



The carrot and the stick in online reviews: determinants of un-/helpfulness voting choices

Filipe Sengo Furtado¹ · Thomas Reutterer¹ · Nadine Schröder¹ 

Accepted: 26 April 2021
© The Author(s) 2021

Abstract

With increasing volumes of customer reviews, ‘helpfulness’ features have been established by many online platforms as decision-aids for consumers to cope with potential information overload. In this study, we offer a differentiated perspective on the drivers of review helpfulness. Using a hurdle regression setup for both helpfulness and unhelpfulness voting behavior, we aim to disentangle the differential effects of what drives reviews to receive any votes, how many votes they receive and whether these effects differ for helpful against unhelpful review voting behavior. As potential driving factors we include reviews’ star rating deviations from the average rating (as a proxy for confirmation bias), the level of controversy among reviews and review sentiment (consistency of review content), as well as pricing information in our analysis. Albeit with opposite effect signs, we find that revealed review un-/helpfulness is consistently guided by the tonality (i.e., the sentiment of review texts) and that reviewers tend to be less critical for lower priced products. However, we find only partial support for a confirmation bias with differential effects for the level of controversy on helpfulness versus unhelpfulness review votings. We conclude that the effects of voting disagreement are more complex than previous literature suggests and discuss implications for research and management practice.

Keywords Online review · (Un)helpfulness · Hurdle model · Confirmation bias · Price levels · Asymmetric effects

JEL Classification M31

✉ Nadine Schröder
nadine.schroeder@wu.ac.at

Filipe Sengo Furtado
filipe.sengo.furtado@wu.ac.at

Thomas Reutterer
thomas.reutterer@wu.ac.at

¹ Department of Marketing, WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

1 Introduction

The ongoing rise of e-commerce has led to a fundamental change in the way people seek out for product information. One way of shaping the customer's information search is user-generated content (UGC), which, especially in the form of online reviews, has quickly established itself on the Internet. As many customers trust the product evaluation of unknown consumers, UGC has become an independent product information source and is perceived as being more persuasive and influential than traditional marketing communication means (Bickart and Schindler 2001). According to a recent survey, 93.4% of online consumers read customer reviews to familiarize with an unknown digital retailer (eMarketer 2019). As a study by Bright-Local (2019) shows, 76% of consumers trust online reviews as much as personal recommendations. In this regard, the valence of the review plays an important role as positive reviews tend to increase, and negative to decrease trust in the company (Trustpilot 2019). Overall, consumers show a higher willingness to buy a product which is promoted via reviews (Duan et al. 2008).

However, online reviews only simplify information gathering when they are managed in a way that makes it easy for the consumer to scan and assess them. When products have been online for a long time or gain in popularity, online feedback can easily exceed the mark of 100 product reviews. Such large volumes can hardly be processed by readers (Singh et al. 2017). Additionally, inconsistencies in informative content and in the quality of reviews increase the risk that potential customers take predominant notice of unhelpful reviews that contain irrelevant information, and which will ultimately lead them to abandon the product purchase. To alleviate these challenges, many platforms use a feature called the 'helpfulness vote' which enables subscribers 'to review posted reviews'. As depicted in the example in Fig. 1, customers are asked a simple dichotomous question: "Was this review helpful to you?", and can decide whether to vote "yes" or "no" if they were satisfied or dissatisfied with the review. By subsequently ranking online reviews based on the opinion of the crowd, the helpfulness vote allows the platforms to more efficiently provide its consumers with the information most valuable to them (Baek et al. 2012). The helpfulness feature is not only perceived as an indicator for customers to select relevant entries (Cao et al. 2011) but also minimizes the customers' anticipated purchase risk as it serves as a control measure for quality (Huang et al. 2015).

639 of 665 people found the following review helpful

★★★★★ Home secured.

By M. Wolff on January 25, 2014

I purchased this along with 4 turtles and a rat.

18 years worth of karate lessons later, I finally feel safe to leave my house at night.

5 Comments

Was this review helpful to you?

Yes

No

Fig. 1 Example of helpfulness question

Indisputably, reviews have become an indispensable tool for e-retailers and manufacturers, who utilize the user-generated feedback to track consumers' opinions and to gather insights into purchase experiences which are vital for customer care. Not surprisingly, retailers benefit from increased product sales pushed by reviews and are advised to allocate their budget accordingly (Marchand et al. 2017).

As helpful votes attract more attention from information seekers and thus positively affect website traffic (Qazi et al. 2016), platforms try to feature helpful reviews more prominently and hide rather unhelpful reviews. There has been a lot of research on why customers consider a certain review as helpful (e.g. Yin et al. 2016; Zhou and Guo 2015). We build on this prior research and extend it by contributing along the following three dimensions: (1) we explicitly address differences between helpfulness and unhelpfulness of reviews, (2) we investigate the differential effects of potential drivers of attracting zero or at least one vote and therefore account for the latent propensity to vote for a review, and (3) we consider different predictors and different operationalizations of well-known predictors of online review un-/helpfulness.

Previous research has ignored review unhelpfulness and solely focused on determinants that affect review helpfulness (e.g., Zhou et al. 2020; Zhu et al. 2014). As a result, ignoring what drives unhelpful votes implicitly assumes that what causes helpful votes will reduce unhelpful ones in the same manner. This is in contrast to Mittal et al. (1998) who draw on prospect theory (Kahnemann and Tversky 1979) and show asymmetric effects of negative vs. positive product performances on customer dis-/satisfaction. As the review voting behavior can be seen as a proxy of a user's dis-/satisfaction with a review, it may well be possible that asymmetries exist in what causes helpful and unhelpful voting behavior. This sheds new light on the findings from previous research and calls for a revisit in the context of unhelpfulness. To the best of our knowledge, we are the first to study unhelpful votes and to empirically evaluate the differential effects of helpfulness drivers on unhelpful votes. This way, our study draws a more holistic view on review voting patterns and helps platforms to better identify unhelpful reviews.

Because of the sheer volume of reviews available online, it is possible that many reviews remain unread and hence do not receive any votes (and attention) at all. Since a prospective customer on average reads ten reviews before she feels able to trust a business (BrightLocal 2019), she might not want to vote on all of them. Interestingly, the vast majority of previous research on the determinants of review helpfulness has considered the review voting mechanism as a one-step process by just considering the number of helpful votes (e.g., Zhou and Guo 2015) or the proportion (e.g., Schlosser 2011) as dependent variable. Typically, observational data sets on review helpfulness contain a latent propensity to vote which we cannot observe directly. In order to control for that, we model the number of un-/helpful votes as a two-step process. In particular, we want to identify those review dimensions that lead a review to receive at least one vote, and those that lead reviews to receive multiple votes.

Finally, in our aim to understand the mechanism behind voting, we draw on previous research that has centered on rating deviations (e.g., Yin et al. 2016) and (in-)consistencies within a review (e.g., Zhou and Guo 2015) as predictors of the number

of helpful votes. Still, studies do not agree on the effects of these predictors and some find their hypotheses unsupported. We account for possible asymmetries in these effects in order to understand whether the sign of such deviations moderates their effects. In addition to that, only few studies consider how the price of a product influences the voting behavior (see, e.g., Baek et al. 2012). However, prospective customers might be more involved when dealing with higher prices and hence vote differently. This way, our set of predictors covers research on the three main areas of review helpfulness, namely ratings, text related aspects, and product characteristics.

Our approach enables us to dissect the online review voting behavior by separating and studying the interplay between these three dimensions.

