IMPACT OF ONLINE WORD-OF-MOUTH ON AMAZON RANKS IN SELECTED
PRODUCT CATEGORIES

by

IRINA GRINBERG

A Thesis submitted to the

Graduate School-New Brunswick

Rutgers, The State University of New Jersey

in partial fulfillment of the requirements

for the degree of

Master of Science

Graduate Program in Food and Business Economics

written under the direction of

Dr. Sanjib Bhuyan

and approved by

________________________

________________________

________________________

________________________

New Brunswick, New Jersey

October, 2012
ABSTRACT OF THE THESIS
IMPACT OF ONLINE WORD-OF-MOUTH ON AMAZON RANKS IN SELECTED
PRODUCT CATEGORIES

By IRINA GRINBERG

Thesis Director:
Dr. Sanjib Bhuyan

In this thesis, I address a growing trend in how consumers obtain information prior to making their online purchasing decisions. Specifically, I investigate the relationship between online word-of-mouth (WOM; customer reviews and ratings) and sales, measured by Amazon.com “ranks”, for different product categories. I collect daily data from Amazon.com for twenty products in ten categories from December 8, 2011 to February 29, 2012 in an effort to capture changes in online WOM and how it impacts sales. I expand on the traditional metrics used to measure online WOM and mimic how consumers use review information on Amazon.com to make their purchases there.

As part of my approach, I take into account the endogenous nature of reviews via instrumental variables. Results of this thesis show that in the Electronics category (1) the number of reviews significantly impact its rank, (2) a two-star rating (compared to a one-star rating) versus a four-star rating (compared to a one-star rating) has a greater impact on its rank, (3) Amazon’s Facebook.com “Likes” feature significantly impacts rank in Books, Electronics and Music, and (4) the impact of online WOM is significant for experience goods, but not for search goods.
My findings suggest that online WOM metrics impact consumer purchase decisions through awareness and persuasion. Moreover, it underlines the emerging role of social media in sales outcomes. From a managerial standpoint, these findings suggest that businesses should use ratings, reviews, and social media to build awareness about their products and influence purchase decisions. Future research in online WOM and sales (or rank) outcomes for product categories should control for advertising and product heterogeneity within categories, as well as incorporate social media metrics, review text and length.
Acknowledgement and/or Dedication

I want to acknowledge my thesis advisor, Dr. Sanjib Bhuyan, for his unwavering confidence and encouragement on both a personal and professional level. I am very grateful for his continued support and careful guidance throughout the entire research process. I also want to acknowledge both of my committee members, Drs. Yanhong Jin and Lei Wang. Dr. Jin’s technical expertise helped immensely with the empirical portion of my thesis, and Dr. Wang’s familiarity with the online word-of-mouth literature and her marketing expertise gave my work more perspective and depth. I want to acknowledge Brian Thompson, a Computer Science PhD student at Rutgers University, for his outstanding work in creating the program, “Amazon Info Collector,” which not only made it possible for me to easily collect my data, but allowed my thesis to evolve into something larger than anticipated. Last, but most certainly not the least, I want to acknowledge my family, Zoë Wild, for showing me that the (seemingly) impossible is possible, my childhood friend, Kritee Gujral, and my loving boyfriend, Joshua Javaheri, for their love, support, and firm belief in all I do and wish to achieve. Thank you everyone!
Table of Contents

Abstract of the thesis ................................................................. ii
Acknowledgement and/or Dedication ............................................ iv
List of tables ................................................................................ vi
Chapter I ....................................................................................... 1
   Introduction ............................................................................... 1
Chapter II ...................................................................................... 4
   Literature Review ....................................................................... 4
   2.1 Theoretical Framework ....................................................... 4
   2.2 Empirical Findings .............................................................. 20
Chapter III ................................................................................... 33
   Methodology ............................................................................. 33
   3.1 Hypotheses .......................................................................... 33
   3.2 Data ....................................................................................... 36
   3.3 Model ..................................................................................... 40
   3.4 Econometric Approach ....................................................... 42
Chapter IV .................................................................................... 47
   Results ....................................................................................... 47
   4.1 Descriptive Statistics .......................................................... 47
   4.2 Regression Analysis ............................................................ 50
Chapter V ....................................................................................... 54
   Conclusions ............................................................................... 54
References .................................................................................... 72
Table 1: Description of Extracted Variables ...............................................................37
Table 2: Datasets and Number of Observations ..........................................................39
Table 3: Description of Variables in Model ..................................................................41
Table 4: Description of Variables Used as Instrumental Variables ...............................41
Table 5: Descriptive Statistics for Volume, Valence, Variance and Additional Metrics ..................................................................................................................49
Table 6: Effect of WOM Volume on Category Rank ......................................................50
Chapter I

Introduction

In general, consumers seek information about a product prior to making a purchase decision. With the internet becoming a greater part of personal and professional life, online user reviews are an increasingly important source for discovering product quality (Zhu and Zhang 2010). For example, Cone Trend Tracker find in their 2011 survey that 64% of their responders search for consumer reviews prior to purchase (up from 55% in 2010) (Cone Trend Tracker 2011). The literature refers to online reviews as online word-of-mouth (WOM). As a formal definition, online WOM is “peer-generated product evaluations posted on company or third part websites” (Mudambi and Schuff 2010). Online WOM both substitutes and complements traditional WOM interaction (e.g., between friends or family) and business-to-consumer communication (e.g., advertising) (Chevalier and Mayzlin 2006). For example, consumers may rely on traditional or online WOM, advertising, or a combination of these sources to fulfill their product information needs.

The literature suggests that online user reviews influence sales (Godes and Mayzlin 2004; Liu 2006; Duan, Gu and Whinston 2008; Dellarocas, Zhang, and Awad 2007; Chevalier and Mayzlin 2006; Zhu and Zhang 2010; Clemons, Gao, and Hitt 2006; Moe and Trusov 2011); however there is less agreement on which metric, namely volume\(^1\) or valence\(^2\), is most important (Godes and Silva 2011). Considerable attention is given to experience goods, whose dominant attributes are difficult to assess prior to

\(^1\) Number of reviews
\(^2\) Average rating
purchase, over search goods, whose dominant attributes are easier to assess prior to purchase, and analysis is generally limited to single product categories, especially box office movies (Liu 2006; Duan, Gu and Whinston 2008; Dellarocas, Zhang, and Awad 2007). Researchers also differ on their treatment of online WOM as endogenous or exogenous (Duan, Gu, Whinston 2008).

Despite empirical support for the efficacy of online reviews in driving sales, there are still reasons to suspect that its role may be limited. Credibility of online reviews may be discounted as many firms “fake” consumer-to-consumer WOM (Zhu and Zhang 2010; Dellarocas and Narayan 2006) and that highly dissatisfied customers tend to post reviews (Anderson 1998). Reliance on online reviews may vary based on whether a product is experience versus search, and the anonymity of online WOM may make it less credible compared to traditional WOM (Dellarocas 2003).

In an attempt to fill a gap in the literature, this thesis examines the relationship between Amazon ranks³ and online WOM for ten experience and search categories in Amazon’s Hot New Releases⁴ during the regular and holiday season. I examine volume, valence and add Facebook.com Likes⁵ as an additional metric. Unlike previous studies, I use review helpfulness⁶, number of comments for reviews⁷, and variance of ratings as instruments to account for the endogenous nature of volume (Duan, Gu, and Whinston

---
³ Amazon.com provides product ranks, which are based on sales. The number one ranked item is the highest selling item in a given category; rankings appear to range from one to over a million (Chevalier and Mayzlin 2006).
⁴ Hot New Releases on Amazon.com contain best-selling items that are updated hourly.
⁵ Individuals may view the number of people who have liked a product by looking at the number of Facebook.com Likes.
⁶ The number of people who find a review to be helpful expressed as a fraction – for example, “1 of 2 people found the following review helpful.”
⁷ Individuals may read comments posted about a particular review – for example, a review may have 4 comments.
2008; Godes and Mayzlin 2004) using 2SLS with fixed effects. Data is collected daily from December 8, 2011 through February 29, 2012. The specific objectives are as follows:

1. Address which online WOM metric, or combination of metrics, is most important in influencing rank for different product categories.

2. Address the impact of online WOM on rank for experience versus search categories.

The organization of this thesis is as follows: Chapter 2 reviews the theoretical and empirical literature related to the impact of online WOM on sales, and evidence for the endogeneity of WOM. In Chapter 3, I discuss the methodological approach, which includes the hypotheses, introducing the data, and the empirical model. In Chapter 4, I provide the empirical results and discussion. I conclude in Chapter 5, present limitations, and direction for future research.
Chapter II

Literature Review

This thesis draws on several streams of literature to understand the relationship between online WOM and consumer behavior (i.e., purchase decisions). In the first section, I present the theoretical framework for understanding WOM, and in the second section, I present findings from the empirical literature.

2.1 Theoretical Framework

2.1.1 The Role of WOM as a Predictor

The “herds and information cascades” literature provides theoretical basis for the role of WOM as a behavioral driver (Banerjee 1992, 1993; Bikhchandani, Hirshleifer, and Welch 1992). This literature suggests that individuals optimally choose to follow observable information cues from others in the cascade, but ignore their own set of opinions (“private information”). For example, Bikhchandani, Hirshleifer, and Welch (1992) use informational cascades to explain fads, politics, and peer pressure. While these models do not incorporate actual WOM interaction per se, the dynamic behavior of observational implies that the influence of WOM is time invariant (Godes and Silva 2011).

In the “learning from others” literature, Ellison and Fudenberg (1993, 1995) and McFadden and Train (1996) suggest that the level of WOM and type of interaction impacts consumer behavior in different ways. For example, Ellison and Fudenberg (1995) find that when there is limited WOM interaction, the average payoff from a purchase decision is identical, social learning is efficient (i.e., everyone picks the best choice). Lastly, Mayzlin (2006) studies WOM behavior directly when firms pose as
consumers. In the following paragraphs, I discuss papers from both strands of literature, i.e., (“herds and information cascades” and “learning from others.”)

Banerjee (1992) develops a sequential decision model in which rational decision makers (“agents”) ignore their own information; instead, agents mimic choices made by others, i.e., “benign” herd behavior. In this model, obtaining information (“signals”) is costless, untradeable, and unobserved among agents. Herd behavior is rational because other people have relevant signals for rational agents. In equilibrium, however, Banerjee shows that such behavior is suboptimal; when everyone does what everyone else is doing, even when an individual’s private information suggests something else, an agent’s decision becomes less responsive to his or her own information, and thus less helpful to others. That rational agents ignore their own information suggests that online reviews may be a powerful signal for potential buyers who have limited information about product quality.

Bikhchandani, Hirshleifer, and Welch (1992) show that convergence to an optimal outcome is possible in herd behavior when communication between agents is both credible and costless. They explain that the addition of new information causes interrupted changes in individual behavior in terms of an informational cascade. The informational cascade occurs when an agent’s actions do not depend on his or her private information, which shifts the social equilibrium. Individuals enter the model sequentially, observe previous decisions made by others, and decide to adopt or reject a given behavior. The presence of marginally new information may change the social equilibrium significantly, suggesting a different behavioral pattern, or even that a societal shift is taking place. Thus, cascades can explain both uniform behaviors as well as drastic
societal or economic changes. Bikhchandani, Hirshleifer, and Welch conclude that individuals lacking information or experience will rely on others for information. For this research the Bikhchandani, Hirshleifer, and Welch work gives a solid basis for suggesting that uncertain individuals do rely on others (i.e., WOM) to fulfill their information needs, and that such information may significantly impact final decisions.

In the “learning from others” literature, Ellison and Fudenberg (1993) model social behavior in the adoption of competing agricultural technologies. They assume agents are not fully rational, relying instead on “rules of thumb,” (i.e., learning shortcuts) to make decisions. In the first model (homogenous population), the “rule of thumb” is to only use the most recent information to make a decision. By specifying certain popularity weights, Ellison and Fudenberg show that in the long run, agents choose and benefit from the superior technology. Their research provides theoretical support in considering the importance of “rules of thumb” (potential buyers may view only the most recent Amazon review(s), in my thesis) and popularity (possibly signaled through Facebook.com Likes, in my thesis).

Building on their 1993 study, Ellison and Fudenberg (1995) find that the structure of WOM communication helps determine whether a better technology, product, or practice is adopted. For example, an agent “hears” a “report” on a product from a random sample of other agents. Agents can change their decision based on the most recent report received from others, ignore information, or altogether not seek information in a given period. Ellison and Fudenberg demonstrate two cases: identical and divergent payoffs from purchase decisions. When payoffs are the same, limited WOM generates conformity (market share of a choice converges to a single choice); when there is greater
WOM, diversity occurs (market share changes or is divided). When payoffs are different there may be (1) diversity in choice (2) all agents pick the optimal choice (efficient social learning), or (3) all agents pick a suboptimal choice (inefficient herding). Interestingly, they conclude that efficient social learning occurs when WOM communication is limited among agents.

Mcfadden and Train (1996) show that learning from others helps reduce risk when agents face uncertainty; however, the decision to wait until one can observe others’ experiences implies that utility gains from immediate consumption are delayed. Moreover, because tastes are different, the learning may be incomplete to the extent that preferences do not match. In this rational decision model, social learning is exogenous; the choice to buy a product now, or wait to learn from others is a dynamic optimization (unlike Ellison and Fudenberg 1993, 1995). Additionally, unlike Ellison and Fudenberg (1993, 1995), Mcfadden and Train assume that the learning impact is fixed over time, varies by agent, and involves no search costs. For single product purchases (e.g., books and movies), learning from others is effective when those products satisfy at least fifty percent of the population; the minimum share is lower for repeat purchases than for single purchases. Learning from others is based on agents’ knowledge of decision payoffs; therefore, agents make rational choices, and conformity occurs if the average choice payoff is sufficiently high.

In what McFadden and Train call the “popularity effect,” certain specifications of social learning can lead to a single product (irrespective of quality) capturing the entire market. This is unlike the result in Ellison and Fudenberg (1993, 1995) where behaviors diverge in equilibrium given substantial WOM activity.
To summarize, McFadden and Train (1996) provide a theoretical basis for understanding WOM as a rational behavior choice used to reduce risk. In my research, Amazon.com consumers can access several information sources to inform their buying choice (e.g., reviews and ratings), which to some extent, resemble McFadden and Train’s assumption that agents observe others’ decision payoffs, and Facebook.com Likes may produce a “popularity effect.” Thus, when consumers observe that the ratings, reviews, and/or Likes are sufficiently favorable or high for a product, they are more likely to commit to buying it.

Mayzlin (2006) explore online WOM when firms pose as consumers. Using a game theoretic model, where advertising and WOM are perfect substitutes, they explore if (1) WOM maintains its credibility, and (2) if competing firms invest more in promoting low or high-quality products. The model includes two products of different, but uncertain quality to consumers; consumers rely on online content to tell which of the two products is superior. In spite of promotional chat behavior by firms, Mayzlin (2006) shows that, in equilibrium, rational individuals still pay attention to these chats. Moreover, firms selling inferior quality products will actually engage in more promotional chat because they cannot obtain free publicity from legitimate WOM. Mayzlin (2006) support the efficacy of WOM in impacting consumer behavior, even when firms pose as consumers “faking” such interaction.

