64** (2), 239–259.

——— (2016) "Twitter use in election campaigns: A systematic literature review," *Journal of Information Technology & Politics*, **13** (1), 72–91.

Kepplinger, Hans Mathias (2011) "Theorien der Nachrichtenauswahl als Theorien der Realität," in *Realitätskonstruktionen*: Springer, 47–65.

Kind, Hans Jarle, Marko Koethenbuerger, and Guttorm Schjelderup (2008) "Efficiency enhancing taxation in two-sided markets," *Journal of Public Economics*, **92** (5-6), 1531–1539.

Kinder, Donald R, Mark D Peters, Robert P Abelson, and Susan T Fiske (1980) "Presidential prototypes," *Political behavior*, **2** (4), 315–337.

King, Gary, Jennifer Pan, and Margaret E Roberts (2014) "Reverse-engineering censorship in China: Randomized experimentation and participant observation," *Science*, **345** (6199), 1251722.

Kleibergen, Frank and Richard Paap (2006) "Generalized reduced rank tests using the singular value decomposition," *Journal of econometrics*, **133** (1), 97–126.

Knobloch-Westerwick, Silvia and Jingbo Meng (2009) "Looking the other way: Selective exposure to attitude-consistent and counterattitudinal political information," *Communication Research*, **36** (3), 426–448.

——— (2011) "Reinforcement of the political self through selective exposure to political messages," *Journal of Communication*, **61** (2), 349–368.

Kürschner (2015) *Kürschners Volkshandbuch Deutscher Bundestag 18. Wahlperiode*: Neue Darmstädter Verlagsanstalt.

Lancaster, Kelvin (1990) "The economics of product variety: A survey," *Marketing science*, **9** (3), 189–206.

Larcinese, Valentino, Riccardo Puglisi, and James M Snyder Jr (2011) "Partisan bias in economic news: Evidence on the agenda-setting behavior of US newspapers," *Journal of public Economics*, **95** (9-10), 1178–1189.

Laugesen, Murray and Chris Meads (1991) "Tobacco advertising restrictions, price, income and tobacco consumption in OECD countries, 1960–1986," *British journal of addiction*, **86** (10), 1343–1354.

Leichtmann Research Group, (2013) "On-Demand TV 2013: A Nationwide Study on VOD and DVRs," *URL: https://www.marketingcharts.com/demographics-and-audiences/household-income-38610.*

Lerner, Jennifer S, Ye Li, Piercarlo Valdesolo, and Karim S Kassam (2015) "Emotion and decision making," *Annual review of psychology*, **66**, 799–823.

Lichter, S Robert (2017) "Theories of media bias," *The Oxford handbook of political communication*, 403.

Lieberman, Trudy (2007) "Epidemic," *Columbia Journalism Review*, **45** (6), 38.

Liu, Laura Xiaolei, Ann E Sherman, and Yong Zhang (2014) "The long-run role of the media: Evidence from initial public offerings," *Management Science*, **60** (8), 1945–1964.

Luca, Michael (2016a) "Reviews, reputation, and revenue: The case of Yelp. com," Unpublished.

——— (2016b) "User-generated content and social media," in *Handbook of Media Economics*, **1**: Elsevier, 563–592.

Luski, Israel and David Wettstein (1994) "The provision of public goods via advertising: Existence of equilibria and welfare analysis," *Journal of public economics*, **54** (2), 309–321.

Mathur, Mahima and Amitava Chattopadhyay (1991) "The impact of moods generated by television programs on responses to advertising," *Psychology & Marketing*, **8** (1), 59–77.

Mayzlin, Dina, Yaniv Dover, and Judith Chevalier (2014) "Promotional Reviews: An Empirical Investigation of Online Review Manipulation," *American Economic Review*, **104** (8), 2421–2455.

McChesney, Robert D (2004) *The problem of the media: US communication politics in the twenty-first century*: NYU Press.

McCrary, Justin (2008) "Manipulation of the running variable in the regression discontinuity design: A density test," *Journal of Econometrics*, **142** (2), 698–714.

Mesgari, Mostafa, Chitu Okoli, Mohamad Mehdi, Finn Arup Nielsen, and Arto Lanamaki (2015) ""The sum of all human knowledge": A systematic review of scholarly research on the content of W ikipedia," *Journal of the Association for Information Science and Technology*, **66** (2), 219–245.

