Tapping into the Pulse of the Market: Essays on Marketing Implications of Information Flows

by

Xiaoquan (Michael) Zhang

B.E. Computer Science, Tsinghua University, 1996
B.A. English Literature, Tsinghua University, 1996
M.S. Management, Tsinghua University, 1999

Submitted to the Alfred P. Sloan School of Management
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Management

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 2006

Signature of Author . . . .
Alfred P. Sloan School of Management
May 1, 2006

Certified by . . . .
Erik Brynjolfsson
George and Sandi Schussel Professor of Management
Thesis Supervisor

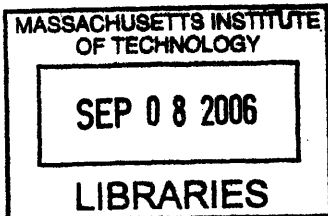Accepted by . . . .
Birger Wernerfelt
Professor of Management Science
Chair, Doctoral Program

Tapping into the Pulse of the Market: Essays on Marketing Implications of Information Flows

by

Xiaoquan (Michael) Zhang

Abstract

As the Internet continues to penetrate consumer households, online marketing is getting increasingly important for firms. By adapting to online strategies, firms are blessed (or doomed) with a plethora of new business models. The information flows created in the process poses both opportunities and challenges for marketers. On one hand, information flows captured online are usually easier to be stored and processed, thus empowering firms to be better informed about the consumers or the market itself. On the other hand, how to use the information flows to make the correct managerial decisions is still a challenging task for managers and academics alike. My dissertation studies the marketing implications of these information flows. Broad as the research question is, my dissertation focuses on specific market settings. I adopt both analytical and empirical methodologies to study information flows in these markets. Overall, this dissertation concludes that information flows can engender new market mechanisms, can provide valuable information of unobservable market forces, and can be created to improve social welfare.

Essay 1: Innovation Incentives for Information Goods

Digital goods can be reproduced costlessly. Thus a price of zero would be economically-efficient for consumers. However, zero revenues would eliminate the economic incentives for creating such goods in the first place. We develop a novel mechanism which tries to solve this dilemma by decoupling the price of digital goods from the payments to innovators while maintaining budget balance and incentive compatibility. Specifically, by selling digital goods via large bundles the marginal price for consuming an additional good can be made zero for most consumers. Thus efficiency is enhanced. Meanwhile, we show how statistical sampling can be combined with tiered coupons to reveal the individual demands for each of the component goods in such a bundle. This makes it possible to provide accurate payments to creators which spurs further innovation. In our analysis of the proposed mechanism, we find that it can operate with an efficiency loss of less than 0.1

Essay 2: Edgeworth Cycles in Keyword Auctions

Search engines make a profit by auctioning off advertisement positions through keyword auctions. I examine the strategies taken by the advertisers. A game theoretical model suggests that the equilibrium bids should follow a cyclical pattern— "escalating" phases interconnected by "collapsing" phases — similar to a pattern of "Edgeworth Cycles" that was suggested by Edgeworth (1925) in a different context. I empirically test the validity of the theory. With an empirical framework based on maximum likelihood estimation of latent Markov state switching, I show that Edgeworth price cycles exist in this market. I further examine, on the individual bidder level, how strategic these bidders are. My results suggest that some bidders in this market adjust their bids according to Edgeworth predictions, while others not. Finally, I discuss the important implications of finding such cycles.

Essay 3: The Lord of the Ratings

Third-party reviews play an important role in many contexts in which tangible attributes are insufficient to enable consumers to evaluate products or services. In this paper, I examine the impact of professional and amateur reviews on the box office performance of movies. I first show evidence to suggest that the generally accepted result of "professional critics as predictors of movie performance" may no longer be true. Then, with a simple diffusion model, I establish an econometrics framework to control for the interaction between the unobservable quality of movies and the word-of-mouth diffusion process, and thereby estimate the residual impact of online amateur reviews on demand. The results indicate the significant influence of the valence measure (ratings) of online reviews, but their volume measure (propensity to write reviews) is not significant once I control for quality. Furthermore, the analysis suggests that the variance measure (disagreement) of reviews does not play a significant role in the early weeks after a movie's opening. The estimated influence of the valence measure implies that a one-point increase in the valence can be associated with a 4-10% increase in box office revenues.

Thesis Committee:
Erik Brynjolfsson, George and Sandi Schussel Professor of Management, Chair
John D.C. Little, Institute Professor, Professor of Management Science
Chrysanthos (Chris) Dellarocas, Assistant Professor of Information Systems

Acknowledgments

Contents

A Novel Mechanism for Providing Innovation Incentives for Digital Goods

Doctoral Dissertation Essay 1

Xiaoquan (Michael) Zhang

MIT Sloan School of Management

Cambridge, MA

December, 2004

This paper is based on joint work with Erik Brynjolfsson.

A Novel Mechanism for Providing Innovation Incentives for Digital Goods

Abstract

Digital goods can be reproduced costlessly. Thus a price of zero would be economically-efficient for consumers. However, zero revenues would eliminate the economic incentives for creating such goods in the first place. We develop a novel mechanism which tries to solve this dilemma by decoupling the price of digital goods from the payments to innovators while maintaining budget balance and incentive compatibility. Specifically, by selling digital goods via large bundles the *marginal* price for consuming an additional good can be made zero for most consumers. Thus efficiency is enhanced. Meanwhile, we show how statistical sampling can be combined with tiered coupons to reveal the individual demands for each of the component goods in such a bundle. This makes it possible to provide accurate payments to creators which spurs further innovation. In our analysis of the proposed mechanism, we find that it can operate with an efficiency loss of less than 0.1% of the efficiency loss of the traditional price-based system. It is not surprising to find that innovation incentives in our mechanism are improved relative to the zero-price approach often favored by content consumers. However, it is surprising to find that the incentives are also substantially better than those provided by the traditional system based on excludability and monopoly pricing which is often favored by content owners. The technology and legal framework for our proposed mechanism already exist, the key issues of implementing it are organizational.

1 Introduction

Digital goods are different. Unlike other goods, perfect copies, indistinguishable from the original, can be created at almost zero cost. With the advent of the Internet, mobile telephony, satellite communications, broadband and related technologies, these goods can be distributed to almost anyone in the world at nearly zero cost as well. Furthermore, when a person consumes a digital good, he or she doesn't reduce the stock for anyone else.

This should be a virtual nirvana. Yet, ironically, low cost digital goods are seen as a mortal threat to the livelihoods of many individuals, companies and even whole industries. For example, digitized music has been blamed for an 8.0% decline in music CD sales in 2005 again, and from 2000 to 2005, the units shipped has shrunk more than 25%[1]. The availability of digital music is said to threaten the incentives for innovation and creativity itself in this industry. It has engendered a ferocious backlash, with thousands of lawsuits, fierce lobbying in Congress, major PR campaigns, sophisticated digital rights management systems (DRMs), and lively debate all around.

Music is not the only industry affected. Software, news, stock quotes, magazine publishing, gaming, classified ads, phone directories, movies, telephony, postal services, radio broadcasting, and photography are just a few of the other industries that are also in the midst of transformation. It's said to be difficult to predict the future, but a few predictions can be made with near certainty about the next decade: the costs of storing, processing and transmitting digital information will drop by at least another 100-fold and virtually all commercial information will be digitized. While our colleagues in computer science, both in academia and industry, deserve much praise for this, it is incumbent upon information systems researchers to understand the business, social and economic implications of these changes. Unfortunately, these implications have proven far less predictable[2]. What's more, we should go beyond prediction and seek to develop methods for maximizing the benefits from technological innovations while minimizing the costs.

---

[1] Source: RIAA. http://www.riaa.com/news/newsletter/pdf/2005yrEndStats.pdf

[2] Although, as noted by Sorenson and Snis (2001), and Lyman and Varian (2004) among others, we can predict with some confidence that there will be an increasing need for, and existence of, computer supported codified knowledge and information, and the concomitant institutions for managing this information.

Two schools of thought have dominated the debate on the economics of digital goods. One school stresses the benefits of the traditional market system. Clear property rights allow creators to exclude users from access to their creations. Users who wish to benefit from a creation must therefore pay the creator. This payment in turn assures that a) the goods go to those individuals with the highest value for the good and b) that the creator has incentives to continue to create valuable goods. This system has been pretty successful in modern market-based economies. To many people, it seems natural to apply the same principles to digital goods, typically via a some combination of law (e.g. the Digital Millennium Copyright Act), technology (e.g. DRMs) and social education.(e.g. the software industries ongoing anti-piracy public relations efforts).

Another school of thought thinks this approach is all wrong. "Information wants to be free" some of them argue. More formally, the point can be made that since digital goods can be produced at zero marginal cost, the textbook economic principle of efficiency: "price equals marginal cost" demands that price should never be greater than zero. After all, society as a whole is only made worse off if a user is excluded from access to a digital good which could have been provided without reducing the consumption of anyone else. While appealing, this approach begs the question of how to provide incentives for the goods creators. While some creators might continue to create for the shear joy of it, for indirect economic benefits such as enhancing their reputation or competency, or out of altruism, economic systems and business models which rely solely on these motivations have historically not fared as well as those which provide more tangible rewards to innovators and creators.

Thus, the debate can be thought of as over which group should be impaled on the two horns of the dilemma: should users be deprived of goods which cost nothing to produce or should creators be deprived of the rewards from their creations? Either approach is demonstrably suboptimal (Lessig 2004, for example). It would seem impossible to have both efficiency and innovation when it comes to digital goods. Improving one goal appears to be inextricably intertwined with hurting the other goal.

In this paper, we argue there is a third way. In particular, we develop and analyze a method for providing optimal incentives for innovation to the creators of digital goods. We show that it is possible to decouple the payments to the innovators from the charges to consumers while

14

still maintaining budget balance. In this way, we can slice the Gordian knot and deliver strong incentives yet unhindered access to the goods for almost all interested consumers. In fact, we find that our system actually provides *better* incentives for innovation than the traditional price system, even when bolstered by powerful DRMs and new laws to enhance excludability and thus monopoly power.

We argue that it is misguided to try to force the old paradigm of excludability onto digital goods without modification. Ironically, DRMs and new laws are often used to strip digital goods of one of their most appealing, and economically-beneficial attributes. At the same time, we take seriously the need to reward innovators financially if we wish to continue to encourage innovation and creativity.

The essence of our mechanism is to (1) aggregate a large number of relevant digital goods together and sell them as a bundle and then (2) allocate the revenues from this aggregation to each of the contributors to the bundle in proportion to the value they contribute, using statistical sampling and targeted coupons. We do this in a way which is fully budget-balancing and which provides accurate incentives for innovation with efficiency losses as small as 0.1% of the traditional price system.

Large digital collections are increasingly common as Internet content moves from free to fee and as new forms of digital content, such as satellite radio, emerge. Consider XM radio, Cable TV, AOL content, Rhapsody music, *Consumer Reports* reviews, JSTOR academic articles and last but not least, Microsoft Office software.

Bundling has been analyzed in some depth in the academic literature (McAfee, McMillan and Whinston 1989), including a cluster of articles specifically focusing on the bundling of digital information goods (Bakos and Brynjolfsson 1999, Bakos and Brynjolfsson 2000, and the references therein). A key finding from the literature is that in equilibrium, very large bundles will provide content that is accessible to the vast majority of the consumers in the relevant market. It will not be profitable to exclude (via pricing) any consumers except those with very low valuations for all the goods in the bundle. Thus, bundling can dramatically increase economic efficiency in the allocation of information goods to consumers.

Our paper focuses on the second part of the mechanism, which involves designing a system for allocating revenues from such a bundle. This is necessary because by its very nature, bundling destroys the critical knowledge about how much each of the goods in the bundle are valued. Did I subscribe to XM radio for the classical music or for jazz in the bundle? How much did I value each of these components? Unlike for unbundled goods, my purchase behavior for the bundle does not reveal the answers to these questions, creating a problem when it comes time to reward the creators and providers of the component goods. Surveys, usage data and managerial "instinct" can all help, but none is likely to be anywhere as accurate as a true price-based system. Our mechanism re-introduces prices, but only for a tiny fraction of consumers. For instance, only 1000 consumers out of several million would face any prices for individual goods, typically via special coupons. This allows us to get accurate, unbiased assessments of value but because the vast majority of consumers do not face any non-zero price for individual goods, they incur virtually no inefficiency. Specifically, 99.9% of users have access to any given good as long as their value for that good is greater than zero and their values for all other goods in the bundle are not simultaneously extremely low.

The academic literature related to this part of our analysis is quite sparse. Some of the closest research is the work on a monopolist facing an unknown demand curve (Aghion, Bolton, Harris and Jullien 1991) where it is shown that the seller can experiment by pricing to different buyers sequentially and updating the price accordingly. Some of the works on optimal market research is also relevant (Jain, Mahajan and Muller 1995).

We are not aware of any systems which fully implement *both* part of our mechanism, although bits and pieces are used in various industries and applications. For instance, as noted above, there are many examples of bundling for digital goods. Revenue allocation similar to our approach is more difficult to find. However, the American Society of Composers, Authors and Publishers (ASCAP) does seek to monitor the consumption of its members' works and distribute its revenues to each creator in rough proportion to this consumption. However, they have no direct price data, and thus must work under the implicit assumption that all songs have equal value to each listener.

Thus, our paper both introduces a novel mechanism and rigorously analyzes it, finding that it is technically feasible and that it can dominate any of the approaches debated thus far. Barriers to

diffusion and assimilation of this approach are likely to include overcoming knowledge barriers and some measure of organizational and institutional learning. Our analysis is meant to be a first step in addressing these obstacles. Notably, if this innovation succeeds, it should actually increase the pace of *future* innovations by improving incentives for the creation of useful digital goods. At a minimum, a broader discussion of this type of approach should change the terms of the existing debate about business models for digital goods.

The paper is organized as follows. Section 1 describes the basic assumptions and derives the asymptotic properties of massive bundling of information goods. Section 2 introduces the problem of revenue distribution in bundling and characterizes the different types of solutions to this problem. Section 3 shows that the traditional way of distributing revenue does not provide a socially desirable innovation incentive for goods. Section 4 proposes a mechanism to solve the revenue distribution problem and gives the convergence properties. It is shown that our proposed mechanism can induce the correct innovation incentives. Section 5 discusses practical issues of using the mechanism in the real world, and Section 6 concludes with a brief summary and some implications.

2 A Basic Model of Bundling

Our goal here is to provide a theoretical framework to which we refer in later sections.

We consider a market with many providers of digital goods and many potential buyers. Digital goods are assumed to have a reproduction cost of zero.[3] If these goods are sold separately, then any price greater than zero will be socially inefficient. Some consumers (e.g. those with valuations less than the price but greater than zero) will be excluded from consuming the good even though it would be socially beneficial for them to have access to it. This is commonly called deadweight loss. In this section, we briefly show how bundling can radically eliminate this inefficiency, albeit at the cost of introducing a different problem involving incentives for

---

[3]In our models, the condition of zero marginal cost is important. Digital goods typically satisfy this assumption easily. However, Bakos and Brynjolfsson(1999) showed that the main results for bundling continue to hold even for small marginal costs.

innovation.

Suppose a monopolistic bundler connects the producers and the buyers by designing an optimal pricing and revenue distribution policy to maximize the bundler's profit. Each buyer has (at most) unit demand for any of the information goods. Suppose a buyer's valuations of the goods in the bundle are independent draws from a random variable V in the range normalized to $[0, 1]$, and that the random variable has a cumulative distribution function $F(v)$, whose corresponding probability density function is $f(v)$. In other words, a buyer's value for one good (e.g. a Britney Spears song) is independent of his value for an unrelated good (e.g. a news story about a local fire). At a price of p, the demand will be $Q(p) = Prob(v > p) = 1 - F(p)$, yielding revenue of $\pi(p) = p[1 - F(p)]$. This implies that the inverse demand curve is $p(q) = F^{-1}(1 - q)$, and the bundler's problem is to solve:

$$\pi^* = \max_p \{p \cdot (1 - F(p))\} \tag{1}$$

Taking first order condition, we have $\dfrac{\partial \pi}{\partial p} = (1 - F(p) - p \cdot \dfrac{\partial F(p)}{\partial p}) = 0$, which can be rearranged to:

$$\frac{p^* \cdot f(p^*)}{1 - F(p^*)} = 1 \tag{2}$$

For the monopolistic bundler, it turns out that her profit maximizing decision is not difficult. As shown in Bakos and Brynjolfsson (1999), the bundler's job is to find the optimal price for the sum of many random variables ($S_n = \sum_{i=1}^{n} v_i$). By the law of large numbers, it is easier to find an optimal price for the sum S_n than for individual goods v_i, because the distribution variance of S_n is decreasing as n becomes large.

In particular, it can be shown (Lemma 1 and Lemma 2 in the appendix) that for non-negative random variables, the expected value of the random variable V can be written as

$$E[X] = \int_0^\infty [1 - F(x)] \, dx \tag{3}$$

Interestingly, this expression can be linked directly to the area under the demand curve. When

18

price is v, demand is given by $Q(v) = 1 - F(v)$, so the area under the demand curve is just $\int_0^\infty Q(v)dv = \int_0^\infty [1 - F(v)] = E[V]$.

As shown by Bakos and Brynjolfsson (1999), in equilibrium, the profit maximizing price will be set low enough so that virtually all consumers interested in any of the goods in the bundle will buy the whole bundle (even if they use only a small fraction of its components). For instance, most PC users buy Microsoft Office, even if they don't use all its applications, or even all of the features of the applications that they do use. While there may be anti-competitive implications to this fact (see Bakos and Brynjolfsson, 2000), such bundling does give the socially desirable result of dramatically reducing the deadweight loss because very few consumers are excluded from using any of the bundled goods in equilibrium. In essence, once consumers purchase the bundle, they can consume any of the goods in the bundle at zero marginal cost. Thus, when the cost of (re)producing the goods is close to zero, bundling provides close-to-optimal allocation of goods to consumers (Bakos and Brynjolfsson, 1999).

However these benefits comes at a major cost. Bundling inherently destroys information about how each of the component goods are valued by consumers. Is the bundle selling because of the fresh sounds of a new artist or due to the lasting appeal of a traditional favorite? Without this information, it is impossible to allocate revenues to the providers of content in a way that accurately encourages value creation. Selling goods individually would automatically solve this problem, but as discussed above, individual sales create enormous inefficiencies because they exclude some users with positive value from access to the good.

Accordingly, the remainder of the paper studies the question of how to provide the correct rewards to content providers, and thereby give them financial incentives to create content.

3 The Revenue Distribution Problem

Bundling strategies help sellers to extract more consumer surplus. If one single seller can not provide enough numbers of information goods, it is worthwhile to have one content aggregator to negotiate with multiple sellers to offer a bundle of information goods from multiple sources.

The ideal revenue distribution mechanism would be one which somehow determined each good's demand curve, and distributed the revenue among the content providers in proportion to the social value of each good to all consumers. This value can be calculated by integrating the area below each good's demand curve. Various mechanisms used to derive demand curve proposed in the literature all fail here because bundle pricing does not automatically provide a way to observe the market's response to a price change of individual goods.

If the benefits created by each good cannot be observed or calculated, then a host of inefficiencies may result. First, the content providers may not have enough incentives to produce creative products, and consumers will eventually be harmed. Second, without a good signal of consumers' preference, content providers may not produce the content that best fit the consumers' taste. Third, in any effort to overcome these problems, the content producers may force the potential bundler to adopt other strategies such as pay-per-view (the case of iTunes). However, such strategies re-introduce the deadweight loss problem discussed at the beginning of section 2.

In the following subsections, we discuss the costs and benefits of several ways to distribute revenue to address this challenge, culminating with our proposed approach.

3.1 Payment determined by number of downloads

In the context of digital information goods, it is natural to assume that the seller may be able to observe the number of times for each good is accessed. This gives us the following solution.

If one is willing to assume that the number of accesses signals popularity, and popularity is a measure of value, we can infer the value by the number of accesses. Traditionally, this scheme is broadly used in the market of digital goods such as music, movie, TV shows, and software. For example, each episode of *Friends* gets about 29 million viewers per week, which is far more than most other TV shows; as a consequence, each of the six stars gets paid $1.2 million per episode, which is far more than most other TV actors.

More formally, suppose we have n goods in the bundle, the price for the bundle is B. Also suppose there are m buyers of the bundle, each represented by j ($j = 1, ...m$), then the total bundle revenue

is $R = B \cdot m$. We assume the system can record the number of downloads of buyer j for good i: d_{ij}, then the provider of content i should be paid:

$$revenue_i = B \cdot \sum_{j=1}^{m} \frac{d_{ij}}{\sum_{k=1}^{n} d_{kj}} = R \frac{d_i}{\sum_{k=1}^{n} d_k}.$$

This method is extremely easy to implement. In fact, the last equation implies that the bundler does not even have to keep record of all the downloads made by the m buyers[4], she can simply record d_i, the number good i has been downloaded.

This method is powerful in the context when all the goods are approximately equal in value. If goods differ in value (bundling very cheap "Joke-A-Day" with more expensive "Forrester Research Report"), then pricing based on number of downloads is misleading. (The Joke-A-Day may be downloaded more times than the Forrester Research Report, but aggregate value of the latter may be much higher to consumers).[5]

3.2 Payment determined by downloads combined with a stand-alone price

Number of downloads itself is not a good measure of consumer valuation in many cases. Assuming there also exists a stand-alone price for every information good in the bundle, and assuming these prices are all fair prices, we can then derive an improved mechanism to distribute the revenue.

Consider the market introduced in subsection 2.1, suppose each item i ($i = 1,..,n$) in the bundle also has a stand-alone price p_i.

Building on the equation from subsection 2.1, an improved way to distribute the revenue is through the following formula:

$$revenue_i = B \cdot \sum_{j=1}^{m} \frac{p_i d_{ij}}{\sum_{k=1}^{n} p_k d_{kj}} = R \frac{p_i d_i}{\sum_{k=1}^{n} p_k d_k} \tag{4}$$

---

[4]Since no j appears in the final term.

[5]Another problem with this method is that it gives dishonest content providers a way to distort the values by manipulating the number of downloads of their own content. This has been a problem, for instance, with some advertising-supported content where prices are based on thousands of impressions recorded (Hu 2004).

which says that the revenue to distribute to content provider i should be a proportion of the total revenue, and the proportion is determined by the sum of each consumer's valuation of good j.

This method has the advantage of being more precise comparing to the previous solution. Indeed, if "Joke-A-Day" is sold separately, its price will probably be much lower than that of "Forrester Research Report". The disadvantage of this method is that a fair and separate price may not always be readily available. If the distribution of revenue is set according to this method, and when bundling becomes a major source of revenue, there are rooms for content providers to misrepresent the stand-alone price. Furthermore, this approach implicitly assumes that the value from each good is proportional to the stand-alone price. However, this will only be true if the price paid by the marginal consumer of each goods is proportional to the average price that would be paid by all consumers of that good, for all goods[6].

3.3   Other Mechanisms

In William Fisher's new book (Fisher 2004), he explores various solutions to the music piracy problem brought about by the new peer-to-peer technology. Specifically, he proposes to replace major proportions of the copyright and encryption-based models with a "governmentally administered reward system", and he correctly points out that what we really need is not the number of downloads, but the "frequency with which each recording is listened to or watched" (i.e. the real value to consumers). Fisher's proposal is similar to the Nielsen TV sampling approach, and he proposes to implement special devices to estimate the frequency of each recording is listened to. He also suggests that the frequency should be multiplied by the duration of the works, and that consumer's intensity of enjoyment (obtained through a voting system) should be taken into consideration to make more precise estimates of the valuations.

This proposal, if carried out, should be superior to the current practice taken by ASCAP (and BMI, SESAC, etc.) to compensate music producers, and it comes very near to our ideal of learning consumers' valuations and distribute money accordingly; but it also suffers from several problems. First, different from Nielson TV sampling, people may use different devices to enjoy

---

[6]Barro and Romer (1987) explore how similar proportionalities can explain a number of pricing anomalies.

the same digital content. For example, a song can be played with an MP3 player in the car, a CD player in the home entertainment system, or a DVD drive on a computer. Second, as shown in the public goods literature, votes are not reliable because individual hidden incentives may induce voters to misrepresent their true values. In essence, the Fisher approach still does not provide a reliable, incentive-compatible way to determine the true value of each good to consumers.[7]

4 Innovation Incentives

Before moving on to propose our mechanism to solve the revenue distribution problem, we will look at another related issue in this section. We show how innovation incentives are severely limited by the traditional pricing mechanism.

In particular, we show that, contrary to common belief, the traditional price system based on excludability does not provide correct innovation incentives to producers. Our subsequently proposed couponing mechanism not only solves the revenue distribution system, but also can be a socially desirable way to promote innovation for digital goods.

Suppose that the seller can invest in trying to create an innovation which improves consumers' valuations of her digital good. The investment can be in the form of improving product quality, functionality or educating users to use the product more effectively. We now give a closer look at the innovation incentives of the seller.

4.1 Uniform enhancement

Suppose the innovation can increase each consumer's valuation by δ, this is equivalent to moving the demand curve upward by δ.