The social network literature also supports the impact of WOM on consumer behavior (Brown and Reingen 1987). This literature emphasizes tie strength (a measure of social relation and frequency of contact, e.g., a friend versus acquaintance) in moderating the impact of WOM (Granovetter 1973) and homophily (the likeness in
individuals based on age, sex, education, and social status) (Rogers 1983). People tend to interact with like-minded individuals, and strong social ties imply a stronger level of homophily (Granovetter 1973). Thus, when tie strength and homophily is strong, WOM has a greater impact on behavior. In the online environment, tie strength and homophily may be based on cues such as gender, age, professional background; however these cues are often missing, hidden, or purposefully false. Thus, assessment based on homophily is often not possible, leaving consumers to rely on other cues (Grewal, Gotlieb, Marmorstein 1994). In my thesis, however, the presence of Facebook.com Likes may improve perceived homophily when consumers see their friends Liking products on their social media pages.

To summarize, the “herd behavior and informational cascades” literature suggests that WOM as a result of social networks is “contagious and persistent” (Liu 2006) in its impact on consumer behavior. While models on “herds and information cascades” do not incorporate WOM per se, the dynamic pattern of observational learning implies that the influence of WOM is time invariant. That is, in the cascade, the nth individual’s observation of individual (n – 1) has an identical impact on his decision, as does the observation of person(n + t – 1), where t is any number of players entering a cascade, on individual (n + t) (Godes and Silva 2011).

The “learning from others” literature suggests that peer learning is a rational choice that reduces risk (McFadden and Train 1996) and that its impact on behavior depends on the extent of WOM. On the one hand, Ellison and Fudenberg (1993, 1995) find that when choice payoffs differ, learning from others may result in divergent market shares, a superior choice picked by all agents, or inefficient learning. When payoffs are
the same, limited WOM produces efficient learning. On the other hand, according to McFadden and Train (1996), given substantial WOM activity, learning can lead to a single (even, a poor quality) product capturing the entire market. When firms pose as consumers using WOM, there is no change in consumer behavior (Mayzlin 2006). Lastly, homophily and tie strength are important in moderating the influence of WOM (e.g., a high level of homophily implies a stronger influence of WOM on behavior) (Brown and Reingen 1987).

2.1.2 The Role of WOM as an Outcome

There is comparatively less theoretical basis for what actually drives online WOM engagement; however the literature acknowledges that WOM is endogenous (Godes and Mayzlin 2004; Duan, Gu, and Whinston 2008). Anderson (1998) develops a utility based model to understand the relationship between satisfaction and WOM engagement, and several researchers shed light on psychological reasons driving WOM behavior (e.g., Sundaram, Mitra, and Webster 1998; Hennig-Thurau et al. 2004). Researchers also focus on the dynamics of ratings themselves; they find that average ratings fall over time (Li and Hitt 2008; Goes and Silva 2011; Wu and Huberman 2008), but disagree on why this occurs. Li and Hitt (2008) argue for self-selection, while Wu and Huberman (2008) and Godes and Silva (2011) argue for a motivation-based theory. While modeling the ratings environment is not the focus of this research, it is a viable area for future inquiry. In the following paragraphs, I discuss the literature on WOM as an outcome of previous behavior.
Anderson’s (1998) utility based model has a U-shaped function representing negative and positive WOM. He shows that increasingly positive or negative experiences with a product should increase the marginal utility of engaging in WOM; thus, activity is expected to be greater for customers whose product-experiences tend to the positive or negative polarities. Given that there is diminishing marginal utility to increasing customer satisfaction, the function is also asymmetrical. That is, diminishing returns to higher levels of customer satisfaction implies that decreases in satisfaction always drives use of WOM more compared to increases in satisfaction.

Several researchers focus on unusual advertising as the driving force behind WOM engagement (Bayus 1985; King and Tinkham 1990). Others emphasize that consumers are more likely to use WOM with others when their opinions make them appear more intelligent (Dichter 1966; Sundaram, Mitra, and Webster 1998; Hennig-Thurau et al. 2004). For example, Sundaram, Mitra, and Webster (1998) find that positive WOM is used to enhance self-confidence through helping others; on the other hand, negative WOM is used to reduce anxiety, vent, or seek advice. Similarly, Hennig-Thurau et al. (2004) point out that individuals’ desire to enhance their self-worth along, with their concern for others, is important in driving online WOM interaction. Engel, Blackwell, and Miniard (1993) discuss that the propensity to use WOM increases in the purchase of controversial products as consumers try to reduce doubt.

Wu and Huberman (2008) point out that the order of opinions (i.e., the “ratings environment”) is most important in explaining lower average ratings. They show that people are more likely to use WOM when they anticipate their comments will matter. That is, prior to deciding to add a review, an individual will consider its impact versus the
cost of submission. A review’s impact magnifies the greater its deviation from the average opinion or rating, and decreases with the number of reviews already existing. Drawing on information cascades, Wu and Huberman explain that people behave differently than the observed behavior (i.e., the observed average rating) in the cascade.

On the other hand, Li and Hitt (2008) argue that online product reviews exhibit self-selection bias, which explains the decline in average ratings. Self-selection bias happens when early buyers have different preferences about product quality than those who purchase a product later; it also suggests that some users self-select to post early leaving those who are less likely to post. Li and Hitt’s findings imply that due to differences in preferences, user reviews may be biased signals of unobserved product quality. They show that early book reviews are systematically positively biased, and that consumers do not discount early ratings as signals of product quality. Godes and Silva (2011) study both ordinal and temporal effects of ratings and offer an alternative explanation for consumer ratings behavior. They find support for the motivation-based theory offered by Wu and Huberman (2008), but find inconsistencies explaining the self-selection theory offered by Li and Hitt (2008).

In bridging social network analysis to the online environment, Brown, Broderick and Lee (2007) show that a website behaves as a “social proxy” for tie strength. That is, the relationship between an individual and a website is strengthened when a website is more interactive and personalized. Homophily is evaluated based on the closeness of a user’s psychological characteristics to a website (e.g. emotional closeness), and credibility is evaluated in terms of perceived website and contributor credibility. Credible sources are those that have greater expertise and trustworthiness (Kelman 1961; Buda and
Zhang 2000); expertise and trustworthiness have been found to be positively associated with consumer brand attitude, intentions, and actual behaviors (Brown, Broderick and Lee 2007). For my research, this suggests that reviews rated as very helpful may be perceived as more reliable than less helpful reviews. Thus, in accounting for the endogeneity of reviews, I use the reviews’ perceived helpfulness.

The existing theoretical literature suggests that online WOM is an outcome of previous behavior, i.e., (1) it is endogenous; (2) satisfaction and WOM are related, and in particular, the use of WOM is more likely when dissatisfaction is high (Anderson 1998); (3) WOM engagement depends on product experience (i.e., positive versus negative) and psychological factors, including enhancing self-image and altruism (Hennig-Thurau et al. 2004; Sundaram, Mitra, and Webster 1998); (4) the likelihood to post a review is related to motivation, which increases the more dissimilar an opinion is to the observed reviews (Wu and Huberman 2008) and declines as individuals self-select to post a review early (Li and Hitt 2008); (5) the declining trend in the average ratings across products can be attributed to either the ratings environment (Wu and Huberman 2008) or self-selection bias (Li and Hitt 2008), and support for the ratings environment (Godes and Silva 2011) is found thus far; lastly, (6) trustworthiness and expertise of reviewers is important in mediating the impact of WOM on consumer behavior.

In my thesis, I draw on Wu and Huberman (2008) to support the use of variance of ratings as an instrument for the volume of reviews. Individuals are more likely to post a review when they believe that their comments will matter, comparing its potential impact to the cost of submission. When reviews are sufficiently different, there may be less incentive to post a review. While Wu and Huberman specifically compare a new
rating’s impact to how different it is to the average; however, use of the average rating as an instrument would be problematic as it is likely to be correlated with product sales. I also use lagged volume to capture Wu and Huberman’s finding that individuals are less likely to post a review when volume of reviews is sufficiently high. I draw on Brown, Broderick and Lee (2007) to justify the use of review helpfulness as an instrument for the volume of reviews. That is, helpful reviews will be relied on more than comparatively less helpful ones, and will reduce the incentive to post additional reviews. As far as comments on reviews, these may be positive or negative. Thus, it is not clear how a greater number of comments may impact the number of reviews posted, or their perceived credibility.

2.1.3 Experiments and Surveys

Traditional empirical literature on WOM is experimental and survey oriented. Reingen et al. (1984), for example, find that sorority women living close to each other report similar brand preference, whereas those living further away from each other report dissimilar preferences. Drawing on Granovetter (1973) theory on social ties, they measure brand taste similarity in terms of whether women lived in the sorority house. The study, however, is unable to rule out that reported preferences may due to the fact that women of similar tastes also tend to live in the same place. In a survey study, Richins (1983) focuses on WOM as an outcome of consumer behavior. Richins (1983) identifies three variables that correlate with the propensity to spread negative WOM: nature of dissatisfaction, perceptions of blame for the dissatisfaction, and perception of speediness of retailer response.
Foster and Rosenzweig (1995) infer learning from others through WOM in the adoption of high yield varieties. In prescription drug adoption by physicians, Van den Bulte and Lilien (2003) find evidence for social influence in driving prescription drug choice by doctors. They emphasize the importance of distinguishing between awareness and evaluation phases; the awareness phase is related to advertising effects, and the evaluation phase is related to social influence.

In an online experiment, Senecal and Nantel (2004), examine the influence of recommendation sources on product choice. The experiment includes 487 subjects looking at three websites, four recommendation sources, and two products. Using generalized estimating equations, they find that subjects viewing recommendation systems choose recommended products twice as often as subjects who do not view recommendations. Moreover, recommendations for experience goods are more influential on product choice than for search goods.

Schlosser (2005) finds that individuals posting reviews are influenced by existing reviews. Specifically, she shows that in trying to appear indiscriminate when reading a negative review, individuals actually post a more negative review. Moe and Trusov (2011) suggest that Schlosser’s (2005) explanation may be applicable to the downward tendency of ratings that Li and Hitt (2008) and Godes and Silva (2011) observe. Kumar and Benbasat (2006) conduct an experiment using filtered Amazon.com data. They find that the presence of reviews positively impacts the perceived usefulness of a website. Consistent with homophily, Forman, Ghose, and Wiesenfeld (2008) show that reviews have a greater impact on sales when reviewer identity (e.g., name or location) is revealed.
These surveys and experiments suggest that WOM influences consumer behavior. People rely on WOM more for experience goods than search goods (Senecal and Nantel 2004), individuals posting reviews are influenced by existing posts (Schlosser 2005), and reviews are important in improving perceived website quality (Forman, Ghose, and Wiesenfeld 2008).

2.1.4 Understanding Search and Experience Goods

Traditional models in information economics assume that, for either buyers or sellers, the optimum level of search for any product occurs when the cost of the search is equal to the expected marginal return (Stigler 1961; Bei and Widdows 2004). Obtaining information can be costly in terms of time spent for search, and trade-offs exist between perceived cost and benefits of additional search activity (Stigler 1961). In his work on the economics of information and advertising, Nelson (1970) builds on that idea, classifying what he calls search and experience goods. Later, he modifies this differentiation, noting that goods can be seen as having a mix of search and experience qualities (Nelson 1974).

Nelson (1974) categorization rule is as follows: a good is classified as a search good when full information for “dominant” product attributes are known before purchase; a good is classified as an experience good when either (1) full information on the product’s main features cannot be assessed without direct purchase and experience and/or (2) search for information on these dominant attributes is more costly than experiencing the product. Credence attributes are added to include goods whose features are difficult to assess even after purchase and use (Darby and Karni 1973).
Classification of goods into either search or experience is not wholly unanimous. Cameras (Nelson 1970), natural supplement pills (Weathers, Subhash, and Wood 2007), toys, furniture, sporting equipment, and footwear (Franke, Huhmann, and Mothersbaugh 1996; Zeithaml and Bitner 2003; Nelson 1970), as well as cell phones (Bei and Widdows 2004; Mudambi and Schuff 2010) are classified as search goods. Wine (Klein 1998), music (Bhattacharjee et al. 2006), books (Chevalier and Mayzlin 2006), video games (Zhu and Zhang 2010; Mudambi and Schuff 2010), movies (Duan, Gu, and Whinston 2008; Liu 2006), and MP3 players (Weathers et. al. 2007) are considered experience goods in the literature. Other experience goods include DVD players, household and kitchen appliances, GPS devices (Franke, Huhmann, and Mothersbaugh 2004; Girard, Korgaonkar, and Silverblatt 2003). The literature is less conclusive on classification of office machines and computer hardware/software.

Weathers, Subhash, and Wood (2007) add that if a consumer has to explore a product further than simply reading information, such a good is an experience good. Moreover, taking into account sampling costs helps to classify products into search versus experience. Consumers will always sample a good when sampling through purchase is less costly than search. When product search is more costly, this product will be an experience good. As an example, cars are considered to be search goods, despite the fact that important attributes, such as car performance on longer trips and in the snow, are experiential in nature; this is because it is less costly to search for information (e.g., through WOM, test driving) than to sample the car through buying (Klein 1988).

Researchers note that the Internet may be changing the traditional landscape of search versus experience goods. They point out that because the Internet allows
consumers to access product information easily and learn from one another, product attributes become searchable (Alba et. al 1997; Klein 1998). Customers benefit from the Internet’s reduced cost of obtaining and sharing information (Morton and Silva-Risso 2006) and new ways to experience products prior to purchase (e.g., interactive media) (Lynch and Ariely 2000). As Klein (1998) illustrates, websites selling wine can give consumers superior information about the wine (e.g., information about flavors, expert knowledge, and customer reviews) versus wine shopping offline. Thus, by reading product reviews, consumers can “experience” products prior to actual purchase.

Nevertheless, researchers support the continued use of Nelson (1974), i.e., focusing on the dominant attribute in classifying goods (Huang, Lurie, and Mitra 2009).

In decision-making and information search, studies emphasize that search constitutes both physical and cognitive costs for the consumer (Johnson, Bellman, and Lohse 2003). Bettman, Johnson, and Payne (1993) and Lynch and Ariely (2000) point out that cognitive processing levels change depending on different information types and structures (e.g., website format). Cognitive processing is expressed in terms of a “time per acquisition” metric. These researchers also explain that consumers tradeoff information search and cognitive expenditure with accuracy (Johnson and Payne 1985). The availability of decision and comparison tools (Todd and Benbasat 1991) and ratings (Poston and Speier 2005) reduces physical and cognitive costs, ultimately improving the purchase decision process. Differences in search versus experience attributes can alter how consumers process information. While search attributes are more straightforward (e.g., price, color, shape, dimensions) and involve less cognitive processing, processing information on experience attributes is more taxing, and may involve looking at customer
reviews (Weathers, Subhash, and Wood 2007). WOM information for experience attributes is also more subjective, implying consumers must obtain information from multiple sources for final product evaluation (Coupey 1994).

Integrating these findings, Huang, Lurie, and Mitra (2009) confirm that assessing search versus experience attributes requires different levels of cognitive effort. They show that use of “communication mechanisms,” such as consumer reviews is significant for experience goods, but not for search goods. From a theoretical perspective, the researchers emphasize the continued relevance of Nelson (1974) experience versus search classification based on dominant attributes. They add that in the online setting, the type of information is sought, and how it is processed, may be more important than ability to assess product quality prior to purchase. One major limitation of the Huang, Lurie, and Mitra study is that the researchers were unable to pinpoint whether differences in experience versus search goods are due to differences in the kind of information accessed, the way the information is presented by the seller, or the way consumers process the information.

