

ESSAYS ON SOCIAL EFFECTS AND SOCIAL MEDIA

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Tong Bao

January 2011

© 2011 Tong Bao

## ESSAYS ON SOCIAL EFFECTS AND SOCIAL MEDIA

Tong Bao, Ph. D

Cornell University 2011

Two significant phenomena emerge from recent internet development: consumers are influenced by social network; and consumers engage in consumption and production of user-generated content. This dissertation studies social influence and social media.

In Chapter 1, we study how summer internship employer choices of MBA students at a major university are influenced by the choices made by their fellow students. We develop a simultaneous model of each individual's choice as a function of other students' choices. Our model of interdependence in decision making is structural and equilibrium-based. Also, the model is general enough to allow both positive and negative effects of average group choices on any individual's decision. The structure of our data enables us to identify endogenous social effects separately from exogenous or correlated effects. Specifically, in our data we see each student making choices about whether or not to apply for each job opening; exogenous and correlated effects do not vary in this sample and therefore endogenous effects are identified. We employ a two-stage procedure to address the endogeneity of choices: we estimate empirical choice probabilities in the first stage, and taste parameters for employer attributes and peer influence in the second stage. Our results show that as expected, students prefer jobs with strong employer attributes (e.g. high salary, large firm size). In addition, students are influenced by their peers' choices. However, in contrast to previous studies, we find negative (rather than positive) social effects. That is, strong attributes also make an internship employer less attractive, leading to a lower

choice probability relative to cases of zero or positive social effects. This negative social effect is consistent with congestion, i.e. students are aware that a good internship will attract the interest of more students, thus lowering the odds of getting it. We find that these negative social effects are stronger for students with more work experience and stronger GMAT scores. While positive social effect leads to concentration of choices, negative social effect helps prevent concentration.

In chapter 2, we analyze how large content-sharing websites operate for companies like Google and Yahoo. A content website provider needs to understand content users to achieve different objectives. Consumers searching content take sampling probability as given in deciding consumption, and producers are motivated by endorsement. Sampling probability is a key policy instrument. Endorsement may explain why a small number of producers generate most content. Individual behaviors alone cannot explain genesis and persistence of sampling probability and endorsement. Two distinguishing features of content—being free and non-rival preclude application of celebrated market equilibrium theory. We develop a content equilibrium from first principles. Consumer and producer can be compatible, and their interaction gives rise to endogenous sampling probability and endorsement. Inequality arises: higher quality producers always earn more endorsement and produce more content; and higher quality content is easier to find. Content system is optimal for consumer welfare despite inequality. Content system possesses a self-organization property to find equilibrium from other less desired states. We use this framework to show policy conflict may arise due to content firm's multiple identities.

## BIOGRAPHICAL SKETCH

Tong (Tony) Bao is a PhD candidate in Marketing at the Johnson Graduate School of Management at Cornell University. He received an undergraduate degree from Shanghai Jiaotong University in Shanghai in 1995, and a master degree from Simon Fraser University in Vancouver in 1998, both of which are in engineering. Tony Bao previously worked as a design engineer and product manager at JDS Uniphase Corp. and Solitan Ltd. His research interests include marketing applications of emerging internet development, applied econometrics, and industrial organization.

## DEDICATION

To Hillary and Vanessa

And to my parents: Jiping Bao, Qinhua Yao, Zhenghong Han, and Suzhen Chen

“Little Red Berries”

by Hillary Han

Little red berries, my baby's most favorite,

everytime we walk out, my baby cries:

"mama, gimme little berries."

Then her tips of fingers

painted with fruit colors.

Little berries, red, blue and yellow,

give her joys walking over path,

spring n' summer.

Now winter's come,

my baby says "mama, no more berries!"

“Yes,” I say, “but they will be back!”

## ACKNOWLEDGMENTS

I have the great fortune to cross the paths of two extraordinary professors: Sachin Gupta and Vrinda Kadiyali. They along with other faculty members in Marketing Department at the Johnson School provided me the privilege of coming to Cornell. They took the risk of becoming my advisors, and have devoted their precious time in advising my researching and teaching. I am proud to be part of academic lineage of two great researchers. Sachin's research in discrete choice model is the bread and butter of consumer choice, and Vrinda was working on structural models long before it becomes the vogue. These two strains of ideas are the foundation for this dissertation. The main theme—studying social network and social media can be interpreted as a continuation of discrete choice model and structural model. An individual is influenced by her environment. And environment is shaped by a group of individuals. This dissertation develops new equilibrium theory and applies existing equilibrium theory to empirical analysis to explore the interplay between micro behavior and macro outcome and its managerial implications. I am grateful for the opportunity to work on important issues of my time. And I must thank Sachin and Vrinda for their intellectual openness, curiosity, and vision that allow me to freely explore new areas. I benefited from numerous discussions with them, and from their painstaking efforts to improve my dissertation. Being around them for so long, research is only one part of my learning process. They are great examples of all around educators, mentors, and human beings. Hillary and I believe Vrinda is a role model for every female professional, as well as for every male professional, who juggles between a demanding career and raising children. I admire Sachin's integrity and strength. They together have showed me work ethics, ability to work with colleagues and students, and capacity of empathy. The last two years are a trying time for my family. They played a pivotal role in helping me and my family going through