---

[7]The public goods mechanism design literature seeks to provide a remedy to the voter misrepresentation problem. Specifically, the Vickrey-Clarke-Groves (VCG) mechanism can be shown to induce truth-telling by all participants. However, it has two fatal flaws. First, it is not budget-balancing - significant inflows (or net penalties) are generally needed. Second, it is quite fragile. Each participant must believe that all other participants are truth-telling or he will not tell the the truth himself. Accordingly, while VCG design is intriguing in theory, it is rarely, if ever, seen in practice.
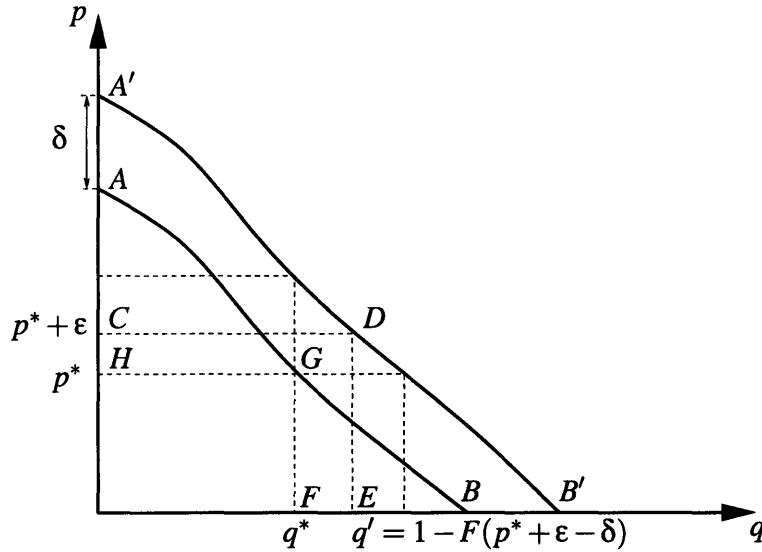
Figure 1: Upward shift of demand curve

When the demand is shifted upward, the monopolistic seller will be charging a new price $p' = p^* + \varepsilon$ that maximizes her profit. With this innovation, she can expect to gain some extra profit than before indicated by the area $CDEFGH$ in Figure 1. In the figure, although what the seller has created to the society is the area between the two demand curves, in the traditional price system, she gets paid by area $CDEFGH$, which is not exactly matched with her effort. To the society, this innovation on one hand reduces consumer surplus, but on the other hand also reduces the deadweight loss to a certain extent, so the overall social welfare effect is mixed.

When the demand is shifted upward, if the monopolistic seller charges a higher price of $p^* + \delta$, she will keep selling the optimal quantity q^*, or alternatively, she could keep charging the optimal price p^* and sell to more people (the demand will be $q' = 1 - F(p^* - \delta)$ now). We next show, in Lemma 3, that both strategies lead to the same expected profit for the seller.

Lemma 3: Marginally, the innovative monopolist seller can charge a higher price or enjoy a increased demand, and the two strategies are equivalent in terms of expected profit.

Note that Lemma 3 is a natural result implied by the optimality of p^*, we will be using this result in the next sections. From the figure, the seller should be charging a new price at $p^* + \varepsilon$, $0 < \varepsilon < \delta$, which is strictly better than either p^* or $p^* + \delta$.

24

4.2 Targeted innovation

We have assumed above that the innovation can uniformly increase consumers' valuations of all types. Now we look at the case that innovation can only affect a small subset of consumers' valuations. In particular, the innovation may be less significant so that only some consumers with valuation near some $\tilde{v}$ are affected. For instance, a software developer could invest in adding features which would i) make satisfied users of its product even more satisfied, ii) increase the value to consumers whose values were just below the market price, turning them into buyers, or iii) features which would increase the value of non-buyers but not enough to turn them into buyers. Suppose that the developer has a finite budget and can only pursue one of these three types of innovations. Even if innovations of type i) or iii) might create more value, the traditional price system will only provide incentives for innovation ii).
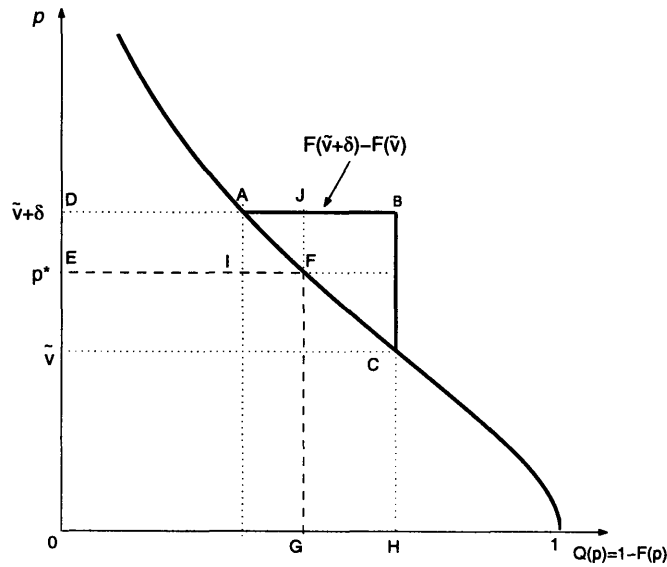
Figure 2: Social benefit/loss of seller innovation.

More formally, the total potential social value of the good equals the area under the demand curve. If the seller makes a targeted innovation for some consumers with valuation $\tilde{v}$, the social gain of the innovation is thus denoted by the area ABC in Figure 2. When δ is small, $\triangle ABC \approx \frac{1}{2}\delta[F(\tilde{v}+\delta) - F(\tilde{v})] \approx \frac{1}{2}\delta^2 f(\tilde{v})$.

25

We shall need the following technical assumption to get a well-behaved demand curve.

Assumption: $F(v)$ is twice continuously differentiable with $F(0) = 1, F(1) = 1, f(v) > 0 \ \forall v > 0$, and $\frac{1}{1-F(v)}$ is strictly convex for $v \in (0,1)$.

This assumption is implied by log-concavity of $1 - F(v)$, which itself is implied by log-concavity of $f(v)$. The intuitive meaning of the assumption that the random variable V has a log-concave density function is that it has a unique global maximum. Note that this assumption implies that the profit function $p \cdot [1 - F(p)]$ is concave.

Log-concavity property is frequently assumed in the economics literature. See, for example, Laffont and Tirole (1988) in the context of games with incomplete information; Baron and Myerson (1982) in the context of theory of regulation; Myerson and Satterthwaite (1983), among others, in the context of auction theory. It is also well known that log-concavity of the density function implies the notions of IFR (increasing failure rate), and NBU (new better than used) in survival and reliability analysis literature (Barlow and Proschan 1975). Bagnoli and Bergstrom (1989) give a good review. In our context, log-concavity is sufficient to guarantee that solutions are unique and well-behaved.

Given any innovation that increases some consumers' valuation by δ, there exists some $\bar{v}$, such that the seller is indifferent between carrying out the innovation and not carrying out the innovation. For the seller,

$$(\bar{v}+\delta)[1 - F(\bar{v})] = p^*[1 - F(p^*)] \tag{5}$$

and solving for $\bar{v}$, we have two values, $\bar{v}_L$ and $\bar{v}_H$[8], such that the optimal price $p^* \in (\bar{v}_L, \bar{v}_H)$ satisfies (5). Also, $\forall \ v \notin [\bar{v}_L, \bar{v}_H]$, it must be that $(v+\delta)[1 - F(v)] < p^*[1 - F(p^*)]$, so the seller has no incentive at all to innovate for consumers with valuations outside the range $(\bar{v}_L, \bar{v}_H)$. This is very intuitive, in the traditional market, if the seller sells goods to consumers with valuation higher than $\bar{v}_H$, it makes no sense to increase their valuations further because that will only contribute to consumer surplus, and the seller will not be able to extract the added value. Similarly, for the potential consumers with lower valuations (lower than $\bar{v}_L$, to be precise), the seller will not take

---

[8]This is a direct consequence of the assumption of log-concave density function. Here we omit a formal proof of the existence and uniqueness of $\bar{v}_L$ and $\bar{v}_H$, which can be easily derived with the fixed point theorem.

the effort to innovate because they will not be converted to consumers. For small δ, the range $(\bar{v}_L, \bar{v}_H)$ is very small, and even in this range, innovation may not be socially desirable.

For consumers with valuation in the range $(\bar{v}_L, \bar{v}_H)$, one can look at three distinct cases:

(1) The good region: $v \in [\bar{v}_L, p^* - \delta)$

In this case, the seller would want to charge a price $p = v + \delta$ and earn profit $\pi = (v + \delta)[1 - F(v)]$. By Lemma 3, $\pi > (\bar{v}_L + \delta)[1 - F(\bar{v}_L)] > p^*[1 - F(p^*)]$. So the seller prefer to lower the price from p^* to $p = v + \delta$, and earn a higher profit. The reduction in price has two socially desirable effects. First, the consumer surplus is increased. For people with valuation in the range $(v + \delta, p^*)$, they are no longer excluded from accessing the good; and for people with valuation in the range $(p^*, +\infty)$, they can each enjoy an increase in consumer surplus of $\Delta CS = p^* - (v + \delta)$. Second, deadweight loss is reduced, the change in deadweight loss is $\Delta DWL = [F(p^*) - F(v + \delta)](v + \delta)$. The reduction in deadweight loss is composed of two parts: first, for people with valuation in the range $(v + \delta, p^*)$, apart from the increase in consumer surplus, there is also reduction in deadweight loss due to the fact that their demand is satisfied, second, for people with valuation in the range $(v, v + \delta)$, the innovation increases their valuation, and they are no longer excluded from purchasing the good.

(2) The bad region: $v \in [p^*, \bar{v}_H]$

In this case, the seller innovates for people with valuation just higher than the optimal price. By lemma 1, we know it is worthwhile for her to increase the price to $v + \delta$, there are two socially undesirable effects associated with this. First, for consumers originally having a valuation above $v + \delta$, they each lose consumer surplus by $\Delta CS = v + \delta - p^*$. Also, for people with valuation in the range $(v, v + \delta)$, although their valuation is increased due to the innovation, they no longer enjoy a surplus now. Second, for people with valuation in the range (p^*, v), they can no longer afford to buy the good now, so there is an increase in deadweight loss.

(3) The ugly region: $v \in [p^* - \delta, p^*)$

This case has mixed effects. On one hand, there are socially desirable effects as in 1, and on the other hand, there are also socially undesirable effects as in 2. For people with valuation higher than $v + \delta$, they suffer a reduction in consumer surplus by $\Delta CS = v + \delta - p^*$. In Figure 2, the loss is

indicated by the area $ADEI$. For people with valuation in the range $(p^*, v+\delta)$, due to innovation, they have a higher valuation now, but due to the increased price, they no longer enjoy a surplus (the area AIF). For people with valuation in the range (v, p^*), their valuation is increased to $v+\delta$, but again, the seller gleans all the surplus due to innovation. A socially desirable side effect is that the dead-weight-loss is reduced because these group of people are able to use the product now. The area $FGHC$ indicates the social gain from reduced deadweight loss. In total, consumer surplus is hurt by the area $ADEF$, deadweight loss is reduced by the area $FGHC$, and the seller enjoys the extra value created by innovation indicated by area ABC.

Thus, in the traditional price mechanism, the seller has too little incentive to create innovations that mainly benefit consumers with very low or very high valuations. Once you are locked into a software provider, don't expect her to put a lot of effort into innovating to address your needs.

To see the socially wasteful incentive of innovation in the traditional price system, consider the case of the consumers' valuations near the optimal price. For example, if the seller takes an effort to innovate and increases the valuation for some consumers from p^* to $p^* + \delta$, then her gain is $\delta[1 - F(p^*)]$. The ratio of her gain over her contribution is $incentive\_ratio_{Traditional} = \delta[1 - F(p^*)]/[\frac{1}{2}\delta^2 f(p^*)] = 2\frac{1-F(p^*)}{\delta f(p^*)} = \frac{2p^*}{\delta}$, and $lim_{\delta \to 0} incentive\_ratio_{Traditional} = \infty$. This is a very shocking result, as shown above in case (2), the innovation for people whose valuation is just above the optimal price (the bad range) will bring about socially undesirable effects such as reduced consumer surplus and increased deadweight loss, yet this is exactly the range where it is most attractive for the sellers to innovate.

5 Couponing Mechanism

As discussed in Section 3, the ideal way to provide correct incentives is to learn consumers' valuations for each good and make corresponding payments. Since bundling itself obscures consumers' valuations for individual goods, here we propose a mechanism to derive the demand curve for each good by issuing targeted coupons to a small, but statistically representative, sample of consumers. Our mechanism is substantially different from the traditional use of coupons

as a marketing method to price discriminate consumers. Instead, coupons in this mechanism is similar to the price experiments suggested in the optimal pricing literature.

Suppose the monopolistic bundler offers a bundle of information goods to a group of consumers. In order to derive the demand curve for one of the components, she could choose $m \cdot n$ representative consumers and issue each of them a coupon, where n is the number of price levels covering the range of the valuations, which we call "coupon levels" (one simple way to get these levels is to offer coupon values from $\frac{1}{n}\bar{V}$ to $\frac{n-1}{n}\bar{V}$ where $\bar{V}$ is the upper bound of consumer valuations for this good), and m is the number of coupons to be offered for each of the price levels, which we call "sample points" (there will be m consumers who can receive a coupon with face value $\frac{i}{n}\bar{V}$, $i = 1,...n - 1$). While $m \cdot n$ is large enough to make statistically valid inferences, it is nonetheless a very small fraction (e.g. 1/1000 or less) of the total set of consumers buying the bundle.

If a consumer receives a coupon with face value $\tilde{v}$, then he can either choose to ignore the coupon and enjoy the complete bundle or choose to redeem the coupon and forfeit the right to use the indicated component. So upon observing the consumer's action, the bundler can learn whether his valuation of the component is higher or lower than the face value of the coupon. Aggregating the m consumers' valuations will give the bundler a good estimate of demand at that price, summarizing the results for the n coupon levels, the bundler can plot a fairly accurate demand curve, and the area under the demand curve is the social valuation for the particular good. Using the same method for all the components, the bundler can learn the social valuation of each of the goods in the bundle. She can then distribute the revenue among the content providers according to their share of the total valuation. Let R be the total revenue from selling bundles, and v_i be the social value of the component i in the bundle, content provider of i should be paid

$$revenue_i = R\frac{v_i}{\sum_{j=1}^{N} v_j} \tag{6}$$

where N is the total number of content providers.

This method compares favorable to the traditional price mechanism. The traditional price mechanism subjects 100% of consumers to the inefficiency of positive prices. However, only data from a small fraction of consumers are needed to get extremely accurate estimates of the value created

and contributed by each good. The greater precision obtained by increasing the sample declines asymptotically to zero while the cost for subjecting each additional consumer to a positive price remains just as high for the last consumer sampled as the first one. When balancing the costs and benefits, the optimal sample size is almost surely less than 100%. Secondly, the proposed couponing mechanism actually provides a more accurate estimate of the overall demand curve than any single-price traditional system. Because multiple different prices for coupons are offered, a much more accurate overall picture of demand can be obtained than simply revealing the demand at a single price, as conventional prices do. As discussed in section 3, this has large and important implications for dynamic efficiency and innovation incentives.

One can also compare our couponing mechanism with the well-known Vickrey-Clarke-Groves (VCG) mechanism. Unlike VCG, our couponing mechanism does not give us exact valuations for each consumer. However, in general, approximate demand functions of the components will suffice, and by increasing the sample size, the accuracy can be made almost arbitrarily precise. Our couponing mechanism is superior to the VCG mechanism in several ways. (1) Truth-telling is a robust and strong equilibrium in the couponing mechanism, in the sense that each consumer simply compares his valuation with the coupon's face value, he is not required to assign correct beliefs on all other people's votes. (2) In the VCG, if one respondent misreports his value (due to irrationality or due to error), the consequence may be very severe for the rest of the people. Furthermore, coalitions of consumers can "game" the VCG to their advantage. However, in the couponing mechanism, the effects on others from a consumer's misreport are minimal. (3) The couponing mechanism is fully budget balancing, unlike the VCG. (4) The couponing mechanism is more intuitive than the VCG for real world problems.

The following proposition asserts that the Couponing Mechanism indeed gives us correct demand curve estimations in expectation.

PROPOSITION 1: For any one of the components in the bundle, given a large number of randomly chosen respondents and levels of coupons, the above mechanism gives an empirical demand function $\hat{Q}(p) = 1 - \hat{F}_V(p)$ that arbitrarily approximates the true demand function $Q(p) = 1 - F_V(p)$.

Proposition 1 gives an asymptotic result, we run simulations to see the effectiveness of this mechanism.
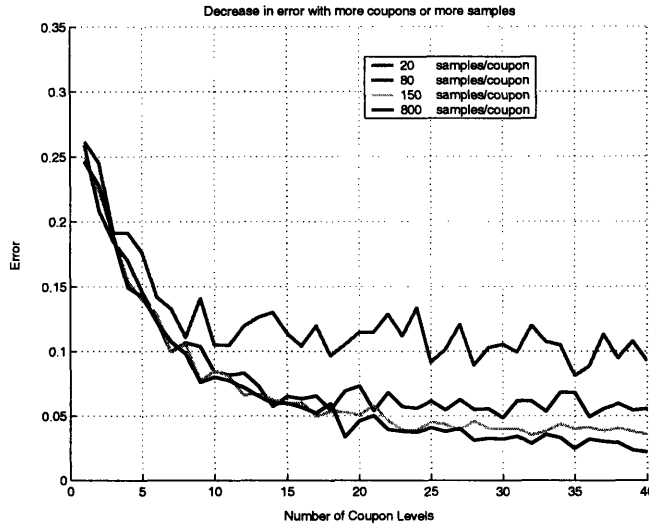
Figure 3: Simulation Results for the Couponing Mechanism

The use of coupon mechanism gives us empirical estimates of the inverse demand curves for each of the distributions, and we define the error rates to be the percentage differences between the area under the empirical demand curve and the area under the true demand curve. Figure 3 shows the result of the couponing mechanism applied to the uniform distribution (other distributions yield similar figures). We see that error rate is declining with more coupon levels and with more sample points for each coupon value. It is remarkable that with just 20 coupon levels, the error rate can be as low as 5%. Adding more sample points for each coupon value also helps to improve the precision. For example, with 40 coupon levels, sampling 20 consumers for each coupon level (for a total of 800 respondents) gives us an error rate of 10%, and sampling 80 consumers improves the error rate to be near 5%. From the error rate curves, we can also see that when sampling 20 consumers, adding coupon levels more than 10 does not improve the precision significantly; also, when sampling 80 consumers, adding coupon levels more than 15 does not improve the precision significantly. This observation tells us that we have to add coupon levels and sampling points simultaneously in order to achieve the best result estimating the social values of goods. Error rate converges toward 0 more quickly/slowly for fatter/thinner demand curves (the ones with a

higher/lower expected value). In our simulations, for some demand curves, with just 5 coupon levels and 20 sample points (for mere 100 respondents), the coupon mechanism can give us an error rate below 0.1%. Thus, sampling just 100 consumers can provide almost as accurate an estimate of demand as sampling all the consumers of the good, which could be in the millions.

The deadweight loss is proportionately smaller, too. Consumers who cash-in the coupon forgo access to the corresponding good, which creates a deadweight loss (unless the consumer's value was exactly zero). For such a consumer, this decision is analogous to facing a market price, with similar costs, benefits and overall incentives. However, in contrast to the traditional price approach, the Coupon Mechanism only subjects a fraction of consumers to this decision, so only a fraction choose not to buy, and the total deadweight loss is a fraction at large.

This mechanism can be used to solve the revenue distribution problem discussed in section 3, and we will show next, with a few propositions, that this mechanism can also help to avoid the innovation incentive issues arising in traditional price systems.

Consider the uniform enhancement introduced in section 3.1. When the demand is shifted upward, the seller can get paid virtually the full amount of the extra valuation it created for the consumers. Let the original profit be $\pi = E[V] = \int_0^\infty [1 - F(v)]dv$, she can now earn $\pi' = \delta + \pi$. We will show in Proposition 2 that the seller' innovation incentive in the bundling scheme is higher than that in the traditional market.

PROPOSITION 2: If an innovation can increase consumers' valuations uniformly higher, the proposed mechanism gives the producer strictly greater incentives of innovation than does the traditional market mechanism.

For targeted innovation introduced in section 3.2, it is obvious that the seller does not care how high or how low the targeted consumers' valuation is because she is paid according to the area under the demand curve. Combining bundling with couponing can provide balanced incentives for all three types of innovations, leading the developer to pursue any innovations whose expected benefits exceed expected costs.

In Figure 2, no matter where $\tilde{v}$ is , the reward to the seller is the area ABC, so she will not discriminate against consumers with low or high valuations. This brings us to Proposition 3.

32

PROPOSITION 3: If an innovation can increase only some consumers' valuations, the traditional price system does *not* provide correct incentives for the producer to innovate for people with relatively high or relatively low valuations. In contrast, the proposed mechanism always gives the producer socially desirable level of incentives to innovate.

Similarly, for the case of the proposed mechanism, the ratio of the expected return over the social contribution is $incentive\_ratio_{Bundling} = \frac{1}{2}\delta^2 f(\tilde{v}) / \left(\frac{1}{2}\delta^2 f(\tilde{v})\right) = 1$, which is fair. So we have the following proposition:

PROPOSITION 4: The traditional market gives the producer too high an incentive to innovate where it is most harmful to the social welfare, and no incentive elsewhere; the proposed mechanism induces the producer to make socially desirable innovation efforts.

6 Discussion

Throughout this paper, we assumed the more general case that the demand curves of different goods look different. If the demand curves are all the same, or at least all parallel to each other, there can be easier mechanisms to distribute the revenue while ensuring to keep the innovation incentives of producers. When the goods all have similar social values (the areas under the demand curves are the same), Equation (6) becomes $revenue_i = R\frac{v_i}{\sum_{j=1}^{N} v_j} = R\frac{q_i v}{\sum_{j=1}^{N} q_j v} = R\frac{q_i}{\sum_{j=1}^{N} q_j}$, where q_i denotes the number of times that good i is consumed, and the payment to content provider i is solely determined by the number of downloads. Interestingly, even if the social values are not similar, as long as the demand curves of different goods have similar shapes, we can still use the number of downloads as a sufficient statistic to derive the correct revenue distribution rule. For example, imagine the simplest case that we have linear demands with slope k; different goods have different intercepts, but they are all parallel to each other. Once we observe the total quantity consumed for each good, we know the social value created by this good is just a quadratic function of this number. So, in the spirit of equation (6), we have:

$$revenue_i = R\frac{v_i}{\sum_{j=1}^{N} v_j} = R\frac{q_i^2}{\sum_{j=1}^{N} q_j^2}.$$

33

This paper contributes to establishing a more efficient approach to create, distribute and consume digital goods. The theoretical foundation proposed here is just the first step toward this goal; in order to build viable business models, we need to address some practical issues to be discussed below.

In this paper, couponing has been analyzed solely as a mechanism for revealing existing demand, not for influencing it. Of course, in practice, couponing may also be viewed as a form of advertising that increases demand. If it increases demand more for some goods, and not for others, then the estimated values may be biased in a non-uniform fashion. There is a related, more conspicuous problem: due to the heterogeneity in people's tastes, some goods are surely downloaded less than some others (consider a Forrester report, maybe only tens out of millions of consumers would want to download it), if we do not offer enough sampling points, there will be a bigger error in estimating demand for these less popular goods. It turns out that both issues can be easily addressed by a practice we call "passive couponing". Under "passive couponing" regime, only those who downloaded a good will be offered a coupon for that good. After downloading, the consumer learns all the product characteristics, so the informative role of couponing as advertising is no longer valid. For goods downloaded by the majority of people, we can choose a small fraction out of them to offer coupons, and for goods downloaded only by a few, we may offer coupons to most or all of them. In either case, subsequent access to that good, or similar goods, can be restricted for consumers who prefer to redeem the coupon instead. By discriminating coupons offered to different types of goods, we can get a better overall estimate of the specific demands[9].

In previous sections, we avoided the issue of duration of contracts. It is likely to be unnecessary to permanently block from access to a good for consumers who redeem the corresponding coupon. Temporary blockage will generally suffice. We can put this question into the context of subscription-based business models. Suppose the bundle is to be paid by month (e.g. $20/month), then for time-critical information goods (e.g. news, stock quotes, etc.), we can offer the coupons

---

[9]What if a good is only downloaded by one consumer? First of all, in this case, this good is not important in the bundle, the bundler can exclude it in the future. Second, the bundler can offer this consumer a different coupon in each period with the face value determined by a random draw. Within some periods of sampling, the bundler can still extract the true value, the math works exactly the same as in the proof of proposition 1. It can also be easily shown that there is no incentive for the consumer to mis-report his value in each period.

by month, too (e.g. "Take this $1 coupon and sacrifice CNN news for the next month"). For those less time-critical information goods (e.g. music, software updates, etc), we can offer the coupons by longer periods (e.g. "Take this $10 coupon and give up downloading Madonna for the next whole year")[10].

What if the valuations are not independent as assumed in the paper? If two goods are substitutes, offering a coupon for one of them will only help us to estimate the incremental value that it brings to the bundle, and this is also true for the other good, so we will be paying less for the two creators than the value they bring into the bundle. For complements, we overestimate total value of the goods. First of all, non-independence will only affect the estimated share of contributions of each content provider, so the payment to each of the creators will be changed, but the benefit of innovation incentives will not be affected. Second, if we can identify clusters of goods that are substitutes or complements to each other, we can offer coupons for individual clusters and use the proposed mechanism to estimate the share of contribution by each cluster. This will ensure that a cluster of content providers will be paid a fair overall payment. Within a cluster, each individual content provider can be paid according to the estimated share of incremental value they bring to the cluster.

7 Conclusion

Revolutionary technologies often engender innovations in business organization. The digitization of information is no exception. We seek to advance the debate on how best to allocate digital goods and reward their creators by introducing a novel mechanism and analyzing its implications. Our approach eliminates the marginal cost of consuming digital information goods for the vast majority of consumers via massive bundling. For very large aggregations, this preserves most of the static efficiency which could be achieved with a zero price policy. However, in the long run, the more important issue is how to create incentives for ongoing innovation. Indeed, our living

---

[10]The mechanism proposed here may not work as well with information goods whose value is time-invariant (e.g. Symphony No. 9 by Beethoven). Depending on the nature of the DRM system used, once someone downloads a copy of the work, there may be no point in offering coupons because the consumer might not need to download any more copies in the future.

standards, and those of future generations, depend far more on continuing innovation than on simply dividing up the existing set of digital goods. In this area, the proposed mechanism shows particular promise. We find that our approach can provide substantially better incentives for innovation than even the a traditional monopoly price system bolstered by artificial excludability (e.g. via DRMs, laws, etc.). In particular, the traditional price system, in which each good is sold for a specific price with the proceeds going to the monopolist creator, focuses virtually on incentives on a very narrow band of consumers - those just on the margin of buying. In fact, the price system provides *too* strong incentives for innovations that help this narrow group of consumers. Rents transferred to the creator from such innovations exceed the social benefits. In contrast, our approach, using statistical sampling and couponing, can provide incentives which are nearly optimal for every type of innovation.