The remainder of this article is structured as follows: In the next section, we provide a brief overview of existing literature and position our research in this field, which helps us to derive our hypotheses. In Sect. 3 we present our model and briefly discuss the empirical setting we are investigating. Section 4 contains the results of our analyses. We conclude with a discussion of our results and explore the implications of our study.

2 Prior research on review helpfulness

Previous literature has looked into different aspects of review helpfulness voting behavior. Especially review-related aspects like star rating (e.g., Mudambi and Schuff 2010), review depth (e.g., Kuan et al. 2015) or sentiments (e.g., Siering and Muntermann 2013) have been of interest. However, such simple metrics are inapt to capture the specific contexts reviews are read and ultimately voted as helpful or unhelpful. The latter very likely depends on the phase of a customer journey or the stage of a customer's product search process. For example, the perception of a review at a later stage of product search (i.e., when a customer already has acquired some product information) is affected by comparing it with some prior beliefs, which in turn affects the helpful- vs. unhelpful-ness judgement (and the willingness to make this judgement public by voting). Furthermore, there may also be conflicting elements within a review as opposed to previously read reviews that might affect a user's decision to vote on the review. We discuss these aspects of conflicts below by referring to confirmation bias and those capturing the consistency of review content. As a third group of predictors, we include the price of a brand. Because we expect users of higher priced brands to be more involved and to investigate reviews more thoroughly we also expect them to vote differently.

Using these three groups of predictors, Table 1 summarizes how our study differs from previous research. All of the studies have focused on helpful votes and completely ignored effects on unhelpful votes. To the best of our knowledge, this finding still holds if we extend the focus beyond our set of determinants and take into account all previous studies on review voting behavior. Studies focusing on the (assumed) helpfulness of product reviews either model the number of helpful votes or their proportion among all votes. Interestingly, most studies employ a one-step approach when it comes to explaining helpfulness of a product review. Finally, the research contexts and data sources used to analyze reviews differ. Whereas some

Table 1 Related studies on review helpfulness

Studies	Dependent variable		Operationalization		Determinants		Modeling		Research context
	Type of votes				Confirmation bias	Consistency of review content	Pricing	Two-Step	
	Helpful	Unhelpful							
Danescu-Niculescu-Mizil et al. (2009)	✓		Proportion		✓				Books on Amazon
Otterbacher (2009)	✓		Number				✓		Various product categories on Amazon
Schlosser (2011)	✓		Proportion		✓				Experiment
Baek et al. (2012)	✓		Number		✓		✓		Various products on Amazon
Zhu et al. (2014)	✓		Number		✓		✓		Hotels in San Francisco on yelp
Zhou and Guo (2015)	✓		Number			✓			Restaurants in 5 major US cities on yelp
Yin et al. (2016)	✓		Proportion		✓		✓		Apps from the Apple store
Wang et al. (2020)	✓		Number				✓		Hotels in NYC on tripadvisor
Zhou et al. (2020)	✓		Number			✓			Product types on Amazon
This study	✓	✓	Number		✓		✓	✓	Tablet PC reviews on Amazon for the first 2 weeks after submission

studies have a rather narrow scope and intensively analyze one particular product category offered on Amazon (e.g., Danescu-Niculescu-Mizil et al. 2009), others take a more generalized perspective by considering 24 different product types (e.g., Zhou et al. 2020). Relative to prior work in this field, our research is the only one that studies both helpful and unhelpful voting behavior as a two stage-process and considers all three types of determinants which we will discuss below in more detail. By extracting the information from the reviews only within the first two weeks after they were posted, we ensure that all of the reviews have approximately the same age. This procedure serves to reduce “winner takes all dynamics” which often occur among review voting (see Sect. 3.3). The domain of our research covers reviews on tablet PC offered at Amazon.

2.1 Determinants

In the following, we briefly summarize previous research that relates to the three groups of determinants of review helpfulness highlighted in Table 1, namely confirmation bias, consistency of review content, as well as price of the underlying product or product category. In doing so, we also derive a set of hypotheses emerging from the literature.

2.1.1 Confirmation bias

According to confirmation bias, individuals prefer information that is in line with their initial beliefs (Nickerson 1998). Using an experimental setting, Cheung et al. (2009) were the first to investigate confirmation of prior belief and found it to be positively influencing the review’s credibility. As many e-retailers make the distribution of ratings easy to access on the product introduction page, the average rating per product may support customers in forming their initial beliefs about a product. Consequently, we can assess confirmation bias with rating disagreement, i.e., the interplay of a focal review’s rating with the products’ average rating relates to helpfulness (Yin et al. 2016). We will refer to positive disagreement as the extent to which a review deviates positively from the average rating of the product. Naturally, negative disagreement refers to the extent to which the review deviates negatively from the average rating.

A lot of studies have addressed how a disagreement of a specific review’s rating from the product’s average rating affects helpfulness of a review: Baek et al. (2012) find disagreement to have a negative effect on helpfulness votes. These effects are moderated by product type and price, with the effect being stronger for experience and low-priced goods, respectively. Zhu et al. (2014) find that higher disagreement reduces the positive effects of online attractiveness and reviewer expertise respectively on the number of helpful votes.

The effects of disagreement may also be moderated by the controversy (i.e., the variance) of the products’ ratings and the valence of disagreement. Generally, as Pan and Zhang (2011) pointed out, controversy can lead to an elevated uncertainty about the validity of any specific review. If every review points in the same direction, users

tend to deem each individual review as more helpful, as opposed to when reviews lead to diverging conclusions.

Based on controversy and the valence of disagreement, Danescu-Niculescu-Mizil et al. (2009) only find support for the conformity theory if controversy is very low. If, on the other hand, the level of controversy is high, readers consider a review more helpful if it disagrees, especially in a positive way.

Interestingly, Yin et al. (2016) find that controversy reduces negative impact of rating disagreement, which is in contrast to the findings by Danescu-Niculescu-Mizil et al. (2009). However, Yin et al. (2016) did not take into account the sign of the disagreement. These diverging findings suggest that it may be useful to disentangle positive and negative disagreement.

Against the background of the mixed evidence in previous findings, we base our hypotheses on the most frequent findings which are supported by confirmation bias. Voters will form their opinions based on the elements available to them and the average star rating of a product is one of the most salient elements. Therefore, the more a review disagrees with the average star rating, the less helpful we expect it to be. Simultaneously, these initial impressions are stronger when the reviews available are unanimous, and weaker when they are controversial. Therefore, in a highly controversial setting, voters' beliefs will not be as strong and, therefore, weaken the effect of rating disagreement. We posit the following hypotheses:

H1a: (Positive and negative) disagreement reduces the number of helpful votes.

H2a: Controversy has a negative impact on the number of helpful votes.

H3a: Controversy weakens the effect of disagreement on helpful votes.

2.1.2 Consistency of review content

Another set of studies has addressed (in-)consistencies of review content by comparing arguments (Schlosser 2011) or sentiments (Zhou and Guo 2015; Zhou et al. 2020) to other review specific characteristics.

In general, sentiments can vary from truly positive through neutral to extremely negative. Thus, polarity is used to extract the emotional value a person gains from a purchased product (Wilson and Hoffmann 2009). Siering and Muntermann (2013) find that reviews with a stronger positive sentiment are more likely to get helpful votes. Their results are further moderated by product type such that reviews with negative sentiment polarity positively influence the number of helpful votes for experience goods, whereas consumers interested in search goods prefer positively framed reviews. Zhou and Guo (2015) studied the effect of sentiment polarity on helpfulness using data from restaurant reviews on Yelp.com. They observed a negativity bias, which is in line with Siering and Muntermann (2013) given that restaurant visits are experience goods. The existence of a moderating impact of product type on the effect of sentiment polarity is challenged by the results of Baek et al. (2012), who find that the number of negative words in a review increase the review's helpfulness both for search and experience goods.