While it is simpler to categorize individual goods as either experience or search, problems arise for this thesis when attempting to label whole categories as either experience or search given the heterogeneity of products within a category. Some categories are broad and contain individual products that may be experience, search, or both. For example, Electronics contains a router (search good) as well as an Ipod Touch (experience good). Similarly, Health & Personal care contains baby pampers (search good) and therapeutic electronic devices (experience good). Because the literature is silent on many of the specific products sold in these categories, it may be argued that
categories have both experience and search qualities. Nevertheless, I draw on Nelson’s (1974) definition, focusing on products’ dominant attributes in a given category to label categories as experience versus search. Continued relevance of Nelson’s (1974) experience versus search categorization is supported in Huang, Lurie, and Mitra (2009). Thus, I consider Appstore for Android, Books, Music, Movies & TV, Videogames, Electronics, and Grocery & Gourmet Food to be experience categories, and Health & Personal Care, Home & Garden, and Office Products to be search categories.

2.2 Empirical Findings

2.2.1 Overview

The existing empirical literature covers online WOM in various categories, namely, books, box office movies, TV shows, craft beer, videogames, beauty products and travel. Books and box office titles remain the most popular categories due to data availability; however these categories exhibit short-lived and predictable life-cycle patterns (Moe and Trusov 2011). This suggests the need to study the impact of online WOM in different product categories. Two main metrics of online WOM are generally measured: valence and volume (Chevalier and Mayzlin 2006). Valence is an average rating (Chevalier and Mayzlin 2006; Clemons, Gao, and Hitt 2006; Dellarocas, Zhang, and Awad 2007; Duan, Gu, and Whinston 2008) used to capture the positivity or negativity of ratings (Godes and Mayzlin 2004; Chevalier and Mayzlin 2006; Liu 2006). Volume is the number of review posts. Variance, also called dispersion, is measured as statistical variance (Clemons, Gao, and Hitt 2006) or entropy⁸ (Godes and Mayzlin 2004) define dispersion as the entropy of conversations across Usenet discussion boards. As an example, variance is largest, and the entropy is
2004); it is used to a far lesser extent as predictor of sales. The literature is unclear in regards to which of these metrics is most important. When volume is significant, it signals WOM’s awareness role, i.e., spreading information (Godes and Mayzlin 2004; Dellarocas, Zhang, and Awad 2007; Zhu and Zhang 2010), and when valence is significant, it signals WOM’s persuasive role, i.e., influencing decisions (Liu 2006; Chatterjee 2001).

Previous studies vary on their treatment of WOM as endogenous or exogenous. By treating WOM as exogenous (Dellarocas, Zhang, and Awad 2007) researchers do not account for the theoretical basis of WOM as an outcome of previous consumer behavior (Dellarocas and Silva 2011; Godes and Mayzlin 2004). Moreover, measurement of online WOM in a cross-sectional (Chatterjee 2001; Liu 2006) versus panel context (Duan, Gu, and Winston 2008) matters. By using cross-sectional data, researchers cannot account for product heterogeneity, nor explain whether observed sales differences are related to unobserved product quality, or due to online WOM (Duan, Gu, and Winston 2008).

In this thesis, I tailor the traditional metrics of volume and valence more closely to what Amazon.com consumers see on a product webpage, and incorporate Facebook.com Likes as part of that effort. I also control for the endogeneity of online WOM volume using comments on reviews, helpfulness of reviews, and variance as instruments.

smallest, when all posted reviews are in a single newsgroup. On the other hand, entropy is largest, and variance is smallest, when posted reviews are spread equally across newsgroups when there is at least one post
2.2.2 The Role of WOM as a Predictor

This section focuses on the empirical work relevant to the impact of online WOM on product sales. Chevalier and Mayzlin (2006) investigate the impact of customer reviews on relative books sales at Amazon.com and Barnesandnoble.com (bn.com). They collect book information on price, promised shipment time; from the 500 most recent reviews, they extract the star rating, posted date, and rank. Lower rank on both websites implies higher sales, and more stars imply a higher rating. First, Chevalier and Mayzlin examine their data in a cross-sectional context. Overall, for both websites, the valence improves sales. For example, they find that if a book receives one, two or three star rating on Amazon.com, rank on Amazon.com increases (sales fall) holding rank at bn.com constant, but that four or five stars ratings decreases rank (sales increase).

Breaking down valence into fraction of one and five star reviews, Chevalier and Mayzlin (2006) find that five-star reviews improve sales, and one star ratings decrease sales on Amazon.com. For bn.com, the fraction of five-star ratings is surprisingly insignificant and carries the wrong sign, but the fraction of one-star reviews is significant, suggesting that consumers at bn.com discount the credibility of highly rated books. They also find that review length is positive and statistically significant in improving sales on Amazon.com. In the second part of their analysis, the researchers use a difference-in-difference approach to account for unobserved fixed effects for the two websites and produce similar results: positive reviews increases relative sales at that website, and the impact of one-star reviews exceeds the impact of five-star reviews.

To summarize, Chevalier and Mayzlin (2006) find that the relative books sales at Amazon.com and bn.com are related to differences in volume and valence across both sites. However, they are unable to show that retailers can profit from reviews – it may be
that reviews simply shift sales around in a website. In this thesis, I incorporate the metrics measured in Chevalier and Mayzlin, with the exception of review length. Different from Chevalier and Mayzlin, I use cumulative variable formulations for volume and valence to more closely mimic what consumers see on an Amazon product webpage.

In TV shows, Godes and Mayzlin (2004) link the dispersion of WOM across online communities to Nielsen viewership ratings using a diffusion model. They analyze the data using the truncated and full sample with rating as the dependent variable. In the truncated sample, the positive coefficient on the lagged entropy variable, suggests that greater dispersion in WOM is associated with higher future ratings early in a TV show’s life cycle. Thus, higher dispersion means that more people become informed about a TV show across online communities. Using the full sample, Godes and Mayzlin (2004) compare early versus late WOM. Similar to the first analysis, they find that the impact of entropy on ratings diminish over time, but argue that the dynamic nature of WOM is still strongly supported – dispersion in one period and its relationship to ratings is related to dispersion in the following period. Using Seemingly Unrelated Regressions (SUR) with rating, review post, and lagged entropy as the dependent variables, Godes and Mayzlin find that while volume is insignificant in the truncated fixed-effects model using valence data, it becomes significant in the SUR in a later period. They conclude that volume more fully captures information about past ratings.

To conclude, Godes and Mayzlin (2004) emphasize the managerial importance of tracking online WOM early in a product’s life cycle across different communities online. At a given volume, a greater number of individuals find out about a TV show, implying greater dispersion, or reach, across communities. Godes and Mayzlin caution that their
data may be prone to sample selection bias, emphasizing the importance of controlling for confounding factors, such as advertising, and using metrics beyond volume. In this thesis, I incorporate measures beyond volume and capture early product life cycle patterns.

Liu (2006) examines WOM dynamics in Yahoo!Movies prerelease message boards on motion picture revenues to assess awareness versus persuasive effects. Following previous studies on box office movies, the author studies the effects and antecedents of WOM for movies – the opening week, the weeks following the opening, and aggregate box office revenues. Data on pre-release and post-release movies for the five month period between May and September 2002, in addition to public data on box office revenues, production budget, and other related film-variables is used. Because consumer attitudes do not predict actual behavior perfectly (Ajzen and Fishbein 1980), the role of valence is unclear in his view. In terms of the antecedents of movie WOM, Liu identifies five variables: genres, Motion Picture Association of America (MPAA) ratings, star power (whether the cast has major Hollywood stars), critics’ reviews, and consumer WOM in the previous week.

Using multiple regression with the log of box office revenue as the dependent variable, Liu (2006) finds that the volume of online WOM has significant explanatory power for aggregate and weekly box office revenue, especially during the early weeks after the release of film. The primary antecedent of WOM after movie release is the volume of prior week’s WOM is evidence that WOM is endogenous; however he treats WOM as exogenous using a single-equation ordinary least squares estimator. Liu (2006) cautions that the analysis relies on correlations between variables at a given point in time
to deduce how WOM impacts box office sales. Thus, conclusions about causality cannot be made. Nonetheless, their results emphasize the need for managers in the industry to track online WOM, especially early on in movie’s life cycle, and that online WOM is useful in forecasting film revenues.

Duan, Gu, and Whinston (2008) assess the persuasive and awareness effect of online reviews via simultaneous equation and GMM using panel data. Obtained from Yahoo!Movies.com, Variety.com, and BoxOfficeMojo.com, their data consists of user review’s yahoo ID, post data, overall grade, grade for movie plot, acting, visual, review length, daily gross revenues, rank, average revenue per theatre, and number of participating theatres for 71 movies between July 2003 and May 2004. Following previous studies in this industry, Duan, Gu, and Whinston jointly estimate daily revenue and daily review post in the first (opening) week and in the second week following a movie’s release. Using a 3SLS approach, their study shows that valence impacts movie sales in the short-run, but not in the long-run when controlling for product heterogeneity. Thus, there is weaker evidence for the persuasive effect of online WOM. However, confirming Liu (2006), they find that the volume of WOM is significant in predicting box office sales, supporting its awareness role. Their research is the only one in the field, to my knowledge, to employ pure panel data. In my thesis, I evaluate the persuasive versus awareness role for different product categories, while controlling for the endogenous volume of reviews. This will be discussed in Section 3.4.

Dellarocas, Zhang, and Awad (2007) propose a hazard function based on the Bass Diffusion Model\(^9\) to investigate the effect of WOM on consumption incidence of

\(^9\) The Bass Diffusion Model is a non-linear model developed by Frank Bass used to describe how new products are adopted as an interaction between users and potential users (Bass 1969).
entertainment goods and forecast box office sales. Since entertainment goods typically exhibit declining post release marketing, they investigate sales growth patterns. Dellarocas, Zhang, and Awad employ a rich data set from Yahoo!Movies (all titles released in 2002), BoxOfficeMojo (to obtain weekly movie sales data and marketing expenses), and Hollywood Reporter (a measure of an actors’ “bankability”), and individual level demographic information. Their results show that early volume of online reviews is an important predictor of early motion picture sales. Additional exogenous information such as pre-release marketing, critical reviews, and theatre availability also significantly contribute to forecasting box office sales accurately.

Unlike previous studies, Zhu and Zhang (2010) use the psychological choice model to study how product and consumer characteristics moderate the impact of online reviews on product sales in videogames. In their model, an influencer (online reviews) is moderated by the environment as well as contextual factors (consumer and product attributes); the interplay of these variables produces the response (final purchase decision). In their case, the individual characteristic is consumer internet experience, and the product characteristic is popularity. They obtain monthly sales and videogame data for PlayStation 2 and Xbox from NPD Fun Group from October 2000 to 2005; monthly review data is obtained from GameSpot.com, a widely popular hub for video videogame reviews, for every videogame between March 2003 and October 2005.

Like Chevalier and Mayzlin (2006), they use a difference-in-difference approach, in addition to a two-stage nested logit demand model, to show the decision process in videogame purchasing. Zhu and Zhang (2010) find that the valence, volume, and the
coefficient of variation\textsuperscript{10} of reviews impact sales of less popular games, and conclude that online WOM is plays an “information role” (similar to Duan, Gu and Whinston’s (2008) “awareness role”) in unpopular game titles among players with greater internet experience.

Zhu and Zhang (2010) bridge contradictory findings in Chevalier and Mayzlin (2006) in books sales as well as Duan, Gu, and Whinston (2008) and Zhang and Dellarocas (2006) in box office movies. While Duan, Gu, and Whinston (2008) point out that mixed findings in the literature on online WOM are due to data limitations and treatment of WOM, Zhu and Zhang (2010) suggest that future studies consisting of different product types can mitigate these discrepancies. For example, reviews for products that are typically purchased online will have a greater impact on sales than products typically purchased offline, where online versus offline is a product characteristic in the psychological choice model (Zhu and Zhang 2010). My thesis incorporates their suggestion insofar as studying a diverse set of product categories.

Moe and Trusov (2011) measure the impact of social dynamics of ratings on subsequent ratings behavior and product sales using weekly retail sales for 500 beauty products from December 2006 to 2007. They model the arrival of posted ratings and separate the effects of social dynamics on ratings from a baseline ratings behavior (a consumer’s “socially unbiased” evaluation). The ratings arrival is modeled within each star level (five-star scale) as five separate, but interrelated hazard processes. In this way, Moe and Trusov take into account the timing and valence simultaneously, and are able to

\textsuperscript{10} The coefficient of variation is the ratio of the standard deviation of user ratings to the mean rating. Thus, the coefficient of variation measures how much disagreement there is about a videogame. While high variation is risky and potentially rewarding, low variation signals a “safe bet” to potential videogame purchaser.
compute the ratings metrics with and without social influence. Observed ratings are
deconstructed into a baseline ratings component, social dynamics component, and error
component. Finally, they model sales as a function of these components, incorporating
recent work in opinion dynamics by Li and Hitt (2008) and Godes and Silva (2011).

It is important to note that in Moe and Trusov (2011), price does not change over
time, and that the variation in the arrival of ratings does not depend on variation in
previous sales. While this may be true of mature product categories, it is likely not true of
newer products, given evidence that WOM is endogenous. To conclude, they find that
ratings dynamics have an impact on sales through their valence, but not through volume
or variance.

In craft beer, Clemons, Gao, and Hitt (2006) analyze online reviews using
multiple regression, with the sales growth rate as the dependent variable, to evaluate the
effectiveness of two theoretical product differentiation strategies: hyper-differentiation\textsuperscript{11}
and resonance marketing\textsuperscript{12}. Unlike previous literature treating product reviews as a
measure of product quality, which then influences sales, Clemons, Gao, and Hitt use
online WOM to study the influence of product positioning on the sales growth rate. They
explain that with highly informed consumers, firms providing more differentiated
products should have higher sales growth rates than firms who offer relatively less
differentiated products. Using data from April 2000 to July 2004 of 281,868 ratings for
1,159 craft brewers from 6,212 reviewers on Ratebeer.com, they find that the valence and

\textsuperscript{11} Hyper-differentiation theory says that because firms can produce nearly anything a consumer may want, product variety increases (Clemons, Gao, and Hitt 2006)
\textsuperscript{12} Resonance marketing suggests that buyers will respond strongly to products which answer to their needs and wants (Clemons, Gao, and Hitt 2006).
variance is positively and significantly correlated with the sales growth rate. This result supports resonance marketing.

Interestingly, Clemons, Gao, and Hitt (2006) find that while higher ratings are an apt predictor of increasing future sales, poor ratings are not an apt predictor of decreasing future sales. They suggest that firms may benefit from making new products that elicit more reviews in general, as well as more positive reviews. Similar to Chevalier and Mayzlin (2006) and Dellarocas, Zhang, and Awad (2007), they find that higher average ratings are correlated with higher craft beer sales. However, unlike Chevalier and Mayzlin (2006), who find that the fraction one-star ratings impact sales rank more than five-star ratings on bn.com, Clemons, Gao, and Hitt finds the opposite. They explain that while premium beer commands a higher price versus sub-premium beer, prices for best sellers and average quality books are similar. For example, a consumer who believes a certain beer is excellent is a better predictor of future sales, versus a consumer who believes a beer to be average. Secondly, because craft beer is purchased repeatedly, those who give high ratings are most often repeat buyers. Thus, higher ratings are more important in predicting sales than lower ratings.

To summarize, Clemons, Gao, and Hitt (2006) find that the valence and variance are correlated with the sales growth rate. The researchers emphasize that product context matters when studying online WOM, arguing that dealing with highly differentiated, high-priced, or repeat-purchase products (unlike, for example, books and movies), calls for a reevaluation of traditional online WOM metrics.