In summary the mechanism we introduce,

- has orders of magnitude less inefficiency than the traditional price system,

- is budget balancing, requiring no external inflows of money,

- works with existing technology and existing legal framework,

- requires no coercion and can be completely voluntary for all parties, since it is fully incentive compatible,

- doesn't assume that innovators will continue innovate even without financial rewards,

- can be implemented and run in real-time, and

- is scalable to very large numbers of goods and consumers (in fact, works better for larger numbers),

Our approach also has weaknesses and challenges. First of all, since massive bundling is a component of the mechanism, our approach only works for digital goods. As long as the marginal cost is not close to zero, the benefit of using bundling to ensure social efficiency will be invalid. Second, this mechanism may not work for all information goods. In order to have the mechanism

functioning, the contents must be updated regularly and consumers should have a good estimate of the expected value of the future contents. If consumers can subscribe once and download all contents and sign off the service, the mechanism will be useless[11]. Compared to giving away all digital goods for free, our approach will exclude a small number of consumers and create some inefficiency as a result. More importantly, our approach does require the creation of new business institutions or models, which is never easy (Fichman and Kemerer 1999). Specifically, an entity is needed to manage the statistical sampling and couponing, analyze the resulting data, and allocate payments to the content owners accordingly. Near misses for this type of entity already exist. For instance, ASCAP does much the same thing already for broadcast music, but without accurate price information. Nielsen and similar organizations provide usage information, but again without accurate price information. There are organizations which regularly collect and distributed large sums of money to member companies based on various algorithms. The Federal Deposit Insurance Corporation which does this for banks is one example. Some cooperatives are also run this way. Last but perhaps not least, the government regularly makes these types of transactions(Kremer 1998). However, it should be stressed, that our mechanism does not require any government role since all of the participants (consumers, content creators, bundlers) have incentives to participate completely voluntarily. This stands in contrasts to the proposal by Fisher (2004) or the varied proposals to change copyright or other laws.

By offering this new framework and analysis, with a new set of opportunities and challenges, we hope to lay the foundation for future research on the critical question of providing incentives for innovation in the creation of digital content and implementing mechanisms to deliver that content to consumers efficiently.

We expect that the next 10 years will witness a scale of organizational innovation for creating and distributing digital goods surpassing even the remarkable pace of the last 10 years. New coordination mechanisms, such as the innovation incentive approach described and analyzed in this paper will flourish. With a proactive attitude toward technology-enabled organizational innovation, we believe that academia can speed this process by framing the issues, and by providing

---

[11]Modern digital rights management technology may help to alleviate the problem. Some services allow the subscribers to download the music but require them to log on at least once per month to varify the status of subscription, if a subscription gets expired, the content will no longer be accessed.

tools, "cookbooks", repositories and analyses.

References

Aghion, P., Bolton, P., Harris, C. and Jullien, B.: 1991, Optimal learning by experimentation, *Review of Economic Studies* **58**, 621–654.

Bagnoli, M. and Bergstrom, T.: 1989, Log-concave probability and its applications, *Working Paper* .

Bakos, J. and Brynjolfsson, E.: 1999, Bundling information goods: Pricing, profits and efficiency, *Management Science* **45**(12), 1613–1630.

Bakos, J. and Brynjolfsson, E.: 2000, Bundling and competition on the internet: Aggregation strategies for information goods, *Marketing Science* **19**(1), 63–82.

Barlow, R. and Proschan, F.: 1975, *Statistical Theory of Reliability and Life Testing, Probability Models*, Holt, Rinehart, and Winston, New York, NY.

Baron, D. and Myerson, R.: 1982, Regulating a monopoly with unknown costs, *Econometrica* **50**, 911–930.

Barro, R. and Romer, P.: 1987, Ski-lift pricing, with applications to labor and other markets, *American Economic Review* **77**, 875–890.

Fichman, R. and Kemerer, C.: 1999, The illusory diffusion of innovation: An examination of assimilation gaps, *Information Systems Research* **10**(3), 255–275.

Fisher, W.: 2004, *Promises to Keep: Technology, Law, and the Future of Entertainment*, Stanford University Press.

Hu, Y.: 2004, Performance-based pricing models in online advertising, *Working Paper, Sloan School of Management, MIT.* .

Jain, D. C., Mahajan, V. and Muller, E.: 1995, An approach for determining optimal product sampling for the diffusion of a new product, *The Journal of Product Innovation Management* **12**, 124–135.

Kremer, M.: 1998, Patent buyouts: A mechanism for encouraging innovation, *Quarterly Journal of Economics* **113**(4), 1137–1167.

Laffont, J.-J. and Tirole, J.: 1988, Dynamics of incentive contracts, *Econometrica* **55**(5), 1153–1175.

Lessig, L.: 2004, *Free Culture: How Big Media Uses Technology and the Law to Lock Down Creativity*, Penguin Press.

Lyman, P. and Varian, H. R.: 2004, How much information?, School of Information Management and Systems, University of California Berkeley. Retrieved from http://www.sims.berkeley.edu/research/projects/how-much-info-2003/ on April 27, 2004.

McAfee, R., McMillan, J. and Whinston, M.: 1989, Multiproduct monopoly, commodity bundling, and correlation of values., *Quarterly Journal of Economics* **104**, 371–383.

Myerson, R. and Satterthwaite, M.: 1983, Efficient mechanisms for bilateral trading, *Journal of Economic Theory* **28**, 265–281.

Sorensen, C. and Snis, U.: 2001, Innovation through knowledge codification, *Journal of Information Technology* **16**, 83–97.

Appendix: Proofs

Proof of Proposition 1:

We prove proposition 1 in two steps. First, we show that for each price level, the mechanism offers a consistent estimate of the true demand at that level. Second, we show given enough price levels, the demand curve can be arbitrarily closely approximated.

For one particular component in the bundle, the seller first chooses the number n of coupon levels, then, for each coupon level, sends m coupons to m randomly chosen consumers. For a coupon with face value $\tilde{v}$ for the component, the respondent will take it only if he has a valuation lower than $\tilde{v}$. The probability of the coupon getting accepted is $Prob(V \leq \tilde{v}) = F_V(\tilde{v})$. We now define indicator variables $Y_1, ..., Y_m$ where Y_i is 1 if the coupon with face value $\tilde{v}$ is accepted by the ith consumer, and Y_i is 0 if otherwise. We have $Y_k = \begin{cases} 1 & if X_k \leq \tilde{v} \\ 0 & if X_k > \tilde{v} \end{cases}$, where $k = 1, ..., m$. Note that $Prob(Y_k = 1) = Prob(X \leq \tilde{v}) = F_V(\tilde{v})$, and $Prob(Y_k = 0) = Prob(X > \tilde{v}) = 1 - F_V(\tilde{v})$. For all the m people to whom we sent coupon $\tilde{v}$, we know the number of acceptance is $a_m = \sum_{j=1}^{m} Y_j$. Define $\hat{F}_V(\tilde{v}) = \dfrac{a_m}{m}$ as the empirical cdf at $\tilde{v}$, which gives the result of the experiments telling us what percentage of people accepts the coupon $\tilde{v}$. We can show the expected value of the empirical cdf is the true unknown cdf.

$$E[\hat{F}_v(\tilde{v})] = E[\frac{a_m}{m}] = \frac{E[a_m]}{m} = \frac{m \cdot E[Y]}{m} = E[Y] = 0 \cdot Prob(Y = 0) + 1 \cdot Prob(Y = 1) = F_V(\tilde{v}).$$ That completes the step 1.

Next consider the interval between any neighboring coupon's value levels. For explanatory purpose, we now assume that the seller sets equi-distance intervals on the value range $[0,1]$, that is, the coupon values are $0, \dfrac{1}{n}, ... \dfrac{n-1}{n}$. Our result does not rely on this assumption, it holds as long as the distances are all weakly shrinking when adding more coupon levels.
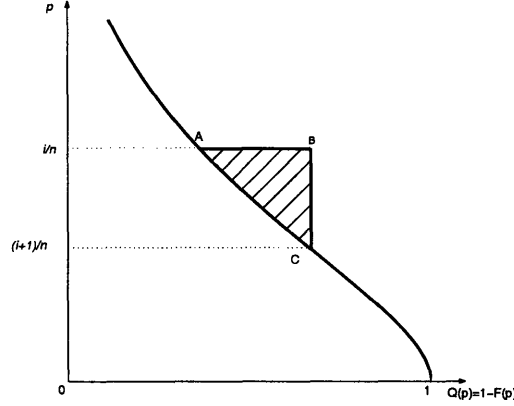
Figure A1. The upper bound of error in estimating demand

For neighboring coupon levels $\frac{i}{n}$ and $\frac{i+1}{n}$, the seller may estimate points A and C from step 1. She can simply connect the estimated points to approximate the demand curve between the two points. Since the demand curve is monotonically decreasing from 1 to 0, when estimating the area below the demand curve, the triangle ABC is the upper bound for the error. The area of ABC is $\Delta ABC = \frac{1}{2}(\frac{i+1}{n} - \frac{i}{n})(\hat{F}(\frac{i}{n}) - \hat{F}(\frac{i+1}{n}))$. We know $\hat{F}(\frac{i}{n}) - \hat{F}(\frac{i+1}{n}) \leq 1$, and given the assumption that $F_V(x)$ is continuously differentiable. We have $lim_{n \to \infty}(\hat{F}(\frac{i}{n}) - \hat{F}(\frac{i+1}{n})) = 0$, so we have $lim_{n \to \infty}\Delta ABC = \frac{1}{2}\left(lim_{n \to \infty}\frac{1}{n}\right) \cdot \left(lim_{n \to \infty}(\hat{F}(\frac{i}{n}) - \hat{F}(\frac{i+1}{n}))\right) = 0$, which says that when n is large enough, the error in estimation will converge to 0.

Q.E.D

PROOF OF PROPOSITION 2:

In Figure 1, the demand curve is moved upward by δ, we need to show that the area between the two demand curves is larger than the area $CDEFGH$. We first show that the new optimal price can not be out of the range $(p^*, p^* + \delta)$. Suppose, for contradiction, that charging a price $\tilde{p} > p^* + \delta$ gives a higher profit than charging p^*(or equivalently $p^* + \delta$, due to Lemma 3), then mapping this back to the original demand curve tells us that charging a little bit higher than p^* can give us a higher profit, which can not be true since p^* is the optimal price in the original demand curve. Using the same argument, we can show that the new optimal price can not be lower than p^*. So we have, for the new optimal price, $p' = p^* + \varepsilon \in (p^*, p^* + \delta)$, or equivalently, $0 < \varepsilon < \delta$.

Next, we only need to show that the increased profit from the traditional price mechanism is lower than what the couponing mechanism can provide.

The area between the demand curve is

$$A'B'BA = \int_0^\infty [1 - F(p - \delta)]dp - \int_0^\infty [1 - F(p)]dp = \delta + o(\delta)$$

where $o(\delta)$ is defined as $\text{Lim}_{\delta \to 0} \frac{o(\delta)}{\delta} = 0$. The area $CDEFGH$ can be calculated as

$$
\begin{aligned}
CDEFGH &= (p^* + \varepsilon)[1 - F(p^* + \varepsilon - \delta)] - p^*[1 - F(p^*)] \\
&= p'[1 - F(p' - \delta)] - p^*[1 - F(p^*)] \\
&= (p' - p^*)[1 - F(p^*)] + p^*[F(p^*) - F(p' - \delta)] + (p' - p^*)[F(p^*) - F(p' - \delta)] \\
&= \varepsilon[1 - F(p^*)] + (p^* + \varepsilon)[F(p^*) - F(p^* + \varepsilon - \delta)] \\
&= \varepsilon - \varepsilon F(p^*) + p^* F(p^*) + \varepsilon F(p^*) - p^* F(p^* + \varepsilon - \delta) - \varepsilon F(p^* + \varepsilon - \delta) \\
&= p^*[F(p^*) - F(p^* + \varepsilon - \delta)] + \varepsilon[1 - F(p^* + \varepsilon - \delta)] \\
&= p^* I + \varepsilon J
\end{aligned}
$$

where $I \equiv F(p^*) - F(p^* + \varepsilon - \delta)$, and $J \equiv 1 - F(p^* + \varepsilon - \delta)$.

Since $0 < \varepsilon < \delta$, we have $\text{Lim}_{\delta \to 0} \varepsilon = 0$, $\text{Lim}_{\delta \to 0} \frac{\varepsilon^2}{\delta} = 0$ and $\text{Lim}_{\delta \to 0}(\delta - \varepsilon) = 0$, so

$$\text{Lim}_{\delta \to 0} \frac{I}{\delta - \varepsilon} = \text{Lim}_{\delta \to 0} \frac{F(p^*) - F(p^* + \varepsilon - \delta)}{\delta - \varepsilon} = f(p^*).$$

Since $p' = p^* + \varepsilon$ is the new optimal price, it must satisfy the optimal condition given in equation (2), so we must have:

$$\frac{p' f(p')}{1 - F(p' - \delta)} = 1$$

where the term $F(p' - \delta)$ corresponds to the shifted demand curve.

42

Substituting for J, we have $J = 1 - F(p^* + \varepsilon - \delta) = 1 - F(p' - \delta) = p'f(p') = (p^* + \varepsilon)f(p^* + \varepsilon - \delta)$. By continuity, we also know that $\text{Lim}_{\delta \to 0} f(p^* + \varepsilon - \delta) = f(p^*)$,so we can write

$$
\begin{aligned}
\lim_{\delta \to 0} \frac{CDEFGH}{\delta} &= \lim_{\delta \to 0} \frac{p^* I + \varepsilon J}{\delta} \\
&= \lim_{\delta \to 0} \frac{p^*[F(p^*) - F(p^* + \varepsilon - \delta)] + \varepsilon[1 - F(p^* + \varepsilon - \delta)]}{\delta} \\
&= \lim_{\delta \to 0} \frac{p^*(\delta - \varepsilon)f(p^*) + \varepsilon(p^* + \varepsilon)f(p^*)}{\delta} \\
&= \lim_{\delta \to 0} \frac{p^* \delta f(p^*) + \varepsilon^2 f(p^*)}{\delta} \\
&= p^* f(p^*)
\end{aligned}
$$

It is now obvious that $\text{Lim}_{\delta \to 0} \dfrac{CDEFGH}{\delta} = p^* f(p^*) = 1 - F(p^*) < 1 = \dfrac{A'B'BA}{\delta} = \text{Lim}_{\delta \to 0} \dfrac{A'B'BA}{\delta}$, this completes our proof that the area $CDEFGH$ is smaller than the area between the two demand curves.

LEMMA 1. Given any nonnegative random variable Y with finite mean(i.e. a random variable for which $F_Y(y) = 0$ for $y < 0$, and $E[Y] < \infty$), $\lim_{y \to \infty} yP(Y \geq y) = 0$.

Proof of Lemma 1: First note that, $yP(Y \geq y) = y(F_Y(\infty) - F_y(y)) = y \int_y^\infty dF_Y(z) \leq \int_y^\infty z \, dF_Y(z)$, the last inequality is due to the fact that y is a lower bound for all z when z goes from y to infinity.

Next, from the definition of mean, we know $E[Y] = \int_0^\infty z \, dF_Y(z) = \int_0^y z \, dF_Y(z) + \int_y^\infty z \, dF_Y(z)$. Taking the limit, we have

$$
\lim_{y \to \infty} \int_y^\infty z \, dF_Y(z) = \lim_{y \to \infty}(E[Y] - \int_0^y z \, dF_Y(z)) = E[Y] - E[Y] = 0 \tag{7}
$$

So we have $\lim_{y \to \infty} yP(Y \geq y) \leq \lim_{y \to \infty} \int_y^\infty z \, dF_Y(z) = 0$.

Q.E.D

The above lemma enables us to write the expected value of a random variable in a very enlightening way.

LEMMA 2. For a nonnegative random variable X, $E[X] < \infty$, the expectation can be written in the following form:

$$E[X] = \int_0^\infty [1 - F_X(x)]\,dx \qquad (8)$$

Proof of Lemma 2: By definition,

$$E[X] = \int_0^\infty x f_x(x)\,dx = \int_0^\infty x\,dF_x(x)$$

we now define $G_X(x) = 1 - F_X(x)$, then we have $E[X] = \int_0^\infty x\,d(1 - G_X(x)) = -\int_0^\infty x\,dG_x(x)$.

Using integration by parts:

$$
\begin{aligned}
-\int_0^\infty x\,dG_X(x) &= -x G_X(x)\big|_0^\infty + \int_0^\infty G_X(x)\,dx \\
&= -(lim_{x\to\infty} x G_X(x) - 0 \cdot G_X(0)) + \int_0^\infty G_X(x)\,dx \\
&= \int_0^\infty G_x(x)\,dx = \int_0^\infty [1 - F_X(x)]\,dx
\end{aligned}
$$

Q.E.D

Proof of Lemma 3: We need to compare $\pi_p = (p^* + \delta)[1 - F(p^*)]$ and $\pi_q = p^*[1 - F(p^* - \delta)]$, and show that as $\delta \to 0$, they are equal.

Equivalently we need to show: $\lim_{\delta\to 0}(p^* + \delta)[1 - F(p^*)] = \lim_{\delta\to 0} p^*[1 - F(p^* - \delta)]$, which is

$\lim_{\delta\to 0} \frac{F(p^*) - F(p^* - \delta)}{\delta} = \frac{1 - F(p^*)}{p^*} \Leftrightarrow f(p^*) = \frac{1 - F(p^*)}{p^*}$, which is true due to equation (2), the optimality condition.

Q.E.D

Finding Edgeworth Cycles in Online Advertising Auctions

Doctoral Dissertation Essay 2

Xiaoquan (Michael) Zhang

MIT Sloan School of Management

Cambridge, MA

July, 2005

Finding Edgeworth Cycles in Online Advertising Auctions

Abstract

Search engines make a profit by auctioning off advertisement positions through keyword auctions. I examine the strategies taken by the advertisers. A game theoretical model suggests that the equilibrium bids should follow a cyclical pattern— "escalating" phases interconnected by "collapsing" phases — similar to a pattern of "Edgeworth Cycles" that was suggested by Edgeworth (1925) in a different context. I empirically test the validity of the theory. With an empirical framework based on maximum likelihood estimation of latent Markov state switching, I show that Edgeworth price cycles exist in this market. I further examine, on the individual bidder level, how strategic these bidders are. My results suggest that some bidders in this market adjust their bids according to Edgeworth predictions, while others not. Finally, I discuss the important implications of finding such cycles.

1   Introduction

Online advertising spending has rapidly grown over the last few years, from 1999's $3.5 billion to 2005's $12.9 billion[1]. Among the different types of online advertising, sponsored search, where positions in search engine's result pages are sold to advertisers through auctions (thus also referred to as "keyword auctions"), has grown most significantly and is widely credited for the revitalization of the search engine business [2]. The revenue of the sponsored search market increased from a few million dollars in 2000 to more than $6 billions in 2005. Sponsored search marketing has become one of the fastest developing advertising media in history.

Different from the "banner ad" model that is based on a payment mechanism of CPM (Cost-Per-1000-iMpressions), the industry standard in these keyword auctions is a performance-based payment model[3], and different search engines adopt various auction mechanisms. For example, Yahoo!, the biggest online keyword auction broker, allocates positions to advertisers purely based on their bidding amount. Advertisers pay their own bid, as in a first price auction. (Yahoo! now also allows bidders to choose a "proxy bidding" option, where each advertiser pays only one cent above the next highest bidder's bid. This is very similar to a second price auction.) Google, the other industry leader, adopts a different mechanism, where the ad positions are determined by bidder's bids and click-through rates jointly.

Comparing to the traditional auctions studied in the literature, these keyword auctions have the following distinctive characteristics :

- These auctions are held in a dynamic environment in infinite time horizon. That is, there is no ending in the auctions, as long as there is a demand for showing advertisements for the keywords. Any bidder may join or exit the auction at any time, and both advertisers' bids and their rankings can be updated in real time.

---

[1] Source: Jupiter Research (Elliott and Scevak 2004), Internet Advertising Bureau, http://www.iab.net/resources/ad_revenue.asp

[2] http://www.economist.com/displaystory.cfm?story_id=1932434.

[3] Also referred to as CPC (Cost-Per-Click) or PPC (Pay-Per-Click). The search engine gets paid only if an advertisement generates a click-through (or referral).

- In traditional auctions, bids can only go in one direction (e.g. English: up; Dutch: down), but in keyword auctions, bidders can change the bid in any direction they feel like.

- In keyword auctions, the advertisers agree to pay the search engine *ex ante* for the leads (click-throughs). The actual payment is determined by their own bids, as well as other bidders' bids at the time of the actual click-throughs.

- There is an inexhaustible supply of items in keyword auctions, any bidder is guaranteed to have a position, thus there is no "loser" in this market.

- Whenever someone changes his bid, others' valuations to the affected positions will be changed. So bidders' valuations keep changing throughout the auctions.

In practice, it is not clear whether there is a straightforward strategy for advertisers to follow. In a survey conducted by Jupiter Research, it is found that "the overwhelming majority of marketers have misguided bidding strategies, if they have a strategy at all"(Stein 2004). Among the 538 surveyed search marketers, the most popular strategies are "Historical data" (29%), "Top Three Listing" (28%), and "None" (22%). Only 7% advertisers selected "percentage of the cost of the good", which sounds a more reasonable strategy in light of the auction theory. This is surprising, given the vast amount of advertisers competing in such auctions[4], and the associated high stakes [5]. The same survey also found that marketers without a strategy were less successful in their campaigns.

Then do most of the bidders really act irrationally in these auctions? What are the optimal bidding strategies for a rational advertiser? What kind of bidding price and ranking outcome are implied by the optimal strategies?

To model this phenomenon, one-shot games are powerless. First, in online advertising auctions, bidders move sequentially, it is very rare to observe two bidders submitting bids at the same time. Second, every bidder's decision at any time is dependent on other bidders' current bids, if some bidder modifies the bid, it is very likely to induce a chain reaction of changing bids; this is further

---

[4]A popular keyword like "flower" or "car" in Yahoo! attracts hundreds of advertisers to compete. (www.overture.com). Google now has more than 325,000 CPC advertisers.

[5]Some keywords, such as "teleconferencing" or "mesothelioma" can be as expensive as $50 - $90 per click.

complicated by entry. Third, mixed strategy argument in one-shot games does not apply here, suppose at some point of time, all bidders submitted bids according to a mixed strategy, in the very next second, there will be incentives for some of the bidders to change the bids to obtain higher profit[6].

In Zhang and Feng (2005)'s model, bidders are allowed to update their bids sequentially. Different from a classical sequential model, where each player's strategy in a certain period is based on the whole history up to that period, their model incorporates a Markovian setting, where each bidder only reacts to the state variables that are directly payoff-relevant. In that model, bidders learn their competitors' valuations after a sufficiently long period of time and adopt Markov Perfect Nash Equilibrium (MPE) bidding strategy. They find that such a bidding strategy does not produce a stable outcome in prices and ranks. Instead, an "escalating" phase is observed in which each bidder increases his bid by only the smallest increment above his competitor's to reach a better position. A "collapsing" phase follows when one bidder can no longer afford the costly price war and drops his bid to a lower level, then the other bidder drops his bid accordingly and a new round of price war begins.

The cyclical pattern of bidding prices described above is similar to the Edgeworth Cycle in the literature modeling duopolistic price competition, where two duopolists with identical marginal cost undercut each other's price in an alternating manner. In Edgeworth (1925), the theory is a disequilibrium theory in which expectations are inconsistent or irrational. Maskin and Tirole (1988a, 1988b) relaxed the capacity constraint assumption and studied equilibrium outcomes of dynamic price competition.

In this paper, I will test the claims of the theory established in Zhang and Feng (2005), and show empirical evidence of the existence of the cycles.

In the literature, despite the long history of the discussion of dynamic updating of prices in response to competitor's price change (for example, Brock and Scheinkman (1985); Davidson and Deneckere (1986);Benoit and Krishna (1987), etc.) , there has been little empirical evidence

---

[6]Something unique in the online keyword auctions is that payoffs are not realized immediately after the submission of bids. Take the game "matching pennies" as an example, after the two players take actions, at least one of them will want to change the action, in our context, he can change the bid in less than a second. When he does so, his competitor will find it worthwhile to change the bid, too. This results in the conundrum of disequilibrium.

to support the existence of Edgeworth Cycles. In experimental settings, Isaac and Smith (1985), and Kruse, Rassenti, Reynolds and Smith (1994) reported evidence of Edgeworth price patterns in posted offer duopoly experiments. In a recent paper, Choi and Messinger (2005) find that most players in a series of experiments do not randomize from period to period, and under their symmetric treatment condition, the Edgeworth cycle hypothesis fits the data best.

Edgeworth cycles are rarely observed in the real world. So far, they have been reported to exist only in the Canadian retail gasoline market (Eckert 2002, Eckert and West 2004, Noel 2003, Noel 2004).

There are a few papers studying how search engines should auction off the advertisement positions. Feng, Bhargava and Pennock (forthcoming) examine the performances of various ranking mechanisms used by different search engines using computational simulations, and finds that depending on the correlation between advertisers' willingness to pay and their relevance (quality), both Google's and Overture's ranking mechanisms can out-perform each other under specific conditions. Weber and Zheng (2005) develop a two-stage model to search for a suitable mechanism for search engines to sell referrals to firms, they found that it is generally not optimal for search engines to base the ranking of advertisers purely on bid amounts. In the appendix of their paper, they show empirical evidence of firms' "unorthodox" bidding behavior in mixed strategies. They used the word unorthodox to describe the inconsistencies between the bidders' correlated bids and their theoretical prediction of independent bids of mixed strategies. My paper zooms in to study the causes and effects of the correlation between bids and the associated cyclical pattern.

The paper proceeds as follows. In section 2, I will introduce the setting of the model and explain the theory to be tested. In section 3, I describe the dataset to be used in the study that contains the records of bids submitted by the advertisers. I develop an empirical framework characterizing the price cycles in section 4 based on maximum likelihood estimation of latent Markov state switching, and report the results in section 5. In section 6, I discuss individual bidder-level strategies and show that MPE may not be played by all players. Finally, section 7 concludes.