We follow the majority of the results for search goods and hence assume:

H4a: Positive sentiment polarity increases the number of helpful votes.

However, we can conclude that there is much discrepancy regarding the effect of sentiment polarity on perceived review helpfulness. Review polarity is undeniably important, but there remains a number of open questions concerning this topic. We argue that applying polarity in the context of interactions with other variables might provide clarification. Consequently, we are interested to see how polarity interacts with the star ratings of a product and how this interaction influences the perceived helpfulness and unhelpfulness of reviews.

Lak and Turetken (2014) found that star ratings and polarity scores are often in agreement. However, it is not always valid to assume that a star rating and the written content of a review measured by a polarity score go in the same direction. Information inconsistencies of review content can easily occur. Hence the possibility that a review contains a positive rating and a negative review content, or vice versa, cannot be ruled out entirely. In such situations, it is necessary to think of the combined influences on the readers' attitudes towards the review and their voting behavior. In the following, we refer to the concept of consistency of review content by describing it as the extent to which a review's quantitative aspect (star rating) is in agreement with its qualitative aspect (sentiment polarity).

Please note that the topic of inconsistent information provision is not new in consumer research. Zhou et al. (2020) study whether the text and the sentiment in title and body are consistent, respectively. Zhou and Guo (2015) discovered a strong interaction effect of rating and sentiment for long reviews. This sounds reasonable, as it is easy to imagine that a review which is consistent in its rating and text valence is assessed as more trustworthy (Tsang and Prendergast 2009). Likewise, reviews that lack consistency make the reviewer appear less competent and persuasive (Schlosser 2011). We thus assume:

H5a: Consistent reviews are considered more helpful.

2.1.3 Price

Generally, the product price appears to be one variable which researchers have studied in less detail within the context of online review helpfulness. Even if a study addresses price as a determinant, only some studies measure it with the price of the product or service (Baek et al. 2012; Otterbacher 2009; Yin et al. 2016) whereas others rely on some proxy (Wang et al. 2020; Zhu et al. 2014). This is especially surprising since price is one of the primary motivators of people to shop online, as reported in a UPS survey on online searching and selecting behavior of customers (Gupta 2017). In addition, price also serves to determine the (perceived) risk involved with a purchase. It is hence not surprising, that consumers engage more with expensive products (Laurent and Kapferer 1985). In line with that, Liu et al. (2019) find that consumers tend to pay attention to reviews of higher priced product categories and for which they still require some information. Otterbacher (2009) finds a positive correlation of price with helpfulness. Baek et al. (2012) relate to Petty and Cacioppo's (1986) Elaboration Likelihood Model (ELM) and claim that the price of a product affects whether consumers are using central or peripheral

cues. However, their hypothesis is only partially supported. Their results show that people decide to up vote a review as helpful upon the basis of central cues when purchasing high-priced search goods, and on the basis of peripheral cues when buying low-priced experience goods.

In a more recent application of the ELM, Wang et al. (2020) find that price cues (in terms of money related words in the review text) only positively affect the number of helpful votes for low-class hotels. Another study from the hotel industry uses the number of \$-signs as a price indicator and finds that the price has a moderating effect: For higher priced hotels the positive impact of reviewer online attractiveness is higher whereas the positive effect of reviewer expertise is lower (Zhu et al. 2014). Comparable to Wang et al. (2020), we focus on a single product category and use product price as an indicator for involvement. In our application of ELM in the context of consumer involvement with regard to information processing of online reviews, we link the concept to the motivation to handle product information (Celsi and Olson 1988).

Considering a high-priced product will increase risk perception among consumers wishing to avoid a wrong purchase decision and also to waste money on a product which does not satisfy one's needs, it follows that consumers will be highly motivated to carry out a product information search in order to reduce or even eliminate the perceived risk and to make a better purchase decision in general. Consequently, we expect consumers to be more cognitively involved, and to scrutinize and intensively study the messages to obtain further information (Petty and Cacioppo 1986). In contrast, the consideration of buying a low-price product is connected to less serious negative consequences of a bad purchase decision. Therefore, we expect customers to search less extensively for information. In such cases, consumers take the peripheral route as a consequence of their lack of motivation and the limited efforts they are willing to invest. Because of that we assume

H6a: Low-priced products will obtain more helpful votes than high-priced products.

H7a: Low-priced products weaken the effect of rating disagreement.

2.2 Type of votes

In terms of the overall variable of interest, two groups of studies can be identified: One uses the number of helpful votes (e.g., Baek et al. 2012; Otterbacher 2009; Wang et al. 2020; Zhou and Guo 2015) another stream of literature uses the proportion of helpful votes among all votes as their dependent variable (e.g., Danesu-Niculescu-Mizil et al. 2009; Schlosser 2011). However, common to most of both approaches is to analyze the emergence of review helpfulness as a one-stage process. This implies that all reviews are equally likely to receive votes, which in turn can result in either over- or underestimating the impact some review features may have on the outcome. To the best of our knowledge, Zhu et al. (2014) were the first to account for this bias and acknowledge that some reviews did not receive any votes because they have not been read, while others have not been found helpful, which captures the decision to vote for a review and the conditional helpfulness ratio

Table 2 Overview of research hypotheses

Hypotheses	Expected effect on	
	Helpful votes	Unhelpful votes
H1 (rating disagreement)	H1a: negative	H1b: positive
H2 (controversy)	H2a: negative	H2b: positive
H3 (interaction of disagreement and controversy)	H3a: weaker	H3b: stronger
H4 (positive sentiment polarity)	H4a: positive	H4b: negative
H5 (internally consistent reviews)	H5a: positive	H5b: negative
H6 (low price)	H6a: positive	H6b: negative
H7 (interaction of low price and rating disagreement)	H7a: weaker	H7b: stronger

separately. Similar by idea, but conceptually different we also account for the possibility that some reviews may never receive any votes. As detailed below in Sect. 3.2, we do so by adopting a hurdle model approach (Mullahy 1986; Zeileis et al. 2008).

Our study further differs from previous research by distinguishing between helpful and unhelpful votes to examine whether the two measures are indeed determined by the same underlying logic. In Table 2 we summarize our set of previously discussed hypotheses on the expected effects of review characteristics on helpful votes (H1a–H7a). Given the absence of previous research on unhelpfulness we are assuming a corresponding set of reverse effects on unhelpful votes (H1b–H7b), which we will empirically examine in the following application study.

3 Empirical modeling setup

In this section we introduce how we measure our variables of interest, describe the modeling framework adopted in the subsequent application study, and introduce the empirical setting at hand.

3.1 Operationalization

Table 3 gives an overview of the dependent and independent variables in our study. We chose both stated helpfulness and unhelpfulness of reviews as our dependent variables and estimate two models. Instead of following the approach of other researchers (e.g., Danesu-Niculescu-Mizil et al. 2009; Schlosser 2011) to measure helpfulness as a ratio we change this ratio into a count measure. This transformation results in an offset variable which accounts for the total number of votes. This allows us to avoid some of the shortcomings of the helpfulness ratio, such as over- or underestimating the true (un-)helpfulness of reviews.