this period. This dissertation would not be possible if they are not determined to see me through. I am forever in debt to them.

I have the great fortune to know Professor Dan Huttenlocher and Professor Marty Wells. Dan introduced me the fascinating social network research in computer science. His interdisciplinary knowledge greatly stimulates my research interest. It is such a treat to discuss research with Dan that I regret that I did not knock at his door more often before he became the Dean of CIS College. His enthusiasm on research and humor on about everything is contagious. I miss it.

Marty is not only my committee member, but also Hillary's Chair. I am in constant awe of his encyclopedic knowledge on theoretical statistics, and its applications in biology, law, and medicine. Marty has given Hillary and me extraordinary support over the years. When we were in difficult patches, he was there. His help has made big difference for us. We are forever in debt to him.

I had the opportunity to be Professor Vithala Rao's TA twice. I appreciate his sharing of his vast knowledge of research and teaching with me. And I am grateful for all the support he has kindly given me. I thank all other Marketing faculty members: Alice Isen, J. Edward Russo, Doug Stayman, Young-Hoon Park, Vishal Narayan, and Thomas Manoj for their teaching, mentoring, and support.

I wish to thank several mentors: PK Kannan, Martin Dresner, Bill Gruver, Luya Li, Paul Colbourne, and Berry Keyworth. I wish to acknowledge several professors who have had influences on this dissertation although I do not know them in person. They are Fernando Vega Redondo, Jon Kleinberg, Larry Blume, David Easley, and Leonard Gross.

I am grateful to our PhD field coordinator, Professor Julia D'Souza. She understands what I went through, and provides extraordinary support.

I thank Dave Crandall for his sharing of his passion for mapping the world's photos,

and for his friendship. I thank Nancy Bell for providing great help in my teaching and research. I thank my fellow PhD students, Sungho Park, Qiang Liu, Jin Pyone, Ping Zhao, Guergana Spassova, Anirban Mukherjee, Yu Yu, and Chang-Hee Park for their friendship.

I thank Brenda Wickes, Barbara Jastran, Lori Percival, and Jo Ann Molnar-Kieffer for providing great support to my family.

I wish to thank my sister and brother-in-law, Gracie Hui Han and Jun Pan for their extraordinary love and support. And I thank my parents: Jiping Bao and Qinhua Yao, and my parents-in-law: Zhenghong Han and Suzhen Chen for their unconditional love.

My co-chairs, committee members, colleagues, and friends all have been involved in a person's heroic quest to survive. There is a Talmud saying: "Whoever saves one life saves the world entire." That person is *the* world entire for her child and me.

## TABLE OF CONTENTS

Biographical Sketch.....	iii
Dedication.....	iv
Acknowledgments.....	v
Table of Contents.....	viii
List of Figures.....	xi
List of Tables.....	xii
Chapter 1 Modeling Endogenous Social Effects: An Application to MBA	
Student Summer Internship Employer Choices.....	1
1.1 Introduction.....	1
1.2 Related Literature.....	6
1.3 Model.....	8
1.3.1 Data.....	8
1.3.2 A Model of Binary Choice with Social Effects.....	10
1.3.3 Existence and Multiplicity of Equilibrium Choice Probabilities.....	15
1.4 Identification and Estimation.....	19
1.4.1 Identification.....	19
1.4.2 Estimation.....	21
1.5 Empirical Results.....	24
1.5.1 Model Selection.....	24
1.5.2 Effects of Internship Employer Attributes.....	26
1.5.3 Negative Social Effects.....	28
1.5.4 Counterfactual Analysis and Policy Implication.....	29