2 The Model

Consider n risk-neutral advertisers competing for n positions in a search engine's result page. Denote θ_i as advertiser i's per-click valuation associated with the search engine — this is simply the expected profit he can get from each click or the maximum amount he is willing to pay to receive one click referred by the search engine. θ_i can be regarded as player i's type. In reality, θ_i is determined by two factors: the unit profit of the good, and the conversion rate. The unit profit of the good is a function of the unit price, unit cost and frequency of repeat purchase. The conversion rate is determined by the design of the website, convenience of payment and the attractiveness of the product. Both factors are independent from the advertisement's rank in position.

We can assume each position has a positive expected click-through-rate (CTR) $\tau_j > 0$, that is solely determined by the position, j, of the advertisement. More specifically, the higher the position of an advertisement, the more valuable it is, thus $\tau_{j'} \geq \tau_j \ \forall j' \leq j$[7]. An advertiser i's per-click expected revenue from winning position j can be represented by $\theta_i \tau_j$.

On Yahoo!, advertisers' rankings are determined purely by the rankings of their bids b_i, and each bidder pays the amount of his own bid. The auction is held periodically and can last arbitrarily long. In keyword auctions at Internet search engines, current information technologies not only enable bidders to observe their competitors' bidding prices (or infer their prices from some quick experiments by varying their own bids), but also allow them to update their bids in real time. For simplicity, I consider the competition between 2 bidders. To capture the dynamic feature of this game, I focus on a game that takes place in discrete time in infinitely many periods, with a discount factor $\delta \in (0,1)$. Following Maskin and Tirole (1988a), I assume bidders adjust their bids in an alternating manner. That is, in each period t, only one bidder is allowed to update his bid, and in the next period, only the other bidder is allowed to update. Thus each bidder commits to his bidding price for two periods. This modeling choice of alternating bidding behavior makes sense because in practice, a bidder has no incentive to update his bid if no other bidders update (if a bidder changes his bid in consecutive periods without other bidders updating, then at least

---

[7]It is a standard assumption in the literature (Breese, Heckerman, and Kadie 1998), and is frequently confirmed by industry reports.

one of the bids is not optimal).

Bidder i's current period payoff function $\pi(b_i, b_{-i})$ can be described as:

$$\pi(b_i, b_{-i}) = \begin{cases} (\theta_i - b_i)\tau_1 & \text{if } b_i > b_{-i} \\ (\theta_i - b_i)\tau_2 & \text{if } b_i \leq b_{-i} \end{cases}$$

then it can be shown that

Proposition 1 *For a sufficiently fine grid (ε is sufficiently small), there exist threshold values $\bar{b}_i$, $\underline{b}_{-i}$, and $\bar{b}_{-i}$, and probabilities $\sigma \in [0,1]$ and $\mu \in [0,1]$, such that bidder i's equilibrium strategy can be expressed as:*

$$b_i = R_i(b_{-i}) = \begin{cases} b_{-i} + \varepsilon & & \text{if } r \leq b_{-i} < \underline{b}_{-i} \\ \left. \begin{array}{ll} b_{-i} + \varepsilon & \text{with probability } \sigma(\delta) \\ \bar{b}_{-i} & \text{with probability } 1 - \sigma(\delta) \end{array} \right\} & \text{if } b_{-i} = \underline{b}_{-i} \\ \bar{b}_{-i} & & \text{if } \underline{b}_{-i} \leq b_{-i} < \bar{b}_{-i} < \bar{b}_i \quad (1) \\ \left. \begin{array}{ll} b_{-i} + \varepsilon & \text{with probability } \mu(\delta) \\ r & \text{with probability } 1 - \mu(\delta) \end{array} \right\} & \text{if } b_{-i} = \bar{b}_i \\ r & & \text{if } b_{-i} > \bar{b}_i \end{cases}$$

where r is the minimum required reserve price, and ε is the smallest increment.

Proof of this proposition is in Appendix C (see also Zhang and Feng (2005)). Proposition 1 describes bidders' equilibrium bidding strategy, and it is characterized by the process of bidders submitting small increments of bids until some threshold value is reached, and one of the lower-valued bidders will drop the bid back to the reserve price r.

This cyclical price pattern can be easily inferred from bidders' equilibrium strategy. Beginning with the smallest possible price r, the two players will wage a price war (outbidding each other by ε) until the price $min\{\underline{b}_1, \underline{b}_2\}$ is reached, (without loss of generality, suppose $\theta_1 \geq \theta_2$). Then bidder 1 will jump to bid $\bar{b}_2$, which is the highest bid that bidder 2 can afford to get rank 1. Now bidder 2 can no longer afford the costly competition, and will be better off dropping back to bid

r and obtaining the second position. Consequently bidder 1 should follow this drop to bid $r + \varepsilon$, where he remains at the first position and pays much less (than $\bar{b}_2$). Now bidder 2 has incentives to outbid bidder 1 again, then a new round of price war begins, and so on.

3 Data

To examine the interesting equilibrium outcome of keyword auctions, and to learn how strategic the bidders in this market are, I examine real world bidding patterns with a dataset obtained from Yahoo!. This dataset contains the complete bidding history of one keyword in the year 2001[8]. Each time someone submitted a new bid, the system would make a record on the bidder's ID, the date/time, and the bid value. A total of 1800 records have been included in the dataset. There were a total of 49 bidders who have submitted at least one bid during this period. Among whom, there were 7 bidders who submitted more than 100 bids. These most active bidders were mostly competing for the first position. Summary statistics are given in Table 1.

Variable	Mean	Std.Dev	Median	Minimum	Maximum
Bid Price ($)	0.4011	0.3440	0.39	0.05	4.97
Time in Market (days)	120.6712	137.2615	58.9979	0.7083	364.7438
Number of Bids Submitted by Firms(#)	36.6939	60.8749	9	1	245
Time Between Consecutive Bids (hours)	4.7732	8.6469	1.6333	1/60	88.2000
Time Between Consecutive Bids Submitted by the Same Firm (days)	2.3254	6.7256	0.7083	1/(60*24)	98.0806

Table 1: Summary Statistics

In this paper, I have a few empirical objectives. First, I would like to identify the pattern of cycles in this market as predicted by the theory. Edgeworth cycles are characterized by the unique interchanging phases of bids going up and down. To constitute an Edgeworth cycle, the speed of price going up and going down should be noticeably different. In the original Edgeworth cycle context, the price should go down slowly (undercutting phase), then suddenly go back to the

---

[8]The data was collected by Overture, which got acquired by Yahoo! in 2003.

top (relenting phase). In this paper, the Edgeworth cycle should have a flipped structure: the bids should go up little by little (escalating phase), then drop suddenly to the bottom (collapsing phase). Confirming the existence of the cycles has important theoretical implications. As a theoretical construct, Edgeworth Cycles have been discussed since the 1920s. Yet not until the late 1980s, has the rigorous equilibrium concept (Markov Perfect Equilibrium) been used (Maskin and Tirole 1988a, Maskin and Tirole 1988b) to describe it. Noel (2003) concludes that "I am not yet aware of similar Edgeworth-like cycles currently in other product markets." So finding empirical evidence of Edgeworth Cycles in the keyword auction market enriches the literature by giving one more empirical support to this concept. More importantly, this will help to explain why retail gasoline market is so unique to display cyclical price patterns. A second objective of my study is to characterize the cycles so that we can learn something about the behavior of the bidders. Given the existing auction mechanism, it is interesting to see how strategic the bidders are. This will also provide useful managerial insights for the practitioners (the advertisers and the search engines).
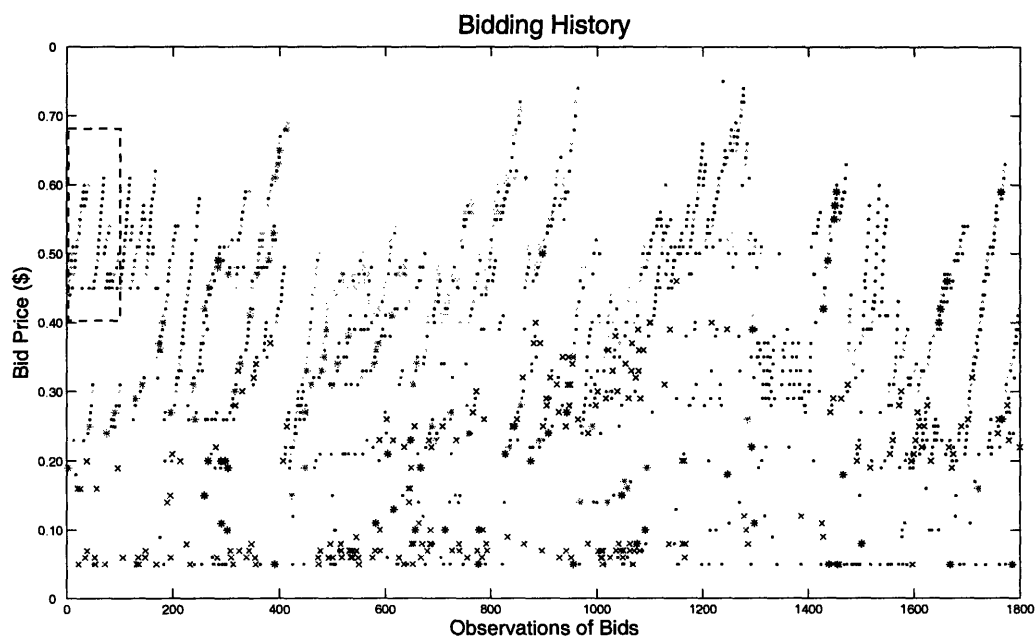

Figure 1: Bidding History

Figure 1 shows the complete bidding history. Each colored symbol represents a bidder, I am able to show the nine most active bidders in different symbols (e.g. blue dot, red dot, blue cross, red cross, purple star, etc.) and the rest bids are shown as black dots. In the figure, we can visually identify the price cycles going through the escalating phases and the collapsing phases.

From the raw data, I reconstruct the complete history of the rankings. In the process, I have to make a few reasonable assumptions. Since the dataset starts from Jan 1, 2001. I do not have information about how many bidders were competing for this keyword before my first observation. But given the fact that we are only interested in the responses of bidders to changes in others' bids, and that lower (higher) bids do not affect people's behavior competing for higher (lower) positions, I can safely take the first observed bid as the first bid entering the market[9]. A second data issue is that I do not observe who exits the market and when. The possible effect of ignoring exiting from the market is that I might mistakingly view a high bid as dominating the market for a long period before someone makes a higher bid. To alleviate this problem, I calculate an exit time for each bidder by "forcing" the bidder to quit the market one day after his last bid is observed[10]. I repeated my study with and without enforcing the exit time, and the results do not qualitatively differ.

Zooming in the dashed rectangle in Figure 1 on the facing page, we can get Figure 2 on the next page and have a closer look at the price cycles in the first 100 bids. Figure 3 on the following page gives the rank history of these 100 bids. Each instance of two lines crossing each other represents a change in ranking. Among the total of 1849 bids[11], about 1/3 of the bids change the highest rank, about 1/2 of the bids change the first two ranks.

To reject the claim that bidders were playing mixed strategies when they submitted the bids (Weber and Zheng 2005), I ran a few non-randomness tests. The rationale behind this is that a mixed-strategy argument would imply independent bids over time. A summary of these tests is given in Appendix B.

---

[9] If someone places the highest bid before the period (for example, placing $10 upfront, and never changes it within the sampling period), it does not affect my analysis, because I can still characterize the competition by observing people competing for the second position.

[10] The median time between a same bidder's consecutive bids is 17 hours. 70% of time, a bidder submits the next bid within one day.

[11] The original 1800 bids together with the 49 artificial exiting bids for the bidders.
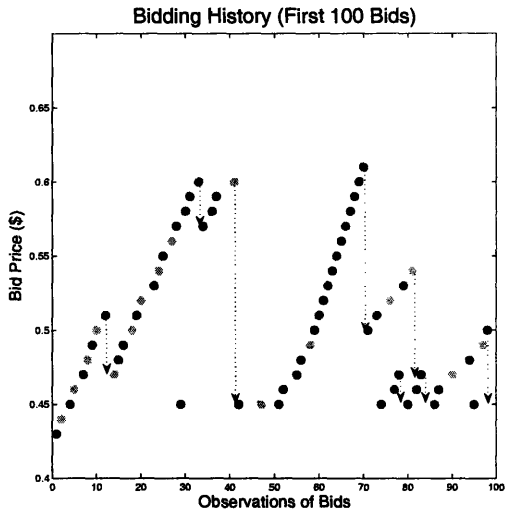
Figure 2: Bidding History Zoomed
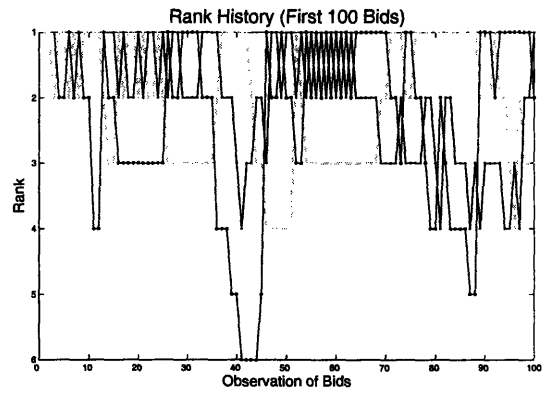
Figure 3: Rank History

Test	Parameter	μ	σ^2	Z	Pr
M-Test	$N = 530$	881.32	20.97	16.75	<0.0001
Sign Test	$P = 530$	422.5	70.42	12.81	<0.0001
Runs Test	$R = 881$	752.60	18.83	6.82	<0.0001
Mann-Kendall	$MK = 140592$	0	6.06×10^7	5.70	<0.0001

Table 2: Tests of Non-Randomness

The results of non-randomness tests are reported in Table 2. All tests can significantly reject the null hypothesis that the bids were drawn from a random sample.

In order to examine the level of autocorrelation, I applied Box-Jenkins (ARIMA) technique on the time series of the bids. Consistent with the non-randomness tests, the white noise test is strongly rejected with Chi-square value 686.35 (DF:6, Pr<.0001). The partial autocorrelation coefficient is 0.3778 for lag-1, but quickly diminishes as the order of lags becomes larger. The fast decay rate of autocorrelation function suggests that this series may be stationary, this result is confirmed by augmented Dickey-Fuller test and Phillips-Perron test, both tests reject the null hypothesis that there exists a unit root with a significance level of 0.0001. These results are consistent with the empirical observation of Weber and Zheng (2005), where they report a cross-

correlation coefficient of 0.65.

It is interesting to note that if I take a snapshot of the bids every two weeks and adopt the methodology of Iyer and Pazgal (2003), I would incorrectly reach the conclusion that these bids are the result of advertisers playing the game with mixed strategies. However, as suggested by the tests above and the analysis in the following sections, mixed strategy argument is not a plausible one to study this phenomenon.

4 A Markov Switching Model

In this section, I adopt the empirical strategy of Markov switching regression to examine the cycles. Markov switching regression was proposed by Goldfeld and Quandt (1973) to characterize changes in the parameters of an autoregressive process. From the shocking cyclical trajectory of the bids shown in figures in section 3, it is very tempting for me to assign one of two states (E - for escalating state; C - for collapsing state) for each observation of the bids and directly estimate the parameters with a discrete choice model. The use of the Markov switching regression, however, gives me a few advantages over other potential empirical strategies. First, serial correlation can be incorporated into the model, the parameters of an autoregression are viewed as the outcome of a two-state first-order Markov process. Second, when the price trajectory is not as regular as displayed by my data, Markov switching regression can help us to identify the latent states[12]. This eliminates the need to subjectively assign dummy variable values for the states, making it less dependent on human interference. Third, from the estimation process, I can easily derive the Markov transition matrix, the parameter estimates can be directly used to study the validity of the theory.

Formally, consider a two-state ergodic Markov chain shown in Figure 4, with state space $S(S_t) = \{e, c\}$, where $s_t = e$ represents the escalating phase, and $s_t = c$ represents the collapsing phase, and $t = 1, ..., T$. The process $\{S_t\}$ is a Markov chain with the stationary transition probability

---

[12]The smoothing technique I use in this paper was first proposed in Cosslett and Lee (1985). Hamilton (1989) further developed it. This technique is similar to Kalman filter which uses a time path to infer about an unobserved state variable. The nonlinear filter introduced in Hamilton (1989) draws inference about a discrete-valued unobserved state vector.
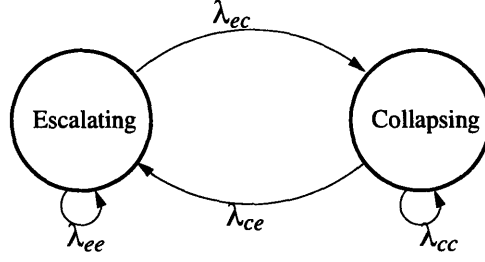
Figure 4: Markov Switching Probabilities

matrix $\Lambda = (\lambda_{ij})$, where

$$\lambda_{ij} = Pr(s_t = j | s_{t-1} = i), \qquad i, j \in \{e, c\} \tag{2}$$

This gives us a total of 4 transition probabilities: $\lambda_{ee}, \lambda_{ec}, \lambda_{cc}, \lambda_{ce}$, and we also have $\lambda_{ij} = 1 - \lambda_{ii}$, for $i \in \{e, c\}$ and $j \neq i$.

The ergodic probabilities for an ergodic chain is denoted π. This vector π is defined as the eigenvector of Λ associated with the unit eigenvalue; that is, the vector of ergodic probabilities π satisfies: $\Lambda \pi = \pi$. The eigenvector π is normalized so that its elements sum to unity: $1'\pi = 1$. For the two-state Markov chain studied here, we can easily derive,

$$\pi = \begin{bmatrix} \lambda_e \\ \lambda_c \end{bmatrix} = \begin{bmatrix} (1 - \lambda_{cc})/(2 - \lambda_{ee} - \lambda_{cc}) \\ (1 - \lambda_{ee})/(2 - \lambda_{ee} - \lambda_{cc}) \end{bmatrix} \tag{3}$$

After defining the latent states, we can write the following model:

$$y_t = \begin{cases} \alpha_e + \mathbf{x}_{et}\beta_e + \varepsilon_{et} & \text{if } s_t = e \\ \alpha_c + \mathbf{x}_{ct}\beta_c + \varepsilon_{ct} & \text{if } s_t = c \end{cases}, \qquad t = 1, ..., T, \tag{4}$$

where y_t is the bid submitted at time t, and $\mathbf{x}_{s_t}$ ($s_t \in \{e, c\}$) is a vector of explanatory variables. The error terms ε_{s_t} are assumed to be independent of $\mathbf{x}_{s_t}$. Following the standard approach, I assume $\varepsilon_{et} \sim N(0, \sigma_e^2)$, and $\varepsilon_{ct} \sim N(0, \sigma_c^2)$. Notice that for each period, the regime variable (Markov state s_t) is unobservable.

To save space, I relegate the detailed derivation of the maximum likelihood estimation procedure to Appendix A.

If the transition probabilities are restricted only by the conditions that $p_{ij} \geq 0$ and $\sum_{j=1}^{N} p_{ij} = 1$, for all i and j, Hamilton (1990) has shown that the maximum likelihood estimates for the transition probabilities satisfy

$$\hat{\lambda}_{ij} = \frac{\sum_{t=2}^{T} Prob(s_t = j, s_{t-1} = i | \mathbf{Y}_T; \hat{\theta})}{\sum_{t=2}^{T} Prob(s_{t-1} = i | \mathbf{Y}_T; \hat{\theta})}, \tag{5}$$

where $\mathbf{Y}_T$ is a vector containing all observations obtained through date t, $\hat{\theta}$ denotes the full vector of maximum likelihood estimates. The meaning of equation (5) is that the estimated transition probability $\hat{\lambda}_{ij}$ is the number of times state i have been followed by state j divided by the number of times the process was in state i. I estimate the transition matrix based on the smoothed probabilities.

Equipped with the estimates of the transition probabilities, I can derive very intuitive parameters to characterize the cycles. The expected duration of a typical price war (or cease fire) can be calculated with[13]:

$$E(\text{duration of phase } i) = \sum_{k=1}^{\infty} k \lambda_{i,i}^{k-1} (1 - \lambda_{i,i}) = \frac{1}{1 - \lambda_{i,i}} \tag{6}$$

where $\lambda_{i,i}, i \in \{e, c\}$ are the transition probabilities.

Thus the expected duration of a cycle is

$$E(\text{duration of a cycle}) = \sum_{s_t \in \{e,c\}} \frac{1}{1 + \lambda_{s_t, s_t}} = \frac{1}{1 + \lambda_{ee}} + \frac{1}{1 + \lambda_{cc}}.$$

I rely on an EM, or expectation-maximization algorithm to carry out the maximum likelihood estimation of parameters. The EM algorithm was originally developed in Dempster, Laird and Rubin (1977). It proceeds by iteratively maximizing the current conditional expectation of the log likelihood function, where parameters are replaced by their expectation prior to maximization. Given estimates from the maximization step (known as the M-step), the new expected values

---

[13]For details, please refer to Gallager, Chapter 4 (1996).

are computed and substituted in the likelihood function (called the E-step). An advantage of the EM algorithm is that each step is guaranteed to increase the likelihood function value. The EM algorithm is criticized for its slow convergence rate. Indeed, this algorithm takes a few orders of magnitude more time than a simple model based on a dummy variable to distinguish the regimes, but the previously stated advantages of using it justifies the extra time.

5   Estimation and Results

Following model (4), I consider the following specification:

$$y_{i,t} = \begin{cases} \beta_{e0} + \beta_{e1} y_{i,t-1} + \beta_{e2} r_{i,t} + \beta_{e3} r_{i,t-1} + \varepsilon_{et} & \text{if } s_t = e \\ \beta_{c0} + \beta_{c1} y_{i,t-1} + \beta_{c2} r_{i,t} + \beta_{c3} r_{i,t-1} + \varepsilon_{ct} & \text{if } s_t = c \end{cases} \tag{7}$$

where $y_{i,t}$ is the bid submitted by bidder i at time t, and $r_{i,t}$ is the achieved rank[14] after submission at time t. The model estimates two sets of parameters corresponding to the two underlying states, and I estimate the underlying time sequence of states with a Markov chain. I use maximum likelihood estimation to determine the parameters and the states simultaneously as detailed in the previous section. We need to be careful with the notation here because $y_{i,t-1}$ refers to the bid submitted by bidder i before time t, so $y_{i,t-1}$ does not necessarily refer to the bid submitted at time $t - 1$. Alternatively, $y_{i,t-1}$ can be thought of as the "current price" of bidder i at time $t - 1$. $r_{i,t-1}$ refers to the rank of bidder i at time $t - 1$. To illustrate, suppose there are only two bidders in the market, and at time $t = 1$, bidder 1 submits a bid of \$0.40 to reach rank one, then bidder 2 came to outbid him with price \$0.41. Now, at time $t = 3$, suppose bidder 1 bids \$0.42, we have $y_{1,3} = 0.42$, and $y_{1,3-1} = 0.40$, which was submitted in time $t = 1$. We also have $r_{1,3} = 1$, and $r_{1,3-1} = 2$. So, each time period, the current bidder makes a decision about the bid to be submitted based on the information of his last bid, his previous rank and the rank he wants to achieve after the bid.

The estimation procedure detailed in the last section will help us to calculate the latent state for

---

[14]Note that a "high" rank has a "low" numerical value — the highest rank is associated with a rank $r = 1$.

each of the bids, along with the estimated probabilities, I can also get parameter estimates for

$\beta_e \equiv (\beta_{e0}, \beta_{e1}, \beta_{e2}, \beta_{e3})$ and $\beta_c \equiv (\beta_{c0}, \beta_{c1}, \beta_{c2}, \beta_{c3})$.

The Markov switching regression based on Hamilton's algorithm turns out to be highly significant with the estimates reported in Table 3.

	Regime 1 (escalating)	Regime 2 (collapsing)
$\hat{\beta}_0$	0.16037	0.42510
(Constant)	(7.16)	(29.15)
$\hat{\beta}_1$	0.81296	-0.15285
(Previous Bid)	(43.17)	(-4.97)
$\hat{\beta}_2$	-0.03223	-0.01823
(New Rank)	(-53.32)	(-12.51)
$\hat{\beta}_3$	0.02775	-0.02238
(Previous Rank)	(30.24)	(-17.91)

$$R^2 = 0.91$$
t-statistic reported in parentheses, all parameters significant at .01 level

Table 3: Markov Switching Estimates

It is estimated that in regime 1 (escalating phase), β_1 is positive, so the bid will be increased in each step. The higher the previous rank a bidder has, the less likely he will be bidding higher, and the higher the new rank is, the higher he will be bidding. In regime 2 (collapsing phase), the sign of β_1 is negative, suggesting that a bidder would want to reduce the bid when the price is getting higher. The higher the previous rank a bidder has, the more likely he will be bidding higher (to keep the current rank[15]), and the higher the new rank is, the higher he will be bidding (after all, each bidder is seeking for the best position to introduce the collapsing phase).

So far, I have found evidence supporting the existence of the cycles, and the parameter estimates clearly show the distinctive properties of the two regimes. My next objective is to characterize the cycles by applying the results of Markov Chain theory.

From equation (13), I am able to use the smoothed probabilities to obtain the latent state of each bid. By examining the change of states, I can calculate the transition matrix as shown in Table 4.

---

[15]These findings help me to construct the rule of predicting Edgeworth prices in the next section.

λ_{ij}	e	c
e	0.9540	0.0460
c	0.7889	0.2111

Table 4: Transition Matrix

From equation (3), I can also obtain the limiting unconditional probabilities:

$$\pi = \begin{bmatrix} \lambda_e \\ \lambda_c \end{bmatrix} = \begin{bmatrix} (1-\lambda_{cc})/(2-\lambda_{ee}-\lambda_{cc}) \\ (1-\lambda_{ee})/(2-\lambda_{ee}-\lambda_{cc}) \end{bmatrix} = \begin{bmatrix} 0.945 \\ 0.055 \end{bmatrix}$$

So in the long run, about 94.5% states are in the escalating phase, and about 5.5% states are in the collapsing phase.

By (6), I also know

$$E(\text{duration of an escalating phase}) = \frac{1}{1-\lambda_{ee}} = 21.74,$$

and

$$E(\text{duration of a collapsing phase}) = \frac{1}{1-\lambda_{cc}} = 1.27,$$

which says that a typical escalating phase lasts about 22 periods, and a typical collapsing phase lasts about one period. Therefore, we can infer that a typical cycle lasts about 23 periods.