Morton, Rebecca B and Kenneth C Williams (2010) *Experimental political science and the study of causality: From nature to the lab*: Cambridge University Press.

Myers, ML, C Iscoe, C Jennings, W Lenox, E Minsky, and A Sachs (1981) "Federal Trade Commission staff report on the cigarette advertising investigation, May 1981 (public version)," *Division of Advertising Practices, Federal Trade Commission.*

Nagler, Matthew G (1993) "Rather bait than switch: Deceptive advertising with bounded consumer rationality," *Journal of Public Economics*, **51** (3), 359–378.

Niederdeppe, Jeff and Dominick L Frosch (2009) "News coverage and sales of products with trans fat: effects before and after changes in federal labeling policy," *American Journal of preventive medicine*, **36** (5), 395–401.

Nielsen Media Research, (2013) "Free to move between screens. The Cross-Platform Report," *URL: https://www.warc.com/content/paywall/article/nielsen/free-to-move-between-screens-the-cross-platform-report/99366.*

Niven, David (2001) "Bias in the news: Partisanship and negativity in media coverage of presidents George Bush and Bill Clinton," *Harvard International Journal of Press/Politics*, **6** (3), 31–46.

Noam, Eli M (2016) *Who owns the world's media?: media concentration and ownership around the world*: Oxford University Press.

OFCOM, (2011) "Regulating the quantity of advertising on television," *https://www.ofcom.org.uk/data/assets/pdf-file/0021/19083/advertising-minutage.pdf.*

O'Neill, Deirdre and Tony Harcup (2009) "News values and selectivity," in *The handbook of journalism studies*: Routledge, 181–194.

Oreskes, Naomi (2004) "The scientific consensus on climate change," *Science*, **306** (5702), 1686–1686.

Owen, Donald B (1962) *Handbook of statistical tables*: Addison-Wesley Publication Company.

Pager, Devah (2007) "The use of field experiments for studies of employment discrimination: Contributions, critiques, and directions for the future," *The Annals of the American Academy of Political and Social Science*, **609** (1), 104–133.

Panagopoulos, Costas and Donald P Green (2008) "Field experiments testing the impact of radio advertisements on electoral competition," *American Journal of Political Science*, **52** (1), 156–168.

Pang, Bo and Lillian Lee (2008) "Opinion mining and sentiment analysis," *Foundations and Trends® in Information Retrieval*, **2** (1–2), 1–135.

Peitz, Martin and Tommaso M. Valletti (2008) "Content and advertising in the media: Pay-tv versus free-to-air," *international Journal of industrial organization*, **26** (4), 949–965.

Perrin, Andrew J and Stephen Vaisey (2008) "Parallel public spheres: Distance and discourse in letters to the editor," *American Journal of Sociology*, **114** (3), 781–810.

Persson, Torsten and Guido Enrico Tabellini (2005) *The economic effects of constitutions*: MIT press.

Pew Research Center, (2013) "The State of the News Media 2013. An Annual Report on American Journalism," *URL: https://www.pewresearch.org/topics/state-of-the-news-media/.*

Prat, Andrea and David Strömberg (2013) "The political economy of mass media. Advances in Economics and Econometrics," in *Applied Economics: Tenth World Congress*, **50**, 135.

Puglisi, Riccardo and James M Snyder (2016) "Empirical studies of media bias," in *Handbook of media economics*, **1**: Elsevier, 647–667.

Raeymaeckers, Karin (2005) "Letters to the editor: A feedback opportunity turned into a marketing tool: An account of selection and editing practices in the Flemish daily press," *European Journal of Communication*, **20** (2), 199–221.

Rauch, Ferdinand (2013) "Advertising expenditure and consumer prices," *International Journal of Industrial Organization*, **31** (4), 331–341.

Reagle, Joseph and Lauren Rhue (2011) "Gender bias in Wikipedia and Britannica," *International Journal of Communication*, **5**, 21.

Renfro, Paula Cozort (1979) "Bias in selection of letters to editor," *Journalism Quarterly*, **56** (4), 822–826.

Reuter, Jonathan (2009) "Does advertising bias product reviews? An analysis of wine ratings," *Journal of Wine Economics*, **4** (2), 125–151.