Controversy reflects the degree of disagreement among existing reviewers of a given product. We operationalize it as the standard deviation of the ratings provided in the reviews for the same product that antecede the review at hand. If past ratings of past reviews for the same product denote a high standard deviation, we assume a

Table 3 Overview of variables

Category	Name	Operationalization/comments
Dependent variable	Helpful votes	The number of "yes" answers provided to the question "Was this review helpful to you?"
	Unhelpful votes	The number of "no" answers provided to the question "Was this review helpful to you?"
Confirmation bias	Controversy	Standard deviation of the ratings provided in the reviews for the same product that antecede the review at hand
	Negative disagreement	Focal review's rating – average rating for the same product category if the focal review deviates negatively, 0 otherwise
	Positive disagreement	Focal review's rating – average rating for the same product category if the focal review deviates positively, 0 otherwise
Pricing	Negative disagreement X controversy	Interaction of controversy and negative disagreement
	Positive disagreement X controversy	Interaction of controversy and positive disagreement
	Low price	1 if price of the table is below 500*, 0 otherwise
	Low price X negative disagreement	Interaction of low price and negative disagreement
Consistency of review content	Low price X positive disagreement	Interaction of low price and positive disagreement
	Positive sentiment	1 if occurrences of positive words – occurrences of negative words > 0, 0 otherwise
	Positive sentiment X negative opinion	Interaction of positive sentiment and negative disagreement > 0
Control variables	High number of reviews	1 if the product being reviewed received more than 690 reviews*, 0 otherwise. This variable controls for the popularity of the product
	Log (word count)	Number of words in a single review. Similar to the study by Mudambi and Schuff (2010), this variable can account for review depth
	Review age	Number of days since the review was posted. This variable intends to control for exposure, i.e. the longer a review has been online, the more opportunities it has to receive more votes. According to Yin et al. (2016) it has a significant impact on helpfulness
	Days since first review	Number of days since the first review was posted, which serves as a proxy for product age
	log (Sales rank)	To control for the success of a product

Table 3 (continued)

Category	Name	Operationalization/comments
	Verified purchase	Provides the information on whether the reviewer has actually bought the reviewed product via amazon.com or not. It can account for reviewer credibility
	Reviews by reviewer	Number of reviews that the reviewer has posted so far
	log (Reviewer ranking)	Reviewer's ranking in "Amazon's Top Customer Reviewers" program
	Real name	1 if reviewer used her/his real name when submitting the review
	Location disclosed	1 if reviewer provided her/his geographic location
Offset	(Total votes)	Helpful + unhelpful votes

*We decide on the threshold by looking at the distributions of the variables. In both cases (price and number of reviews), we observe bi-modal distributions and use as threshold the value that separates the two distributions

high level of controversy. Conversely, if past reviewers were more homogenous in the ratings they provided, we see it as a sign of low controversy regarding the product's quality.

To measure how a specific review's rating disagrees from the average ratings for the same product, we will look at positive absolute deviation and negative absolute deviation separately. Herein, our operationalization differs noticeably from other studies (e.g., Yin et al. 2016). While consistency across reviews has so far mostly been measured as the absolute rating deviation, our approach enables us to check for a positive–negative asymmetry, which would also not have been possible using the signed difference between a rating and its average.

To determine whether a review was positive, neutral or negative, we extract the polarity score, a dictionary-based metric of the sentiment in a review. For this purpose, we apply an algorithm proposed by Rinker (2013) and available in the R package *qdap*, which utilizes a sentiment dictionary created by Hu and Liu (2004). It deducts the occurrences of negatives words from those of positive words to compute an overall polarity score for each review. Besides analyzing naturally evaluative words (e.g., “great”), this augmented sentiment analysis considers both valence shifters (e.g., “not”) and amplifiers (e.g., “very”). This allows the algorithm to identify a sentence such as “I am not satisfied with the product” as a negative, rather than positive sentence due to the valence shifter “not”. We will use the polarity score to test the concept of review consistency. Building on our polarity scores, we then measure consistent review content via the interaction between polarity and rating disagreement.

Furthermore, we create a binary variable *price*. We split the unique prices at the 75th percentile to distinguish between high-priced and low-priced products, which corresponds to a price cut at \$500. Everything below this reference point is considered a comparatively affordable product and everything above an expensive product. Please note that the price variable reflects information concerning different tablet brands.

We further interact our pricing variable with the disagreement of a focal review to assess how pricing effects may change depending on the consistency across reviews. Finally, we also take into consideration a set of control variables to account for further characteristics of a review.

3.2 Model

As already mentioned, we assume that voting for a review follows a two-step approach, i.e., reviews must have been displayed and read in the first place. Based on our individual level assumptions, a consumer then decides whether and how to vote on a review. We hence examine reviews regardless of whether they were evaluated or not. Although we do not directly observe the process assumed on the individual level, the outcome is reflected in our final dataset which evinces many zero observations for the number of un-/helpfulness votes. Implementing a hurdle regression (Mullahy 1986; Zeileis et al. 2008) allows us to handle both zero-inflation and over-dispersion, another common issue with count data. As many reviews do not receive

any votes at all, these data sets often contain more zero observations than a classic count model can account for. In addition, the hurdle model captures two different processes regarding the voting behavior, which a classic count model would ignore. First, we establish whether a certain review receives any helpful or unhelpful votes, which is captured by the hurdle model's right-censored binary part (i.e., zero component of the model) via a negative binomial distribution. This way, we can account for the latent propensity to vote, which we cannot observe directly. If the propensity becomes positive, a review was able to attract at least one un-/helpful vote. In such a case we can model the number of votes with a truncated count model in the second process. In this case, the hurdle model estimates the expected number of un-/helpful votes conditional on the probability of receiving a vote as established in the binary part (i.e., count component of the model). This count component follows a left-truncated negative binomial distribution and assesses the actual level of (un)helpfulness of a review. The formal specification of the hurdle model is as follows:

$$f_{Hurdle}(y;x, \beta, \gamma) = \begin{cases} f_{zero}(0; x, \gamma) & \text{if } y = 0 \\ \frac{(1-f_{zero}(0; x, \gamma))f_{count}(y;x, \beta)}{(1-f_{count}(0; x, \beta))} & \text{if } y > 0 \end{cases} \quad (1)$$

3.3 Data

To illustrate our approach, we model review (un-)helpfulness by using publicly available data provided by Wang et al. (2014). The dataset includes information on different tablet brands (such as, e.g., Kindle or Apple) available on amazon.com and their corresponding customer reviews and review information. We will however not consider all reviews as the context of review voting is prone to "winner takes all" dynamics, with very skewed distributions of votes across reviews. In particular, most reviews receive very little votes and a few reviews receive the vast majority of votes. This occurs in part due to the nature of the review management system applied by amazon, and that is not unique to this company. Reviews that receive more helpful votes tend to be displayed more prominently, which enables them to receive even more votes, perpetuating their position at the top (Liu et al. 2007). Thus, we restrict our analyses on users' reactions towards a review within the first two weeks after its submission, which reduces our sample to reviews posted between January 26 and July 04, 2012. This sample selection is supported by findings by, e.g., Yin et al. (2016), who have shown the natural tendency for reviews to receive fewer votes the older they get. Our novel approach to only consider reactions to a review that occurred within the first 2 weeks after posting also allows us to reasonably control at least one aspect of the early bird bias. By looking exclusively at the 2-weeks time frame of a review's existence, we prevent the exponential development of votes for helpful reviews that were posted early on from disproportionately influencing our results. To further account for these biases, we also include three additional variables in our model: high number of reviews, $\log(\text{Sales rank})$, and review age. To our knowledge, previous studies have ignored these sources of bias.

Because of this selection and further data cleaning, such as removing cases for which the number of helpful votes exceeded those of total votes, we arrive at a sample of 12,547 reviews. Table 4 briefly summarizes our variables and their correlation structure. As almost all correlations are well below the absolute value of 0.7, the correlation matrix does not indicate serious collinearities.