1.6	Conclusion .....	32
	References.....	34
Chapter 2 An Equilibrium Analysis of Online Social Content-sharing Websites		
	.....	38
2.1	Introduction .....	38
2.2	Related Literature .....	48
2.3	Basic Model of Content Equilibrium.....	52
2.3.1	Content and Individual .....	52
2.3.2	Individual Behavior and System Dynamic .....	55
2.3.3	Content Equilibrium .....	61
2.4	Endogenous Endorsement and Sampling Probability .....	70
2.4.1	Baseline Results .....	70
2.4.2	Other Forces.....	72
2.5	Extending Basic Model .....	77
2.5.1	Multiple Qualities.....	77
2.5.2	Bounds for Nonlinear Demand .....	78
2.5.3	Dependence between Consumption and Production .....	80
2.5.4	Content Quality Ranking.....	84
2.6	Managing Content System.....	85
2.6.1	Central planner.....	86
2.6.2	Profit maximization .....	86
2.7	Conclusion .....	88
	References.....	91
Appendix A.....		95

Appendix B.....	97
-----------------	----

## LIST OF FIGURES

Figure 1 Existence and Multiplicity of Choice Probability as a Function of Marginal Social Effects.....	18
Figure 2 Impact of Salary Under Negative, Zero, and Positive Social Effects .....	30
Figure 3 Inequality in Production (X-axis shows number of photos, and y-axis shows number of producers. Both are in log scale).....	41
Figure 4 Inequality in Production (X-axis shows number of photos, and y-axis shows number of producers. Both are in log scale).....	42
Figure 5 Content System Dynamics .....	61
Figure 6 Consumption Time by High Quality User as a Function of Consumption Time by Low Quality User.....	83
Figure 7 Endorsement Ratio as a Function of Consumption Time by Low Quality User.....	84

## LIST OF TABLES

Table 1 Summary Statistics .....	10
Table 2 Model Fit .....	25
Table 3 Interaction Effects of Student Demographics with Internship Employer Attributes .....	27
Table 4 Main Effects of Internship Employer Attributes on Probability of Choice ...	27
Table 5 Interaction Effects of Student Demographics with Social Influence .....	29
Table 6 Summary Statistics .....	42
Table 7 Notation and Variable Explanation .....	53

# **Chapter 1 Modeling Endogenous Social Effects: An Application to MBA Student Summer Internship Employer Choices**

## **1.1 Introduction**

Modeling consumer choice is a cornerstone of research in marketing. Marketers have developed sophisticated choice models where consumer choice is a function of firm actions, product characteristics over time, a consumer's own past choices, consumer heterogeneity etc. More recently, marketers have explored how consumer choices can be functions of choices made by other consumers (e.g., Yang and Allenby 2003, Yang et al. 2006, Manchanda et al. 2008, Van den Bulte and Joshi 2007, Hartmann forthcoming). Adding to this stream of literature, we study how MBA student choices of summer internship employers are influenced by the choices made by their fellow students. Ubiquity of information and technology has made it easier for individuals to influence, and to be influenced by, choices of other individuals. While our application is to MBA summer internship employer choices, the model we propose is general enough to be applicable to various technology-mediated environments where marketers can expect more opportunities for interdependencies in decision making.

Social effects may lead to similar observed behaviors among group members. But not all such similarities are due to social effects. Manski (1993) provides the following classification of reasons for observed interdependencies. First, individual choices might be directly influenced by the choices of individuals around them (e.g. in our case, MBA students' job choices are directly influenced by their fellow students' choices). Manski calls this interdependency the endogenous social effect, and we model this phenomenon in our paper. Second, individual choices might vary with exogenous characteristics of the group being studied (e.g. the socio-economic

composition of MBA students; these have also been termed contextual effects in the sociology literature). Choices of individuals with similar characteristics can be similar and hence be mistaken for endogenous effects. Finally, individual choices in a group are correlated with choices of others in the group because they face similar environments (e.g. MBA students have all taken the same required classes and might choose to apply to similar firms based on their professors' recommendations). Correlated effects in the data can also be mistaken for endogenous effects. Endogenous and exogenous effects both describe the social environment of decision makers, whereas correlated effects capture non-social phenomena that cause interdependent choices.

Distinguishing between different reasons for interdependence of individual decisions is important. Consider an intervention wherein potential employers or business-school administrators provide more detailed information to a certain group of students to enable better decision making. The impact of this detailed information provision will be felt in two ways. First, it will affect the decisions of the students who receive direct communication from employers or school administrators. Second, employer/administrator intervention will also have a “multiplier”<sup>1</sup> effect given other students' decisions will then be affected via endogenous effects. Exogenous and correlated effects do not lead to such multiplier effects. Therefore, researchers run the risk of misestimating the impact of an intervention if they do not isolate the endogenous effects.