In tourism, Ye et al. (2011), find that traveler ratings have a significant impact on the number of published reviews on Ctrip.com, an online China-based travel agency.
Unlike the previous literature discussed in this section, Ye et al. use the number of published reviews as a proxy for sales, and collect review volume, text, posted date, and rating as well as hotel description on Ctrip.com from February 2007 through January 2008. They find that positive ratings increase the number of bookings, and that variance in opinions does not necessarily result in fewer bookings. This latter result suggests that travelers consider overall opinion in the presence of extremely positive or negative ratings.

Previous findings suggest that online WOM has an impact on sales; however, studies differ in metrics measured, how those metrics are computed, and in econometric approaches used. While volume and valence are consistently used in the literature, their impact on sales is inconsistently supported. Moe and Trusov (2011) and Clemons, Gao, and Hitt (2006) find support for valence, while Liu (2006) and Duan, Gu, and Whinston (2008) find greater support for volume. Zhu and Zhang (2011) and Chevalier and Mayzlin (2006) find that both volume and valence are significant. The variance metric, however, is inconsistently used in the literature. In some cases variance is not at all measured (Chevalier and Mayzlin 2006; Liu 2006; Duan Gu and Whinston 2008; Dellarocas, Zhang and Awad 2007). In other cases, researchers find that ratings dynamics impact the arrival of future ratings through variance (Moe and Trusov 2011), or in the case of highly differentiated or repeat-purchased items, variance is correlated with the sales growth rate (Clemons, Gao, and Hitt 2006), or has no significant impact on hotel bookings (Ye et al. 2011). This suggests that the variance metric may be better understood through its impact on the number of reviews, not sales directly. In this thesis, I use variance (along with comments on reviews, helpfulness of reviews, and lagged
volume) as an instrument capturing how the distribution in ratings impacts the number of reviews.

### 2.2.3 The Role of WOM as an Outcome

In their discussion on problems associated with measuring online WOM, Godes and Mayzlin (2004) emphasize that WOM is endogenous. That is, WOM is a result of past consumer behavior. Anderson’s (1998) U-shaped function explaining satisfaction and WOM communication is supported empirically in the context of firm customer-initiated contacts\(^\text{13}\) (CICs) (Bowman and Narayandas 2001; Dellarocas and Narayanan 2006). In explaining satisfaction resulting from a CIC, Bowman and Narayandas (2001) find that perceived brand quality, level of disconfirmation of expectation about the contact, and perceived fairness of manufacturers’ provision of information (e.g., brochures and samples) are important. Confirming Anderson (1998), they find that loyal customers are more likely to use WOM when they are highly dissatisfied.

Liu (2006) and Duan, Gu, and Whinston (2008) show that the impact of WOM volume in a previous period increases current WOM volume (“buzz effect”); however, the effect becomes insignificant in analysis of the second week of their datasets. Similarly, Godes and Mayzlin (2004) find that the impact of WOM, measured by lagged entropy, on TV show rating falls over time.

In box office movies, Dellarocas and Narayanan (2006) focus on consumer propensity to use WOM and provide important implications for viral marketing campaigns. Successful viral marketing implies that “spreading the word” produces

\(^{13}\text{Customer-initiated-contact occurs when customers directly initiate conversation about products with manufacturers.}\)
multiplicative effects in creating product awareness and interest. They introduce a new metric called “density” to measure the ratio of the total number of people who rated a product over the number of people who purchased a product in a given period; this metric approximates the conditional probability that a buyer will post a rating online, and can be viewed as an additional measure of marketing campaign success (e.g., similar to click-through rates). Their data consists of movies titles and corresponding critic and user ratings extracted from Yahoo!Movies released in 2002, and sales, budget, and marketing expense data from BoxOfficeMojo. Using a logit multiple regression model on the weekly density, they find (1) support for Anderson (1998) observing a U-shaped relationship between perceived quality and ratings density, (2) a negative relationship between movie availability and density of ratings, and (3) greater disagreement in critic reviews leads to higher density of ratings. They conclude that the antecedents to traditional WOM can be applied to online WOM (e.g., as in Bayus 1985; Sundaram, Mitra, and Webster 1998).

In this thesis, I take into account the “buzz effect” using the lag of Cvolume. I do not include lagged rank in my model because of its correlations with cumulative WOM variables, which already capture past WOM. This will be discussed in more detail in the Results section.
Chapter III

Methodology

3.1 Hypotheses

This section formulates the hypotheses that fulfill my objectives: **Objective 1** is to address which online WOM metric, or combination of metrics, is most important in influencing category rank. **Objective 2** is to address the impact of online WOM on rank for experience versus search categories.

**Objective 1** is addressed in **H1-H3. Objective 2** is addressed in **H4.** Based on literature on “herds and information cascades” (Banerjee 1992 and Bikchandani, Hirshleifer, and Welch 2003), “learning from others” (McFadden and Train 2003) and previous empirical literature, the following hypotheses specify a positive impact of volume on sales – i.e., negative impact on rank (Duan, Gu, and Whinston 2008, Liu 2006, Dellarocas, Zhang, and Awad 2007, Chevalier and Mayzlin 2006, Zhu and Zhang 2010) and a positive impact of valence on sales – i.e., negative impact on rank (Chevalier and Mayzlin 2006, Dellarocas, Zhang, and Awad 2007, Zhu and Zhang 2010, Ye et al. 2011, Moe and Trusov 2011). The literature also suggests that when valence is significant it signals the persuasive role of WOM, and when volume is significant, it signals its awareness role.

Different from previous studies focusing on experience goods with predictable product life cycle patterns (e.g., box office movies and books), this thesis includes different experience and search categories. In addition to testing the effect of valence and volume, I incorporate Facebook.com Likes, and detailed information on reviews posted, which are used as instruments to account for the endogeneity of online WOM volume.
Formulations of variables were chosen to most closely capture what consumers actually see on Amazon.com: consumers may look at the overall review content for a product, and/or they may look at only the most recent review. Thus, I measure volume at the cumulative and daily level. While cumulative volume gives a summary about a product’s review content, daily volume represents the most current review activity. The literature suggests that users may use “rules of thumb” (Ellison and Fudenberg 1995) (e.g., pay attention to recent reviews), or may look at historical ratings information. The detailed hypotheses are given below.

**H1**: *Volume has a negative impact on category rank.*

**H1a**: *Cumulative volume has a negative impact on category rank.*

**H1b**: *Daily volume has a negative impact on category rank.*

On a webpage for a product sold on Amazon.com, consumers can see a breakdown of all previous star ratings given to a product in a horizontal bar chart form (Appendix, Figure 1). Thus, for valence, the focus is on the cumulative rating, represented by the percentage of 1, 2, 3, 4, or 5 stars received.

**H2**: *Cumulative valence has a negative impact on category rank.*

Amazon.com’s integration with Facebook.com, and its potential impact on sales has not been previously explored by academics. When clicking on the Facebook.com Like icon, individuals can share the item via e-mail (e.g., by posting on their own Facebook page or a friends’ page, or share on Twitter.com). Previous literature on “herds and information cascades” and “learning from others” suggest that Facebook.com Likes
may influence behaviors, and in particular, by signaling popularity (Ellison and Fudenberg 1995). While it is beyond the scope of this study to show how many people actually post their product Like information to their social media websites (e.g., Facebook.com or Twitter), social network theory further suggests that Facebook.com Likes will increase homophily (Rogers 1983), thereby improving perceived tie strength (Granovetter 1973) and credibility (Kelman 1961), and negatively impact rank (increase sales). Amazon consumers can see the total number of Facebook.com Likes given at any point in time.

Thus,

**H3:** The cumulative number of Facebook.com Likes has a negative impact on category rank.

The information economics literature suggests the consumers will use online WOM to inform buying decisions, but that reliance on online WOM will differ based on product types (Nelson 1974; Klein 1998; Huang, Lurie, and Mitra 2009). The empirical literature finds evidence for the impact of online WOM on product sale in books (Chevalier and Mayzlin 2004), movies (Duan, Gu, and Whinston 2008; Liu 2006; Dellarocas, Zhang, and Awad 2007), TV shows (Godes and Mayzlin 2004), beauty products (Moe and Trusov 2011), and videogames (Zhu and Zhang 2010). Drawing on Nelson (1974), consumers should rely more on WOM information for product categories with more experiential qualities (Appstore for Android, Books, Music, Movies & TV, Videogames, Electronics, Grocery and Gourmet Food) than for with those with more search qualities (Health & Personal Care, Office Products, and Home & Garden). Differences in how online WOM impacts sales should be mediated by consumer ability
to assess product quality prior to purchase (Nelson 1970, 1974), information needs, and level of cognitive processing (Huang, Lurie, and Mitra 2009). I hypothesize the following:

**H4:** Online WOM (volume, valence, and Facebook.com Likes) has a greater impact on category rank for experience categories than for search categories.

### 3.2 Data

The data used for this thesis was collected from Amazon.com’s Hot New Releases section for ten product categories: Books, Music, Movies & TV, Videogames, Appstore for Android, Health and Personal Care, Office Products, Grocery and Gourmet Food, and Home and Garden from December 8, 2011 through February 29, 2012. Information on twenty products was collected for each category using a computer program, “Amazon Info Collector,” written in Java. Every day, at approximately 9:30AM, I ran “Amazon Info Collector” using the Command Prompt in Windows. Once I input the product URLs, the program extracted the product ID, Amazon price, rank, volume of reviews, Facebook Likes (named “Dataset 1,” Table 1). Data was collected daily in an effort to capture frequent changes in online WOM metrics, price, and rank.

At the end of the collection period, the product ID, reviewer ID, rating by reviewer, posted date, number of people who say a review is helpful, and total number of people who indicate whether a review is helpful or not, was extracted (named “Dataset 2,” in Table 1). In the final, merged data set, date and product ID allows matching of online WOM information to the appropriate product and reviewer (named “Merged Dataset” in Table 1).
Table 1 summarizes and defines all the variables extracted. On the left hand side of the divider, I present all the variables that “Amazon Info Collector” extracted daily for each product. On the right hand side of the divider, I present all the review-related variables that were extracted at the end of the collection period. I use day \( t \) to identify that a variable is extracted on a given day; this notation will reappear in the empirical model in Section 3.3.

**Table 1: Description of Extracted Variables**

<table>
<thead>
<tr>
<th>Variable</th>
<th>Dataset 1</th>
<th>Dataset 2</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Product ID</strong></td>
<td>A unique identification code given to Amazon products.</td>
<td>A unique identification code given to Amazon reviewers.</td>
</tr>
<tr>
<td><strong>Date accessed</strong></td>
<td>The time, calendar month, day, and year product information was collected from Amazon.</td>
<td>The calendar month, day, and year a reviewer posted a review.</td>
</tr>
<tr>
<td><strong>Price</strong></td>
<td>Amazon price for a product on day ( t )</td>
<td>Rating</td>
</tr>
<tr>
<td><strong>Rank</strong></td>
<td>Daily main-category rank for a product on day ( t ), where 1 is the highest selling item in a given category.</td>
<td>Helpful</td>
</tr>
<tr>
<td><strong>NumofReviews</strong></td>
<td>The number of reviews posted for a product on day ( t ).</td>
<td>Out_of</td>
</tr>
<tr>
<td><strong>NumofLikes</strong></td>
<td>The number of Facebook.com Likes posted for a product on day ( t ).</td>
<td>NumofComments</td>
</tr>
</tbody>
</table>

Attention was given to capturing the “essence” of the product category. For example, in the Health and Personal Care category, the Hot New Releases section lists the top 20 “new” products in that category. For this category, the top 20 products (at the time of data collection) were solely baby pampers. It is problematic to call a category “Health and Personal Care” for the purposes of this research when observations are only
from one product type (i.e., baby pampers). Thus, I use the primary sub-category filter to ensure that a product category is appropriately represented. For example, in the Movies & TV section, there are three main sub-categories Blu-Ray, DVD, and Amazon Instant Video. “Amazon Info Collector” could not read the website features in the Instant Video web pages, so this sub-category was excluded. Because twenty products are collected in total for each category, ten products were selected in ranked order from the Blu-Ray sub-category and from the DVD sub-category. If the price or sales rank was missing, the product was skipped, and the next product down the list was used instead. The same logic was applied to the remaining product categories.

Due to the nature of Amazon.com, the final data set underwent changes beyond my control. Amazon products temporarily “lose” their price and/or sales rank, or become unavailable. For products that become “unavailable,” Amazon.com often indicates that “we don’t know when or if this item will be back in stock.” Such products may “come back” within several days, or remain unavailable for an extended period of time. Thus, some items in Dataset 1 could be missing. In the case that a URL stopped working in the middle of the collection period, extraction of items in Dataset 2 was not possible. Lastly, a product category name changed from Home and Garden to Home and Kitchen on January 29, 2012; however there were no major disturbances in the data following this change.

Table 2 describes the datasets used and the associated number of observations. The “Merged Dataset” consists of Dataset 1 and Dataset 2 merged on product ID and date. In Dataset 1, I dropped observations where rank and price were missing, product URLs stopped working, products with no reviews, and products belonging to inconsistent
category memberships. A total of 27 products were dropped in Dataset 1 (“Cleaning Step,” in Table 2). In Dataset 2, I dropped a single product which had the same reviews as another, very similar product (“Cleaning Step,” in Table 2). In the Merged Dataset, I dropped missing variables in Dataset 1 again because data was missing for three days in the collection period. Thus, ratings, comments, or helpfulness given on such days, produce missing information for variables in Dataset 1, and by extension missing variables in the Merged Dataset.

Table 2: Datasets and Number of Observations

<table>
<thead>
<tr>
<th>Data</th>
<th>Raw</th>
<th>Cleaning Step</th>
<th>Final</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Dataset 1</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dataset 1</td>
<td>16200</td>
<td>16200</td>
<td>13406</td>
</tr>
<tr>
<td>Dataset 2</td>
<td>14643</td>
<td>13939</td>
<td>5208</td>
</tr>
<tr>
<td><strong>Merged data</strong></td>
<td></td>
<td>25529</td>
<td>13406</td>
</tr>
</tbody>
</table>

**Dataset 1**
- Price: 15393, 13406, 13406
- Rank: 15960, 13406, 13406
- Number of Likes: 16074, 13406, 13406
- Number of Reviews: 16074, 13406, 13406

**Dataset 2**
- Rating: 14644, 13939, 5208
- Helpful: 8603, 8356, 3072
- Out_Of: 14644, 8356, 3072
- Number of Comments: 14644, 13939, 5208

**Merged Dataset**
- Price: -
- Rank: -
- Number of Likes: -
- Number of Reviews: -
- Rating: -
- Helpful: -
- Out_Of: -
- Number of Comments: -

Note: Dataset1 and Dataset 2 were merged to create the Merged Dataset using their common ProductID and date posted.
3.3 Model

I specify a two stage least squares estimator (2SLS) where rank is the dependent variable. I present the reduced form model in Eq. 1 and address my hypotheses. Asterisks on Dvolume and Cvolume indicate that these variables are instrumented. For **Objective 1**, H1a says that the cumulative volume has a negative impact on category rank, so $\alpha_2 < 0$. H1b says that the daily volume has a negative impact on category rank, so $\alpha_3 < 0$. H2 says that the cumulative valence has a negative impact on category rank, so $\alpha_5 < 0$, $\alpha_6 < 0$, $\beta_7 < 0$, and $\alpha_8 < 0$. H3 says that the cumulative Facebook.com Likes have a negative impact on category rank, so $\alpha_4 < 0$. For **Objective 2**, H4 indicates that reliance on online WOM will be greater for experience categories than for search categories. I expect the same sign and greater significance in $\alpha_2$ through $\alpha_8$ for experience categories (Appstore for Android, Books, Music, Movies and TV, Videogames) than for search categories (Home & Garden, Office Products, Health & Personal Care).

\begin{align*}
(1) \quad \text{Rank}_{itj} &= \theta_t + \alpha_1 \text{Price}_{itj} + \alpha_2 \text{Cvolume*}_{itj} + \alpha_3 \text{Dvolume*}_{itj} + \\
&\quad \alpha_4 \text{clikes}_{itj} + \alpha_5 \text{CPstar2}_{itj} + \alpha_6 \text{CPstar3}_{itj} + \alpha_7 \text{CPstar4}_{itj} + \\
&\quad \alpha_8 \text{CPstar5}_{itj} + \alpha_9 \text{Weekend}_{itj} + \alpha_{10} \text{Month}_{itj} + \mu_i + \epsilon_{itj}
\end{align*}

Detailed descriptions for the independent variables used in the model are listed in Table 3. Based on assumptions and relevant statistical tests outlined in the following section, I specify Dhelpful, Dcomment, Cvariance, and lag of Cvolume as instruments for the endogenous variable, Cvolume. Similarly, I specify Dhelpful, Dcomment, Cvariance, and lag of Dvolume as instruments for the endogenous variable, Dvolume. Variable
descriptions for instruments Cvolume and Dvolume are listed in Table 4. Product fixed
effects, denoted by $\mu_i$, are introduced in the model to capture time invariant product
attributes and to control for product heterogeneity in a given category.