4 Results

Before presenting the effect of our estimates for the predictors of our juxtapositions (i.e., receiving at least one vote vs. changes in the number of votes and helpful vs. unhelpful voting patterns), we examine whether the data support our assumption of a hurdle process. The reviews in our data set receive on average 1.65 (1.76) (un-) helpful votes. Whereas there are some outliers (with more than 100 un-/helpful votes), more than one third of the reviews did not even receive one single vote. On the unhelpful side, the proportion of reviews without votes is even higher (more than 50%). To test, whether we indeed need a hurdle model, we perform a Wald test, which tests the pairwise equality between all coefficients from the two components and hence the necessity of the zero component (see Zeileis et al. 2008). For both models, we find evidence for the zero component which argues in favor of a hurdle model ($\chi^2 = 485.539$ and $\chi^2 = 2236.961$, concerning the helpful and unhelpful models respectively). Furthermore, both hurdle models are superior to classic count models in predicting the observed zero counts, which further supports our hurdle model assumption. Model evaluations based on the Akaike Information Criterion provide evidence that assuming a negative binomial distribution is superior or at least on par (Fabozzi et al. 2014) with other hurdle model specifications.¹

4.1 Results of the hurdle regressions

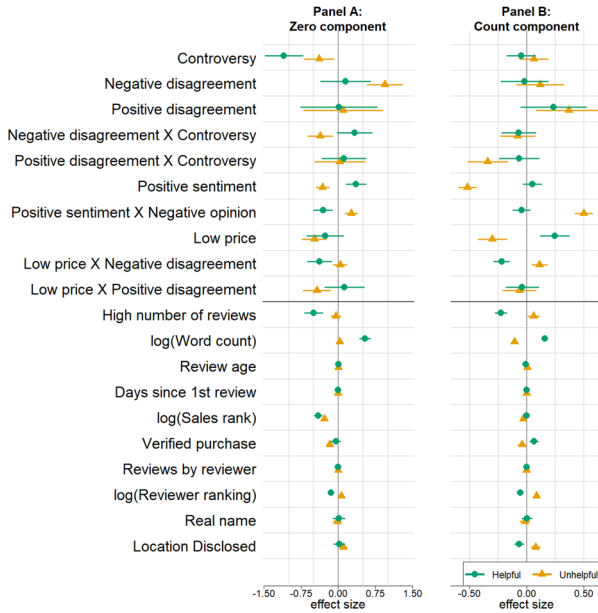
We plot the coefficient estimates (for the coefficients and significance levels see Table 5 in Appendix 1) of our two hurdle models with helpful and unhelpful votes as our dependent variables in Fig. 2. Panel A displays the results of the zero component of the models for both helpful and unhelpful votes, whereas the right panel displays the results of the count component of the models.

Before evaluating our set of hypotheses, we compare how our determinants affect the zero component opposed to the count component of the two hurdle models. Whenever a coefficient in the zero component and the count component is significant, they usually point in the same direction for all the predictors. Take, for

¹ We thank the two anonymous reviewers for drawing our attention to different model candidates. In particular, we tested the following models: classic count models based on Poisson and negative binomial distribution, respectively. Regarding the hurdle models we considered the following four models: binomial distribution for the zero component and Poisson and negative binomial distribution for the count components, respectively, Poisson for zero and count component, and negative binomial distribution for zero and count component.

Table 4 Descriptive statistics and correlations

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
Mean	1.337	0.547	0.540	0.830	0.929	0.533	4.564	14.965	152.685	1.969	0.694	7.130	14.459	0.219	0.358	
St. Dev	0.193	0.958	0.533	0.375	0.256	0.499	0.927	1.959	107.025	2.580	0.461	35.155	1.958	0.414	0.479	
Min	0	0	0	0	0	0	0	12	0	0	0	1	1.792	0	0	
Median	0	0	0	0	0	0	0	12	0	0	0	1	1.792	0	0	
Max	2.828	3.545	3.200	1	1	1	8.081	18	824	12.018	1	2,250	16.567	1	1	
1 Controversy	1															
2 Negative disagreement	0.109	1														
3 Positive disagreement	0.139	-0.578	1													
4 Positive sentiment	-0.088	-0.472	0.349	1												
5 Low price	0.052	-0.003	-0.009	0.005	1											
6 High number of reviews	0.003	-0.106	0.070	0.103	0.295	1										
7 log(Word count)	-0.036	0.085	-0.148	-0.030	-0.070	-0.181	1									
8 Review age	-0.014	-0.027	0.021	0.013	-0.013	0.031	-0.006	1								
9 Days since 1st review	0.106	0.087	-0.032	-0.067	-0.044	-0.245	-0.064	-0.014	1							
10 log(Sales rank)	0.021	0.111	-0.082	-0.132	-0.222	-0.816	0.122	-0.026	0.292	1						
11 Verified purchase	0.015	-0.115	0.095	0.079	0.042	0.062	-0.176	0.007	-0.017	-0.075	1					
12 Reviews by reviewer	-0.037	-0.022	0.002	0.015	-0.013	0.002	0.095	-0.003	-0.009	-0.006	-0.014	1				
13 log(Reviewer ranking)	0.095	0.177	-0.077	-0.106	0.036	0.098	-0.397	0.003	0.042	-0.068	0.044	-0.248	1			
14 Real name	-0.031	0.009	-0.032	0.007	-0.017	-0.019	0.126	-0.009	-0.014	0.009	-0.011	0.082	-0.208	1		
15 Location disclosed	-0.039	-0.014	-0.023	0.003	-0.018	-0.026	0.203	-0.009	-0.025	0.008	-0.035	0.081	-0.313	0.430	1	



Note: Marks ● and ▲ relate to parameter estimates of the helpful and unhelpful model, respectively. – indicates the 95% confidence interval.

Reading example: Controversy has a significantly negative effect (-1.093) on the probability of getting at least one helpful vote (p -value <0.05).

Fig. 2 Predictor effects on helpful and unhelpful votes

example, low price which has a negative effect on the probability of unhelpful votes (zero component) and the number of unhelpful votes (count component) at the same time. However, we also note that not all predictors show a significant effect in both equations. The differences in coefficients show that it is beneficial to consider the review voting behavior as a two-stage process, since the determinants for a review to get many votes are not necessarily the same that separate reviews that obtain no votes from those that obtain one vote or more. Take for instance the level of controversy and its effect on the number of helpful votes. While controversy has a negative effect in the zero component it has no significant effect in the count components. Furthermore, sizes of the coefficients differ across the two model components. Hence, predictors may affect the probability to vote and the number of un-/helpful votes differently, at least to some extent.

4.2 Comparing helpful to unhelpful voting behavior

In this subsection we aim to show how our groups of predictors affect the helpful vs. unhelpful voting behavior based on the estimates visualized in Fig. 2. We will accept our hypothesis as supported whenever the respective coefficient is significant with

a p -value < 0.05 and points to the hypothesized direction in at least one of the two model components.

4.2.1 Confirmation bias

We start by analyzing the effects of rating disagreement on review helpfulness. We assess H1a with negative and positive disagreement, respectively. In contrast to previous studies (e.g., Yin et al. 2016), disagreement (positive or negative) does not show a significant effect neither in the zero nor in the count component of the model (thus, H1a is rejected). We find a negative effect of controversy for the zero component of the hurdle model (H2a supported). Further, we do not find evidence for the interaction of controversy and negative or positive rating disagreement (H3a is rejected).