An important distinction within endogenous social effects is whether these effects are positive or negative. With positive social effects, an agent's utility from a product or activity increases when peers make the same choice. Such positive interaction can lead to concentration of choices on particular alternatives, e.g., in

---

<sup>1</sup> A formal definition of social multiplier can be found in Glaeser and Scheinkman (2003) .

consumer purchases (Yang and Allenby 2003, Hartmann forthcoming), in TV viewing (Yang et al. 2006), and in drug prescription behavior (Manchanda et al. 2008, Van den Bulte and Joshi 2007). One common explanation for positive social effects is conformity. Less discussed but equally important is negative social interaction wherein an agent's utility decreases in peers' choices. One explanation for negative interaction is status seeking. Consumers purchase exclusive goods to achieve prestige (Amaldoss and Jain 2003). Another explanation is congestion (Manski 2000 calls it constraint interaction). Multiple agents share a resource, and their choices collectively determine the availability of the resource. Constraint interaction is especially plausible in our context where students might fear getting crowded out of popular jobs. Therefore they might not apply to popular jobs they would have found attractive in the absence of congestion effects. Of course, popular jobs are likely to have features that are attractive (e.g. higher salary) and therefore, any student will have to balance job attractiveness with the likelihood of greater competition from other students for this job. Positive and negative social interactions are not mutually exclusive. When a diner considers a busy restaurant, he may both be attracted by the popularity while being concerned about table availability. So the estimated social interaction inferred from empirical data is the net effect of both positive and negative interactions.

The distinction between positive and negative endogenous social effects is an important one from a policy perspective. With negative endogenous effects, a prestigious firm might get fewer applicants than it would without negative effects. Therefore, the firm needs to make a more concerted effort to seek out the most desirable applicants and convince them to apply. Econometrically, a researcher may over-estimate (under-estimate) the effect of product characteristics in the presence of positive (negative) social interaction if the measurement does not account for social

interactions and allow for both positive and negative social interactions in the model.

Our empirical setting of internship employer choices of MBA students at a leading university is well-suited for the study of endogenous social effects. Most MBA programs have various planned avenues for social interactions within (and outside) the student body (e.g. happy hours, club activities, etc.). There are several benefits from such social interactions. First, through such interactions, students acquire industry and company-specific information from their peers with experience in these industries and from peers looking for similar jobs. Second, such interactions help students establish career contacts. This is especially true for students who switch careers (e.g. scientists wanting to become marketers) or even switch firms (e.g. moving from a small market research company to a larger company). Given our empirical application is the summer internship search, and these internships often play a central role in determining students' full-time job offers, social interactions where information and contacts can be gleaned are likely to be important. In other words, social interactions in the context of summer internship search are consequential, given the significant economic consequences of making a sound choice. Therefore, it appears plausible that social interactions lead to the existence of endogenous social effects in this choice setting; a complete model of student internship choice should account for these.

As we will explain in detail in Section 1.3, our model builds on and extends existing empirical literature in the following ways. We model endogenous social effects in a multi-person setting, without placing any restrictions on who might be influenced by whom, instead modeling any one person's decision as potentially being influenced by all others' decisions. To do this, we have to relax the assumption of perfect information about other individuals' preferences, an assumption that is commonly made in extant literature. This assumption also makes our model applicable to decision-making settings where decision-makers are unlikely to have full

information on other decision-makers, and where decision dependencies are more than two-person wide. Our model is general enough to allow for both positive and negative social effects, and allow for heterogeneity in the intensity of social effects across students. Our methodological contribution is to empirically implement Blume (1993) and Brock and Durlauf's (2001) theoretical models of endogenous social effects. We employ a two-stage procedure (Manski and Wise 1983, Hotz and Miller 1993, Bajari et al. forthcoming) to estimate a system of simultaneous choice equations for all students, where each student's choice is a function of others' choices. Importantly, in our data we are able to identify the endogenous social effects separately from the two confounding effects discussed previously. Our empirical setting involves a single group in which exogenous and correlated effects do not vary in the data.

Substantively, we find that endogenous social effects exist and vary by student characteristics. As expected, strong internship employer attributes (e.g. high salary, large firm size) attract the students. However, in contrast to previous studies, we find negative (rather than positive) social effects. That is, strong attributes also make an internship employer less attractive via the social effect, leading to a lower choice probability relative to cases of zero or positive social effects. This negative social effect is consistent with congestion, i.e. student choices for an internship collectively determine the likely availability of the internship to any student. Stronger students (higher GMAT scores, more work experience) have larger negative social effects than other students.

The rest of the paper is organized as follows. In section 1.2, we describe the literature in this area. In section 1.3, we describe the data and present the proposed model. In section 1.4, we discuss model identification and the estimation procedure. In section 1.5, we present empirical results. Section 1.6 concludes.