This is a very strong piece of evidence for the existence of Edgeworth cycles, in the escalating phase, the bids go up little by little, but in the collapsing phase, the bids drop significantly in one period.

For the recorded bids, the distribution of duration between consecutive bids is heavily skewed to the right. If I use the median of 93 minutes as the duration of the periods[16], one cycle would typically last 2139 minutes, or roughly 1.5 days.

Table 5 gives further evidence about the bidding behavior in the two phases. Using the smoothed probabilities calculated from the Markov switching regression, I confirm that the behavior is

---

[16]The average is 325 minutes per bid, which is heavily biased by a few outliers.

$\Delta y_{i,t} \equiv y_{i,t} - y_{i,t-1}$	mean	stdev	median
Escalating Phase	0.06	0.061	0.02
Collapsing Phase	-0.28	0.129	-0.21

Table 5: Summary of change in bids in the two phases

consistent with the story of two phases. In the escalating phase, a typical bidder bids two cents higher than his previous bid (thus, one cent above his competitor's), and in the collapsing phase, a typical bidder reduces the bid by about 20 cents down from his last bid.

These results give strong support to the Edgeworth Cycle argument: on average, each cycle lasts 23 periods with 21.74 escalating periods and 1.27 collapsing periods. In each escalating period, a bidder outbids the competitor by only one cent; but in a collapsing period, a bidder drops the bid by more than 20 cents.

6 Individual Level Strategies

The results presented in section 5 suggest that Edgeworth Cycles are the dominant pattern in this market. These results give the impression that all bidders in the market are perfectly rational and adopt the MPE strategy. It is very hard, however, to believe that all bidders in this market are rational and submit correct bids all the time. For example, in Figure 1 on page 56, the bidders represented by the blue dots or the red dots are obviously competing according to the MPE strategy, but bidder represented by the red cross submitted bids within a narrow band between $0.05 and $0.10. There is no sign of his bids getting triggered by others' bids. This red-cross bidder, at least, was not a bidder following the optimal strategy suggested by the theory.

I now turn to examine how well the prediction of MPE strategy stack up against the data for individual bidders.

One difficulty we are facing to do an individual level analysis is that we can not observe each individual's private value for each of the advertisement positions. Without this piece of information, the theoretical model can not give predictions about the timing for regime switching. To address this issue, I now derive the information from the data. One way to do this is to take a

market player's perspective and examine when other bidders dropped their bids at certain prices, those prices can serve as a proxy for their maximum willingness to pay. The problem with this approach is that for any bidder, he has different values for different ranks, and over time, his value might be changing. So I adopt an alternative approach in this section: instead of trying to predict something unobservable, I take the decision of the next state as given, and focus on examining the decision of choosing the next bid. In other words, from the dataset, for each bidder, we can observe the decision of increasing or decreasing the current bid, and we can use this information to calculate the next bid according to the strategy suggested by MPE. This approach is used in the experimental economics literature for studying Edgeworth Cycles (Kruse, Rassenti, Reynolds, and Smith 1994, Choi and Messinger 2005); of course, in an experimental setting, the researcher knows the private values. The model below examines how well the predicted strategy fits with the data for each of the bidders. Presumably, it will fit better for those bidders who follow the MPE strategy more closely.

I have the following specification:

$$y_{i,t} - y_{i,t-1} = \beta_0 + \beta_1(y_{i,t}^E - y_{i,t-1}) + \beta_2(y_{i,t-1}^E - y_{i,t-2}) + \varepsilon_{i,t} \tag{8}$$

where $y_{i,t}$ is the bid submitted by bidder i at period t, $y_{i,t-1}$ is the bid amount of bidder i at period $t-1$. On the right hand side, $y_{i,t}^E$ is the predicted Edgeworth price for bidder i in period t, based on the rivals' prices in period $t-1$. $y_{i,t-1}^E$ and $y_{i,t-1}$ are defined similarly. I assume $\varepsilon_{i,t} \sim N(0, \sigma^2)$.

There are many ways to obtain the predicted Edgeworth price $y_{i,t}^E$ for bidder i in period t. The simplest method adopts the following rule:

1. If the next bid is increasing (going into an escalating phase):

    Check the rank r of bidder i for period $t-1$, find the bidder whose rank is $r-1$ in the rank table;

    Obtain the bid $y_{r-1,t-1}$ of the bidder with rank $r-1$ in period $t-1$;

    Set the $y_{i,t-1}^E = y_{r-1,t-1} + 0.01$.

2. If the next bid is decreasing (going into a collapsing phase):

Check the rank r of bidder i for period $t - 1$, find the bidder whose rank is $r + 1$ in the rank table;

Obtain the bid $y_{r+1,t-1}$ of the bidder with rank $r + 1$ in period $t - 1$;

Set the $y^E_{i,t-1} = y_{r+1,t-1} + 0.01$.

A more complicated method calculates predictions based on "generalized Edgeworth cycle" patterns:

Find out the rank r achieved by the next bid (note that in the auctions, this rank is determined by the bidder before he submits the bid):

Find the bidder whose rank is $r + 1$ in the rank table;

Obtain the bid $y_{r+1,t-1}$ of the bidder with rank $r + 1$ in period $t - 1$;

Set the $y^E_{i,t-1} = y_{r+1,t-1} + 0.01$.

Yet another way to calculate predictions is to utilize the smoothed state probabilities calculated in Section 5. But since the implicit assumption of using smoothed state probabilities is that all bids follow the Edgeworth cycle pattern, the result is less meaningful to study individual level strategies. In this paper, for simplicity, I report the results derived from the first method; the second method yields qualitatively similar results.

Equation (8) captures the extent to which actual prices adjusted to predicted Edgeworth prices. The form makes it possible to incorporate both complete and immediate adjustment and partial or zero adjustment to the predicted Edgeworth price.

To ensure analysis validity, I only include those bidders who submitted at least 10 bids; among the 49 bidders who competed for this keyword, 21 bidders submitted more than 10 bids. The least active bidder submitted 12 bids, and the most active bidder submitted 245 bids.

For 11 (or about 1/2) bidders, the regression was significant at the $p < 0.05$ level, suggesting that about half of bidders were using the MPE strategy (I will call these bidders the strategic ones henceforth). On average, the bidders adopting MPE strategy submitted more bids (mean:116.54) than the other bidders (mean:42.1) during the sampling period[17]. Interestingly, the mean bid sub-

---

[17]A bidder can benefit greatly by responding quickly to the competitors' bids, since he can enjoy to stay in a state he prefers for a longer period of time.

mitted by the strategic bidders is 0.3663, and it is significantly less than the mean bid submitted by the non-strategic bidders whose mean bid is 0.5422 — these firms being more strategic does not necessarily mean that they have higher valuations! One possible explanation is that if a firm has a much higher value than the competitors, it will not engage in the price wars[18]. Another explanation can be that firms with higher per-click valuation also have higher time value, thus unable to modify the bids as often. Yet another explanation may be that these less strategic firms tend to bid higher than the level that is best for them, thus unnecessarily paying more than what they should. From the search engine's point of view, however, these less strategic bidders are very valuable to extract more profit, not only because they tend to pay more, but also because they force the strategic bidders to submit unnecessarily higher bids.

For model (8), I run four tests to study the adjustment to predicted prices for each of the bidders, and I summarize the findings below in Table 6.

Null Hypothesis	Reject Null*
(1) $\beta_1 = \beta_2 = 0$	11/21
(2) $\beta_0 = 0$; $\beta_1 = 1$; $\beta_2 = 0$	14/21
(3) $\beta_1 = 0$	15/21
(4) $\beta_2 = 0$	4/21
Coefficient Estimates When All Subjects Are Pooled**:	
β_0	0.00495 (0.00506)
β_1	0.28116 (0.02753)
β_2	-0.11024 (0.02838)

$R^2 = 0.07$; 1680 observations
*Reject Null: Number of subjects whose behavior allows rejection of the null hypothesis at the 0.05 level
**Standard errors are in the parentheses.

Table 6: Analysis of Adjustment to Predicted Edgeworth Prices

In the first test, the null hypothesis is that both parameter estimates β_1 and β_2 are zero. Since β_0 has non-significant estimates, this is the same F test for the significance of the model; therefore we identify the same 11 firms that are more strategic. The null hypothesis of the second test is that the actual price adjusts completely ($\beta_1 = 1$) and immediately ($\beta_2 = 0$) to the predicted Edgeworth

---

[18]In Zhang and Feng (2005), they give specific conditions under which a bidder may not engage in competition in Edgeworth Cycles.

price. The null hypothesis is rejected for 14 firms, suggesting that a considerable number of firms do not adjust the price completely and immediately. In the third test, I test the null hypothesis that there is completely no adjustment, and we reject this hypothesis for 15 firms, suggesting that for most firms, there is a certain level of adjustment in response to the predicted adjustment to Edgeworth prices. In the last test, the null hypothesis is that when a firm adjusts the price, it does so immediately, and we can reject this null hypothesis for four firms. So the one-period lagged predicted Edgeworth adjustment does not have a significant effect on current price adjustment for most bidders.

I also report the estimation results for model (8) with all records pooled together in Table 6. Interestingly, these results are highly consistent with those reported in Kruse et al. (1994), and Choi and Messinger (2005)[19], even though both previous papers are experimental and studied different kinds of dynamic pricing games. The estimated coefficients of the pooled model suggest that in the aggregate level, firms partially adjust the prices according to the Edgeworth predictions.

7   Conclusion

Selling advertisements with keyword auctions is one of the most successful online business models and has attracted a lot of discussion. Much of the existing literature has been focusing on the strategy of the search engines. Little attention, however, has been given to the advertisers who rely critically on the market and aggregately contribute to the success of this business model. This paper empirically studies the dynamic bidding behavior in online keyword auctions, and the evidence in this paper suggests that advertisers in this market adopted different strategies when they submitted the bids to these auctions. Overall, the group of strategic players dominated the market, and we can identify price cycles and characterize them with "escalating" phases and "collapsing" phases.

My empirical analysis supports that the price competition in the keyword auction market is con-

---

[19]Kruse et al. (1994) reports $\beta_1 = 0.291$ with standard error 0.010, and $\beta_2 = -0.174$ with standard error 0.010. The model $R^2 = 0.158$ with 4616 observations. Choi and Messinger (2005) reports $\beta_1 = 0.234$ with standard error 0.010, and $R^2 = 0.131$.

sistent with the MPE prediction. Individual level analysis of bidders suggests that there is hetero-geneity across the bidders in terms of level of sophistication in choosing a strategy.

In this paper, I identify the Edgeworth price cycles pattern in an "unexpected" market, and by verifying the theoretical predictions, I find evidence to support the theoretical construct of the pattern.

This finding also has important practical implications. The advertising industry has always favored big companies. Super Bowl, the American National Football League's championship game, charges 2.4 million dollars for a 30-second spot. Consequently, the biggest sponsors are usually Automotive and Motion Picture industries. In 2005's final game, they each won 330 seconds of commercial time. Beer and Soft Drinks followed with 180 seconds. However, the U.S. economy is by no means dominated by giant corporations. Fully 99 percent of all independent enterprises employ fewer than 500 people. According to the U.S. Small Business Administration (SBA), these small enterprises account for half of the GDP and hire 52 percent of all U.S. work-ers; they are also a major source of innovation. Small enterprises not only face the difficulty of advertising budget, the variety of products and services they offer also makes it hard to choose suitable advertising media. Sponsored search solves both problems elegantly: on one hand, the cost of sponsored search is considerably cheaper than other means of marketing[20], and on the other hand, the relevance of the advertisements in sponsored search is by definition very high. The inception and development of the sponsored search marketing created huge opportunities for small enterprises to find a venue to compete with big players and the possibility to develop into larger enterprises themselves. The associated value of innovation incentives and product/service variety are practically immeasurable. This study gives a rigorous analysis of the optimal bidding strategies for the advertisers, providing the bidders a concrete strategy to follow. The framework used in this study will also prove to be useful for search engines to have a better understanding of the bidders and improve the design of advertising mechanisms.

This work leaves a number of questions unexplored. First, I only looked at the bidding history of one keyword, it would be interesting to compare across keywords to see how different keywords can have different cyclical patterns. Second, it is interesting to examine whether other equilibrium

---

[20]On average, it only costs $0.50 for a firm to get a click-through from the search engines.

outcomes can be observed in this market, again, data from more keywords are needed. Third, since Yahoo! enabled the feature of "proxy bidding", many advertisers have started to use this feature, some others also started using third-party tools to implement more complicated bidding strategies, the impact of these new features should be further examined.

References

Benoit, J., and V. Krishna, 1987, "Dynamic duopoly: Prices and quantities,," *Review of Economic Studies*, 54, 23–46.

Breese, J. S., D. Heckerman, and C. Kadie, 1998, "Empirical analysis of predictive algorithms for collaborative filtering," in *Proceedings of the Fourteenth Conference on Uncertainty in Artificial Intelligence*, pp. 43–52.

Brock, W., and J. Scheinkman, 1985, "Price setting supergames with capacity constraints," *Review of Economic Studies*, 52, 371–382.

Choi, S., and P. R. Messinger, 2005, "Edgeworth promotional cycles in the loyal/switcher game," *Working Paper*.

Cosslett, S. R., and L.-F. Lee, 1985, "Serial Correlation in Discrete Variable Models," *Journal of Econometrics*, 27, 79–97.

Davidson, C., and R. Deneckere, 1986, "Long-run competition in capacity, short-run competition in price and the cournot model," *Rank Journal of Economics*, 17, 404–415.

Dempster, A. P., N. M. Laird, and D. B. Rubin, 1977, "Maximum Likelihood from Incomplete Data via the EM Algorithm," *Journal of the Royal Statistical Society*, B39(1), 1–38.

Eckert, A., 2002, "Retail price cycles and response asymmetry," *Canadian Journal of Economics*, 35(1), 52–77.

Eckert, A., and D. S. West, 2004, "Retail Gasoline Price Cycles across Spatially Dispersed Gasoline Stations," *The Journal of Law and Economics*, 47, 245–273.

Edgeworth, F., 1925, "The pure theory of monopoly," *Papers Relating to Political Economy*, 1, 111–142.

Elliott, N., and N. Scevak, 2004, "Online advertising through 2009," Online Advertising Vol. 3, Jupiter Research.

Feng, J., H. K. Bhargava, and D. Pennock, forthcoming, "Implementing Sponsored Search in Web Search Engines: Computational Evaluation of Alternative Mechanisms," *Informs Journal of Computing*.

Gallager, R., 1996, *Discrete Stochastic Processes*. Kluwer Academic Publishers, Norwell MA.

Goldfeld, S. M., and R. E. Quandt, 1973, "A Markov Model for Switching Regressions," *Journal of Econometrics*, 1, 3–16.

Hamilton, J., 1990, "Analysis of time series subject to changes in regime," *Journal of Econometrics*, 45(1-2), 39–70.

Hamilton, J. D., 1989, "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57(2), 357–384.

Isaac, R. M., and V. L. Smith, 1985, "In search of predatory pricing," *Journal of Political Economy*, 93, 320–345.

Iyer, G., and A. Pazgal, 2003, "Internet Shopping Agents: Virtual Co-Location and Competition," *Marketing Science*, 22(1), 85–106.

Kruse, J., S. Rassenti, S. Reynolds, and V. Smith, 1994, "Betrand-Edgeworth competition in experimental markets," *Econometrica*, 62(2), 343–371.

Maskin, E., and J. Tirole, 1988a, "A Theory of Dynamic Oligopoly, I: Overview and Quantity Competition with Large Fixed Costs," *Econometrica*, 56(3), 549–569.

————— , 1988b, "A Theory of Dynamic Oligopoly, II: Price Competition, Kinked Demand Curves, and Edgeworth Cycles," *Econometrica*, 56(3), 571–599.

Noel, M., 2003, "Edgeworth Price Cycles: Evidence from the Toronto Retail Gasoline Market," Working Paper, University of California at San Diego.

————— , 2004, "Edgeworth Price Cycles, Cost-based Pricing and Sticky Pricing in Retail Gasoline Markets," Working Paper, University of California at San Diego.

Stein, G., 2004, "The Search Game," Paid Search Vol. 6, Jupiter Research.

Weber, T., and Z. Zheng, 2005, "A Model of Search Intermediaries and Paid Referrals," *Working Paper, Stanford University and University of California at Riverside*.

Zhang, X. M., and J. Feng, 2005, "Price cycles in online advertising auctions," *Proceedings of the 26th International Conference on Information Systems (ICIS), Dec. Las Vegas, NV.*

Appendix A: Maximum Likelihood Estimation Procedure

Following Hamilton (1989), in the spirit of Markov process, I confine my analysis here to the cases where the density function of y_t depends only on finitely many past values of s_t:

$$f(y_t|s_t, S_{t-1}, Y_{t-1}) = f(y_t|s_t, s_{t-1}, ..., s_{t-k}, Y_{t-1}) \tag{9}$$

where k is a finite integer, and $S_t = \{s_t, s_{t-1}, ..., s_0\}$ denotes the complete history of the states till time t. The corresponding conditional likelihood is $Prob(s_t, s_{t-1}, ..., s_{t-k}|Y_{t-1})$.

From (2), we also have

$$Prob(s_t|S_{t-1}, Y_{t-1}) = Prob(s_t|s_{t-1}) = \lambda_{s_{t-1}, s_t}.$$

The conditional likelihood can be iteratively calculated by the following equations:

$$
\begin{aligned}
&Prob(s_t, ..., s_{t-k}|Y_{t-1}) \\
&= \sum_{s_{t-k-1}\in\{w,s\}} Prob(s_t|s_{t-1}, ..., s_{t-k-1}, Y_{t-1}) \cdot Prob(s_{t-1}, ..., s_{t-k-1}|Y_{t-1}) \\
&= \sum_{s_{t-k-1}\in\{w,s\}} \lambda_{s_{t-1}, s_t} \cdot Prob(s_{t-1}, ..., s_{t-k-1}|Y_{t-1}) \\
&= \begin{cases} \lambda_{s_{t-1}, s_t} \cdot Prob(s_{t-1}, ..., s_{t-k}|Y_{t-1}), & k > 0 \\ \sum_{s_{t-1}\in\{w,s\}} \lambda_{s_{t-1}, s_t} \cdot Prob(s_{t-1}|Y_{t-1}), & k = 0 \end{cases}
\end{aligned}
\tag{10}
$$

So given the conditional likelihood for any period $t - 1$, we can use equation (10) to calculate the conditional likelihood for period t.

The term on the right hand side can be computed by

$$
\begin{aligned}
&Prob(s_t, ..., s_{t-k}|Y_t) \\
&= \frac{f(y_t|s_t, ..., s_{t-k}, Y_{t-1}) \cdot Prob(s_t, ..., s_{t-k}|Y_{t-1})}{\sum_{s_t\in\{w,s\}} \cdots \sum_{s_{t-k}\in\{w,s\}} f(y_t|s_t, ..., s_{t-k}, Y_{t-1}) \cdot Prob(s_t, ..., s_{t-k}|Y_{t-1})} \\
&= \frac{f(y_t|s_t, ..., s_{t-k}, Y_{t-1}) \cdot Prob(s_t, ..., s_{t-k}|Y_{t-1})}{f(y_t|Y_{t-1})}
\end{aligned}
\tag{11}
$$

Given initial value $Prob(s_1, s_0, s_{-1}, ..., s_{-(r-1)}|Y_o)$, we can calculate $Prob(s_t, ..., s_{t-k}|Y_{t-1})$ iteratively by (10) and (11). Hamilton (1989) does not give a detailed discussion about how to decide the initial value, and finding the initial value is indeed quite challenging computationally. Instead of using Hamilton's subiteration technique, I adopt an easier approach. The trick is to use Markov property again to assume that

$$Prob(s_{-j}|s_{-j-1}, s_{-j-2}, ..., Y_0) = Prob(s_{-j}|s_{-j-1}) \equiv \lambda_{s_{-j-1}, s_{-j}}, \text{ for } j = 0, 1, 2, ...$$

So the initial condition is simplified to

$$Prob(s_1, s_o, ..., s_{-(r-1)}|Y_0) = \lambda_{s_0, s_1} \cdot \lambda_{s_{-1}, s_0} ... \lambda_{s_{-(k-1)}, s_{-(k-2)}} \cdot Prob(s_{-(k-1)}|Y_0) \qquad (12)$$

where, I only need to estimate the 2 values for the $Prob(s_{-(k-1)}|Y_0)$ term. Since they are only initial values for an iterative algorithm, there are many ways to pick them. For example, I can assume that they are two arbitrary constants that sum up to 1. Another possible choice, and the one selected for this study, is to use equation (3) to derive the stationary probabilities λ_e and λ_c, and set $Prob(s_{s_t}|Y_0) = \lambda_{s_t}$ for $s_t \in \{e, c\}$.

So far, the density function of y_t has been simplified to

$$f(y_t|Y_{t-1}) = \sum_{s_t \in \{e,c\}} ... \sum_{s_{t-r} \in \{e,c\}} f(y_t|s_t, ..., s_{t-k}, Y_{t-1}) \cdot Prob(s_t, ..., s_{t-k}|Y_{t-1})$$

and the likelihood function for the ML estimation is

$$\prod_{t-1}^{T} f(y_t|Y_{t-1}).$$

The filtering probability $Prob(s_t|Y_t)$ is derived from the iterative process by

$$Prob(s_t|Y_t) = \sum_{s_{t-1} \in \{e,c\}} ... \sum_{s_{t-k} \in \{e,c\}} Prob(s_t, s_{t-1}, ..., s_{t-k}|Y_t).$$

Based on the filtering probabilities, I can infer about the complete history of the past states

75

(smoothing). For an arbitrary $j=1,...,r$, I can calculate

$$Prob(s_{t-j}|Y_t) = \sum_{s_t\in\{e,c\}} ... \sum_{s_{t-j+1}\in\{e,c\}} \sum_{s_{t-j-1}\in\{e,c\}} ... \sum_{s_{t-k}\in\{e,c\}} Prob(s_t,s_{t-1},...,s_{t-k}|Y_t). \quad (13)$$

The set of smoothed probabilities I use in this study is calculated from equation (13) as

$$Prob(s_t|Y_T) = \sum_{s_{t-1}\in\{e,c\}} ... \sum_{s_{t-k}\in\{e,c\}} Prob(s_t,s_{t-1},...,s_{t-k}|Y_T)$$

where $Prob(s_t,s_{t-1},...,s_{t-k}|Y_T) = Prob(s_t,s_{t-1},...,s_{t-k}|Y_t) \prod_{j=t+1}^{T} \dfrac{f(y_j|s_t,...,s_{t-k},Y_{j-1})}{f(y_j|Y_{j-1})}$. These smoothed

probabilities identify the regime each y_t belongs to. I infer that y_t is in state e (escalating) if $Prob(s_t = e|Y_T) > 0.5$, and similarly for state c.[21]

Appendix B: Tests of Non-Randomness

I ran a few tests to confirm the visually striking non-randomness. They are all similar in that they study the structure of runs of submitted bids. All tests have the null hypothesis that the bids are independently, identically distributed random quantities.

M-Test:

This test exploits results of geometric distribution. I first calculate the median, M, of the sample of n observations. Then assign values to the observations according to the following rule: $\tilde{b}_i =$
$\begin{cases} 1 & if\ b_i > M \\ 0 & otherwise \end{cases}$. Let N be the number of runs of consecutive zeros and ones. The mean and

variance of N among n observations are: $\mu_N = \frac{2m(n-m)}{n} + 1$; $\sigma_N^2 = \frac{2m(n-m)[2m(n-m)-n]}{n^2(n-1)}$, where m

is the number of occurrences of zeros. The statistic $Z_M = \frac{N-\mu_N}{\sqrt{\sigma_N^2}}$ should be standard normal under

null hypothesis.

---

[21]This is just an arbitrary decision rule. Based on the dataset, most of the states are determined with smoothed probabilities near 0.9.

Sign Test (P-Test):

This test uses results of binomial distribution. Take first differencing of the bids to get $\Delta b_i = b_i - b_{i-1}$, $\forall i \geq 2$, discard zeros. Suppose we are left with N non-zero values and P positive values, then the mean and variance of number of positive values is: $\mu_P = \frac{N}{2}$; $\sigma_P^2 = \frac{N}{12}$. The statistic $Z_P = \frac{P - \mu_P}{\sqrt{\sigma_P^2}}$ should be standard normal under null hypothesis.

Wald-Wolfowitz Test (Runs Test)

This is similar to both the M-Test and the Sign Test, I take first differencing of the bids to get $\Delta b_i = b_i - b_{i-1}$, $\forall i \geq 2$, and discard zeros. Let R be the number of runs of consecutive positive numbers or negative numbers, n be the total number of observations, and m be the number of occurrences of positive numbers. We can calculate: $\mu_R = \frac{2m(n-m)}{n} + 1$; $\sigma_R^2 = \frac{2m(n-m)[2m(n-m)-n]}{n^2(n-1)}$, and the test statistic $Z_R = \frac{R - \mu_R}{\sqrt{\sigma_R^2}}$ should be standard normal under null hypothesis.

Mann-Kendall Test

Define $MK = \sum_{i=2}^{n} \sum_{j=1}^{i-1} sign(b_i - b_j)$, where $sign(a) = \begin{cases} 1 & if\ a > 0 \\ 0 & if\ a = 0 \\ -1 & if\ a < 0 \end{cases}$, then we have $\mu_{MK} = 0$;

$\sigma_{MK}^2 = \frac{n(n-1)(2n+5)}{18}$, and under the null hypothesis, the test statistic $Z_{MK} = \frac{MK - \mu_{MK}}{\sqrt{\sigma_{MK}^2}} \sim Normal(0,1)$.

Appendix C: Proof of Proposition 1 (Zhang & Feng, 2005)

Proof of Proposition 1.

- First we show for what conditions these parameters $(\underline{b}_i, \overline{b}_i, \sigma, \mu)$ must satisfy to constitute an MPE.