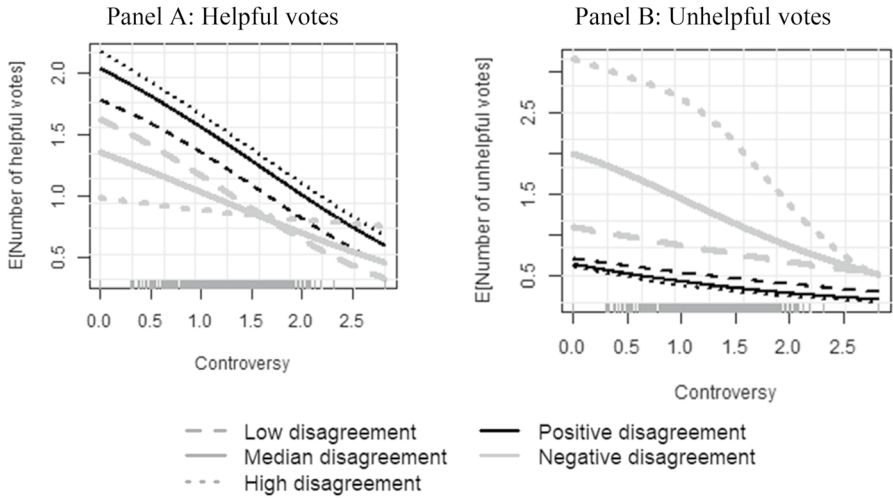
Moving to the effects on unhelpfulness, we find that positive and negative disagreement lead to more unhelpful votings, as the coefficient for negative (positive) disagreement is significant for the count (zero) equation, respectively (H1b supported). For controversy we observe a negative effect for the zero component (thus, H2b is rejected). Finally, both interaction terms of rating disagreement with controversy have a significant negative effect in at least one of the two model components and hence reduce instead of increase the effect of rating disagreement (H3b rejected).

We next construct marginal effect plots based on the overall expectation. This way, we are able to also control for the many interactions considered in our model and to ease interpretation of the coefficients. In the following plots, the vertical axis refers to the expected number of helpful votes or the expected number of unhelpful votes; the horizontal axis refers to a (continuous) explanatory variable, and the different lines represent different levels of another explanatory variable that interacts with the variable on the horizontal axis. As the interaction variable is continuous, we present the 5th, the 50th and the 90th percentiles of that variable, naming these low, median, and high. In doing so, we vary the determinants of interest between low, medium, and high levels and simultaneously fixing the remaining predictors.

The two panels of Fig. 3 show how the expected number of helpful votes (Panel A) and unhelpful votes (Panel B) are affected by different levels of controversy and disagreement.

Looking at the left panel we find that positive disagreement leads to more helpful votes than negative disagreement, as illustrated by the fact that the three lines corresponding to positive disagreement are higher than the three lines corresponding to negative disagreement. The only exception occurs when the level of controversy is high (i.e., above 2). This finding is consistent with the one from Danescu-Niculescu-Mizil et al. (2009), which lead them to move away from the theory of conformity. This plot one more time shows the negative effect of controversy on the number of helpful votes which is consistent with findings by Pan and Zhang (2011).

Turning to unhelpful voting patterns, negative disagreement makes the reviews receive more unhelpful votes. For high levels of negative disagreement, we see by far the highest expected number of unhelpful votes. Revisiting H2b, we can one



Note: The rug at the bottom of the plots shows the distribution of controversy.

Fig. 3 Marginal effects of controversy and positive/negative disagreement on the expected number

more time observe that in the presence of controversy among reviews the number of unhelpful votes decreases, which rejects the hypothesis. It seems that readers will be less certain about what to believe in and abstain from punishing if there is a lot of controversy among the reviews.

From this observation we conclude that helpful votes are generated differently than unhelpful votes. For the helpful voting behavior, we do not find a significant impact of rating disagreement, which suggests that confirmation bias is not an issue. This is in contrast to studies by, e.g., Baek et al. (2012) or Yin et al. (2016). However, these authors only looked at disagreement in general, whereas we differentiate between positive and negative disagreement, respectively. In fact, readers may value positive disagreement more positively than negative disagreement. It is important to note however, that controversy among the reviews' ratings reduces the number of helpful votes, which shows that readers have difficulties assessing the quality of reviews if opinions diverge. Regarding unhelpful voting behavior, people seem to punish reviews that disagree negatively from their prior belief, especially if they disagree negatively. One possible explanation could be the confirmation bias: People reading a tablet review are likely to have a good impression of the product. This suggests that they would punish reviewers who disconfirm their prior beliefs. As we do not find this effect in the helpful voting behavior, it seems that people punish more easily than reward (with a helpful vote if the review disagrees positively). Let us draw your attention one more time to controversy among reviews. Interestingly we find that increasing controversy does not lead to more punishing behavior among readers. In fact, increasing levels of controversy seem to work in the same manner as in the helpful voting behavior. It seems that readers having a harder time to evaluate the reviews which moves them away of casting a vote.

4.2.2 Effects of consistency of review content and price

For the remaining determinants, we again start with the effect on the number of helpful votes. From Fig. 2 we observe that reviews with an overall positive sentiment have a positive effect in the zero component of the hurdle model, which is in line with previous research (e.g., Siering and Muntermann 2013) and thus supports H4a. Reviews that are inconsistent, i.e., have an overall positive sentiment paired with a negative disagreement, have a negative effect in the zero component, which is in line with Zhou and Guo (2015) and thus gives us reason to support H5a.

Regarding the unhelpful voting pattern, we find strong evidence that reviews phrased with negative sentiments are prone to receive unhelpful votes (H4b supported). Similar to the helpfulness side, we also find evidence for H5b which means that inconsistent reviews are generally more prone to receiving unhelpful votes.

The results from both models suggest that people who are reading a review are guided by the tonality, i.e., sentiment of the text. Whenever this is positive, it increases the number of helpful or decreases the number of unhelpful votes. It is however important that the sentiment is in line with the rating of a review which should not disagree negatively as this may confuse readers.

We finally assess the effect of price and its interactions with rating disagreement on un-/helpfulness. According to Fig. 2, a focus on a low-priced product has a positive effect on helpful votes in the count component (H6a supported). In contrast, we find that low prices paired with negative disagreement have a negative impact on review helpfulness, whereas positive disagreement among low price products has no effect on the number of helpful votes (H7a is partly supported).

Regarding unhelpful voting behavior, a low price has a negative effect in both components of the hurdle model (H6b supported). Regarding moderation effects, we once again observe mixed findings depending on the sign of the rating disagreement. Negative disagreement seems to slightly increase the number of unhelpful votes for low priced products, as suggested by the positive significant effect in the count component. On the other hand, positive disagreement reduces the number of unhelpful votes, as suggested by the negative significant effects of both the zero and count components. This provides partial support for H7b.

The support of our hypotheses H6a and H6b suggests that readers generally take a less critical stance if prices are low. This is in line with findings from Wang et al. (2020). In such situations, readers tend to face less serious consequences if they make a wrong purchasing decision. Therefore, readers' motivation on elaborating the purchase decision is lower. However, reviews for low priced products are not immune to the effects of disagreement. If the review disagrees negatively, they can still expect fewer helpful votes and more unhelpful votes. If the review disagrees positively, it does not seem to affect helpful votes, but it reduces the number of unhelpful votes.

Finally, we have mixed findings for our control variables: Some neither show a significant impact on helpful or unhelpful voting behavior (e.g., real name), others (e.g., log(Word count)) have a positive impact on helpful and a negative impact on the number of unhelpful votes. Still, we also observe asymmetric effects in this

group of variables as $\log(\text{Sales rank})$ decreases the number of helpful as well as the number of unhelpful votes.