## 1.2 Related Literature

In this section, we examine research in economics and marketing on modeling social effects. Social effects have been used in economics to explain a variety of phenomena, including crime rates in neighborhoods (Glaeser et al. 1996) and army desertions (DePaula 2009).

As noted, the literature recognizes that these effects can be positive, (e.g. conforming behavior (Jones 1984)), or negative (e.g. status seeking (Frank 1985)). Akerlof (1997) describes conformity as individuals moving towards each other in a social space. Benefits that individuals receive from conformity include love, friendship, and monetary payoffs. Individuals may distance themselves from others to avoid negative consequence of closeness, such as jealousy and envy of friends, relatives, and neighbors. Another seminal paper in this area is Blume (1993), who presents one of the earliest social interaction models built on individual-level decision processes. His model draws from mean-field theory in statistical mechanics (see Stanley 1971 for a description). In a system with many interacting individuals, the effects of the system on an individual can be captured by average behavior of the group (mean field). Blume (1993) and Brock and Durlauf (2001, 2002) apply mean field theory to a discrete choice model. They assume that economic agents form expectations on group behavior, and such expectation is consistent with actual outcome.<sup>2</sup>

In one of the early papers in this stream in marketing, Yang and Allenby (2003) examine whether there are interdependencies in consumers' automobile purchase decisions. Using data on purchases of midsize cars, they specify a consumer's choice as a function of car prices, features of cars, and income. They use

---

<sup>2</sup> A related model is supermodular game model (Topkis 1979, Milgrom and Roberts 1990). For an example of using supermodular game solution in social effects, see Krauth (2006).

a spatial error model to allow the error terms in the utility function to be correlated across individuals. The strength of the correlation depends on geographic or social distance (e.g. similar demographics) between individuals. They find that geographic distance explains the correlation better than demographic distances. Yang et al. (2006) examine spouses' television program choices. They model a viewer's utility as a linear function of the partner's utility by using a spatial autoregressive model. An important distinction between these two papers and ours is that we are able to identify endogenous social effects separately from exogenous and correlated effects. A further point of contrast is that our model is equilibrium-based, as we will discuss in further detail in section 1.3.

Manchanda et al. (2008) consider social effects in the adoption of a new product. They model physicians' adoption of a new drug as a function of average prescription behavior of the peer group in the previous time period, assuming a unidirectional social effect. In other words, a logical condition is imposed in the model that earlier adopters influence those who have not yet adopted, and any reciprocal influence is irrelevant because the early adopters have already made their choices. Nair et al. (2006) also model physician peer effects. A unique feature of their data is that physicians self-identify their social networks. This allows the authors to assume that opinion leaders influence others, but are not influenced by others. This unidirectionality or asymmetry of social effects is similar to the Manchanda et al. (2008) formulation. Our formulation, wherein any student's choices can influence and be influenced by others' choices is different from both the Nair et al. and Manchanda et al. formulations. Our model is therefore applicable to situations where self-identification of social networks is unavailable, or where the researcher believes that latent networks exist in addition to self-identified ones.

The paper closest to ours is Hartmann (forthcoming) who studies positive

endogenous social effects in a two-person golf-partner pair setting and finds evidence for endogenous effects. Our model is similar in that it focuses on endogenous social effects and can be interpreted as the equilibrium outcome of a game. However, in contrast with Hartmann, our model applies to multi-person settings, allows for both positive and negative endogenous social effects, and relaxes the assumption of perfect information about everyone else's tastes.

### **1.3 Model**

In our application to MBA students' summer internships application choices, we treat the entire first-year class applying for summer internships as the relevant group that potentially influences the decisions of each student. While each individual in the class can belong to multiple sub-groups for the purpose of internship application process (e.g. clubs for specific career paths, ethnic/gender-based social networks, etc.), we do not have *a priori* information that would allow us to choose particular sub-groups as being especially relevant to the decision being modeled. Furthermore, the class size of 221 students is relatively small compared with several other top-ranked MBA programs, making it reasonable to assume that social interaction might occur among the entire class<sup>3</sup>. As the model is specified for our data context, we first discuss the data and subsequently present the model.