Table 3: Description of Variables in Model

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rank&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Rank for product i on day t in category j</td>
</tr>
<tr>
<td>Price&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Price for product i on day t in category j</td>
</tr>
<tr>
<td>CVolume&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Number of reviews posted for product i until day t in category j</td>
</tr>
<tr>
<td>DVolume&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Average number of reviews posted for product i on day t in category j</td>
</tr>
<tr>
<td>Likes&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Number of Facebook.com Likes received for product i until day t in category j</td>
</tr>
<tr>
<td>Cstar&lt;sub&gt;1itj&lt;/sub&gt;</td>
<td>Percentage of 1 star received for product i until day t in category j</td>
</tr>
<tr>
<td>Cstar&lt;sub&gt;2itj&lt;/sub&gt;</td>
<td>Percentage of 2 star received for product i until day t in category j</td>
</tr>
<tr>
<td>Cstar&lt;sub&gt;3itj&lt;/sub&gt;</td>
<td>Percentage of 3 star received for product i until day t in category j</td>
</tr>
<tr>
<td>Cstar&lt;sub&gt;4itj&lt;/sub&gt;</td>
<td>Percentage of 4 star received for product i until day t in category j</td>
</tr>
<tr>
<td>Cstar&lt;sub&gt;5itj&lt;/sub&gt;</td>
<td>Percentage of 5 star received for product i until day t in category j</td>
</tr>
<tr>
<td>Weekend</td>
<td>Dummy variable for weekend, where weekend = 1, otherwise = 0</td>
</tr>
<tr>
<td>Month</td>
<td>Dummy variable for month, where December = 1, January = 2, and February = 3</td>
</tr>
</tbody>
</table>

Table 4: Description of Variables Used as Instrumental Variables

<table>
<thead>
<tr>
<th>Variable</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.CVolume&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>One day lag of Cvolume</td>
</tr>
<tr>
<td>L.DVolume&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>One day lag of Dvolume</td>
</tr>
<tr>
<td>CVariance&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Variance of ratings for product i until day t in category j</td>
</tr>
<tr>
<td>DComment&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Average number of comments for product i reviews given on day t in category j</td>
</tr>
<tr>
<td>Dhelpful&lt;sub&gt;itj&lt;/sub&gt;</td>
<td>Average helpfulness of product i reviews given on day t in category j</td>
</tr>
</tbody>
</table>
3.4 Econometric Approach

I use a two-stage least squares (2SLS) approach with fixed effects because it is an unbiased and consistent instrumental variable (IV) estimator. It is also robust against multicollinearity and misspecification, and has small sample properties that surpass other estimators (SAS, Simultaenous Equations). In the following paragraphs, I explain the theory behind IV and 2SLS. I also give a brief explanation of the data structure and relevance of fixed effects. Finally, I explain the appropriate tests needed for IV estimation to be valid.

In a regression equation (2), when \( y_2 \) is endogenous, \( y_2 \) and \( u \) are correlated, it produces biased and inconsistent estimator of \( \beta_1 \). A new variable, called an instrumental variable can be introduced to fix this problem so long as it fulfills certain assumptions.

From Wooldridge (2002, p. 496):

\[
y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1
\]  

also called the structural equation, where \( y_1 \) is endogenous and correlated with \( u_1 \), and \( z_1 \) is exogenous and not correlated with \( u_1 \); \( y_2 \) is suspected to be correlated with \( u_1 \).

Given that \( E(u_1) = 0 \), \( Cov(z_1, u_1) = 0 \) and \( Cov(z_2, u_1) = 0 \) the method of moments approach can be used to solve for the IV estimators.

The assumption of partial correlation can be stated as expressing the endogenous explanatory variable as a linear function of the exogenous variables and error term.

\[
y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + v_2 ,
\]  

also called the reduced form equation, where, by definition, \( E(v_2) = 0 \), \( Cov(z_1, v_1) = 0 \) and \( Cov(z_2, v_2) = 0 \). The identification condition along with the previous assumptions
on the structural equation is that $\pi_2 \neq 0$; thus after partialling out $z_1$, $y_2$, and $z_2$ remain correlated (see Wooldridge 2002, p. 497 for the general case).

The 2SLS method for a single endogenous variable can be easily extended to multiple endogenous variables. Expanding on (2), suppose that there are now two exogenous variables, $z_2$ and $z_3$, which are excluded to fulfill the exclusion restriction (assumes $z_2$ and $z_3$ are uncorrelated with the error $u_1$). The best IV is the linear combination that is most highly correlated with $y_2$, given by the reduced form for $y_2$,

$$y_2 = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3 + v_2,$$  \hspace{1cm} (4)

where we build on previous assumptions, $E(v_2) = 0$, $Cov(z_1, v_1) = 0$, $Cov(z_2, v_2) = 0$ and $Cov(z_3, v_3) = 0$. Then, the best IV is the linear combination $z_j$, called $y_2^*$:

$$y_2^* = \pi_0 + \pi_1 z_1 + \pi_2 z_2 + \pi_3 z_3,$$  \hspace{1cm} (5)

For the IV to not be perfectly correlated with $z_1$, either $\pi_2$ or $\pi_3$ should not equal zero. So, (5) breaks into two parts – the first piece is $y_2^*$, which is the part of $y_2$ that is uncorrelated with the error term, and the second piece is $v_2$, which is endogenous (i.e., correlated with the error term). Since the population parameters are unknown, we can estimate the reduced form by OLS regression of $y_2$ on $z_1$ and $z_j$ to find the fitted values:

$$\hat{y}_2 = \hat{\pi}_0 + \hat{\pi}_1 z_1 + \hat{\pi}_2 z_2 + \hat{\pi}_3 z_3,$$  \hspace{1cm} (6)

After obtaining $\hat{y}_2$, it can be used as the IV for $y_2$. Thus, the 2SLS estimator is obtained in two stages. Stage 1: run the regression in (6) and obtain the fitted values $\hat{y}_2$. Stage 2: do an OLS regression of $y_1$ on $\hat{y}_2$ and $z_1$. The 2SLS estimates are different from OLS estimates as $\hat{y}_2$ is used in place of $\hat{y}_1$. Put differently, 2SLS gets rid of the $y_2$ correlation with the error $u_i$ prior to doing the OLS regression. Plugging in $y_2 = y_2^* + z_2$ into (4) we get
\[ y_1 = \beta_0 + \beta_1 y_2 + \beta_2 z_1 + u_1 + \beta_1 v_2 , \tag{7} \]

where the composite error \( u_1 + \beta_1 v_2 \) has a mean equal to zero and is uncorrelated with \( y_2^* \) and \( z_1 \) (Wooldridge 2002, p.499-500).

In the multiple endogenous explanatory case we consider the model:
\[ y_1 = \beta_0 + \beta_1 y_2 + \beta_2 y_3 + \beta_3 z_1 + \beta_4 z_2 + \beta_5 z_3 + u_1, \tag{8} \]

where \( E(u_i) = 0 \) and \( u_1 \) is uncorrelated with \( z_1, z_2, \) and \( z_3 \). The endogenous variables are \( y_2 \) and \( y_3 \), and each may be correlated with \( u_1 \). In order fulfill the identification requirement to estimate (8), it is necessary to have at least two exogenous variables that are not included in (8), but that are correlated with endogenous variables \( y_2 \) and \( y_3 \).

While this condition is necessary for identification (order condition), it is not necessarily sufficient. The order condition states that we need at least as many exogenous variables as there are included endogenous variables in the structural equation. The sufficient condition for identification is called rank condition (Wooldridge 2002, p.502- 503).

Estimation using fixed effects is generally used when \( a_i \) is a parameter to be estimated, rather than an outcome of a random variable (random effects). Since I am tracking the same products in ten categories over time, it suggests that fixed effects are an appropriate method. I use the dummy variable regression approach for each product in a given category; estimates obtained using the dummy variable regression gives the same estimates that would be obtained using time demeaned data (Wooldridge 2002, p. 466). One complication of my dataset is that it is unbalanced; however, the dummy variable regression makes the appropriate adjustment for the missing observations (Wooldridge 2002, p.468).
As mentioned previously, I consider what consumers actually see on Amazon.com – both the cumulative number of reviews (Cvolume) and the most recent review (Dvolume). I use Cvariance, Dhelpful, and Dcomment as instrumental variables for Cvolume and Dvolume, and add a lagged endogenous variable to Cvolume and Dvolume.

Thus,

\[(9) \quad Cvolume^* = \hat{\alpha}_1 L. Cvolume + \hat{\alpha}_2 Cvariance + \hat{\alpha}_3 Dhelpful + \hat{\alpha}_4 Dcomment + u_i \]

\[(10) \quad Dvolume^* = \hat{\alpha}_1 L. Dvolume + \hat{\alpha}_2 Cvariance + \hat{\alpha}_3 Dhelpful + \hat{\alpha}_4 Dcomment + u_i \]

Using the xtivreg2\(^{14}\) package in Stata Version 10.1, I tested two important criteria for IVs: overidentification test (Hansen J statistic), where the null hypothesis is that the instruments are valid, and weak identification test (Cragg- Donald Wald F- Statistic / Kleibergen-Paap rk Wald F statistic), where the null hypothesis is that the instruments are weak using the Stock-Yogo weak identification critical values (Stock and Yogo 2005).

Cragg – Donald Wald F-statistic is valid when errors are assumed to be independent identically distributed (iid),\(^{15}\) while Kleibergen-Paap rk Wald F statistic is valid when errors are not iid.

All categories pass the overidentification test (i.e., fail to reject the null hypothesis), except Home and Garden, which is likely due to relatively few product observations. Appstore for Android, Grocery & Gourmet Food, and Office Products had too few product observations to enter regression analysis. All categories entering the

\(^{14}\) Xtivreg2 is a Stata package written by Mark E. Schaffer that implements IV/GMM estimation of the fixed-effects and first-differences panel data models with possibly endogenous regressors (Stata Manual, xtivreg2).

\(^{15}\) Independent identically distributed (iid) implies that each observation is an independent draw from a fixed, ("stationary") probabilistic model (Clauset 2011).
model pass the weak identification test, except for Books and Electronics. When errors are iid, Books passes the weak identification test (i.e., reject the null hypothesis that instruments are weak), and Electronics weakly passes the weak identification test, but both still pass the overidentification test (i.e., fail to reject the null hypothesis). I report results with the dropped iid assumption in the Results section.

I check for data stationarity using the Fisher type unit root test for panel data, which is appropriate for unbalanced panel data. It combines the p-values from N independent unit root tests; the null hypothesis is that all series are non-stationary against the alternative that at least one series in the panel is stationary (Stata Manual, xtfisher). I reject the null hypothesis that all series are non-stationary at the 1% significance level.

As mentioned previously, some data in my datasets are missing simply because of how Amazon.com is structured (e.g., products “lose” price/rank during the collection period). Thus, six out of the ten enter the regression analysis due to insufficient observations in the following categories: Appstore for Android, Office Products, Grocery & Gourmet Food, and Home & Garden.
Chapter IV

Results

4.1 Descriptive Statistics

Descriptive statistics are shown in Table 5 for each online WOM variable by category. In the table, I use the general name for the online WOM metrics (volume, valence, and variance) as it is specified in the literature, and underneath each general name, I give the variables used in my model. Cvolume, is highest for the Electronics category (490.26) and lowest for Appstore for Android (61.61). This means that the Electronics category received the highest review volume, while Appstore for Android received the lowest review volume. Instead of using the average rating, I use cumulative percentage of stars received, which is more disaggregated, and thus may better capture valence. The percentage of 5-stars received is over 50% for all categories, with the exception of Appstore for Android (31.6), so overall ratings are positive. The largest average Cvariance is in Videogames (1.81), and the smallest in Grocery & Gourmet Food (0.53). Thus, variability in the rating is relatively low, which is consistent with the fact that overall ratings are positive.

Additional metrics used in this study are Facebook.com Likes, Glikes, and variables about reviews, Ghelpful and Gcomment Average Facebook.com Likes are highest for the Movies & TV category and lowest for Appstore for Android. Reviews are rated most helpful by Amazon voters in Office Products, and least helpful in Videogames. The number of comments on reviews is highest in Books and lowest in Office Products.
I provide a correlation matrix by product category for each variable (Appendix, Figure 2). One possible correlation may be between price and number of reviews, ratings, and/or Facebook.com Likes. For example, as the number of reviews increases, demand for a given product may increase, which would increase its price. The correlation coefficients between price and online WOM, however, are overall less than 0.5, which suggests that such a relationship is weak. Exceptions include: in Electronics, Clikes (0.739); in Grocery & Gourmet Food, Clikes (-0.579); in Movies & TV, Creview (0.6016) and Clikes (0.5492); in Videogames, Clikes (0.5223) (Appendix, Figure 2). While lagged rank (L.rank) may be correlated current rank, cumulative metrics are also correlated with L.rank, so L.rank is not included in the model specification. In other words, including L.rank would be redundant, given that cumulative metrics already capture past online WOM behavior.
Table 5: Descriptive Statistics for Volume, Valence, Variance and Additional Metrics