If we contrast these findings, we see one more time that disagreement's effect on helpfulness and unhelpfulness is more complex than previous literature suggests. As a rule of thumb, positive disagreement has no influence on helpful votes, but it increases the number of unhelpful votes. For low priced products, this positive disagreement seems to lead to fewer unhelpful votes whereas a negative disagreement lower the number of helpful votes and increases the number of unhelpful votes. The latter effect supports the confirmation bias theory.

5 Discussion and conclusions

This article contributes to the existing literature by contrasting two separate dimensions of the online reviews' environment. One of these dimensions refers to whether readers deem a review helpful, as opposed to unhelpful; the other refers to the differentiation between receiving at least one vote versus the number of votes received, conditional on having received at least one vote. We focused on three groups of determinants, namely confirmation bias, consistency of review content, and price. We extend the finding of previous research in the following way: First, we assume that voting for a review follows a two-step approach. Due to several reasons, such as the sheer bulk of reviews and limited time or cognitive capability from the side of the consumers, not all reviews will be read. Consequently, some reviews will not get any votes. By using a hurdle regression model, we control for this two-step process and split the voting process into two parts, the probability that a certain review will get at least one vote (zero component of the model), and given that it received at least one vote, the number of votes a review may get (count component of the model).

Using data on amazon customer reviews for tablet PC brands (Wang et al. 2014) we summarize the most important findings of our study on the drivers of review (un-) helpfulness are as follows:

- First, we find asymmetry between the driving forces of helpfulness and unhelpfulness in a sense that not all determinants affect the number of helpful votes in the same way they affect unhelpful votes. This especially holds for controversial reviews, which tend to translate into lower number of votes in general. This is an important finding as previous studies only looked into the effects on the number of helpful votes and implicitly assumed the effects to hold for unhelpful voting behavior as well (with an opposite sign). Our findings suggest that readers are less likely to cast their (un)helpfulness votes in the presence of divided opinions.
- Second, rating disagreement has different effects on helpful vs. unhelpful votes. Whereas, in general, we do not find any effects on the number of helpful votes if a review disagrees, readers tend to punish the review if it is in conflict with initial beliefs about the product. We can ascribe the latter to confirmation bias. As it

is only present among the unhelpful votes, readers obviously tend to punish more easily than to reward.

- Third, reviews written in a positive tonality are favored by readers. However, if the positive sentiment polarity is not consistent with rating disagreement, it will cause the review to get fewer helpful votes and more unhelpful ones.
- Finally, we find that price levels of the underlying products (which we consider as an indicator for perceived risk associated with the respective product) determine the amount of helpful and unhelpful votes. In general, we find that readers of reviews on lower priced products seem to take a less critical stance which results in more helpful votes and less unhelpful votes.

Our study has some important implications for e-retailers. Because of the asymmetric characteristics of some of the review aspects, focusing on the number of helpful votes is not enough. To better understand the overall voting patterns, a careful inspection of what drives helpful as well as unhelpful votes is important. Take for instance an e-retailer who only considers results drawn from the helpful model. If the review is written in a positive tone, i.e., positive sentiment polarity, the e-retailer may correctly assume that this will stimulate helpful votes and hinder unhelpful ones. She may hence consider placing it rather prominently on her platform. If the review however relates to a high price product, such a review may receive fewer helpful votes than a low price review would get, as users pay more attention when prices (and therefore risks) are higher. Because the effects are opposite for the effects on unhelpful votes she may correctly assume that such a review will provoke unhelpful votes more easily than reviews for lower priced products. Hiding these reviews will hence reduce the number of unhelpful votes. Interestingly, reviews posted in a product category with a lot of controversy will not get helpful votes easily. Consequently, our e-retailer may be tempted to hide these reviews as she may be fearing a low number of helpful votes and a higher number of unhelpful votes. However, we find that controversy shows asymmetric effects on helpful vs. unhelpful votes. Because of that, platform owners would not need to punish those reviews completely by moving them to the back of their review collection.

Overall, if a review provider wishes to make adequate recommendations to its users, it should find out which of its users would find a review helpful and which would not. Our study does not provide the necessary tools to evaluate such strategies, but it opens the discussion for such possibilities. The existing body of literature dedicated to UGC knows very little about unhelpful reviews and our study aims at taking the first steps towards that direction.

From a theoretical standpoint, our study unifies previous literature in explaining a widely studied phenomenon. Several studies, most of which we have reviewed in Sect. 2, posit well-founded arguments to explain the effects of rating disagreement

on helpfulness. However, the methodological approaches employed did not allow researchers to find conclusive evidence to support the theoretical frameworks they suggested. The confirmation bias, for instance, can only be truly observed by differentiating between helpful and unhelpful votes, since the consequences of a confirmation bias are more visible in readers' punishing behavior (unhelpfulness votes), than in their rewarding behavior (helpful votes).

Our study also has some limitations as it is concerned with data from amazon only. In addition, our results have been derived for search goods, i.e., laptop brands. Whether similar patterns also hold for other industries remains an interesting avenue for further research. For example, it would be interesting to explore if similar findings can be derived in a setting with more hedonic products and/or service categories. It might also be worth noting that the effects of rating disagreement are very hard to capture, because there is no way of knowing whether readers compare the current rating to the overall average. Future studies where the unit of observation is the reader who rates reviews, as opposed to the reviews, could build on our findings. In this regard, we need to concede that some of our discussion on the reported hypotheses tests could be subject to alternative explanations. However, to further investigate these and related claims we call for additional studies, ideally in controlled laboratory or even field conditions.

Finally, another interesting avenue for future research relates to the type of count models. Our approach analyzes helpful and unhelpful votes separately. A bivariate count regression on the other hand would assume the two dependent variables to be correlated (as both make up the total number of votes a review receives). On the other hand, we find that some predictors affect the two dependent variables differently which gives us reasons to believe that positive voting patterns might indeed be different and to some extent independent of negative voting patterns. In a similar manner, an alternative approach would be to use a model which captures the decision to vote in the first place (assuming that positive and negative patterns are comparable) and then to provide either a positive or negative vote.

Appendix 1: Coefficient estimates of the hurdle models

See Table 5.

Table 5 Coefficient estimates for the two hurdle models

Predictors	Helpful hurdle model		Unhelpful hurdle model	
	Zero	Count	Zero	Count
Constant	2.931***	-0.687***	0.042	-1.355***
Controversy	-1.093***	-0.048	-0.383**	0.063
Negative disagreement	0.150	-0.017	0.940***	0.119
Positive disagreement	0.015	0.235	0.103	0.368
Negative disagreement x Controversy	0.333*	-0.069	-0.359***	-0.079
Positive disagreement x Controversy	0.114	-0.064	0.034	-0.339
Positive sentiment	0.357***	0.050	-0.315***	-0.517***
Positive sentiment x Negative opinion	-0.307***	-0.045	0.265***	0.500***
Low price	-0.260	0.246***	-0.481***	-0.300***
Low price x Negative disagreement	-0.379***	-0.219***	0.038	0.113***
Low price x Positive disagreement	0.126	-0.039	-0.429***	-0.062
High number of reviews	-0.497***	-0.225***	-0.042	0.061**
log(Word count)	0.543***	0.160***	0.033*	-0.105***
Review age	0.005	-0.006	0.004	0.008**
Verified purchase	-0.048	0.065***	-0.173***	-0.039**
log(Sales rank)	-0.403***	0.001	-0.281***	-0.029***
Reviews by reviewer	-0.001*	-0.0005***	0.001***	0.001***
log(Reviewer ranking)	-0.148***	-0.053***	0.064***	0.086***
Real name	0.015	0.002	-0.021	-0.014
Location Disclosed	0.021	-0.067***	0.107***	0.078***
Days since 1st review	-0.002***	-0.0001	-0.0005***	-0.00002
Observations	12,547		12,547	
Log Likelihood	-16,333.020		-14,706.890	

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

Acknowledgements We are grateful to Radoslaw Karpienko for discussions and contributions to an earlier version of this paper. In addition, we thank Magdalena Breitrainer and Nathalie Neureither for further assistance with the study.