#### **1.3.1 Data**

In most leading business schools, a number of companies recruit for summer internship positions on campus each year. These internships are heterogeneous in many ways. For example, they differ by industry, size, compensation, etc. In most

---

<sup>3</sup> An extension of our model is to relax the assumption of the entire class being the relevant social group, and instead allow for endogenous group formation (e.g., Zanella 2007). We discuss this in the concluding section of the paper.

two-year MBA programs, students work as interns in their first summer, and apply for permanent jobs in the second year. We collect data on resume submissions for internships in 2006 by the first-year class at a major business school. Recruitment is formally organized through the Career Management Center (CMC). The CMC posts all available internship opportunities in its database. Each student then chooses whether or not to submit a resume to each posted internship employer.

Descriptive statistics of the 221 students in the first-year class are in Table 1. Demographic information includes gender, nationality, and work experience. The class profile reflects a diverse student body with about 30% female students, 26% international students, and average work experience of around 5 years. We report two sets of indicators of academic performance: GPA, and GMAT scores on verbal, quantitative, and analytical writing sections. Students have higher quantitative scores than verbal scores on average, and the standard deviations for the two are about the same. The variation in student characteristics suggests the need to model potential heterogeneity in student tastes.

One hundred and eight firms recruit interns through CMC.<sup>4</sup> For each firm, each student makes a binary decision- whether or not to submit a resume. Therefore, we consider each firm as a separate choice observation with each student's choice set including the decisions to submit a resume or not. We include the fifty eight publicly traded firms in our sample since their information is publicly available.<sup>5</sup> Using CMC's categorization, the fifty-eight internship positions included in our sample belong to three industry categories: manufacturing, consulting, and finance. Since we have only four consulting firms in the data, we reclassify all firms in the dataset into two categories: finance vs. non-finance, by including the four consulting firms in the non-

---

<sup>4</sup> Some firms offer multiple internship positions, and some students also get internships through off-campus channels.

<sup>5</sup> With regard to the estimation of social effects and students' choice behavior, we believe the omission of privately held firms from our sample should not create a bias.

finance category. Thirty-five percent of internship positions are in the finance category. We include as a covariate average salaries received by previous graduates who joined a recruiting company in a permanent position. Annual sales of the recruiting company serve as a proxy for company size. We observe large heterogeneity in company size.

Table 1 Summary Statistics

Variable	Mean	Standard Deviation
<i>Student Characteristics</i>		
Gender (1=male, 0=female)	0.70	0.46
Nationality (1=domestic, 0=international)	0.74	0.44
Work experience (in number of years)	4.7	2.5
GMAT verbal score	37.73	4.70
GMAT quantitative score	44.57	4.64
GMAT Analytical writing score (AWA)	4.86	0.69
GMAT total score	674.5	49.70
GPA	3.42	0.31
<i>Company Characteristics</i>		
Industry (1=finance, 0=non-finance)	0.35	0.48
Annual Salary of previous full-time recruits (in \$1000)	93.26	8.95
Annual sales (in million \$)	45.75	51.83

### 1.3.2 A Model of Binary Choice with Social Effects

In this section, we describe a binary choice model with endogenous social

effects for the data discussed previously. Let  $I, J$  denote the total numbers of MBA students and summer internship employers. Each individual (a student in our empirical application)  $i \in \{1, \dots, I\}$  decides whether to apply to a summer internship  $j \in \{1, \dots, J\}$ . Let  $a_{ij} \in \{0, 1\}$  denote the binary decision with 1 indicating the decision to apply, and 0 otherwise. The action  $a_{ij}$  reveals the latent utility  $u_{ij}$  that an application to internship  $j$  brings to individual  $i$  such that

$$a_{ij} = \begin{cases} 1 & \text{if } u_{ij} > 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

Let  $A = \{\{0, 1\} \times \dots \times \{0, 1\}\}$  denote the set of actions of all  $I$  individuals. Let  $a_j = (a_{1j}, \dots, a_{Ij}) \in A$  denote the joint actions of all individuals with respect to internship employer  $j$ . Let  $a_{-ij}$  denote the joint actions of all individuals except  $i$  with respect to internship employer  $j$ . Let  $X_j \in R^K$  denote  $K$  attributes of employer  $j$ . Let  $\xi_j \in R$  denote an employer-specific variable. Then the latent utility is specified as follows:

$$u_{ij} \equiv \beta_0 + \beta_{0i} + \xi_j + X_j \beta_i + \frac{\gamma_i}{I-1} \sum_{l \neq i} a_{lj} + \epsilon_{ij} \quad (2)$$

where  $\beta_0$  denotes an overall intercept term,  $\beta_{0i}$  denotes an individual  $i$  specific intercept term,  $\beta_i \equiv (\beta_{i1}, \dots, \beta_{iK})$  represents coefficients for the attributes of internship employers,  $\gamma_i$  is the coefficient for endogenous social effects, and  $\epsilon_{ij}$  is a private utility shock. The private utility shock  $\epsilon_{ij}$  is unknown to the other individuals or the researchers. Let  $\Theta \equiv (\beta_0, \beta_{0i}, \beta_i, \gamma_i)$  denote the parameter set. The covariates  $X_j$  are assumed to be observed by individuals as well as researchers and are included in the model to control for differences in attractiveness of internship employers. The employer-specific effects  $\xi_j$  represent employer characteristics that are common knowledge to all individuals, but unobserved to researchers. These might include, for

instance, reputation of employers for making permanent job offers upon successful completion of the summer internship. Note that the individual specific intercept  $\beta_{0i}$  includes any time cost of preparing and submitting a resume.