  Without loss of generality, consider $\theta_1 \geq \theta_2$. Let $V_i(b_{-i})$ denote bidder i's valuation if (1) he is about to move; (2) the other firm just bid b_{-i}. Define $V_1(r) = \overline{V}_1$, and $V_2(r + \varepsilon) = \overline{V}_2$. There are two cases:

77

1. Case 1: $\underline{b}_2 - r = (2t+1)\varepsilon$.

$$\begin{aligned}
\overline{V}_1 = \ & V_1(r) = (\theta_1 - r - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_1 - r - 3\varepsilon)(\tau_1 + \delta\tau_2) + \dots \\
& + \delta^{2t}(\theta_1 - \underline{b}_2)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_1 - \overline{b}_2)(1 + \delta)\tau_1 + \delta^{2t+4}\overline{V}_1
\end{aligned} \tag{14}$$

$$\begin{aligned}
\overline{V}_2 = \ & V_2(r + \varepsilon) = (\theta_2 - r - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_2 - r - 4\varepsilon)(\tau_1 + \delta\tau_2) + \dots \\
& + \delta^{2t}(\theta_2 - \underline{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_2 - r)(1 + \delta)\tau_2 + \delta^{2t+4}\overline{V}_2
\end{aligned} \tag{15}$$

2. Case 2: $\underline{b}_2 - r = 2t\varepsilon$.

$$\begin{aligned}
\overline{V}_1 = \ & V_1(r) = (\theta_1 - r - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_1 - r - 3\varepsilon)(\tau_1 + \delta\tau_2) + \dots \\
& + \delta^{2t-2}(\theta_1 - \underline{b}_2 + \varepsilon)(\tau_1 + \delta\tau_2) + \delta^{2t}(\theta_1 - \overline{b}_2)(1 + \delta)\tau_1 + \delta^{2t+2}\overline{V}_1
\end{aligned} \tag{16}$$

$$\begin{aligned}
\overline{V}_2 = \ & V_2(r + \varepsilon) = (\theta_2 - r - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_2 - r - 4\varepsilon)(\tau_1 + \delta\tau_2) + \dots \\
& + \delta^{2t-2}(\theta_2 - \underline{b}_2)(\tau_1 + \delta\tau_2) + \delta^{2t}(\theta_2 - r)(1 + \delta)\tau_2 + \delta^{2t+2}\overline{V}_2
\end{aligned} \tag{17}$$

Now given $\overline{V}_1$ and $\overline{V}_2$, how is $\overline{b}_2$ determined? $\overline{b}_2$ is the upper bound bid that bidder 2 will submit for position 1. Given $\theta_1 \geq \theta_2$ and bidder 1 will bid ε more than bidder 2 when the current price is smaller than $\overline{b}_1$, if the current price is $\overline{b}_2 - \varepsilon$, bidder 2 should be better off outbidding bidder 1 by ε than dropping back to r:

$$\begin{aligned}
& (\theta_2 - \overline{b}_2)(\tau_1 + \delta\tau_2) + \delta^2(\theta_2 - r)(1 + \delta)\tau_2 + \delta^4\overline{V}_2 \\
> \ & (\theta_2 - r)(1 + \delta)\tau_2 + \delta^2\overline{V}_2
\end{aligned} \tag{18}$$

If the current price is $\overline{b}_2$, bidder 2 should be weakly better off by dropping back to r than outbidding bidder 1 by ε. Thus,

$$\begin{aligned}
& (\theta_2 - r)(1 + \delta)\tau_2 + \delta^2\overline{V}_2 \\
\geq \ & (\theta_2 - \overline{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_2 - r)(1 + \delta)\tau_2 + \delta^4\overline{V}_2
\end{aligned} \tag{19}$$

Thus $\bar{b}_2$ should satisfy both Eq. (18) and Eq.(19). More specifically, $\bar{b}_2$ is determined by Eq.(19) with an equal sign. It is obvious that $\bar{b}_2 < \theta_2$. This makes sense because bidder 2 won't bid more than θ_2 to win the first position and gain zero expected profit, since by dropping to bid r, he can receive the second position and make positive profit. This can be shown by reorganizing Eq.(19) as $(\theta_2 - \bar{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) = (1 - \delta^2)(\theta_2 - r)(1 + \delta)\tau_2 + \delta^2\bar{V}_2 > 0$, thus $\theta_2 - \bar{b}_2 - \varepsilon > 0$. When bidder 2 feel indifferent between overbidding bidder 1 by ε at $\bar{b}_2$ and dropping to r, he can play mixed strategy at $\bar{b}_2$. Suppose he plays $\bar{b}_2 + \varepsilon$ with probability μ and r with probability $1 - \mu$. To be consistent with the strategy specified in proposition 1, bidder 1 should feel weakly better (indifferent when $\bar{b}_1 = \bar{b}_2$) by over bidding bidder 1 by ε than dropping back to bid r, thus

$$\mu[\delta(\theta_1 - \bar{b}_2 - 2\varepsilon)(\tau_1 + \delta\tau_1) + \delta^3\bar{V}_1] + (1 - \mu)[\delta\bar{V}_1] \geq \delta(\theta_1 - r)(1 + \delta)\tau_2 + \delta^3 V_1(r + \varepsilon) \quad (20)$$

where when $\underline{b}_2 - r = (2t + 1)\varepsilon$,

$$
\begin{aligned}
V_1(r + \varepsilon) &= (\theta_1 - r - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_1 - r - 4\varepsilon)(\tau_1 + \delta\tau_2) + \dots \\
&\quad + \delta^{2t}(\theta_1 - \underline{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^{2t+4}\bar{V}_1
\end{aligned}
$$

It can be shown that $\bar{V}_1 > (\theta_1 - r)(1 + \delta)\tau_2 + \delta^2 V_1(r + \varepsilon)$ by construction (otherwise bidder 2 is better off always staying at the second position). μ is determined by Eq.(20): when θ_1 is sufficiently bigger than θ_2 such that $\bar{b}_1 \geq \bar{b}_2 + \varepsilon$, there exist at least one $\mu \in [0, 1]$ satisfying Eq.(20), since $(\theta_1 - \bar{b}_2 - 2\varepsilon)(\tau_1 + \delta\tau_1) + \delta^3\bar{V}_1 > (\theta_1 - r)(1 + \delta)\tau_2 + \delta^3 V_1(r + \varepsilon)$ by construction: bidder 1 is better off by overbidding bidder 2 by ε when the current price is smaller than $\bar{b}_1$; it is also straightforward to show that $\mu \in [0, 1]$ when $\theta_1 = \theta_2$ from Eq.(19), since $(\theta_1 - \bar{b}_2 - 2\varepsilon)(\tau_1 + \delta\tau_1) + \delta^3\bar{V}_1 < (\theta_1 - r)(1 + \delta)\tau_2 + \delta^3 V_1(r + \varepsilon)$ now. This means that when the two bidders have identical valuations, both of them playing mixed strategy when the current price is $\bar{b}_1 = \bar{b}_2$.

Then what condition should $\underline{b}_2$ satisfy? $\underline{b}_2$ is the threshold value, upon which bidder 1 prefers to jump to bid bidder 2's maximum bid to guarantee the first position and end the price war. So when $\theta_1 \geq \theta_2$, and when bidder 2 follows the strategy to overbid bidder 1

by ε when the current price is lower than $\bar{b}_2$ and $\underline{b}_1$, it should be the case that if the current price is $\underline{b}_2 - \varepsilon$, bidder 1 is better off by outbidding ε than jumping to $\bar{b}_2$, thus:

$$
\begin{aligned}
&(\theta_1 - \underline{b}_2)(\tau_1 + \delta\tau_2) + \delta^2(\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^4\bar{V}_1 \\
> \ &(\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^2\bar{V}_1
\end{aligned}
\tag{21}
$$

While if the current price is $\underline{b}_2$, bidder 1 is weakly better off by jumping to $\bar{b}_2$ in the current period, than outbidding ε and jumping in the next period:

$$
\begin{aligned}
&(\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^2\bar{V}_1 \\
\geq \ &(\theta_1 - \underline{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^4\bar{V}_1
\end{aligned}
\tag{22}
$$

Similarly, Eq.(22) with equality determines $\underline{b}_2$ given $\bar{b}_2$, δ and $\bar{V}_1$, since when Eq.(22) is satisfied, Eq.(21) is automatically satisfied. It can be shown that $\underline{b}_2 < \theta_1$, which makes sense because bidder 1 won't bid more than θ_1, let alone jump at a price greater than θ_1. This can be shown by reorganizing Eq.(22) as $(\theta_1 - \underline{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) = (1 - \delta^2)((\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^2\bar{V}_1) > 0$. It can also be shown that $\underline{b}_2$ is smaller than $\bar{b}_2$, otherwise Eq. (21) can not be satisfied, since $\bar{V}_1 \geq (\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^2\bar{V}_1$

For bidder 1 playing mixed strategy at $b_2 = \underline{b}_2$ (suppose he plays $b_2 + \varepsilon$ with probability σ and $\bar{b}_2$ with probability $1 - \sigma$), bidder 2 must be weakly better off (indifferent when $\theta_1 = \theta_2$) by following the strategy specified in proposition 1, than jumping up to bid $\bar{b}_1$ (indifferent when $\underline{b}_1 = \underline{b}_2$), thus:

$$
\begin{aligned}
&\sigma[\delta(\theta_2 - \underline{b}_2 - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^3(\theta_2 - r)(1 + \delta)\tau_2 + \delta^5\bar{V}_2] \\
+ \ &(1 - \sigma)[\delta(\theta_2 - r)(1 + \delta)\tau_2 + \delta^3\bar{V}_2] \\
\geq \ &\delta(\theta_2 - \bar{b}_1)(\tau_1 + \delta\tau_2) + \delta^3(\theta_2 - r)(1 + \delta)\tau_2 + \delta^5\bar{V}_2
\end{aligned}
\tag{23}
$$

When θ_1 is significantly bigger than θ_2 such that $\underline{b}_1 \geq \underline{b}_2 + \varepsilon$, it can be shown that $\delta(\theta_2 - \underline{b}_2 - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^3(\theta_2 - r)(1 + \delta)\tau_2 + \delta^5\bar{V}_2 > \delta(\theta_2 - \bar{b}_1)(\tau_1 + \delta\tau_2) + \delta^3(\theta_2 - r)(1 + \delta)\tau_2 + \delta^5\bar{V}_2$ by construction. It can also be shown that $\delta(\theta_2 - r)(1 + \delta)\tau_2 + \delta^3\bar{V}_2 > \delta(\theta_2 - \bar{b}_1)(\tau_1 + \delta\tau_2) + \delta^3(\theta_2 - r)(1 + \delta)\tau_2 + \delta^5\bar{V}_2$ because by construction, bidder 2 has

no opportunity to reach $\bar{b}_1$ when $\theta_1 > \theta_2$. Then there exist at least one $\sigma \in [0,1]$ when $\underline{b}_1 \geq \underline{b}_2 + \varepsilon$ such that Eq.(23) is satisfied. This implies that bidder 2 is better off by following the strategy specified in Proposition 1 than jumping up bidding. When $\theta_1 = \theta_2$, it is also straightforward to show that $\sigma \in [0,1]$ (because under this condition $\delta(\theta_2 - \underline{b}_2 - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^3(\theta_2 - r)(1+\delta)\tau_2 + \delta^5\bar{V}_2 < \delta(\theta_2 - \bar{b}_1)(\tau_1 + \delta\tau_2) + \delta^3(\theta_2 - r)(1+\delta)\tau_2 + \delta^5\bar{V}_2)$.

From Eq.(22),

$$\underline{b}_2 = \theta_1 - \varepsilon - \frac{(1-\delta^2)[(\theta_1 - \bar{b}_2)(1+\delta)\tau_1 + \delta^2\bar{V}_1(r)]}{\tau_1 + \delta\tau_2}$$

Note that in equilibrium $\underline{b}_2$ should be greater than the reserve price r, thus,

$$\underline{b}_2 = \min\left\{r, \theta_1 - \varepsilon - \frac{(1-\delta^2)[(\theta_1 - \bar{b}_2)(1+\delta)\tau_1 + \delta^2\bar{V}_1(r)]}{\tau_1 + \delta\tau_2}\right\} \tag{24}$$

Similarly, from Eq. (19),

$$\bar{b}_2 = \theta_2 - \varepsilon - \frac{(1-\delta^2)[(\theta_2 - r)(1+\delta)\tau_2 + \delta^2\bar{V}_2]}{\tau_1 + \delta\tau_2}$$

In the same way, $\bar{b}_2$ can be written as

$$\bar{b}_2 = \min\left\{r, \theta_2 - \varepsilon - \frac{(1-\delta^2)[(\theta_2 - r)(1+\delta)\tau_2 + \delta^2\bar{V}_2]}{\tau_1 + \delta\tau_2}\right\} \tag{25}$$

In equilibrium, jump up bidding exists only when $\underline{b}_2 < \bar{b}_2$. This condition is not always satisfied, however. For example, compare $A = (1-\delta^2)[(\theta_2 - r)(1+\delta)\tau_2 + \delta^2\bar{V}_2]$ and $B = (1-\delta^2)[(\theta_2 - r)(1+\delta)\tau_2 + \delta^2\bar{V}_2]$. By Eq. (19), $(1-\delta^2)[(\theta_2 - r)(1+\delta)\tau_2 + \delta^2\bar{V}_2] = (\theta_1 - \bar{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2)$, thus $B = [(\theta_1 - \bar{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2)]$. It can be easily seen that when δ is sufficiently close to 1, $A < B$, thus $\underline{b}_2 > \bar{b}_2$, which means that in equilibrium there is no jump bidding. This also explains why (Maskin and Tirole 1988b) does not consider jump bidding. On the other hand, $\underline{b}_2 < \bar{b}_2$ is possible when δ is sufficiently close to 0, especially when θ_1 is close to θ_2.

- Next we show (R_1, R_2) constitutes an MPE.

Without loss of generality, again assumes $\theta_1 \geq \theta_2$

1. Both advertisers will never bid lower than r, because r is the smallest allowed bid.

2. Advertiser i will never raise his bid to more than $\bar{b}_i + \varepsilon$. To show this, consider bidder 2, compare his payment when bidding $\bar{b}_2 + 2\varepsilon$ to that when he drops to bid r:

$$(\theta_2 - \bar{b}_2 - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_2 - r)(1 + \delta)\tau_2 + \delta^4 \bar{V}_2$$
$$< \quad (\theta_2 - r)(1 + \delta)\tau_2 + \delta^2 \bar{V}_2$$

This "$<$" comes from Eq.(19), since $\theta - \bar{b}_2 - 2\varepsilon < \theta - \bar{b}_2 - \varepsilon$.

3. Advertiser i will never bid between $[\underline{b}_{-i} + \varepsilon, \bar{b}_{-i} - \varepsilon]$. For the example of bidder 1, this can be seen from Eq.(22), and the fact that $\theta_1 - \underline{b}_2 - 2\varepsilon < \theta_1 - \underline{b}_2 - \varepsilon$.

4. For $r < b_{-i} < \underline{b}_{-i}$, no bidder is willing to deviate from bidding $b_{-i} + \varepsilon$ to $b_{-i} + 2\varepsilon$ (thus $b_{-i} + k\varepsilon$ for $k > 2$). To show this, let the current bid price be p, and there are two cases to consider:

    (a) Case 1: $\underline{b}_2 - p = (2t + 1)\varepsilon$.

$$V_1(p) = (\theta_1 - p - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_1 - p - 3\varepsilon)(\tau_1 + \delta\tau_2) + \dots$$
$$+\delta^{2t}(\theta_1 - \underline{b}_2)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^{2t+4}\bar{V}_1$$

$$V_2(p) = (\theta_2 - p - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_2 - p - 3\varepsilon)(\tau_1 + \delta\tau_2) + \dots$$
$$+\delta^{2t}(\theta_2 - \underline{b}_2)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_2 - r)(1 + \delta)\tau_2 + \delta^{2t+4}\bar{V}_2$$

If bidder 1 switches to use $b_i + 2\varepsilon$ for only one period, while bidder 2 keep the original strategy, thus

$$\tilde{V}_1(p) = (\theta_1 - p - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_1 - p - 4\varepsilon)(\tau_1 + \delta\tau_2) + \dots$$
$$+\delta^{2t}(\theta_1 - \underline{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_1 - \bar{b}_2)(1 + \delta)\tau_1 + \delta^{2t+4}\bar{V}_1$$
$$< \quad V_1(p)$$

The "$<$" is obvious because the first t terms in $V_1(p)$ is greater than those in

82

$\tilde{V}_1(p)$. The same works for bidder 2,

$$\begin{aligned}
\tilde{V}_2(p) &= (\theta_2 - p - 2\varepsilon)(\tau_1 + \delta\tau_2) + \delta^2(\theta_2 - p - 4\varepsilon)(\tau_1 + \delta\tau_2) + \dots \\
&\quad + \delta^{2t}(\theta_2 - \underline{b}_2 - \varepsilon)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_2 - r)(1 + \delta)\tau_2 + \delta^{2t+4}\overline{V}_2 \\
&< V_2(p)
\end{aligned}$$

The "$<$" is straightforward because the first t terms in $V_2(p)$ is greater than those in $\tilde{V}_2(p)$.

(b) Case 2: $\underline{b}_2 - r = 2t\varepsilon$.

This case can be worked out accordingly thus we omitted the details here.

Thus neither bidder has incentive to deviate to bid $b_{-i} + 2\varepsilon$.

5. For $r < b_{-i} < \underline{b}_{-i}$, no bidder is willing to deviate to bid $p < b_{-i} + \varepsilon$.

This follows the same logic as (4) thus omitted.

6. For $\underline{b}_{-i} < b_{-i} < \overline{b}_{-i}$, bidder i prefers to jumping to bid $\overline{b}_{-i}$.

This is straightforward from Eq.(21).

7. For $\overline{b}_{-i} < b_{-i} < \overline{b}_i$, bidder i always prefers jumping up the bid and outbidding the other bidder by ε.

This is also straightforward from Eq.(21) and Eq. (22), since $\overline{b}_{-i} < b_{-i} < \overline{b}_i$. Notice that for bidder 2, this is never reached on the equilibrium path.

By all of these claims, (R_1, R_2) constitutes an MPE for this game.

- Lastly we show the existence of the MPE.

Consider an arbitrary $\overline{V}_1$ and $\overline{V}_2 \in (0, \frac{(\theta_i - r - \varepsilon)\tau_1}{1-\delta})^{22}$. Then $\bar{b}_2(\overline{V}_2)$ is defined by Eq. (25). It can be shown that $\bar{b}_2(\overline{V}_2)$ is continuous in $\overline{V}_2$ and is between $[r, \theta_2]$.

Similarly, given $\overline{V}_1$, $\overline{V}_2$ and $\bar{b}_2(\overline{V}_2)$, $\underline{b}_2(\overline{V}_1)$ is defined by Eq. (24), which is continuous in $\overline{V}_1$, and is between $[r, \theta_1]$.

---

[22]From Eq. (14) to Eq.(17), it's straightforward to obtain that both $\overline{V}_1$ and $\overline{V}_2$ are in $[0, \frac{(\theta_i - r - \varepsilon)\tau_1}{1-\delta}]$.

Now consider $\overline{V}_1$ first. Define

$$U_1(p,\overline{V}_1) = (\theta_1 - p - \varepsilon)(\tau_1 + \delta\tau_2) + ... +$$

$$\begin{cases} \delta^{2t}(\theta_1 - \underline{b}_2)(\tau_1 + \delta\tau_2) + \delta^{2t+2}(\theta_1 - \overline{b}_2)(1+\delta)\tau_1 + \delta^{2t+4}\overline{V}_1 & \text{if } \underline{b}_2 - p = (2t+1)\varepsilon \\ \delta^{2t-2}(\theta_1 - \underline{b}_2 + \varepsilon)(\tau_1 + \delta\tau_2) + \delta^{2t}(\theta_1 - \overline{b}_2)(1+\delta)\tau_1 + \delta^{2t+2}\overline{V}_1 & \text{if } \underline{b}_2 - p = 2t\varepsilon \end{cases}$$

$$(26)$$

We must show that there exists a $\overline{V}_1$ such that $\overline{V}_1 = U_1(\underline{b}_2(\overline{V}_1), \overline{b}_2(\overline{V}_2), \overline{V}_1) = \tilde{U}_1(\overline{V}_1, \overline{V}_2)$.

It is easy to show that $U_1(\overline{V}_1, \overline{V}_2)$ is continuous in $\overline{V}_1$. It also can be shown that $U_1(\overline{V}_1, \overline{V}_2)$ is continuous in $\overline{V}_2$, since U_1 is continuous in $\overline{b}_2(\overline{V}_2)$, and $\overline{b}_2(\overline{V}_2)$ is continuous in $\overline{V}_2$. Thus to show that $\tilde{U}_1$ has a fixed point, we only need to show that $\tilde{U}_1$ maps $[0, \frac{(\theta_1 - r - \varepsilon)\tau_1}{1-\delta}]$ into $[0, \frac{(\theta_1 - r - \varepsilon)\tau_1}{1-\delta}]$. Obviously $\tilde{U}_1 \geq 0$ if $\overline{V}_1 \geq 0$. To show that $\tilde{U}_1 \leq \frac{(\theta_1 - r - \varepsilon)\tau_1}{1-\delta}$ if $\overline{V}_1 \leq \frac{(\theta_1 - r - \varepsilon)\tau_1}{1-\delta}$, note that from Eq.(26),

when $\underline{b}_2 - r = (2t+1)\varepsilon$, given $\overline{V}_1 \leq \frac{(\theta_1 - r - \varepsilon)}{(1-\delta)}$,

$$\begin{aligned} \tilde{U}_1(\overline{V}_1) &\leq (\theta_1 - r - \varepsilon)(1 + \delta + ... + \delta^{2t+3}) + \delta^{2t+4}\frac{(\theta_1 - r - \varepsilon)}{(1-\delta)} \\ &= \frac{(\theta_1 - r - \varepsilon)(1 - \delta^{2t+4})}{(1-\delta)} + \delta^{2t+4}\frac{(\theta_1 - r - \varepsilon)}{(1-\delta)} \\ &= \frac{(\theta_1 - r - \varepsilon)}{(1-\delta)} \end{aligned}$$

The same works for the case when $\underline{b}_2 - r = 2t\varepsilon$. It can be shown if define $\tilde{U}_2$ in the same way, that $\tilde{U}_2 \leq \frac{(\theta_i - r - \varepsilon)\tau_1}{1-\delta}$ if $\overline{V}_2 \leq \frac{(\theta_i - r - \varepsilon)\tau_1}{1-\delta}$.

$$\diamond$$

84

The Lord of the Ratings: Is A Movie's Fate Influenced by Professional and Amateur Reviews?

Doctoral Dissertation Essay 3

Xiaoquan (Michael) Zhang

MIT Sloan School of Management

Cambridge, MA

March, 2006

The Lord of the Ratings: Is A Movie's Fate Influenced by Professional and Amateur Reviews?

Abstract

Third-party reviews play an important role in many contexts in which tangible attributes are insufficient to enable consumers to evaluate products or services. In this paper, I examine the impact of professional and amateur reviews on the box office performance of movies. I first show evidence to suggest that the generally accepted result of "professional critics as predictors of movie performance" may no longer be true. Then, with a simple diffusion model, I establish an econometrics framework to control for the interaction between the unobservable quality of movies and the word-of-mouth diffusion process, and thereby estimate the residual impact of online amateur reviews on demand. The results indicate the significant influence of the valence measure (ratings) of online reviews, but their volume measure (propensity to write reviews) is not significant once I control for quality. Furthermore, the analysis suggests that the variance measure (disagreement) of reviews does not play a significant role in the early weeks after a movie's opening. The estimated influence of the valence measure implies that a one-point increase in the valence can be associated with a 4-10% increase in box office revenues.

Keywords: Diffusion Model; Motion Pictures Industry; Online Reviews; Third Party Reviews; Word-of-mouth

1 Introduction

In the summer of 2004, Steven Spielberg released the movie *The Terminal*, a film with a $75 million production budget and $35 million marketing budget and starring Tom Hank and Catherine Zeta-Jones. During its opening weekend, nearly 3000 theaters showed the movie. Despite these factors that would seem to guarantee a big hit, the movie grossed only $77 million, far less than any predictions had suggested. Notably, the opening gross of *The Terminal* was $19 million, comparable with Tom Hanks's other movies (*e.g., The Road to Perdition*, $22 million; *The Green Mile*, $18 million; *You've Got Mail*, $18 million; *Forest Gump*, $23 million). Thus, the lower total gross of *The Terminal*, less than half the average of his other movies ($157 million), suggests evidence of the effects of bad buzz.

In "experience goods" markets, third-party reviews play an important role in consumer evaluations of products or services (Chen and Xie 2005). When tangible attributes (in the context of movies, these might include the budget, marketing, stars, director, and so forth) are not sufficient to function as signals of the utility gains that might be expected from purchase, consumers likely seek third-party opinions to reduce the risk and uncertainty associated with their consumption (Dowling and Staelin 1994). Restaurants, movies, Broadway shows, books, magazine subscriptions, music recordings, car mechanics, online retailers, television shows, hair styling services, lawn maintenance services, and dentists are but a few examples of such experience goods.

In such industries, others' opinions are as important to the sellers as they are to the consumers. However, low-quality producers may be hard to distinguish from high-quality producers of such goods, and if consumers are willing to pay only for expected gains, producers have less incentive to produce high-quality offerings. In return, only low-quality goods get produced and sold for low prices, a problem typical to the "lemons" market (Akerlof 1970).

One way to address the lemon problem is through repeated interactions, but for goods such as movies, books, and shows, repeat purchase is not a feasible option. Others' opinions can come to the rescue in such circumstances. For example, consumers traditionally use *Consumer Reports* (professionals) ratings to learn about the quality of products; more recently, Web sites like BizRate and Epinions maintain scores, based on consumer (amateur) feedback, of most online

stores. In an online setting, the limited dimensions of store characteristics and virtually homoge-neous nature of the products means higher scores translate into higher perceived quality and thus higher profit for the sellers.

Literature shows that both professional (Litman 1983, Eliashberg and Shugan 1997, Holbrook 1999, Basuroy, Boatwright, and Kamakura 2003, Elberse and Eliashberg 2003, among others) and amateur(Godes and Mayzlin 2004, Dellarocas, Awad, and Zhang 2005, Chevalier and May-zlin 2005, for example) reviews play important roles in consumers' decision process, as is sum-marized in Table 1[1].