<table>
<thead>
<tr>
<th>Category Type</th>
<th>N</th>
<th>Volume</th>
<th>Valence</th>
<th>Variance</th>
<th>Additional Metrics in Thesis</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Cvolume</td>
<td>Dvolume</td>
<td>CPstar1</td>
<td>CPstar2</td>
</tr>
<tr>
<td><strong>Experience</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Appstore for Android</td>
<td>18</td>
<td>61.61</td>
<td>1.06</td>
<td>18.58</td>
<td>16.8</td>
</tr>
<tr>
<td></td>
<td></td>
<td>49.05</td>
<td>1.51</td>
<td>15.65</td>
<td>17.29</td>
</tr>
<tr>
<td>Books</td>
<td>789</td>
<td>290.29</td>
<td>3.15</td>
<td>12.55</td>
<td>9.64</td>
</tr>
<tr>
<td></td>
<td></td>
<td>333.31</td>
<td>3.9</td>
<td>12.68</td>
<td>8.29</td>
</tr>
<tr>
<td>Electronics</td>
<td>529</td>
<td>490.26</td>
<td>2.68</td>
<td>12.31</td>
<td>5.84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>833.46</td>
<td>3.43</td>
<td>9.66</td>
<td>5.81</td>
</tr>
<tr>
<td>Grocery &amp; Gourmet Food</td>
<td>44</td>
<td>102.48</td>
<td>0.7</td>
<td>0.98</td>
<td>3.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>91.42</td>
<td>0.79</td>
<td>3.09</td>
<td>15.51</td>
</tr>
<tr>
<td>Movies &amp; TV</td>
<td>732</td>
<td>419.19</td>
<td>2.52</td>
<td>13.74</td>
<td>6.58</td>
</tr>
<tr>
<td></td>
<td></td>
<td>488.03</td>
<td>2.43</td>
<td>12.81</td>
<td>5.62</td>
</tr>
<tr>
<td>Music</td>
<td>429</td>
<td>116.7</td>
<td>1.38</td>
<td>6.25</td>
<td>4.61</td>
</tr>
<tr>
<td></td>
<td></td>
<td>71.05</td>
<td>1.59</td>
<td>6.94</td>
<td>4.34</td>
</tr>
<tr>
<td>Videogames</td>
<td>497</td>
<td>372.21</td>
<td>3.46</td>
<td>14.12</td>
<td>6.66</td>
</tr>
<tr>
<td></td>
<td></td>
<td>243.48</td>
<td>4.8</td>
<td>12.71</td>
<td>6.41</td>
</tr>
<tr>
<td>Search</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Health &amp; Personal Care</td>
<td>189</td>
<td>144.22</td>
<td>2.88</td>
<td>12.6</td>
<td>6.51</td>
</tr>
<tr>
<td></td>
<td></td>
<td>146.57</td>
<td>4.06</td>
<td>16.77</td>
<td>8.33</td>
</tr>
<tr>
<td>Home &amp; Garden</td>
<td>108</td>
<td>49.25</td>
<td>1.1</td>
<td>5.29</td>
<td>5.98</td>
</tr>
<tr>
<td></td>
<td></td>
<td>47.3</td>
<td>1.77</td>
<td>14.99</td>
<td>9.36</td>
</tr>
<tr>
<td>Office Products</td>
<td>36</td>
<td>25.78</td>
<td>0.28</td>
<td>11.35</td>
<td>2.45</td>
</tr>
<tr>
<td></td>
<td></td>
<td>27.71</td>
<td>0.57</td>
<td>15.18</td>
<td>6.01</td>
</tr>
</tbody>
</table>

Note: The un-italicized numbers are averages and the italicized numbers are standard deviation
4.2 Regression Analysis

Looking at Table 6, significant F-statistics on the overall model for each category suggest that the model performs well. The results suggest that the impact of online WOM on category rank is significant in Books (Clikes), Electronics (Dvolume, Clikes, Cpstar2, and Cpstar4), and Music (Clikes). Thus, online WOM improves sales outcomes in those categories.

Table 6: Effect of WOM Volume on Category Rank

<table>
<thead>
<tr>
<th>Category</th>
<th>Experience</th>
<th>Search</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Books$^a$</td>
<td>Electronics$^a$</td>
</tr>
<tr>
<td>Cvolume</td>
<td>-0.71</td>
<td>-0.15</td>
</tr>
<tr>
<td></td>
<td>0.48</td>
<td>0.09</td>
</tr>
<tr>
<td>Dvolume</td>
<td>-66.21</td>
<td>-15.05*</td>
</tr>
<tr>
<td></td>
<td>46.6</td>
<td>6.18</td>
</tr>
<tr>
<td>Price</td>
<td>74.94</td>
<td>1.67</td>
</tr>
<tr>
<td></td>
<td>40.12</td>
<td>1.83</td>
</tr>
<tr>
<td>Clikes</td>
<td>-0.71**</td>
<td>-1.23*</td>
</tr>
<tr>
<td></td>
<td>0.25</td>
<td>0.54</td>
</tr>
<tr>
<td>Cpstar2</td>
<td>5.17</td>
<td>-11.06**</td>
</tr>
<tr>
<td></td>
<td>13.16</td>
<td>3.42</td>
</tr>
<tr>
<td>Cpstar3</td>
<td>6.1</td>
<td>-3.45</td>
</tr>
<tr>
<td></td>
<td>6.93</td>
<td>5.6</td>
</tr>
<tr>
<td>Cpstar4</td>
<td>-2.81</td>
<td>-4.86**</td>
</tr>
<tr>
<td></td>
<td>4.18</td>
<td>1.63</td>
</tr>
<tr>
<td>Cpstar5</td>
<td>-4.86</td>
<td>-2.39</td>
</tr>
<tr>
<td></td>
<td>6.23</td>
<td>1.16</td>
</tr>
<tr>
<td>Weekend</td>
<td>-103.96**</td>
<td>6.91</td>
</tr>
<tr>
<td></td>
<td>34.4</td>
<td>15.53</td>
</tr>
<tr>
<td>Month1</td>
<td>-523.03**</td>
<td>-13.05</td>
</tr>
<tr>
<td></td>
<td>168.12</td>
<td>46.69</td>
</tr>
<tr>
<td>Month2</td>
<td>-300.06*</td>
<td>33.36</td>
</tr>
<tr>
<td></td>
<td>112.13</td>
<td>47.2</td>
</tr>
<tr>
<td>N</td>
<td>762</td>
<td>527</td>
</tr>
<tr>
<td>F-stat</td>
<td>6.333318</td>
<td>18.04423</td>
</tr>
<tr>
<td>Prob &gt; F</td>
<td>0.0002468</td>
<td>5.71E-07</td>
</tr>
</tbody>
</table>
Notably, we cannot compare online WOM metrics within a category because they are on different scales. For example, Cpstars1 – 5 are on an interval scale, while other metrics are not. Moreover, we cannot compare the magnitude of online WOM metrics across categories because each category rank is based on an unknown product volume. For example, it is an unknown how many music CDs are sold in the Music category.

Overall, Cvolume, the cumulative number of reviews, and Dvolume, the daily number of reviews has the expected negative (however, insignificant) sign in all categories (with the exception of Health & Personal for Cvolume). This means that as the number of cumulative reviews (Cvolume), or the number of daily reviews (Dvolume) increases, ceteris paribus, the category rank falls, i.e., category sales increases.

In the Electronics category, the coefficient on Dvolume is negative and significant at 5% level, confirming H1a; thus, when a product receives a review, ceteris paribus, the rank declines by 15.05 units. We can interpret this in the following way: if a product is ranked 20 in Electronics, and its rank falls by 15 units, this product’s new rank is 5, which implies its sales have increased. While previous empirical studies find a significant impact of volume on sales in box office films (Liu 2006; Duan, Gu, and Whinston 2008) and in videogames (Zhu and Zhang 2010), I find that the volume of reviews has a significant impact in the Electronics category, but not in Movies & TV or Videogames.

Clikes, the cumulative number of Facebook.com Likes, is significant in impacting rank for Books (-0.71 at the 1% significance level), Electronics (-1.23 at the 5% significance level), and Music (-0.14 at the 5% significance level), ceteris paribus,
confirming H3. Previous empirical literature has not examined the influence of social media; however my results suggest that it may be important in influencing consumer choices.

In the Electronics category, the percentage of 2-stars received (Cpstar2), compared to 1-star received (Cpstar1), significantly impacts rank in the expected direction at the 1% significance level: as the percentage of 2-stars received increases by 1 unit, the rank falls by 11.06 points, ceteris paribus. Thus, a product ranked 100 receiving a 2-star versus a 1-star rating will now be ranked at approximately 89 – its sales have increased. Similarly, Cpstar4 negatively impacts rank by 4.86 units, ceteris paribus. The same product ranked 100 receiving a 4-star rating versus a 1-star rating will now be ranked at approximately 95 – its sales have increased, but by less than a product receiving a 2-star rating. Thus, I confirm H2, \( \alpha_5 < 0 \) and \( \alpha_7 < 0 \).

My findings are consistent with the previous literature, which suggests that as the valence increases, sales should increase (Chevalier and Mayzlin 2006; Moe and Trusov 2011; Clemons, Gao, and Hitt 2006; Zhu and Zhang 2010). That Cpstar2 has a larger impact than Cpstar4 on rank suggests that consumers may distrust above average ratings, since overall ratings on Amazon.com are biased upward. This result is consistent with Chevalier and Mayzlin (2006), who find that 1-star reviews impacts relative books sales more than 5-star reviews.

The ability to compare experience versus search categories (which were chosen based on dominant category attributes) (H4) is limited due to insufficient category observations for the search categories. The analysis only includes one search category: Health & Personal Care. I find that online WOM metrics in experience categories only
(Books, Electronics, and Music) impact category rank, while in the search category, Health & Personal Care, online WOM has no significant impact on rank. This is consistent with the literature that suggests that consumers will rely on online WOM more when product quality is unknown prior to purchase (i.e., experience goods).

Significant negative coefficients on Month 1 (December) and Month 2 (January) indicate that rank decreases considerably during the months of December and January compared to February, as expected. For example, in Movies & TV, during the month of December, compared to February, the rank falls by 245.64 units, ceteris paribus. Thus, if a movie or TV show is ranked 300 in the month of December, compared to February, its rank falls to approximately 54 (by 246 units), all else equal. Similarly, rank falls on the weekend compared to a weekday in the Books category by 103.96 units, ceteris paribus. Using the same example, if a book is ranked 300, its rank falls to approximately 154 on a weekend compared to a weekday. Thus, where Month and Weekend is significantly negative implies that category sales increases during the holidays (i.e., December) (compared to non-holiday months, January and February) and on weekends (compared to weekdays).
Chapter V

Conclusions

In this thesis, I set out to (1) understand which metric, or combination of metrics, is most important in impacting Amazon.com rank in different categories and (2) address how online WOM impacts rank in experience versus search categories. Following previous empirical and theoretical literature, I account for the endogeneity of online WOM volume, but use a different econometric approach.

I find support for both volume (number of reviews) and valence (ratings) in the Electronics category; however when a product is rated, receiving a 2-star rating (versus a 1-star rating), has a greater impact on rank than receiving a 4-star rating (versus a 1-star rating). While the previous literature has not looked at the Electronics category specifically, the importance of volume is confirmed in previous studies (e.g., in books, box office movies, and videogames). In book sales, researchers find that lower ratings impact sales more than higher ratings, which suggests that consumers may discount the credibility of highly rated products. Moreover, I find that review helpfulness, comments on reviews, and variance of ratings are, on average, effective instruments for the endogenous nature of WOM volume.

In the Books category, I do not find a significant impact of valence on rank. However, previous literature in books, in a cross-sectional analysis, find that the percentage of one, two, or three stars received decreases sales, and that the percentage of four or five stars increases sales. I also find that Clikes significantly improves book sales, which has not been explored in online WOM literature.
Unlike previous literature in box office movies, I do not find a significant impact of volume on rank in the Movies & TV category. There are several reasonable explanations for that: (a) box office movies and films available on DVD/BluRay are different, (b) there is heterogeneity in movies, and in movies versus TV shows, and/or (c) people may rely on other websites for movie reviews.

As for (a), there is more product uncertainty about newly released films than those on DVD/BluRay, since more people will have seen movies available for home purchase. As for (b), while my dataset contains new movies, some are re-released versions of classic films. This suggests that individuals may rely more on online WOM for new DVD/BluRay releases than for re-released films, since the essential quality of re-released films, prior to purchase, is known (e.g., even if a film is “digitally re-mastered,” the movie is the same). Thus, online WOM may be more important for a newer movie, such as The Help, versus a re-released classic, such as The Lion King (e.g., the newer Lion King may be more pleasing to the eye, but it is essentially the same as the original).

Moreover, there may be differences in online WOM reliance for TV shows versus movies. Lastly, (c) people may rely more on other websites for movie reviews and ratings, such as rottentomatoes.com, imdb.com, or Yahoo!Movies.com. These websites give consumers additional information, including critics’ reviews, which may influence movie-goers’ perceptions and subsequent behaviors.

Unlike the literature studying videogames, I do not find a significant impact of volume or valence on rank in the Videogame category. Similar to the point made in (b), there is considerable heterogeneity in videogames in my dataset. For example, some videogames are geared toward children (e.g., Kirby Mass Attack), while others are geared
toward adults (e.g. Call of Duty: Modern Warfare 3). Reliance on online WOM may be different for these two videogame audiences, and may vary based on game popularity (Zhu and Zhang 2010). The product heterogeneity issue may be relevant for the Music category as well. For example, some music CDs in my dataset are albums released by veteran artists (e.g., Tony Bennett) and others are newer to the scene (e.g., Florence and the Machine). Applying (c) to videogames suggests that videogame users will rely on more popular videogame rating websites, such as ign.com and GameSpot.com. These websites secured the number one and two spots, respectively, based on user reach and site traffic (eBizMBA Rank 2012).

The information economics (Klein 1998; Nelson 1974) and decision-making and information search (Huang, Lurie, and Mitra 2009) suggest that reliance on WOM is greater for experience goods than for search goods, as product quality prior to purchase is difficult to assess and requires greater cognitive effort. I find some support for this: online WOM significantly impacts rank in experience categories (Books, Electronics, and Music), but not in the search category, Health & Personal Care.

The literature suggests two signals to consumers from online WOM; when volume is significant, an awareness effect, and when valence is significant, a persuasive effect. I find evidence for both awareness (significance in Dvolume in Electronics) and persuasive effects (significance in Cpstar2 and Cpstar4 in Electronics). From a managerial standpoint, this is important. It suggests that online WOM builds product awareness and ultimately impacts consumer choices. Thus, businesses should encourage and track online WOM. Moreover, the findings call attention to the importance of adapting to the changing online WOM environment: the significance of Facebook.com
Likes in Books, Electronics, and Music suggests that this metric also influences consumer behavior. Businesses may thus find it worthwhile to use social media to their advantage, and in the case of Facebook.com, encourage consumers to Like products, and share their Likes with others.

There are several shortcomings to my study. One limitation of thesis is the relatively short data collection period (e.g., Chevalier and Mayzlin 2006 collect data from 1998-2002; Duan, Gu, and Whinston 2008 collect data for about a year). Thus, potentially long lasting impacts of online WOM may not be captured in my analysis. Missing data on search categories and on overall information on reviews (ratings, comments on reviews, and helpfulness of reviews) limited my ability to effectively compare experience versus search categories; however extending the data collection period may have mitigated this problem. Moreover, since products that received no reviews are excluded from analysis implies that selection bias may be a limitation; I could account for this by addressing the probability that a product receives a review.

As mentioned previously, product heterogeneity within categories may explain the lack of significant results. While I attempted to group products within categories (e.g., non-classic versus classic films in the Movies & TV category) and introduce interactions with the online WOM metric using these groups, I am not able to show that the results improved. Using data on product popularity may help categorize products within categories and account for the heterogeneity (Zhu and Zhang 2010).

Another limitation is the lack of advertising data on Amazon.com. Consumers may substitute advertising and online WOM for their product information needs. The degree of substitutability between advertising and online WOM will likely vary based on
product type and perceived credibility of both information sources. Since I cannot
directly observe advertising, a time trend may be used as a proxy to capture advertising
intensity by product and category.

If I had sales data, I would have been able to discuss the impact of online WOM
in terms of sales instead of rank. Data interpretation may have been clearer and more
powerful. Moreover, incorporating review length and actual text of reviews may have
better captured online WOM, as it is likely consumers do not just rely on summary
statistics, such as number of reviews or average rating.