The term  $\sum_{l \neq i} a_{lj} / (I-1)$  in equation (2) expresses the average choice of the group except individual  $i$ . Individuals only have access to their own information in the Career Management Center database. So, individuals do not directly observe the average choice of other individuals in their class. Nor do they have information on past years' choices. Instead, they form an expectation on the average choice. Brock and Durlauf (2001) refer to the term  $\gamma_i \sum_{l \neq i} a_{lj} / (I-1)$  as social utility, and  $\beta_0 + \xi_j + \beta_{0i} + X_j \beta_i + \epsilon_{ij}$  as private utility. In Manski's (1993) classification  $\gamma_i \sum_{l \neq i} a_{lj} / (I-1)$  is termed endogenous effects.

We model the heterogeneity in individual taste parameters  $\beta_i$  as a function of individual characteristics such as demographics, work experience, and academic performance. Let  $Z_i \in \mathbb{R}^M$  denote a row vector of these  $M$  covariates. Let  $\bar{Z}_i$  denote a  $K+1$  by  $(K+1) \times M$  matrix. In the  $r^{\text{th}}$  row, put  $Z_i$  between column  $(r-1) \times M + 1$  and column  $r \times M$ . All the other elements in  $\bar{Z}_i$  are zero. The heterogeneous tastes,  $\omega_i \equiv (\beta_i, \gamma_i)$  are specified as

$$\omega_i = \alpha_o + \bar{Z}_i \alpha + \psi_i \quad (3)$$

where  $\alpha_o = (\alpha_{o,1}, \dots, \alpha_{o,K+1})^T$  is a population level intercept for the  $K$  employer attributes and social effects, and  $\alpha \equiv (\alpha_{1,1}, \dots, \alpha_{M,1}, \dots, \alpha_{1,K+1}, \dots, \alpha_{M,K+1})$  denotes population level coefficients. We assume that  $\psi_i$  follows an iid Normal distribution such that  $E(\psi_i) = 0$ ,  $E(\psi_i \psi_i) = \rho_i$ , and  $E(\psi_i \psi_j) = 0$  for  $i \neq j$ .

With regard to the endogenous effects, an alternative approach is to specify them as a function of peer *utilities* (Yang et al. 2006) instead of peer *actions*. The two specifications (peer utilities versus actions) make different assumptions about the information available to decision makers. We make the following assumption about

the information structure of an individual's knowledge of others' utilities: we assume that individual 2's private utility shock is known to individual 1. In our empirical application, the class size of 221 is too large to justify this assumption, hence we assume that  $i$ 's utility shock is indeed private to  $i$  and not known to peers.

We make the following assumption about the private utility shock.

**Assumption 1:** The private utility shock  $\epsilon_{ij}$  is independently and identically distributed across individuals and internships. Let  $F(\epsilon_{ij})$  denote the cumulative density function of  $\epsilon_{ij}$ . We assume that group members know the distribution  $F(\epsilon_{ij})$ . Note that independence in the private shock is a reasonable assumption because we explicitly model correlation in the non-private or social portion of utility.

Next we discuss the individual decision making process. Equations (1) and (2) together define a decision function  $D_i(X_j, \xi_j, \epsilon_{ij} | \Theta, F(\epsilon_{ij})) : \mathbb{R}^K \times \mathbb{R} \times \mathbb{R} \rightarrow \{0,1\}$ , mapping from internship employer attributes (both observed and unobserved to researchers) and  $i$ 's private utility shock, to  $i$ 's action. Note that the peers' utility shocks  $\epsilon_{lj}$  for  $l \neq i$ , are unknown to  $i$ . So they do not enter in to the decision function  $D_i(X_j, \xi_j, \epsilon_{ij} | \Theta, F(\epsilon_{ij}))$ . Recall that endogenous effects are specified as a function of peer choices. The decision function implies that  $i$  does not know her peers' decisions exactly. Instead, by assumption 1, she can estimate the expectation of peer decisions because she knows the distribution  $F(\epsilon_{ij})$ .