Most of the research listed in Table 1 investigates professional reviews in the movie industry. Am-ateur reviews have been less popular study topics because they have not been easily observable or measurable until very recently. In addition, the motion picture industry has served as a study context for various reasons. First, movies offer a very representative example of an experience good, because their quality cannot be assessed easily without consumption, so consumers must incur not only the cost of a movie ticket but also the opportunity cost of time. Second, after a movie has been released, word-of-mouth (henceforth, WOM) plays a crucial role in determining its success or failure, which enables researchers to separate the relevant factors into prerelease (e.g., production budget, marketing cost, stars, directors, genre, MPAA rating, professional critic reviews) parameters and postrelease (number of theaters for each week, competition with other movies, seasonality impacts, and of course WOM) parameters. Third, the ticket prices of movies are generally constant over time and do not vary across movies, making it easier to avoid compli-cated demand/supply analyses . Fourth, movie data sets are easy to obtain, and it is not difficult to compare research results.

With regard to *professional* movie reviews, Eliashberg and Shugan (1997) show that the correla-tion between reviews and demand may be spurious and conclude that critics' reviews can be used only as"predictors" (instead of "influencers") of total box office grosses.

Regarding to *amateur* reviews, it is generally accepted that the reviews can serve as a *proxy* for

---

[1]Extensive literature on movie revenue forecasting models exists, but due to the scope of this research, I do not review these papers here. Readers may find Sawhney and Eliashberg (1996) and the references therein interesting in this regard.

Study	Method	Data	Review	Findings
Litman (1983)	Multiple regression	Movies 1972-1978	Critics	Critics' ratings are a significant factors to explain box office revenue
Sawhney and Eliashberg (1996)	Forecasting Model, generalized Gamma	Movies 1990-1991	Critics	Critics' reviews are positively significant for a number of adopters
Eliashberg and Shugan (1997)	Correlation analysis	Movies 1991-1992	Critics	Critics' reviews are a predictor of box office performance (not an influencer)
Holbrook (1999)	Multiple regression	Movies Pre-1986	Critics	Ordinary consumers and professional critics emphasize different criteria in the formation of their tastes, but the correlation between popular appeal and expert judgments is positive
Basuroy Boatwright, and Kamakura (2003)	Diffusion model	Movies 1992-2001	Critics	Valence of critics' reviews affect diffusion pattern of box office
Elberse and Eliashberg (2003)	Demand/ supply model	Movies 1999	Critics	Less positive reviews correspond to a higher number of opening screens, but more positive reviews mean more opening revenue
Godes and Mayzlin (2004)	Multiple regression	TV Shows 1999-2000	Amateur	Online conversations offer one way to measure word-of-mouth
Reinstein and Snyder (2005)	Difference in differences	Movies 1999	Critics	Critics' influence is smaller than previous studies would suggest but still significant
Dellarocas, Awad, and Zhang (2005)	Diffusion/ forecasting model	Movies 2002	Amateur	Online amateur movie ratings can be used as a proxy for word-of-mouth
Chevalier and Mayzlin (2006)	Difference in differences	Books 2003-2004	Amateur	Online amateur book ratings affect consumer purchasing behavior

Table 1: Previous research related to professional and amateur reviews

WOM, and a model that takes professional or amateur reviews into consideration can provide useful information about the quality of a product. For example, Dellarocas, Awad, and Zhang (2005), using a movie revenue forecasting model, show that with only three days of box office and user and critic ratings data, their model achieves forecasting accuracy levels that a previous model (BOXMOD: Sawhney and Eliashberg 1996) required between two to three weeks of box office data to achieve. Godes and Mayzlin (2004) also conclude that online conversations (in Usenet newsgroups) offer an easy and cost-effective opportunity to measure WOM.

Despite these findings, the *influence* of professional or amateur reviews/ratings on consumer purchasing behavior has not been well established. One of the major difficulties related to estimating the impact of reviews on box office revenue is the endogeneity problem — movies with higher intrinsic quality tend to have better reviews, so it is hard to determine whether the positive review or the high quality of a movie is responsible for its high demand. Although endogeneity is not an issue for revenue forecasting models, it becomes a serious concern whenever the influence of reviews on revenue is to be implied. In the literature, earlier papers (Litman 1983, Litman and Kohl 1989, Sochay 1994, for example) generally do not consider the endogeneity issue, which means the findings of a positive influence of critics' reviews on box office revenue may be spurious. More recent work has addressed this problem using several approaches. Reinstein and Snyder (2005) and Chevalier and Mayzlin (2005) use *"difference-in-differences"* methods to eliminate fixed effects over time and across different critics (Siskel and Ebert) or Web sites (Amazon and Barnes & Noble). Elberse and Eliashberg (2003) propose a *simultaneous-equations model* to address the simultaneity of audience and exhibitor behavior. Elberse and Anand (2005), in examining the causal relationship between movie advertising and revenue, use *panel data analysis* to eliminate the fixed effects of movie quality. Dubois and Nauges (2006), on the basis of an empirical framework proposed by Levinsohn and Petrin (2003), develop a structural *panel data model* to control for and identify the unobserved quality of wines.

Empirical work on the influence of professional or amateur reviews on movie box office performance is sparse. In the case of professional reviews, Eliashberg and Shugan (1997) find no evidence that reviews influence revenue, but both Eliashberg and Shugan (1997) and West and Broniarczyk (1998) quote a Wall Street Journal survey in which "over a third of Americans seek

92

the advice of critics when selecting a movie"[2]. West and Broniarczyk's empirical study also suggests that consumers respond to critic disagreement. Moreover, movie studios consistently quote favorable critics' reviews in their promotions, in the hope of influencing people to watch the movies. These pieces of evidence cast doubt on the "predictor but not influencer" conclusion. In this paper, I briefly revisit this question with new evidence and discuss the role of professional reviews.

With regard to amateur reviews, Chevalier and Mayzlin (2005) is the most relevant to this discussion. They find that users' book reviews on Amazon.com or BarnesandNoble.com can influence the sales of the reviewed books. In the movie context, online amateur reviews, such as those posted on the Yahoo! movies site or RottenTomato.com, are becoming more and more popular. Although no direct verification exists to suggest more people check Yahoo! movie ratings than have previously, the increasing numbers of reviews written for each movie[3] and movie advertisements on the Web site imply indirect evidence. According to the MPAA (http://mpaa.org, *2005 MPA Market Statistics*), its member companies spent $0.84 million on online advertising in 2005 for an average film, compared with $0.35 million in 2001 and $0.24 million in 2002. Given the growing penetration of the Internet in U.S. households[4], online amateur reviews likely play increasingly important roles in helping people make more informed decisions about the movies they want to watch. I investigate whether the influence of these reviews can be identified, notwithstanding the endogeneity problem discussed previously.

This paper proceeds with the following four sections. Section 2 describes the data and measures, and Section 3 briefly discusses the role of professional reviews. In Section 4 , I build a diffusion model for WOM, examine the impact of online reviews on demand, and present some robustness checks. Section 5 concludes.

---

[2] *Wall Street Journal*, "A 'Thumbs Up' Pulls in the Audience," March 25, 1994, B1.

[3] In 2004, the most popular movie received 97,132 ratings on Yahoo! movies, one order of magnitude more than the most popular movie got (8,754) in 2002. The average number of ratings received in these two years are 9,446.87 and 409.28, respectively.

[4] In 2005, the statistics were 70.7 million (63%); in 2001, they were 50.9 million (47%). Source: US Census Bureau, IDC.

2 Data and Measures

The data set consists of movie production information, weekly box office information and Yahoo! movies professional and amateur reviews of nationally released movies between July 4, 2003 and September 10, 2004. A few movies are left out of the final sample because they received too few amateur reviews. The final data set comprises 128 movies.

2.1 Movie and Box Office Data

For each movie, I collect information about the name, release date, distribution studio, production budget, estimated marketing spending, running time (duration), genre[5], MPAA rating[6], and total gross. Weekly box office performance data are also available for these movies. Specifically, I gather the total number of weeks a movie is in the theater, the weekly gross, the number of theaters showing the movie in any week, and the number of competing movies for a movie in any week.

Variable	Median	Mean	Std	Min	Max
Budget (million $)	30	39.16	36.32	0.046	200
Marketing (million $)	15	14.8	12.8	0	50
Running time (min.)	102	105.44	22.47	76	200
Opening gross (million$)	9.13	13.49	14.63	0.03	83.85
Total gross (million $)	31.41	46.61	61.23	0.15	377.02
Film life (weeks)	14	14.6	6.3	4	33
# theaters (1st week)	2456	2009.73	1227	2	3703

Movies: 128

Table 2: Summary Statistics for Movies

The summary statistics in Table 2 are consistent with those published in industry reports (e.g. MPAA theatrical market statistics, *Variety* magazine), which suggests this is a representative

---

[5]Following Yahoo! movie's classification scheme, I designate 12 categories: art/foreign (8 movies), comedy (52), drama (39), romance (26), thriller (17), action/adventure (30), crime/gangster (18), musical/performing (4), kids/family (7), scifi/fantasy (7), western (2), and suspense/horror (15). The numbers in parentheses do not sum to 128 because some movies belong to multiple categories.

[6]There are five rating categories: PG (17), PG13 (63), R (41), NC-17 (1), Unrated (6). There are no G-rated films in the sample.

Variable	Correlation with Overall	Mean	Std	Min	Max
Critics	0.5786	7.1953	1.8403	3	11
Overall	1	9.6774	4.0737	1	13
Story	0.9478	9.5739	3.9918	1	13
Acting	0.9185	9.8969	3.7795	1	13
Direction	0.9445	9.5937	3.9273	1	13
Visual	0.8968	10.0876	3.7477	1	13
Total reviews per film	0.3711	1435	2373	27	18269

Total Reviews: 183690

Table 3: Summary Statistics of the Reviews

sample of movies.

2.2 Review Data

I collected professional and amateur reviews for these movies from Yahoo! Movies[7], which assembles professional reviews from sources like the *Boston Globe, Chicago Tribune,* E! Online, *New York Times,* and so forth. According to the Web site: "Yahoo! converts each critic's published rating into a letter grade. If the critic's review does not include a rating, Yahoo! movies assigns a grade based on an assessment of the review." Amateur reviews are those posted by individual visitors to the Yahoo! movies Web site. For each review, Yahoo! reports the author's Yahoo ID and date of the review. In addition to writing a textual review, amateur reviewers can rate the movie on the basis of four aspects (story, acting, direction, visual) and provide an overall rating.

Table 3 provides the summary statistics of the reviews. For all the scores, I convert letter grades to numerical values, such that an A+ score corresponds to a numerical score of 13, and the lowest possible score,[8] an F, is corresponding to a 1. The average professional score is considerably lower than that of the amateur reviews (thus called critics?), their average grade is a C+ (7.20),

---

[7]http://movies.yahoo.com

[8]In Yahoo! movie ratings, it is not possible to give scores E-, E, or E+, so from F to A+ there are 13 levels.

whereas that of amateurs is a B+ (9.68). However, the high correlation between them (0.5786) indicates that, to some extent, the "expert judgments" and the "public appeal" (Holbrook 1999) agree to each other. The professionals are more consistent than the amateurs, this is reflected by the much smaller standard deviation of professional scores. Professional also are less likely to give extreme scores like A+ or F, but because many of the grades are assigned subjectively by Yahoo!, this may merely reflect Yahoo!'s conservativeness. The scores for the four specific aspects of movies are highly consistent with the overall score, as evidenced by their high correlations. It is interesting to note that amateur reviewers are generally satisfied with the Visual and the Acting aspects of movies, and they are somewhat picky about the Story and the Direction. There are considerable variations (the standard deviation twice as large as the mean) in the total reviews per film, such that the most reviewed movie (*The Passion of the Christ*) received 18,269 reviews, whereas the least (*Buffalo Soldiers*) got only 27 reviews. There is a positive correlation between the valence of reviews and the propensity to review; that is, better movies attract more reviews.

In my subsequent analysis, I match the amateur reviews to the weekly box office data and calculate a cumulative average score for each movie i in week t by summing the overall scores submitted for movie i before week t and dividing it by the number of reviews. Note that this score is the online rating that visitors to the Web site see when they check on movie i on week t. In addition to this valence measure, I calculate volume measures and variation measures. The volume of movie i for week t is measured as the number of reviews posted before week t for movie i, and the variation is a measure of the level of consensus among consumers. I use commonly used dispersion measures (variance, Gini coefficient, coefficient of variation) to evaluate this level of agreement.

3   The Influence of the Professionals

Since the genesis of movies, as early as the silent film era, film criticism has been recognized as an art. After the 1940s, movie criticism became a profession as well. In his new book *American Movie Critics* , Phillip Lopate (2006) calls the period from 1950s to the 1970s "the golden age of movie criticism". Film critics like Pauline Kael and Andrew Sarris were widely read, and

their reviews had the power to cause public stirs. Later, the television program Siskel and Ebert (changed to Ebert & Roeper in 1999) became household names and has been vastly successful for two decades. It is hard to imagine that these reviewers could have been so popular for so long without having an impact on the box office performance of movies. Film studios obviously believe that good reviews from critics help attract viewers, as they routinely quote terms such as "spectacular," "thrilling," "two thumbs up" in their movie advertising blurbs. To get better reviews, studios have been known to "bribe" reviewers by flying them first-class to the studios for a weekend and offering generous gifts. However, if they cannot easily predict reviewers' reactions, studios often refuse to offer critics previews to avoid getting slashed (e.g. *The Avenger*).

Eliashberg and Shugan (1997) study two possible critics' effects: influence and prediction. Whereas "influence" implies an impact, "prediction" suggests that critics' reviews are merely leading indicators that have no significant influence on actual box office revenues. They regress weekly box office revenues on measures of critics' reviews, and find that box office revenues after the fifth week are correlated more strongly with the critics reviews, suggesting in favor of an "influencer" story. Reinstein and Snyder (2005) attempt to exploit the timing of Siskel's and Ebert's reviews to distinguish the influence and prediction effects. They find some weak evidence of an influence effect.

In this section, I replicate Eliashberg and Shugan's (henceforth, ES) study with a new data set, and compare the results with theirs. In Table 4, I report summary statistics for movies and critics' reviews, following the format of Table 2 in ES.

Variable	Median	Mean	Std
Film life in weeks	14	14.6	6.3
Screens (1st week)	2456	2009.73	1227
Box office (1st week)	9,130,000	13,490,317	14,633,000
Cumulative BO	31,410,000	46,613,359.40	61,234,610
Total reviews per film	13	11.89	3.18
Percentage of positive reviews	46.60%	47.30%	0.299
Percentage of negative reviews	21.40%	29.10%	0.278
Number of Observations = 128			

Table 4: Motion Pictures and Critic Summary Statistics

ES's data set contains 56 movies released in 1991-1992; as Table 4 shows, 10 years later, the median film life in weeks is slightly shorter. The number of screens more than doubled, from 1,122 to 2,456, and the first week's box office revenue almost tripled, from $3.6 million to $9 million. However, total gross circa 2004 ($31 million) is only slightly higher than that in 1991 ($28 million), which appears to indicate that the audience has shifted to the earlier weeks of movies' life cycles. Previously, ES read reviews and decided whether each was positive or negative; for my dataset, I simply classify scores higher than B as positive and those lower than C as negative. With this method, the distribution of positive and negative reviews is highly consistent with those in ES.

Week	Multiple R^2	Percent positive	t-statistic	Total Number of Reviews	t-statistic	Total Number of Screens	t-statistic	F-Ratio
	(Adj. R2)	(Std. Coeff.)	(p-value)	(Std. Coeff.)	(p-value)	(Std. Coeff.)	(p-value)	(p-value)
1	0.4958	0.27001	3.83	0.17208	2.5	0.71214	10.62	39.98
	(0.4834)		(0.0002)		(0.0139)		(<.0001)	(<.0001)
2	0.4986	0.31854	4.55	0.18625	2.71	0.68401	10.31	40.44
	(0.4863)		(<.0001)		(0.0076)		(<.0001)	(<.0001)
3	0.5863	0.26788	4.25	0.10831	1.72	0.71879	12.21	56.68
	(0.5759)		(<.0001)		(0.0881)		(<.0001)	(<.0001)
4	0.5801	0.23456	3.65	0.09262	1.45	0.66218	10.99	54.81
	(0.5696)		(0.0004)		(0.1498)		(<.0001)	(<.0001)
5	0.6587	0.13678	2.31	0.12729	2.21	0.72746	13	75.92
	(0.6501)		(0.0224)		(0.0291)		(<.0001)	(<.0001)
6	0.7083	0.13621	2.43	0.136	2.53	0.74462	13.83	92.29
	(0.7007)		(0.0165)		(0.0127)		(<.0001)	(<.0001)
7	0.7128	0.08549	1.45	0.12279	2.22	0.7687	13.49	87.71
	(0.7047)		(0.1496)		(0.0286)		(<.0001)	(<.0001)
8	0.5852	0.05469	0.74	0.17923	2.51	0.67684	9.39	43.74
	(0.5719)		(0.4616)		(0.0139)		(<.0001)	(<.0001)
All	0.4329	0.06839	3.38	0.05485	2.71	0.66102	35.02	409.68
	(0.4318)		(0.0007)		(0.0068)		(<.0001)	(<.0001)

Table 5: Regression results for the percentage of positive reviews

Table 5 shows the result to be compared with Table 4 in ES. The percentage of positive reviews represents a significant variable from the first week, though its significance tapers away slowly to the end of the sixth week. The percentage of positive reviews is most influential in the second week. If the influencer hypothesis is true, it may indicate that during the first week, more people

who tend to love the movie goes, but in the second week, people who are indifferent will be more influenced by the critics. There are of course other possible explanations, for example, the critics are merely predictors, but the WOM process after the first weekend takes up and drives people to the cinema in the second week. This scenario is quite possible because comparing the two datasets, I showed above that people tend to watch the movies in earlier weeks now than in the early 1990s. The decreasing correlation between critics' review and weekly box office revenues over time supports the ES's claim that consumers best remember a review just after they have first read it

Other specifications (not reported) can compare my results with those of ES as well. For example, the percentage of negative reviews and the valence of average critics' reviews might be included as independent variables. The results are similar — measures of critics' reviews are highly correlated with the box office revenues starting from the first week. Care must be taken in interpreting the result though; it is not enough to support the "influencer" claim: the absence of evidence is not the evidence of absence. The early significance of critics' reviews may be a result of the audience's shift to earlier weeks, or simply because online and offline WOM gets communicated faster and easier than in the past. That is, my analysis may weaken ES's result but does not offer definite evidence to reject or accept it. Further analysis therefore is needed obtain a clearer understanding of the role of critics.

4   The Influence of the Amateurs

The fundamental question motivating the construction of the main model is "What is the quality of a movie?" More specifically, with regard to the omitted endogenous quality variable in movie revenue forecasting models, what should be the correct formulation of any measure of quality?

Suppose a movie's quality enters the revenue forecasting model linearly and can be captured by a score (or vector of scores), then a traditional forecasting model can be augmented with such a measure to provide consistent estimates of the impact of both professional and amateur reviews. If quality were controlled for, researchers could examine whether a higher review score induces higher demand and would not be haunted by the endogeneity problem. However, this linearity

assumption implies that, when budget, marketing, reviews, seasonality, and other observable characteristics are controlled for, two movies with the same quality score should bring in same amount of total revenue. Because a movie's quality does not change over time, the two movies' weekly box office revenues also should be equal each week, and the weekly revenue trajectories should be the same. This result does not sound plausible. First, a five-star blockbuster movie might have a totally different revenue trajectory than a five-star sleeper movie. The blockbuster movie might be 100 times more costly to produce than the sleeper movie, but that does not mean the total gross will also be 100 times different, nor that the pattern of revenue will follow the same shape. Second and more important, a movie is a social product, and its quality is subjective and depends highly on the "eye of the beholder". A five-star movie to one person might be a one-star movie to his or her neighbor, and this difference does not mean their tastes always are negatively correlated; both consumers might give five stars to yet another movie.

Therefore, the quality measure must enter the equation in a non-linear way. It also may interact with some observables or non-observables. There does not seem to be a good way to write down a meaningful equation — even if we can write such an equation, it remains unidentifiable because the quality is unobservable.

In a famous Harper's Magazine interview,[9] Mark Gill, senior VP of publicity and promotion for Columbia Pictures, said: "It doesn't matter if the movie doesn't deliver. If you can create the impression that the movie delivers, you're fine. That's the difference between *playability* and *marketability*. When you've got playability, you show it to the critics. If you've got marketability, there's enough there to cheat it or stretch it and make people believe that it really does deliver." This distinction offers one way to think about the quality of a movie beyond a uni-dimensional approach. Because my objective in this section is to estimate the influence of amateur reviews on box office revenue, I focus on the "playability" of a movie and largely take opening weekend revenue as given, which depends heavily on the cast, marketing, production, and so forth (i.e. marketability). Overtime, playability will be reflected in consumers' responses to the movie, and consumers' responses constitute WOM about that movie.

---

[9]Harper's Magazine, "The New Auteurs; Motion Picture Marketing," June, 1993. http://www.ibiblio.org/pub/electronic-publications/stay-free/ml/readings/new_auteurs.html

To understand this reasoning, consider two movies with the same prerelease parameters (i.e. same marketability measures). If the postrelease weekly number of theaters, competition, and seasonal variations are controlled, the different composition of the audience (consumer mix) for each movie should lead to different adoption patterns [10]. Each week, WOM gets reinforced because when people talk about the movie, they will induce the adoption or non-adoption of other people with similar tastes. This underlying process of WOM getting passed on can be modeled as a diffusion, whose pattern reflects the quality of a movie. A better movie will attract more consumers, and their positive WOM will positively influence the adoption positively, until the point of market saturation. As the preceding discussion indicates, WOM is inseparable from the playability of a movie's quality, and the revenue trajectory may signify a movie's quality through the WOM diffusion process.

4.1 A WOM Diffusion Model

Suppose in week t the box office revenue of a movie is R_t, and because the ticket prices in different cinemas are essentially the same, R_t can be used to calculate the number of people, A_t, going to the cinema to view the movie in week t. Hence, $A_t = f(R_t) \doteq \frac{R_t}{p}$, where p is the ticket price. Among these consumers, a proportion $\theta(q) < 1$ will spread positive comments (whether off- or online), where q is a combination of the unobserved quality of the movie and the other time-invariant forces correlated with the quality (e.g., production budget, marketing, star power, critics reviews). There are also naysayers, let $\phi(q) < 1$ represent this portion of the consumers, and the rest $1 - \theta(q) - \phi(q) < 1$ say nothing. Together with other forces, ξ, unrelated to q (but constant over time), θ and ϕ determines the consumers who view the movie in week $t + 1$. Thus,

$$A_{t+1} = g[\theta(q), \phi(q), \xi]A_t = g(q, \xi)A_t \tag{1}$$

Note that equation (1) assumes the number of consumers in week $t + 1$ is solely determined by the number of consumers in week t and the WOM process, I will relax this assumption below.

---

[10]Of course, for typical blockbuster movies like *Star Wars* and *The Lord of the Rings*, the composition of the audience may be quite similar (and thus have similar decay patterns).

Equation (1) can be used to calculate the change in revenue over time,

$$\Delta R_t = R_{t+1} - R_t = p \cdot A_{t+1} - p \cdot A_t = pA_t[g(q,\xi) - 1] = [g(q,\xi) - 1]R_t \qquad (2)$$

Rewriting equation (2) in continuous form and defining $\lambda \equiv g(q,\xi) - 1$, we can get:

$$\frac{d}{dt}R_t = \lambda R_t \qquad (3)$$

where the parameter λ measures the playability (WOM diffusion) of a movie.

Note also that I use $g(.)$ in equation (1) to capture the complex relationship among intrinsic quality, other time-invariant forces, and the WOM process. Although I can not directly observe θ and ϕ, they depend on the quality[11] q, and they should satisfy $\dfrac{d\theta}{dq} > 0$, and $\dfrac{d\phi}{dq} < 0$. It is also reasonable to say, from (1), that $\dfrac{\partial g(\theta)}{\partial \theta} > 0$, and $\dfrac{\partial g(\phi)}{\partial \phi} < 0$; thus, $\dfrac{d\lambda}{dq} = \dfrac{dg(\theta,\phi,\xi)}{dq} = \dfrac{\partial g}{\partial \theta}\dfrac{d\theta}{dq} + \dfrac{\partial g}{\partial \phi}\dfrac{d\phi}{dq} > 0$. Therefore λ, derived from a diffusion process, is an increasing function of q, and can be regarded as a measure of quality. The parameter ξ is a time-invariant variable that captures forces unrelated to the quality of the movie, such as unemployment rates, the availability of a public forum to discuss the movie, and so forth. It does not affect the relationship between λ and q because, by definition, $\dfrac{d\xi}{dq} = 0$.

The solution of (3) is:

$$R_t = R_0 e^{\lambda t} \qquad (4)$$

where R_0 is the initial condition, and in this context, it has a perfect interpretation as the opening week's gross. Equation (4) implies that:

$$R_t = R_{t-1}e^{\lambda} \qquad (5)$$

Starting from a WOM diffusion formulation, I derive an exponential model[12] for movie revenues

---

[11]The parameters θ and ϕ should be movie specific, but because I am examining one movie, I suppress the subscript i in the equations.

[12]Exponential decay models are widely used in similar modeling situations. In similar contexts, Jedidi *et al.*

102

that parsimoniously captures a movie's quality as a decay parameter.

Rearranging (5), and adding the terms of amateur review variables, control variables, and an error term, I obtain the empirical model:

$$log \frac{R_t}{R_{t-1}} = \lambda + \eta t + \sum_{j=1}^{J} \beta_j x_{jt} + \sum_{k=1}^{K} \gamma_k y_{kt} + \varepsilon_t \tag{6}$$

where λ is the fixed-effect quality measure (also a measure of WOM diffusion), η is a time decay factor that is independent of the movie quality, $x_t = (x_{1t}, ..., x_{Jt})$ is a vector of J independent WOM variables calculated for the t-th period, $y_t = (y_{1t}, ..., y_{Kt})$ is a vector of K variables for period t that controls for competition and supply, $\beta = (\beta_1, ..., \beta_J)$ are the parameters to be estimated to indicate the impact of WOM, and $\gamma = (\gamma_1, ..., \gamma_K)$ are parameter estimates of the control variables.