Future research on the impact of online WOM on sales (or rank) in different
product categories should (1) collect data for an extended period of time to better capture
the long term impact of online WOM on sales outcomes, (2) control for product
heterogeneity within categories, (3) incorporate social media metrics (if any), (4) include
review length and review text as additional metrics. Lastly, (5) where advertising data is
unavailable, a time trend should be incorporated as a proxy for advertising intensity by
product and category. Incorporating these suggestions may show that online WOM is
consistently influential in impacting sales (or rank).
Appendix

Figure 1: Screenshot of Online WOM Metrics

![Screenshot of Online WOM Metrics](image-url)

Figure 2: Correlation Matrices of Variables by Category

<table>
<thead>
<tr>
<th>Appstore for Android</th>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clikes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Price</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rank</td>
<td>0.6309</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRank</td>
<td>0.4063</td>
<td>0.7549</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dreview</td>
<td>-0.1563</td>
<td>-0.2145</td>
<td>-0.195</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creview</td>
<td>0.1126</td>
<td>-0.026</td>
<td>-0.0762</td>
<td>0.1518</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clikes</td>
<td>0.1129</td>
<td>-0.2334</td>
<td>-0.2449</td>
<td>-0.0386</td>
<td>0.0994</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar1</td>
<td>0.1693</td>
<td>-0.1041</td>
<td>-0.127</td>
<td>0.0664</td>
<td>0.755</td>
<td>0.0366</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar2</td>
<td>0.0501</td>
<td>0.2518</td>
<td>0.096</td>
<td>0.1817</td>
<td>0.1407</td>
<td>-0.4505</td>
<td>0.0095</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar3</td>
<td>0.1895</td>
<td>0.1726</td>
<td>0.0886</td>
<td>-0.1177</td>
<td>0.1085</td>
<td>-0.2369</td>
<td>-0.0429</td>
<td>-0.1619</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar4</td>
<td>-0.2988</td>
<td>-0.005</td>
<td>-0.1142</td>
<td>0.0023</td>
<td>-0.3602</td>
<td>-0.0579</td>
<td>-0.4244</td>
<td>-0.0685</td>
<td>-0.2441</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cpstar5</td>
<td>-0.1604</td>
<td>-0.1881</td>
<td>0.0314</td>
<td>-0.0697</td>
<td>-0.578</td>
<td>0.4306</td>
<td>-0.5515</td>
<td>-0.428</td>
<td>-0.4972</td>
<td>0.0826</td>
<td>1</td>
</tr>
</tbody>
</table>
### Books

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clikes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.1217</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRank</td>
<td>-0.1307</td>
<td>0.9125</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dreview</td>
<td>0.0523</td>
<td>-0.2452</td>
<td>-0.2361</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creview</td>
<td>0.105</td>
<td>-0.246</td>
<td>-0.2426</td>
<td>0.4693</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clikes</td>
<td>-0.2801</td>
<td>-0.2538</td>
<td>-0.2498</td>
<td>0.26</td>
<td>0.788</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar1</td>
<td>0.2533</td>
<td>-0.0283</td>
<td>-0.0347</td>
<td>-0.0295</td>
<td>-0.0847</td>
<td>-0.1722</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar2</td>
<td>0.2129</td>
<td>-0.1358</td>
<td>-0.137</td>
<td>0.0322</td>
<td>0.0067</td>
<td>-0.0741</td>
<td>0.5771</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar3</td>
<td>0.2617</td>
<td>-0.1305</td>
<td>-0.1358</td>
<td>-0.0332</td>
<td>0.0838</td>
<td>0.0428</td>
<td>0.4169</td>
<td>0.7001</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cpstar4</td>
<td>0.2363</td>
<td>-0.2535</td>
<td>-0.2447</td>
<td>-0.0847</td>
<td>-0.0299</td>
<td>-0.0171</td>
<td>-0.1924</td>
<td>-0.0198</td>
<td>-0.0057</td>
<td>1</td>
</tr>
<tr>
<td>Cpstar5</td>
<td>-0.37</td>
<td>0.198</td>
<td>0.1997</td>
<td>0.0459</td>
<td>0.0282</td>
<td>0.1079</td>
<td>-0.7552</td>
<td>-0.838</td>
<td>-0.7419</td>
<td>-0.2819</td>
</tr>
</tbody>
</table>

### Electronics

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clikes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0543</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRank</td>
<td>-0.0466</td>
<td>0.8638</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dreview</td>
<td>0.4946</td>
<td>-0.1846</td>
<td>-0.1567</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creview</td>
<td>0.3639</td>
<td>-0.1085</td>
<td>-0.0958</td>
<td>0.5856</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clikes</td>
<td>0.7349</td>
<td>-0.2495</td>
<td>-0.2159</td>
<td>0.2311</td>
<td>-0.0105</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar1</td>
<td>-0.1347</td>
<td>0.0309</td>
<td>0.0141</td>
<td>-0.1797</td>
<td>-0.2928</td>
<td>-0.1847</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar2</td>
<td>0.0442</td>
<td>0.0081</td>
<td>-0.0017</td>
<td>-0.1143</td>
<td>-0.2702</td>
<td>0.0325</td>
<td>0.0963</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar3</td>
<td>0.0353</td>
<td>0.1294</td>
<td>0.0721</td>
<td>-0.193</td>
<td>-0.3585</td>
<td>-0.0116</td>
<td>0.1799</td>
<td>0.1466</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cpstar4</td>
<td>0.1604</td>
<td>0.2251</td>
<td>0.3038</td>
<td>-0.1615</td>
<td>-0.4322</td>
<td>0.2372</td>
<td>0.0138</td>
<td>0.0569</td>
<td>0.0623</td>
<td>1</td>
</tr>
<tr>
<td>Cpstar5</td>
<td>-0.0481</td>
<td>-0.1904</td>
<td>-0.2068</td>
<td>0.2865</td>
<td>0.607</td>
<td>-0.0484</td>
<td>-0.6237</td>
<td>-0.4603</td>
<td>-0.4863</td>
<td>-0.6304</td>
</tr>
</tbody>
</table>

### Grocery & Gourmet Food

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clikes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.0314</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRank</td>
<td>0.0162</td>
<td>0.9082</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dreview</td>
<td>-0.1463</td>
<td>-0.1927</td>
<td>-0.1627</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creview</td>
<td>-0.0282</td>
<td>-0.3982</td>
<td>-0.4277</td>
<td>0.1005</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clikes</td>
<td>-0.5789</td>
<td>-0.3861</td>
<td>-0.4075</td>
<td>0.2305</td>
<td>0.6351</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar1</td>
<td>0.3655</td>
<td>0.0003</td>
<td>0.012</td>
<td>-0.0148</td>
<td>-0.2733</td>
<td>-0.1357</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar2</td>
<td>0.096</td>
<td>0.072</td>
<td>0.0729</td>
<td>-0.1423</td>
<td>-0.3227</td>
<td>-0.288</td>
<td>0.1019</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar3</td>
<td>0.0285</td>
<td>0.1203</td>
<td>0.2183</td>
<td>-0.0232</td>
<td>-0.2997</td>
<td>-0.2148</td>
<td>0.043</td>
<td>-0.0817</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cpstar4</td>
<td>0.06</td>
<td>0.1624</td>
<td>0.108</td>
<td>0.0241</td>
<td>-0.499</td>
<td>-0.3089</td>
<td>0.0626</td>
<td>0.0248</td>
<td>0.1182</td>
<td>1</td>
</tr>
<tr>
<td>Cpstar5</td>
<td>-0.1928</td>
<td>-0.1876</td>
<td>-0.2098</td>
<td>0.0806</td>
<td>0.6672</td>
<td>0.4677</td>
<td>-0.3664</td>
<td>-0.5511</td>
<td>-0.5235</td>
<td>-0.6324</td>
</tr>
</tbody>
</table>

### Health & Personal Care

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clikes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>-0.0438</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LRank</td>
<td>-0.0645</td>
<td>0.9146</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dreview</td>
<td>0.272</td>
<td>-0.0689</td>
<td>-0.0971</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Creview</td>
<td>0.4423</td>
<td>-0.2239</td>
<td>-0.3165</td>
<td>0.2789</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Clikes</td>
<td>0.3767</td>
<td>-0.1616</td>
<td>-0.2257</td>
<td>0.3542</td>
<td>0.6895</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar1</td>
<td>-0.0891</td>
<td>0.2625</td>
<td>0.291</td>
<td>-0.2114</td>
<td>0.0438</td>
<td>-0.385</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar2</td>
<td>-0.0407</td>
<td>-0.0186</td>
<td>-0.015</td>
<td>-0.1392</td>
<td>-0.1251</td>
<td>-0.3423</td>
<td>0.4026</td>
<td>1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cpstar3</td>
<td>0.1216</td>
<td>0.002</td>
<td>0.0322</td>
<td>-0.0972</td>
<td>-0.3951</td>
<td>-0.3186</td>
<td>0.0238</td>
<td>0.1452</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>Cpstar4</td>
<td>0.2972</td>
<td>0.1856</td>
<td>0.1529</td>
<td>0.0814</td>
<td>0.0512</td>
<td>0.1002</td>
<td>-0.2763</td>
<td>-0.327</td>
<td>-0.1227</td>
<td>1</td>
</tr>
</tbody>
</table>
### Home & Garden

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clickes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1827</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Movies & TV

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clickes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.1418</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Music

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clickes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0257</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

### Office Products

<table>
<thead>
<tr>
<th>Price</th>
<th>Rank</th>
<th>LRank</th>
<th>Dreview</th>
<th>Creview</th>
<th>Clickes</th>
<th>Cpstar1</th>
<th>Cpstar2</th>
<th>Cpstar3</th>
<th>Cpstar4</th>
<th>Cpstar5</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.5306</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 3: Detailed Results on Effect of WOM Volume on Category Rank Without iid Assumption

Validity of instruments are based on overidentification (Hansen J Statistic) and weak identification (Kleigbergen-Paap rk Wald F statistic/Cragg-Donald Wald F statistic, which are compared to Stock–Yogo critical values. Take for example, the Music category. I look at the Chi-square P-value = 0.2341 given for the Hansen J Statistic. The null hypothesis is that instruments are valid; I fail to reject the null hypothesis at the 5% significance level. For each category, I compare the Kleigbergen-Paap rk Wald F statistic to the Stock-Yogo weak identification critical values. The critical values vary based on how much relative bias/maximal IV size the researcher is willing tolerate (See Stock and Yogo 2005 for further explanation). As an example, take the Movies category. The null hypothesis for the weak identification test is that the IVs are weak. The Kleigbergen-Paap rk Wald F statistic = 1.532, which is smaller than the Stock-Yogo critical values at any level of relative bias or maximal size. Thus, I fail to reject the null hypothesis. The same procedure is done for subsequent categories.

Books

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =        20                    Obs per group: min =        12
avg =      38.1
max =        77

IV (2SLS) estimation
-----------------------
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on myid

Number of clusters (myid) =         20                Number of obs =      762

|               Robust
| Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
### Electronics

**Fixed Effects Estimation**

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of groups</td>
<td>17</td>
</tr>
<tr>
<td>Observations per group</td>
<td>min = 3</td>
</tr>
<tr>
<td></td>
<td>avg = 31.0</td>
</tr>
<tr>
<td></td>
<td>max = 70</td>
</tr>
</tbody>
</table>

**IV (2SLS) estimation**

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on myid

| Number of clusters (myid) | 17 |
| Number of obs            | 527 |
| \( F(11, 16) \)          | 18.04 |
| Prob > F                 | 0.0000 |
| Total (centered) SS      | 8372893.288 |
| Centered R²              | 0.0890 |
### FIXED EFFECTS ESTIMATION

| Robust rank | Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|-------------|-------|-----------|---|-----|----------------------|
| Cvolume     | -1.1474504 | 0.0904694 | -1.63 | 0.123 | -0.3392369 - 0.0443362 |
| Dvolume     | -15.0497 | 6.182782 | -2.43 | 0.027 | -28.15661 - 1.94279 |
| price       | 1.665737 | 1.833026 | 0.91 | 0.377 | -2.220104 - 5.551577 |
| Clikes      | -1.231116 | 0.5361148 | -2.30 | 0.035 | -2.367628 - 0.946033 |
| cpstar2     | -11.06319 | 3.424208 | -3.23 | 0.005 | -18.32218 - 3.804188 |
| cpstar3     | -3.449697 | 5.603419 | -0.62 | 0.547 | -15.32842 - 8.429021 |
| cpstar4     | -4.862038 | 1.6324 | -2.98 | 0.009 | -8.322571 - 1.401505 |
| cpstar5     | -2.390595 | 1.159761 | -2.06 | 0.056 | -4.849178 - 0.679874 |
| weekend     | 6.905326 | 15.53029 | 0.44 | 0.663 | -26.01742 - 39.82807 |
| month1      | -13.04536 | 46.69266 | -0.28 | 0.784 | -112.0294 - 85.93865 |
| month2      | 33.35565 | 47.19786 | 0.71 | 0.490 | -66.69933 - 133.4106 |

Underidentification test (Kleibergen-Paap rk LM statistic): 4.704
Chi-sq(4) P-val = 0.3191

Weak identification test (Cragg-Donald Wald F statistic): 19.384
(Kleibergen-Paap rk Wald F statistic): 20.468

Stock-Yogo weak ID test critical values:
- 5% maximal IV relative bias: 13.97
- 10% maximal IV relative bias: 8.78
- 20% maximal IV relative bias: 5.91
- 30% maximal IV relative bias: 4.79
- 10% maximal IV size: 19.45
- 15% maximal IV size: 11.22
- 20% maximal IV size: 8.38
- 25% maximal IV size: 6.89

NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.

Hansen J statistic (overidentification test of all instruments): 2.131
Chi-sq(3) P-val = 0.5457

Instrumented:
- Cvolume
- Dvolume

Included instruments:
- price
- Clikes
- cpstar2
- cpstar3
- cpstar4
- cpstar5
- weekend
- month1
- month2

Excluded instruments:
- cvariance
- L.Dvolume
- L.Cvolume
- Dhelpful
- Dcomment

Test of overidentifying restrictions:
Cross-section time-series model: xtivreg2 fe robust cluster(myid)
Sargan-Hansen statistic 2.131 Chi-sq(3) P-value = 0.5457

**Health & Personal Care**

**FIXED EFFECTS ESTIMATION**

Number of groups = 15
Obs per group: min = 2
avg = 12.4
max = 57

IV (2SLS) estimation

Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on myid
**Number of clusters (myid)** = 15  
**Number of obs** = 186  
**F(11, 14) =** 3.29  
**Prob > F =** 0.0196  
**Total (centered) SS** = 231285418.7  
**Centered R2 =** -0.6731  
**Total (uncentered) SS** = 231285418.7  
**Uncentered R2 =** -0.6731  
**Residual SS** = 386975104  
**Root MSE** = 1555

| rank | Coef. | Robust Std. Err. | t | P>|t| | [95% Conf. Interval] |
|------|-------|------------------|---|-----|------------------|
| Cvolume | 7.907067 | 8.244986 | 0.96 | 0.354 | -9.776669 - 25.5908 |
| Dvolume | -360.7234 | 476.2428 | -0.76 | 0.461 | -1382.163 - 660.7158 |
| price | 11.86411 | 21.17636 | 0.56 | 0.584 | -33.55467 - 57.28288 |
| Clikes | -14.6258 | 10.45693 | -1.40 | 0.184 | -37.05369 - 7.802078 |
| cpstar2 | 47.91687 | 44.90698 | 1.07 | 0.304 | -48.39902 - 144.2328 |
| cpstar3 | 62.97964 | 66.24928 | 0.95 | 0.358 | -79.11093 - 205.0702 |
| cpstar4 | 125.6426 | 106.6738 | 1.18 | 0.259 | -103.1499 - 354.4351 |
| cpstar5 | 68.4383 | 65.91128 | 1.04 | 0.317 | -72.92734 - 209.8039 |
| weekend | -271.5065 | 266.2024 | -1.02 | 0.325 | -842.4537 - 299.4408 |
| month1 | -1177.336 | 871.8982 | -1.35 | 0.198 | -3047.372 - 692.6998 |
| month2 | -425.6843 | 351.2762 | -1.21 | 0.246 | -1179.097 - 327.7282 |

Underidentification test (Kleibergen-Paap rk LM statistic): 7.000  
Chi-sq(4) P-val = 0.1359

Weak identification test (Cragg-Donald Wald F statistic): 0.413  
(Kleibergen-Paap rk Wald F statistic): 6.460

Stock-Yogo weak ID test critical values:  
5% maximal IV relative bias 13.97  
10% maximal IV relative bias 8.78  
20% maximal IV relative bias 5.91  
30% maximal IV relative bias 4.79  
10% maximal IV size 19.45  
15% maximal IV size 11.22  
20% maximal IV size 8.38  
25% maximal IV size 6.89

NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.