Funding Open access funding provided by Vienna University of Economics and Business (WU).

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Baek H, Joong Ho A, Choi Y (2012) Helpfulness of online consumer reviews: readers' objectives and review cues. *Int J Electron Commer* 17:99–126
- Bright Local (2019) Local consumer review study. <https://www.brightlocal.com/research/local-consumer-review-survey/>. Accessed 10 Jun 2020
- Cao Q, Duan W, Gan Q (2011) Exploring determinants of voting for the 'helpfulness' of online user reviews: a text mining approach. *Decis Support Syst* 50:511–521
- Cheung MY, Luo C, Sia SL, Chen H (2009) Credibility of electronic word-of-mouth: informational and normative determinants of on-line consumer recommendations. *Int J Electron Commer* 13:9–38
- Danescu-Niculescu-Mizil C, Kossinets G, Kleinberg J, Lee L (2009) How opinions are received by online communities: a case study on amazon.com helpfulness votes. In: Proceedings of the 18th international conference on World Wide Web. ACM Press, Madrid, Spain, pp 141–150
- Duan W, Gu B, Whinston A (2008) The dynamics of online word-of-mouth and product sales—an empirical investigation of the movie industry. *J Retail* 84:233–242
- eMarketer (2019) Consumer trust relies heavily on reviews and brand honesty. <https://www.emarketer.com/content/consumer-trust-relies-heavily-on-reviews-and-brand-honesty>. Accessed 10 Jun 2020
- Fabozzi FJ, Focardi SM, Rachev T, Arshanapalli BG (2014) The basics of financial econometrics: tools, concepts, and asset management applications. Wiley
- Gupta N (2017) Data backs amazon adding more perks to prime marketrealist. <https://marketrealist.com/2017/07/data-backs-amazon-adding-more-perks-to-prime/>. Accessed 10 Jun 2020
- Hu M, Liu B (2004) Mining opinion features in customer reviews. In: Proceedings of the National Conference on Artificial Intelligence, San Jose, CA, pp 755–760
- Huang AH, Chen K, Yen DC, Tran TP (2015) A study of factors that contribute to online review helpfulness. *Comput Hum Behav* 48:17–27
- Kahneman D, Tversky A (1979) Prospect Theory: an analysis of decisions under risk. *Econometrica* 47:263–291
- Kuan KKY, Prasarnphanich P, Hui K-L, Lai HY (2015) What makes a review voted? An empirical investigation of review voting in online review systems. *J Assoc Inf Syst* 16:48–71
- Lak P, Turetken O (2014) Star ratings versus sentiment analysis—a comparison of explicit and implicit measures of opinions *IEEE*, 796–805
- Laurent G, Kapferer J-N (1985) Measuring consumer involvement profiles. *J Mark Res* 22:41–53
- Liu J, Cao Y, Lin C-Y, Huang Y, Zhou M (2007) Low-quality product review detection in opinion summarization. In: Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, Association for Computational Linguistics, Prague, Czech Republic, pp 334–342
- Liu X, Lee D, Srinivasan K (2019) Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *J Mark Res* 56:918–943
- Marchand A, Hennig-Thurau T, Wiertz C (2017) Not all digital word of mouth is created equal: understanding the respective impact of consumer reviews and microblogs on new product success. *Int J Res Mark* 34:336–354
- Mittal V, Ross WT Jr, Baldasare PM (1998) The asymmetric impact of negative and positive attribute-level performance on overall satisfaction and repurchase intentions. *J Mark* 62:33–47
- Mudambi SM, Schuff D (2010) Research note: What makes a helpful online review? A study of customer reviews on amazon.com. *MIS Q* 34:185–200
- Mullaly J (1986) Specification and testing of some modified count data models. *J Econ* 33:341–365
- Nickerson RS (1998) Confirmation bias: a ubiquitous phenomenon in many guises. *Rev Gen Psychol* 2:175–220
- Otterbacher J (2009) 'Helpfulness' in online communities: a measure of message quality. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, ACM Press, Boston, Mass, pp 955–964
- Pan Y, Zhang JQ (2011) Born unequal: a study of the helpfulness of user generated product reviews. *J Retail* 87:598–612
- Petty RE, Cacioppo JT (1986) The elaboration likelihood model of persuasion. *Communication and Persuasion, Springer Series in Social Psychology*. Springer, New York, pp 1–24
- Qazi A, Syed KBS, Raj RG, Cambria E, Tahir M, Alghazzawi D (2016) A concept-level approach to the analysis of online review helpfulness. *Comput Hum Behav* 58:75–81

- Rinker TW (2013) qdap: Quantitative discourse analysis package [Computer software manual] Buffalo, New York. Retrieved from <http://github.com/trinker/qdap> (2.2.4)
- Schlosser AE (2011) Can including pros and cons increase the helpfulness and persuasiveness of online reviews? The interactive effects of ratings and arguments. *J Consum Psychol* 21:226–239
- Siering M, Muntermann J (2013) What drives the helpfulness of online product reviews? From stars to facts and emotions. *Wirtschaftsinformatik Proceedings* 103–118
- Singh JP, Irani S, Rana NP, Dwivedi YK, Saumya S, Roy PK (2017) Predicting the ‘helpfulness’ of online consumer reviews. *J Bus Res* 70:346–355
- Trustpilot (2019) Leading factors that increase vs. decrease trust in brands among digital shoppers worldwide. <https://www.emarketer.com/chart/230558/leading-factors-that-increase-vs-decrease-trust-brands-among-digital-shoppers-worldwide-june-2019-of-respondents>. Accessed 10 Jun 2020
- Tsang ASL, Prendergast G (2009) Is a ‘star’ worth a thousand words?: The interplay between product-review texts and rating valences. *Eur J Mark* 43:1269–1280
- Wang X(S), Feng M, Chiang RHL (2014) Database submission—market dynamics and user-generated content about tablet computers. *Mark Sci* 33:449–458
- Wang EY, Fong LHN, Law R (2020) Review helpfulness: the influences of price cues and hotel class. In: Neidhardt J, Wörndl W (eds) *Information and communication technologies in tourism*. Springer Nature, Surrey, pp 280–291
- Wilson TJW, Hoffmann P (2009) Recognizing contextual polarity: an exploration of features for phrase-level sentiment analysis. *Comput Linguist* 35:399–433
- Yin D, Mitra S, Zhang H (2016) Research Note—When do consumers value positive vs. negative reviews? An empirical investigation of confirmation bias in online word of mouth. *Inf Syst Res* 27:131–144
- Zeileis A, Kleiber C, Jackman S (2008) Regression models for count data in R. *J Stat Softw* 27:1–25
- Zhou S, Guo B (2015) The interactive effect of review rating and text sentiment on review helpfulness. In: Stuckenschmidt H, Jannach D (eds) *E-commerce and Web Technologies*. Springer International Publishing, Cham, pp 100–111
- Zhou Y, Yang S, Li Y, Chen Y, Yao J, Qazi A (2020) Does the review deserve more helpfulness when its title resembles the content? Locating helpful reviews by text mining. *Inf Process Manage* 57:1–11
- Zhu L, Yin G, He W (2014) Is this opinion leader’s review useful? Peripheral cues for online review helpfulness. *J Electron Commer Res* 15:267–280

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.