Note that in the specification equation (6), time decay factor η allows the examination of possible changes in the composition of the audience over time, thus the implicit assumption of constant θ and ϕ in equation (1) is relaxed. Two forces may drive the sign of η. If the aficionados see the movie first, followed by the indifferents, there should be a negative η. If, however, people who love the movie induce similar-minded people to go in later weeks (and thus reinforce positive WOM), η should be positive.

The form of (6) is much more flexible than (5), too. Not only the assumptions for constant θ and ϕ are relaxed, the addition of WOM and control variables allows for examination of the postrelease forces in contributing to the changes in revenues over time.

When a movie's revenue decreases over time, the right-hand side of the equation should be negative. Because larger absolute values of λ indicate faster decay, a movie with higher quality should have smaller absolute values of λ. For a sleeper movie, the revenue in some weeks can be larger than that of the previous weeks, and model (6) can accommodate this possibility with positive estimate of the right-hand side.

(1998) use an exponential decay model of market share to identify four clusters of movies, and Eliashberg *et al.* (2000) use one to describe exposure to WOM conversations. Einav (forthcoming) uses exponential decay to control for factors other than seasonality. In the same spirit, a generalized gamma model can be used to model the diffusion process(Sawhney and Eliashberg 1996, Ainslie, Dreze, and Zufryden 2005). More elaborated diffusion models also can be found in the literature, such as in Dellarocas *et al.* (2005).

As stated previously, the WOM effect λ can be regarded as the playability aspect of the quality of a movie, and if online amateur reviews influence people's decisions, this effect should also be captured in λ. The purpose of including the variables x is to estimate the *residual impact* of the online reviews, and the estimate of β then captures the residual influence of online amateur reviews. Included in the vector of x, I have variables to measure the (a) valence, (b) variance, and (c) volume of WOM. To illustrate, consider online amateur reviews as one more component of the consumer WOM process. If it works like other means of WOM, λ will capture the influence. Then, introducing the independent variables of online amateur reviews in the model can provide an estimation of the effects of these reviews on revenue only when changes from week $t-1$ to t cannot be captured by the WOM decay parameter, which this is exactly what I set out to identify as the residual impact of online amateur reviews on box office revenue.

This formulation implicitly (and critically) assumes that, controlling for observables, the decay pattern does not systematically vary with idiosyncratic variations in WOM[13]. I test the validity of this assumption in section 4.2. With this assumption, I can safely use the parameter estimate β to examine the influence of online reviews on the weekly box office performance that is not explained by the general WOM process.

To avoid the potential problem of heteroscedasticity, I conduct White's general test (White 1980) and the Breusch-Pagan test (Breusch and Pagan 1979). Whereas White's test ($p > \chi^2 = 0.0412$, DF=5) rejects the null hypothesis of no heteroscedasticity at the $p < 0.05$ level, the Breusch-Pagan test ($p > \chi^2 = 0.4875$, DF=2) cannot reject the null hypothesis. Estimates of FGLS (Feasible Generalized Least Squares) have almost the same standard error values as those of the ordinary least squares (OLS) and no significance level changes, so OLS standard errors are reported in Table 6. Four specifications (pure decay, WOM valence, WOM variance, and all previous variables plus controls) are estimated. In the pure decay model, the left-hand side variable is regressed on a constant and a variable that indicates the number of weeks since release. Each other model adds one or more variables to the specification. The WOM valence is a cumulative average rating of what consumers view at Yahoo! movies before weekend t. The WOM variance is measured by

---

[13]Note that this is not an unreasonable assumption because the quality of a movie does not change over time.

the inverse of the coefficient of variation (CV)[14] of the averaged amateur ratings[15]. Following the literature (Zufryden 1996, Zufryden 2000, Jedidi, Krider, and Weinberg 1998, Ainslie, Dreze, and Zufryden 2005), competition is measured by the number of new movie releases in week t. I also include the number of theaters showing the movie in week t to control for the possible influence of capacity constraints on the supply of movies. To control for seasonality, I use a different data set that contains the weekly total box office revenues of all movies from January 7, 2000 to December 27, 2002. I mark the holiday weekends (Martin Luther King Day, President's Day, Memorial Day, Labor Day) as anchors, then calculate the three-year average total revenue for each week. To create an index for each week, I normalize the average total revenues, such that the lowest grossing week has a score of 1, and a week with 1.5 times revenue earns a score of 1.5. The WOM volume, a measure of how many people make the effort to write reviews, is not significant in any of the specifications.

The odd numbered columns contain the estimates obtained with five weeks of data for each movie, and the even numbered columns report those obtained with ten weeks of data. Because the table included pooled result for blockbuster and sleeper movies, the movie quality parameter λ changes considerably across specifications. Note that all the five-week estimates of λ are smaller in absolute value than their ten-week counterparts, because the sleeper movies have much greater λ values in earlier than in later weeks. In all columns, λ's are negative and significant. Over a ten-week period, an average movie's quality follows an exponential decay rate of -0.9810, which translates to a reduction of 62.5% in revenue every week. According to the preceding discussion, if a movie's decay rate is less than -0.9810 in absolute value, it is a good movie. When the controls for competition, number of theaters and seasonality are added, the parameter estimates of λ and WOM valence increase. The cumulative average score, valence of WOM, entails a significant force with a positive impact on revenue. In various specifications, the parameter estimates are robust, and they all take values from 0.04 (ten-week models) to 0.09 (five-week model with

---

[14]WOM variance=$1/CV = \frac{1}{\mu/\sigma} = \frac{\sigma}{\mu}$, where σ is the standard deviation, and μ is the mean. The CV is widely used as a measure of reliability, volatility, or risk. Here, it serves as a measure of the level of agreement among consumers about the quality of a movie. This measure will be greater for more controversial movies.

[15]The CV is a dimensionless number, in that it has no unit and thus allows comparison with numbers that have significantly different mean values. This is approach is similar to the use of entropy by Godes and Mayzlin (2004) to measure the dispersion of online conversations while controlling for the total volume of posts. The use of other dispersion measures, such as Gini coefficient and standard deviation, does not qualitatively change the results.

	Pure Decay		WOM Valence		WOM Variance		Controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Movie quality	-0.1882**	-0.5402***	-0.7083***	-0.9107***	-0.4892***	-0.7094 ***	-1.3525**	-0.9810***
	(0.0227)	(0.0455)	(0.1336)	(0.0839)	(0.1539)	(0.0970)	(0.3484)	(0.1842)
Time decay	-0.0909***	0.0096	-0.0926***	0.0079	-0.0942***	0.0054	-0.1299***	-0.0065
	(0.0249)	(0.0071)	(0.0243)	(0.0070)	(0.0241)	(0.0070)	(0.0258)	(0.0091)
WOM Valence			0.0589***	0.0422***	0.0511***	0.0347***	0.0961***	0.0434***
			(0.0112)	(0.0081)	(0.0115)	(0.0082)	(0.0267)	(0.0147)
WOM Variance					-0.3561***	-0.3088***	0.1809	-0.1913*
					(0.1267)	(0.0760)	(0.2163)	(0.1032)
Competition							-0.0124***	-0.0364**
							(0.0254)	(0.0184)
# Theaters							-0.0001***	-0.00006***
							(0.0000)	(0.0000)
Seasonality							0.3014 ***	0.1984***
							(0.0000)	(0.0000)
R^2	0.0255	0.0017	0.0754	0.0265	0.0896	0.0412	0.2001	0.0727
F-Test	13.29	1.82	20.66	14.58	16.6	15.34	18.39	11.96
$Pr > F$	0.0003	0.1772	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Note: Standard errors appear in parentheses

*p<0.10, **p<0.05, ***p<0.01

Table 6: Residual impact of reviews on box office revenue

controls). Therefore, for every one-point increase in the score (e.g., from B+ to A-) in a week, the revenue in the next week will be increased from 4.4% to 10%. In columns (5) and (6) WOM variance seems to be very influential in determining the overall decay rate of movies, but once the control variables are added, it is no longer very significant. The control variable "number of theaters" is significant in the model, but its magnitude is too small for it to be important. As expected, seasonality is an important force, in that movies in "good" weeks usually earn much more than they do in "bad" weeks[16]. When the control variables are added to the model, the five-week variance measure is no longer significant, which likely suggests that in the early release stages of the movies, disagreement in tastes may not be as important a factor in influencing the box office revenues. The estimates for change in consumer mix, η, are all significant in the five-week models but not in the ten-week models, which implies that the aficionado effect is more pronounced in earlier weeks.

[16]For a more elaborated treatment of the issue of seasonality, see Einav (forthcoming).

4.2 Robustness Checks

To verify the robustness of the results, I examine two extensions. The first extension verifies that the online rating scores do not merely capture the idiosyncratic weekly variations in offline WOM. Suppose that, due to television advertising, though the quality of the movie does not change, the perception of its quality varies over time. If the online movie reviews reflect varying perceptions of quality, the parameter estimates for WOM will be spurious. In other words, the WOM may be influenced by other forces, and if both the box office revenues and online ratings are affected by these forces, the parameter estimates will be biased due to the endogeneity problem again. To assess the possible problem of omitted variables, I run model (6) again, but change the WOM valence variable to the averaged amateur rating of reviews posted within one week before weekend t. This idea is based on the observation that "cumulative averages" are what the website visitors actually see, and the "weekly averages" calculated from a short time span are more representative of the possible changes in the perception of the quality (idiosyncratic changes in WOM). If other forces drive the weekly variations in WOM and online scores, substituting cumulative average scores with weekly average scores should provide more significant parameter estimates for the model. If, however, the robustness check regressions show less significant parameter estimates, they offer evidence that idiosyncratic weekly variations in WOM are not a concern.

Table 7 presents the parameter estimates for this new specification. Again, columns (1), (3), (5), (7) report five-week estimates and the rest report ten-week estimates.

In column (3) and (4), the parameter estimates for WOM valence are significant, but when other variables are added, this is no longer true. It is not surprising that WOM valence in the five-week regressions is generally more significant than in ten-week regressions, because the average score in the first week likely is correlated with the other means of WOM that jumpstart the WOM process. But both the diminishing level of significance and the fading magnitude of the parameter estimates for WOM variance over time suggest that idiosyncratic WOM variation moving along with online ratings is not to be worried.

The second extension involves distinguishing between blockbuster and sleeper movies, which

107

	Pure Decay		WOM Valence		WOM Variance		Controls	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Movie Quality	-0.1631*	-0.5261***	-0.5529***	-0.7140***	-0.1039	-0.3942***	-0.5251**	-0.5852**
	(0.0947)	(0.0468)	(0.1248)	(0.0802)	(0.2239)	(0.1015)	(0.2354)	(0.1305)
Time Decay	-0.0954***	0.0081 **	-0.0895***	0.0092	-0.0980***	0.0037	-0.1327***	-0.0057
	(0.0257)	(0.0073)	(0.0250)	(0.0073)	(0.0253)	(0.0073)	(0.0260)	(0.0091)
WOM Valence			0.0411***	0.0204***	0.0193	0.0068	0.0253*	0.0054
(weekly avg)			(0.0098)	(0.0071)	(0.0157)	(0.0074)	(0.0150)	(0.0080)
WOM Variance					-0.5363***	-0.4226***	-0.2747*	-0.3536***
					(0.1683)	(0.0838)	(0.1647)	(0.0889)
Competition							-0.0104	-0.0358**
							(0.0236)	(0.0185)
# Theaters							-0.0001***	-0.00005**
							(0.0000)	(0.0000)
Seasonality							0.3094***	0.2024***
							(0.0537)	(0.0378)
R^2	0.0273	0.0012	0.0563	0.0090	0.0756	0.0326	0.1818	0.0643
F-Test	13.75	1282	14.55	4.76	13.27	11.72	16.33	10.49
$Pr > F$	0.0002	0.2695	<.0001	.0088	<.0001	<.0001	<.0001	<.0001

Note: Standard errors appear in parentheses

*p<0.10, **p<0.05, ***p<0.01

Table 7: Robustness check for idiosyncratic WOM variation

may follow totally different revenue trajectories (and thus have different λ values). Using one model for both types of movies may provide an averaged parameter estimate for the WOM valence.

I estimate models for blockbuster movies and sleeper movies separately. For simplicity, I define a sleeper movie as one whose highest weekly revenue is not obtained in the first weekend[17]. In my sample of 128 movies, 24 are categorized as sleepers. The results of the tests of heteroscedasticity for the blockbuster movies are similar to those for the whole sample, and the standard errors of FGLS estimates are similar to those of OLS. For the sleeper movies, however, some specifications only have fewer than 250 observations and White's test therefore is not reliable for this small sample. MacKinnon and White (1985) raise such concerns about the performance of White's

[17]Note that blockbuster movies are more commonly defined as those with big budgets. They often earn the highest weekly gross during their opening weekends, and their weekly grosses drop sharply over time. Here, I refer to all movies that follow this pattern as blockbusters, so a small production movie that moves quickly out of the cinema might be called a blockbuster.

heteroscedasticity consistent covariance matrix (HCCM) in small samples. They refer to the original White's HCCM as HC0, and proposed three more asymptotically equivalent forms of HCCM, HC1 - HC3. With Monte Carlo simulations, Long and Ervin (2000) recommend HC3 for small samples. Therefore, for the sleeper movies, I will use HC3 to derive the robust standard errors[18].

	Blockbusters (first 5 weeks)				Sleepers (all periods)			
	Pure Decay	Valence	Variance	Controls	Pure Decay	Valence	Variance	Controls
Movie Quality	-0.5041***	-1.0924***	-1.4925***	-1.4381***	0.0245	-0.2755	-0.3086	-0.3332 *
	(0.0628)	(0.1179)	(0.2623)	(0.1816)	(0.0867)	(0.2156)	(0.1434)	(0.1699)
Time Decay	-0.0543***	-0.0512***	-0.0493***	0.0233	-0.0333***	-0.0330***	-0.0327***	-0.0299***
	(0.0170)	(0.0163)	(0.0163)	(0.0187)	(0.0075)	(0.0075)	(0.0074)	(0.0064)
WOM Valence		0.0631***	0.0918***	0.0534***		0.0491***	0.0511**	0.0435**
		(0.0108)	(0.0199)	(0.0203)		(0.0203)	(0.0214)	(0.0225)
WOM Variance			-0.2978*	0.088			0.0381	0.0333
			(0.1745)	(0.1752)			(0.1002)	(0.1327)
Competition				0.0101				-0.0312
				(0.0169)				(0.0319)
# Theaters				0.00015***				0.00002
				(0.0000)				(0.0001)
Seasonality				0.0760**				0.2244***
				(0.0361)				(0.0642)
R^2	0.0256	0.1041	0.1108	0.2374	0.0719	0.0843	0.0846	0.1190
F-Test	10.16	22.36	15.95	17.96	30.85	18.28	12.2	7.88
$Pr > F$	0.0016	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001	<.0001

Note: For blockbusters, standard errors appear in parentheses.

For sleepers, the heterogeneity-adjusted HC3 standard errors appear in parentheses.

*p<0.10, **p<0.05, ***p<0.01

Table 8: Robustness check with blockbusters and sleepers separated

In Table 8, the fixed effect decay parameter for the blockbuster movies is much greater than that for the whole sample; when everything else in the model is controlled for, the average blockbuster movie attracts only 24% of the moviegoers in the previous week. This intrinsic decay parameter is negative for the sleeper movies (except for the pure decay model), but the magnitude is considerably smaller than that of the blockbuster movies. The WOM valence measure

---

[18]Briefly, HC0 takes the form $(X'X)^{-1}X'diag[e_i^2]X(X'X)^{-1}$. HC3 is defined as $(X'X)^{-1}X'diag\left[\frac{e_i^2}{(1-h_{ii})^2}\right]X(X'X)^{-1}$, where $h_{ii} = x_i(X'X)^{-1}x_i'$.

remains at the level of 0.04 to 0.05, though it is slightly higher for blockbuster movies in the first five weeks. This result confirms the previous finding, and suggests that a one-point increase in aggregate amateur ratings will bring about 3.9-4.8% more people in the next week. The level of disagreement associated with a movie is not significant for most specifications, which suggests that WOM variance is not an influential variable in the early stage of a movie's life cycle. The seasonality parameter estimate is much greater (and more significant) for sleeper movies, which suggests that in "good" seasons, demand is indeed higher.

That the parameter estimates of WOM valence are almost the same for blockbuster and sleeper movies gives support to our interpretation of this parameter as a measure of the residual impact of WOM valence. To interpret this parameter, both off- and online WOM must be recognized as contributing to the movie-specific decay rates, and there is no way to separate them from the intrinsic quality of a movie. Thus, the parameter estimated is not the impact of online ratings on box office performance but rather the residual impact, that is, the influence not explained by the decay pattern of intrinsic quality. This result likely explains why the parameter estimates are so consistent across blockbuster and sleeper movies, despite their distinctive decay patterns.

4.3 Panel Results

The preceding analysis derives an average measure of the rate of decay associated with movies' quality. It also has the advantage of using fewer degrees of freedom. To impute a measure of the intrinsic quality for each movie and to further evaluate the impact of WOM, I reformulate model (6) as a panel model.

I write the left-hand side as $L_{it} \equiv log \frac{R_{it}}{R_{i,t-1}}$ and change the time decay parameter η to a fixed effect of time; hence,

$$L_{it} = \lambda_i + \eta_t + \sum_{j=1}^{J} \beta_j x_{jit} + \sum_{k=1}^{K} \gamma_k y_{kit} + \varepsilon_{it} \qquad i = 1,...,N; \ t = 1,...,T_i \qquad (7)$$

where movies are indexed with i, and weeks after release are indexed with t.

Because a movie can be pulled out of the cinema in any week after its release, there are different

110

T_i values for each movie, which makes it an unbalanced panel. From the preceding discussion, it is obvious that λ_i should be treated as a fixed-effects parameter, and its natural interpretation is the quality of the movie i. I also expect η_t to be a fixed time effect.

	Random Two	Fixed One	Fixed Two
Intercept	-0.8883***	-	-
	(0.1336)	-	-
Valence	0.0335***	0.1528**	0.0926***
	(0.0113)	(0.0639)	(0.0312)
Variance	-0.0948	0.0039	-0.0043
	(0.0730)	(0.0839)	(0.0868)
Competition	-0.0273 *	-0.0239	-0.0245
	(0.0149)	(0.0153)	(0.0152)
# Theaters	-0.00005***	-0.00003	-0.00007**
	(0.0000)	(0.0000)	(0.0000)
Seasonality	0.1498***	0.1248***	0.1207***
	(0.0310)	(0.0336)	(0.0342)
R^2	0.0411	0.1882	0.2127

Table 9: Panel data results

To examine the variance/covariance structure of the data set, I conduct a Hausman specification test (Hausman 1978) to compare the fixed versus random effects models. Under the null hypothesis, the individual effects are uncorrelated with the other regressors in the model, both OLS and GLS are consistent, but OLS is inefficient. The Hausman test involves comparing the variance/covariance matrices of OLS and GLS. If the null hypothesis cannot be rejected, a random effects model is favored. The m-statistic for Hausman test is 10.64 with five degrees of freedom, corresponding to $Pr > m = 0.0590$. Therefore, the test is inconclusive and indicates a borderline position between the fixed and random effects models. To confirm the previous findings, I first fit a two-way random effects model using a FGLS procedure — the variance components are estimated in the first stage, then the estimated variance covariance matrix is employed to fit a GLS model. The parameter estimates appear in the first column of Table 9. The result is consistent with that obtained from model (6). The aggregate decay rate is -0.89, which translates to a 58.9% decay across weeks. The parameter estimate for WOM valence is 0.034, with a standard error of 0.0113 ($p = .0033$). I then generate dummy variables for each of movie and run a least square

111

dummy variable (LSDV) regression without intercept. Thus, I can estimate the fixed effects for each movie. The one-way fixed-effects model is highly significant, and the parameter estimate for WOM valence is significant (column 2). Suspecting significant fixed time effects, I also run a two-way fixed-effects model, as reported in the third column. Indeed, fixed time effects are significant for the first three weeks. Controlling for fixed time effects, the parameter estimate for WOM valence is 0.09, which suggests that the WOM residual impact can bring in as much as 9.7% extra revenue. In all three regressions in Table 9, WOM variance is not significant, possibly because with a panel data model formulation, the intercept (in the random effects model) and the movie-specific fixed effects (in the fixed effects models) capture the decay pattern associated with WOM disagreements.

5 Conclusion

This paper uses weekly movie box office data and Yahoo! movies review data to estimate the influence of the professional and amateur reviews on box office revenues. I develop a diffusion model of word-of-mouth, and exploit the weekly changes in revenue to control for the unobservable intrinsic quality and other time-invariant factors of movies. Using this method, I estimate the residual impact of online amateur reviews on box office revenues. Various robustness tests confirm the soundness of the model.

The results suggest that Yahoo! movies' online amateur reviews have a positive and statistically significant influence on other people's decision to watch a movie. Specifically, a one-point increase in the valence of the aggregate score can induce 4-10% more people to attend the cinemas. The volume and variance of the reviews does not play a significant role in the process, especially when competition, availability of screens, and seasonality are controlled.

In the study of the influence of professional reviews, I find that the "critics as influencer of box office gross" story can no longer be rejected.

The finding that positive reviews generate extra revenue does not imply that movie studios should hire people to manipulate online reviews. If in one period, the studio can induce some people to

go to the cinema through manipulation (i.e., introducing disutility to some people), it may generate fierce negative word-of-mouth in all channels and change the λ immediately. The changed word-of-mouth process may well backfire and reduce the potential profit of the studio.

References

Ainslie, A., X. Dreze, and F. Zufryden, 2005, "Modeling Movie Lifecycles and Market Share," *Marketing Science*, 24(3), 508–517.

Akerlof, G., 1970, "The market for lemons: quality uncertainty and the market mechanism," *Quaterly Journal of Economics*, 89, 488–500.

Basuroy, S., P. H. Boatwright, and W. Kamakura, 2003, "Reviewing the Reviewers: The Impact of Individual Film Critics on Box Office Performance," *Working Paper*.

Breusch, T., and A. Pagan, 1979, "A simple test for heteroscedasticity and random coefficient variation," *Econometrica*, 47, 1287–1294.

Chen, Y., and J. Xie, 2005, "Third-party product review and firm marketing strategy," *Marketing Science*, 24, 218–240.

Chevalier, J. A., and D. Mayzlin, 2005, "The Effect of Word of Mouth on Sales: Online Book Reviews," *Working Paper*.

Dellarocas, C., N. Awad, and X. M. Zhang, 2005, "Using online ratings as a proxy of word-of-mouth in motion picture revenue forecasting," *Working Paper*.

Dowling, G. R., and R. Staelin, 1994, "A model of perceived risk and intended risk-handling activity," *Journal of Consumer Research*, 21, 119–134.

Dubois, P., and C. Nauges, 2006, "Identifying the effect of unobserved quality and experts' reviews in the pricing of experience goods: Empirical Application on Bordeaux wine," *Working Paper*.

Einav, L., forthcoming, "Seasonality in the U.S. motion picture industry," *RAND Journal of Economics.*

Elberse, A., and B. Anand, 2005, "Advertising and expectations: The effectiveness of pre-release advertising for motion pictures," *Working Paper.*

Elberse, A., and J. Eliashberg, 2003, "Demand and supply dynamics for sequentially released products in international markets: The case of motion pictures," *Marketing Science*, 22(3), 329–354.

Eliashberg, J., J.-J. Jonker, M. S. Sawhney, and B. Wierenga, 2000, "MOVIEMOD: An implementable decision-support system for prerelease market evaluation of motion pictures," *Marketing Science*, 19, 226–243.

Eliashberg, J., and S. M. Shugan, 1997, "Film Critics: Influencers or Predictors?," *Journal of Marketing*, 61, 68–78.

Godes, D., and D. Mayzlin, 2004, "Using online conversations to study word-of-mouth communication," *Marketing Science*, 23(4), 545–560.

Hausman, J. A., 1978, "Specification tests in econometrics," *Econometrica*, 46, 1251–1271.

Holbrook, M. B., 1999, "Popular Appeal Versus Expert Judgments of Motion Pictures," *Journal of Consumer Research*, 26(2), 144–155.

Jedidi, K., R. E. Krider, and C. B. Weinberg, 1998, "Clustering at the Movies," *Marketing Letters*, 9(4), 393–405.

Levinsohn, J., and A. Petrin, 2003, "Estimating production functions using inputs to control for unobservables," *Review of Economic Studies*, 70(243), 317–341.

Litman, B., and L. Kohl, 1989, "Predicting financial success of motion pictures: The 80's experience," *Journal of Media Economics*, 2, 35–50.

Litman, B. R., 1983, "Predicting the success of theatrical movies: An empirical study," *Journal of Popular Culture*, 17, 159–175.

Long, J., and L. Ervin, 2000, "Using heteroscedasticity-consistent standard errors in the linear regression model," *The American Statistician*, 54, 217–224.

Lopate, P. (ed.), 2006, *American Movie Critics: An anthology from the silents until now*. Library of America.

MacKinnon, J., and H. White, 1985, "Some heteroscedasticity-consistant covariance matrix estimators with improved finite sample properties," *Journal of Econometrics*, 29, 305–325.

Reinstein, D. A., and C. M. Snyder, 2005, "The influence of expert reviews on consumer demand for experience goods: A case study of movie critics," *Journal of Industrial Economics*, LIII(1), 27–51.

Sawhney, M., and J. Eliashberg, 1996, "A parsimonious model for forecasting gross box office revenues of motion pictures," *Marketing Science*, 15(2), 321–340.

Sochay, S., 1994, "Predicting the performance of motion pictures," *Journal of Media Economics*, 7, 1–20.

West, P. M., and S. M. Broniarczyk, 1998, "Integrating multiple opinions: The role of aspiration level on consumer response to critic consensus," *Journal of Consumer Research*, 25(1), 38–51.

White, H., 1980, "A heteroskecastic-consistent covariance matrix estimator and a direct test of heteroskedasticity," *Econometrica*, 48, 817–838.

Zufryden, F., 1996, "Linking advertising to box office performance of new film releases: A marketing planning model," *Journal of Advertising Research*, pp. 29–41.

————, 2000, "New film website promotion and box-office performance," *Journal of Advertising Research*, pp. 55–64.