Hansen J statistic (overidentification test of all instruments): 2.055  
Chi-sq(3) P-val = 0.5611

Instrumented: Cvolume Dvolume  
Included instruments: price Clikes cpstar2 cpstar3 cpstar4 cpstar5 weekend month1 month2  
Excluded instruments: cvariance L.Dvolume L.Cvolume Dhelpful Dcomment

Test of overidentifying restrictions:  
Cross-section time-series model: xtivreg2 fe robust cluster(myid)  
Sargan-Hansen statistic 2.055 Chi-sq(3) P-value = 0.5611

**Movies & TV**

**FIXED EFFECTS ESTIMATION**

| Number of groups | 20 | Obs per group: min | 11 | avg | 35.3 | max | 57 |

IV (2SLS) estimation
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on myid

Number of clusters (myid) = 20                Number of obs = 705
F( 11,    19) = 15.69                Prob > F      = 0.0000

Total (centered) SS     = 11111285.5                Centered R2   = 0.4261
Total (uncentered) SS   = 11111285.5                Uncentered R2 = 0.4261
Residual SS             = 6376377.816                Root MSE      = 97.27

|               Robust |
| Coef.   Std. Err.   t    P>|t|   [95% Conf. Interval] |

Cvolume | -.2845323   .1982594 -1.44   0.168   -.6994941    .1304295 |
Dvolume | -19.05043   9.486312 -2.01   0.059   -38.90551    .804697 |
price   | 5.105064    2.640391  1.93   0.068   -2.4213384   10.63147 |
Clikes  | -.0119188   .0195558 -0.61   0.549   -.0528495    .0290119 |
cpstar2 | 2.038877    3.950697  0.57   0.577   -5.476539    9.554293 |
cpstar3 | .4652688    1.469328  0.32   0.755   -2.610071    3.540608 |
cpstar4 | 2.914251    2.58168  1.13   0.273   -2.489267    8.317769 |
cpstar5 | 1.551619    1.458707  1.06   0.301   -1.50149    4.604729 |
weekend | -4.698802   5.857737 -0.80   0.436   -16.91919    7.601584 |
month1  | -245.6394   46.42248 -5.29   0.000   -342.8028   -148.4761 |
month2  | -144.0374   36.09562 -3.99   0.001   -219.5864   -68.48843 |

Underidentification test (Kleibergen-Paap rk LM statistic): 7.046
Chi-sq(4) P-val = 0.1335

Weak identification test (Cragg-Donald Wald F statistic): 9.191
(Kleibergen-Paap rk Wald F statistic): 5.906

Stock-Yogo weak ID test critical values: 5% maximal IV relative bias 13.97
10% maximal IV relative bias 8.78
20% maximal IV relative bias 5.91
30% maximal IV relative bias 4.79
10% maximal IV size 19.45
15% maximal IV size 11.22
20% maximal IV size 8.38
25% maximal IV size 6.89

NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.

Hansen J statistic (overidentification test of all instruments): 7.661
Chi-sq(3) P-val = 0.0536

Instrumented: Cvolume Dvolume
Included instruments: price Clikes cpstar2 cpstar3 cpstar4 cpstar5 weekend month1 month2
Excluded instruments: cvariance L.Dvolume L.Cvolume Dhelpful Dcomment

Test of overidentifying restrictions:
Cross-section time-series model: xtivreg2 fe robust cluster(myid)
Sargan-Hansen statistic 7.661 Chi-sq(3) P-value = 0.0536

Music

FIXED EFFECTS ESTIMATION

Number of groups = 19                Obs per group: min = 5
Estimates efficient for homoskedasticity only
Statistics robust to heteroskedasticity and clustering on myid

| rank | Robust Coef. | Std. Err. | t | P>|t| | [95% Conf. Interval] |
|------|--------------|------------|---|--------|------------------------|
| Cvolume | -0.4239598 | .7456931 | -0.57 | 0.577 | -1.990603 | 1.142683 |
| Dvolume | -90.49156 | 48.62182 | -1.86 | 0.079 | -192.6422 | 11.65909 |
| price | 20.32939 | 15.35976 | 1.32 | 0.202 | -11.94027 | 52.59904 |
| Clikes | -0.135906 | .053639 | -2.53 | 0.021 | -2.482819 | -0.228892 |
| cpstar2 | 3.741101 | 5.736451 | 0.65 | 0.523 | -8.310735 | 15.79294 |
| cpstar3 | 5.632544 | 4.822073 | 1.17 | 0.258 | -4.498255 | 15.76334 |
| cpstar4 | 5.01105 | 6.244847 | 0.80 | 0.433 | -8.108888 | 18.13099 |
| cpstar5 | 5.174797 | 5.365496 | 0.96 | 0.348 | -6.097691 | 16.44729 |
| weekend | 2.125476 | 13.2362 | 0.16 | 0.874 | -25.68274 | 29.9337 |
| month1 | -228.492 | 67.50791 | -3.38 | 0.003 | -370.3209 | -86.66319 |
| month2 | -107.9827 | 32.51606 | -3.32 | 0.004 | -176.2964 | -39.66896 |

Underidentification test (Kleibergen-Paap rk LM statistic): 6.314
Chi-sq(4) P-val = 0.1769

Weak identification test (Cragg-Donald Wald F statistic): 1.532
(Kleibergen-Paap rk Wald F statistic): 3.858

Stock-Yogo weak ID test critical values:
- 5% maximal IV relative bias: 13.97
- 10% maximal IV relative bias: 8.78
- 20% maximal IV relative bias: 5.91
- 30% maximal IV relative bias: 4.79
- 10% maximal IV size: 19.45
- 15% maximal IV size: 11.22
- 20% maximal IV size: 8.38
- 25% maximal IV size: 6.89

NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.

Hansen J statistic (overidentification test of all instruments): 4.266
Chi-sq(3) P-val = 0.2341

Test of overidentifying restrictions:
Cross-section time-series model: xtivreg2 fe robust cluster(myid)
Sargan-Hansen statistic 4.266 Chi-sq(3) P-value = 0.2341
Videogames

FIXED EFFECTS ESTIMATION

Number of groups = 15  
Obs per group: min = 2  
avg = 31.9  
max = 65

IV (2SLS) estimation

Estimates efficient for homoskedasticity only  
Statistics robust to heteroskedasticity and clustering on myid

Number of clusters (myid) = 15  
Number of obs = 479  
F(11, 14) = 5.72  
Prob > F = 0.0016

Total (centered) SS = 16549253.74  
Centered R2 = 0.1566

Total (uncentered) SS = 16549253.74  
Uncentered R2 = 0.1566

Residual SS = 13957629.79  
Root MSE = 175.5

| rank | Coef. | Robust Std. Err. | t | P>|t| | [95% Conf. Interval] |
|------|-------|------------------|---|-----|------------------|
| Cvolume | -0.7723743 | .3939001 | -1.96 | 0.070 | -1.617206 | .0724574 |
| Dvolume | -0.906108 | 2.085501 | -0.43 | 0.671 | -5.379064 | 3.566848 |
| price | 16.53383 | 9.07275 | 1.82 | 0.090 | -2.925286 | 35.99294 |
| Clikes | 3.4171 | 1.605689 | 2.13 | 0.052 | -0.026761 | 6.860961 |
| cpstar2 | 6.044225 | 10.06216 | 0.60 | 0.558 | -15.53696 | 27.62541 |
| cpstar3 | -0.6140971 | 7.408045 | -0.08 | 0.935 | -16.50277 | 15.27458 |
| cpstar4 | 3.549478 | 7.450251 | 0.48 | 0.641 | -12.42972 | 19.52868 |
| cpstar5 | 10.67011 | 11.79909 | 0.90 | 0.381 | -14.63643 | 35.97665 |
| weekend | -5.619499 | 13.39979 | -0.42 | 0.681 | -34.35919 | 23.12019 |
| month1 | 2.360816 | 74.02972 | 0.03 | 0.975 | -156.4171 | 161.1388 |
| month2 | -49.4862 | 42.63719 | -1.16 | 0.265 | -140.9339 | 41.96149 |

Underidentification test (Kleibergen-Paap rk LM statistic): 3.034  
Chi-sq(4) P-val = 0.5521

Weak identification test (Cragg-Donald Wald F statistic): 26.731  
(Kleibergen-Paap rk Wald F statistic): 1.312

Stock-Yogo weak ID test critical values:  
5% maximal IV relative bias 13.97  
10% maximal IV relative bias 8.78  
20% maximal IV relative bias 5.91  
30% maximal IV relative bias 4.79  
10% maximal IV size 19.45  
15% maximal IV size 11.22  
20% maximal IV size 8.38  
25% maximal IV size 6.89

NB: Critical values are for Cragg-Donald F statistic and i.i.d. errors.

Hansen J statistic (overidentification test of all instruments): 3.730  
Chi-sq(3) P-val = 0.2921

Instrumented: Cvolume Dvolume  
Included instruments: price Clikes cpstar2 cpstar3 cpstar4 cpstar5 weekend month1 month2

Excluded instruments: cvariance LDvolume LCvolume Dhelpful Dcomment
Test of overidentifying restrictions:
Cross-section time-series model: xtivreg2 fe robust cluster(myid)
Sargan-Hansen statistic  3.730  Chi-sq(3)  P-value = 0.2921

Figure 4: Detailed Results on Effect of WOM Volume on Category Rank With iid Assumption for Books and Electronics

Books

FIXED EFFECTS ESTIMATION
------------------------
Number of groups =  20  Obs per group: min =  12
                     avg =  38.1
                     max =  77

IV (2SLS) estimation
----------------------
Estimates efficient for homoskedasticity only
Statistics consistent for homoskedasticity only

| rank | Coef.  | Std. Err. | t     | P>|t|  | [95% Conf. Interval] |
|------|--------|-----------|-------|------|----------------------|
| Cvolume | -0.714379 | 0.5596205 | -1.28 | 0.202 | -1.813034 -0.384276 |
| Dvolume | -66.21163 | 49.97891 | -1.32 | 0.186 | -164.331 31.9077 |
| price   | 74.94077  | 25.73902 | 2.91  | 0.004 | 24.40955 125.472 |
| Clikes  | -0.706765 | 0.2131979 | -3.32 | 0.001 | -1.25322 -0.2882151 |
| cpstar2 | 5.173447  | 8.007169 | 0.65  | 0.518 | -10.54634 20.89324 |
| cpstar3 | 6.1039    | 6.721831 | 0.91  | 0.364 | -7.092496 19.3003 |
| cpstar4 | -2.808961 | 4.43873  | -0.63 | 0.527 | -11.52314 5.905218 |
| cpstar5 | 5.173447  | 8.007169 | 0.65  | 0.518 | -10.54634 20.89324 |
| weekend | -103.9592 | 39.82037 | -2.61 | 0.009 | -182.1351 -25.78325 |
| month1  | -523.0332 | 76.08179 | -6.88 | 0.000 | -672.3787 -373.6878 |
| month2  | -300.0613 | 60.84528 | -4.93 | 0.000 | -419.5136 -180.609 |

Underidentification test (Anderson canon. corr. LM statistic):  10.976
Chi-sq(4) P-val =  0.0268

Weak identification test (Cragg-Donald Wald F statistic):  2.186
Stock-Yogo weak ID test critical values:  5% maximal IV relative bias 13.97
10% maximal IV relative bias  8.78
20% maximal IV relative bias  5.91
30% maximal IV relative bias  4.79
10% maximal IV size  19.45
15% maximal IV size  11.22
20% maximal IV size  8.38
25% maximal IV size  6.89
Sargan statistic (overidentification test of all instruments): 2.798  
Chi-sq(3) P-val = 0.4238

Instrumented: Cvolume Dvolume  
Included instruments: price Clikes cpstar2 cpstar3 cpstar4 cpstar5 weekend month1 month2  
Excluded instruments: cvariance L.Dvolume L.Cvolume Dhelpful Dcomment

### Electronics

**FIXED EFFECTS ESTIMATION**

| Number of groups = 17 | Obs per group: min = 3  
avg = 31.0  
max = 70 |

**IV (2SLS) estimation**

Estimates efficient for homoskedasticity only  
Statistics consistent for homoskedasticity only  

| rank | Coef. Std. Err. | t | P>|t| | [95% Conf. Interval] |
|------|-----------------|---|------|--------------------------|
| Cvolume | -.1474504   | .1137658 | -1.30   | 0.196   | -.3709694   | .0760687 |
| Dvolume | -15.0497   | 6.093284 | -2.47   | 0.014   | -27.02136   | -3.078047 |
| price | 1.665737   | .6579491 | 2.53   | 0.012   | .3730447   | 2.958428 |
| Clikes | -1.231116   | .5662795 | -2.17   | 0.030   | -2.343702   | -.1185298 |
| cpstar2 | -11.06319   | 2.197847 | -5.03   | 0.000   | -15.38136   | -6.74501 |
| cpstar3 | -3.449697   | 2.157542 | -1.60   | 0.110   | -7.688684   | .7892889 |
| cpstar4 | -4.862038   | 1.568925 | -3.10   | 0.002   | -7.944551   | -1.779525 |
| cpstar5 | -2.390595   | 1.345876 | -1.78   | 0.076   | -5.034877   | .2536861 |
| weekend | 6.905326   | 12.29559 | 0.56   | 0.575   | -17.25219   | 31.06284 |
| month1 | -13.04536   | 24.25284 | -0.54   | 0.591   | -60.69563   | 34.60491 |
| month2 | 33.35565   | 16.497 | 2.02   | 0.044   | .943514   | 65.76779 |

**Underidentification test (Anderson canon. corr. LM statistic):** 83.366  
Chi-sq(4) P-val = 0.0000

**Weak identification test (Cragg-Donald Wald F statistic):** 19.384

**Stock-Yogo weak ID test critical values:**
- 5% maximal IV relative bias 13.97
- 10% maximal IV relative bias 8.78
- 20% maximal IV relative bias 5.91
- 30% maximal IV relative bias 4.79
- 10% maximal IV size 19.45
- 15% maximal IV size 11.22
- 20% maximal IV size 8.38
- 25% maximal IV size 6.89

Instrumented: Cvolume Dvolume
Included instruments: price Clikes cpstar2 cpstar3 cpstar4 cpstar5 weekend month1 month2
Excluded instruments: cvariance L.Dvolume L.Cvolume Dhelpful Dcomment
References


Cone Trend Tracker. 2011. *Game Changer: Cone Survey Finds 4-out-of-5 Consumers Reverse Purchase Decisions Based on Negative Online Reviews.* Boston, MA.


—., xtivreg2. “Help for xtivreg2.”