Reuter, Jonathan and Eric Zitzewitz (2006) "Do ads influence editors? Advertising and bias in the financial media," *The Quarterly Journal of Economics*, **121** (1), 197–227.

Riach, Peter A and Judith Rich (2002) "Field experiments of discrimination in the market place," *The economic journal*, **112** (483), F480–F518.

——— (2004) "Deceptive field experiments of discrimination: are they ethical?" *Kyklos*, **57** (3), 457–470.

Richardson, John E and Bob Franklin (2004) "Letters of intent: Election campaigning and orchestrated public debate in local newspapers' letters to the editor," *Political Communication*, **21** (4), 459–478.

Rinallo, Diego and Suman Basuroy (2009) "Does advertising spending influence media coverage of the advertiser?" *Journal of Marketing*, **73** (6), 33–46.

Rozin, Paul and Edward B Royzman (2001) "Negativity bias, negativity dominance, and contagion," *Personality and social psychology review*, **5** (4), 296–320.

Salganik, Matthew (2017) *Bit by bit: Social research in the digital age*: Princeton University Press.

Salop, Steven C (1979) "Monopolistic competition with outside goods," *The Bell Journal of Economics*, 141–156.

Schlenker, Wolfram and Sofia B Villas-Boas (2009) "Consumer and market responses to mad cow disease," *American Journal of Agricultural Economics*, **91** (4), 1140–1152.

Schoenbach, Klaus, Jan De Ridder, and Edmund Lauf (2001) "Politicians on TV news: Getting attention in Dutch and German election campaigns," *European Journal of Political Research*, **39** (4), 519–531.

Seamans, Robert and Feng Zhu (2014) "Responses to entry in multi-sided markets: The impact of Craigslist on local newspapers," *Management Science*, **60** (2), 476–493.

Seufert, Wolfgang and Hardy Gundlach (2012) *Medienregulierung in Deutschland: Ziele, Konzepte, Massnahmen: Lehr-und Handbuch*: Nomos.

Shapiro, Carl (1980) "Advertising and welfare: Comment," *The Bell Journal of Economics*, 749–752.

Shin, Jieun and Kjerstin Thorson (2017) "Partisan selective sharing: The biased diffusion of fact-checking messages on social media," *Journal of Communication*, **67** (2), 233–255.

Shoemaker, Pamela J and Stephen D Reese (1996) *Mediating the message*: White Plains, NY: Longman.

Shoemaker, Pamela J and Timothy Vos (2009) *Gatekeeping theory*: Routledge.

Siegel, Sidney and NJ Castellan (1988) "Nonparametric statistics for the behavioral sciences," *Nonparametric Statistics for the Behavioral Sciences, McGraw-Hill, New York*, 190–222.

Soley, Lawrence C and Robert L Craig (1992) "Advertising pressures on newspapers: A survey," *Journal of Advertising*, **21** (4), 1–10.

Soroka, Stuart and Stephen McAdams (2015) "News, politics, and negativity," *Political Communication*, **32** (1), 1–22.

Soroka, Stuart N (2012) "The gatekeeping function: Distributions of information in media and the real world," *The Journal of Politics*, **74** (2), 514–528.

Spiteri, J (2015) "When Is No News Good News? A Model of Information Disclosure and Commercial Media Bias," in *Working paper*.

Statistisches Bundesamt, (2017) "Ausländische Bevölkerung nach Geschlecht und ausgewählten Staatsangehörigkeiten."

Steiner, Peter O. (1952) "Program patterns and preferences, and the workability of competition in radio broadcasting," *The Quarterly Journal of Economics*, **66** (2), 194–223.

Stigler, George J and Gary S Becker (1977) "De gustibus non est disputandum," *The american economic review*, **67** (2), 76–90.

Stock, James H and Motohiro Yogo (2002) "Testing for weak instruments in linear IV regression."

Stokes, Donald E (1963) "Spatial models of party competition," *American political science review*, **57** (2), 368–377.

Stone, Walter J and Elizabeth N Simas (2010) "Candidate valence and ideological positions in US House elections," *American Journal of Political Science*, **54** (2), 371–388.

Sun, Monic and Feng Zhu (2013) "Ad revenue and content commercialization: Evidence from blogs," *Management Science*, **59** (10), 2314–2331.

Tauber, Peter (2013) "Social Media Leitfaden 2013," *URL:http:// www.petertauber.de/wp-content/uploads/2013/04/SocialMedia-Leitfaden-Peter-Tauber-2013.pdf.*

Tobacco Control Legal Consortium, (2012) *The Federal Trade Commission and Tobacco. Factsheet 2012*: Saint Paul.

Trussler, Marc and Stuart Soroka (2014) "Consumer demand for cynical and negative news frames," *The International Journal of Press/Politics*, **19** (3), 360–379.

Upshaw, James, Gennadiy Chernov, and David Koranda (2007) "Telling more than news: Commercial influence in local television stations," *Electronic News*, **1** (2), 67–87.

Van Dalen, Arjen (2012) "Structural bias in cross-national perspective: How political systems and journalism cultures influence government dominance in the news," *The International Journal of Press/Politics*, **17** (1), 32–55.

Vesterlund, Lise (2016) "Using experimental methods to understand why and how we give to charity," *Handbook of experimental economics*, **2**, 91–151.

Vos, Debby (2014) "Which politicians pass the news gates and why? Explaining inconsistencies in research on news coverage of individual politicians," *International journal of communication.–*, **8**, 2438–2461.

Vos, Debby and Peter Van Aelst (2018) "Does the political system determine media visibility of politicians? a comparative analysis of political functions in the news in sixteen countries," *Political Communication*, **35** (3), 371–392.

de Vreese, Claes H (2017) "Comparative political communication research," *The Oxford Handbook of Political Communication*, 287.

Wahl-Jorgensen, Karin (2007) *Journalists and the public: Newsroom culture, letters to the editor, and democracy*: Hampton Press Cresskill, NJ.

Wakefield, Melanie, Brian Flay, Mark Nichter, and Gary Giovino (2003) "Role of the media in influencing trajectories of youth smoking," *Addiction*, **98**, 79–103.

Waldfogel, Joel (2007) *The tyranny of the market: Why you can't always get what you want*: Harvard University Press.

——— (2017) "How digitization has created a golden age of music, movies, books, and television," *Journal of economic perspectives*, **31** (3), 195–214.

——— (2018) *Digital Renaissance: What Data and Economics Tell Us about the Future of Popular Culture*: Princeton University Press.

Wang, Zhongmin (2010) "Anonymity, social image, and the competition for volunteers: a case study of the online market for reviews," *The BE Journal of Economic Analysis & Policy*, **10** (1), 1–35.

Warner, Kenneth E and Linda M Goldenhar (1989) "The cigarette advertising broadcast ban and magazine coverage of smoking and health," *Journal of Public Health Policy*, **10** (1), 32–42.

Warner, Kenneth E, Linda M Goldenhar, and Catherine G McLaughlin (1992) "Cigarette advertising and magazine coverage of the hazards of smoking a statistical analysis," *New England Journal of Medicine*, **326** (5), 305–309.

Weeds, Helen (2013) "Programme Quality in Subscription and Advertising-Funded Television."

Weitzman, Martin L (1974) "Prices vs. quantities," *The review of economic studies*, **41** (4), 477–491.

Whelan, Elizabeth M, Margaret J Sheridan, Kathleen A Meister, and Beverly A Mosher (1981) "Analysis of coverage of tobacco hazards in women's magazines," *Journal of public health policy*, **2** (1), 28–35.

WHO, (2013) *WHO report on the global tobacco epidemic, 2013: Enforcing bans on tobacco advertising, promotion and sponsorship*: Geneva.

Wiebe, Janyce, Theresa Wilson, Rebecca Bruce, Matthew Bell, and Melanie Martin (2004) "Learning subjective language," *Computational linguistics*, **30** (3), 277–308.

Wilbur, Kenneth C (2008) "A two-sided, empirical model of television advertising and viewing markets," *Marketing science*, **27** (3), 356–378.

Zhang, Xiaoquan (Michael) and Feng Zhu (2011) "Group Size and Incentives to Contribute: A Natural Experiment at Chinese Wikipedia," *American Economic Review*, **101** (4), 1601–1615.

Zinman, Jonathan and Eric Zitzewitz (2016) "Wintertime for Deceptive Advertising?" *American Economic Journal: Applied Economics*, **8** (1), 177–92.