The probability that individual  $i$  chooses to apply to internship employer  $j$  is computed by integrating the decision function over  $F(\epsilon_{ij})$ :

$$P_{ij}(a_{ij} = 1 | X_j, \xi_j) = \int D_i(X_j, \xi_j, \epsilon_{ij} | \Theta, F(\epsilon_{ij})) dF(\epsilon_{ij}) \quad (4)$$

Taking expectation of the utility in Eq. (2) and by the independence in assumption 1, we obtain the expected utility for individual  $i$ :

$$\tilde{u}_{ij} = \beta_0 + \beta_{0i} + \xi_j + X_j \beta_i + \frac{\gamma_i}{I-1} \sum_{l \neq i} P_{lj}(a_{lj} = 1 | X_j, \xi_j) + \epsilon_{ij} \quad (5)$$

where  $\tilde{u}_{ij}$  denotes the expected utility. Expected utility maximization implies

$$P_{ij}(a_{ij} = 1 | X_j, \xi_j, \epsilon_{ij}) = \Pr(\epsilon_{ij} | \beta_0 + \beta_{0i} + \xi_j + X_j \beta_i + \frac{\gamma_i}{I-1} \sum_{l \neq i} P_{lj}(a_{lj} = 1 | X_j, \xi_j) + \epsilon_{ij} > 0) \quad (6)$$

Next we make a distributional assumption on  $\epsilon_{ij}$  to obtain an analytical form for the choice probability.

**Assumption 2:**  $F(\epsilon_{ij})$  is a Type I extreme value distribution, i.e. the density function

$$\text{is } f(\epsilon_{ij}) = \frac{e^{-\epsilon_{ij}}}{(1 + e^{-\epsilon_{ij}})^2}.$$

To simplify notation, we write  $P_{ij}(a_{ij} = 1 | X_j, \xi_j, \epsilon_{ij})$  as  $P_{ij}(a_{ij} = 1)$ . With assumption 2, we obtain the following choice probabilities:

$$P_{ij}(a_{ij} = 1) = \frac{\exp(\beta_0 + \xi_j + \beta_{0i} + X_j \beta_i + \frac{\gamma_i}{I-1} \sum_{l \neq i} P_{lj}(a_{lj} = 1))}{1 + \exp(\beta_0 + \xi_j + \beta_{0i} + X_j \beta_i + \frac{\gamma_i}{I-1} \sum_{l \neq i} P_{lj}(a_{lj} = 1))}, \quad \forall i \in \{1, \dots, I\} \quad (7)$$

The choice probability in Eq. (7) resembles that of a standard logit choice model. The difference is that choice probabilities of peers enter on the right hand side of equation (7). There are  $I$  equations for each internship  $j$ , and they define an implicit function for choice probabilities  $(P_{1j}, \dots, P_{Ij})$ .

The proposed model has a game-theoretic interpretation if we treat the applications to each internship employer  $j$  as a game (in other words, we observe  $J$  games in total). Consider the game played by all individuals applying for a particular internship. We will describe a single incomplete information game or Bayesian game. The strategic form  $\Gamma_j$  of a Bayesian game  $j$  is defined by

$\Gamma_j \equiv \{S, (a_{ij})_{i \in S}, (u_{ij}(a_{ij}, a_{-ij}, \epsilon_{ij}))_{i \in S}, (\epsilon_{ij})_{i \in S}, p\}$  where  $S = \{1, \dots, I\}$  denotes the set of individuals; and  $\epsilon_{ij}$  denotes the type of individual  $i$ ;  $a_{ij}$  denotes the strategy set;  $u_{ij}$  denotes payoff function or utility; and  $p$  denotes joint probability distribution for types. The type  $\epsilon_{ij}$  expresses private information only available to  $i$ . Per assumption 1, the joint probability distribution on  $(\epsilon_{1j}, \dots, \epsilon_{Ij})$  is common knowledge to all individuals. Then an individual  $i$  can compute the expected payoff (5) using the joint probability  $p$ . The choices  $a_{ij}$ s are pure strategies in the Bayesian game in the sense that each  $a_{ij}$  maximizes  $i$ 's expected payoff given peer choices  $a_{-ij}$ .

### 1.3.3 Existence and Multiplicity of Equilibrium Choice Probabilities

Social interaction implies that choices are interdependent. The interdependency is captured by the system of equations (7). In another words, the choice probabilities  $p_{ij}$  are solutions to the equations (7). Note that if social effects do not exist, the equations (7) reduce to  $I$  independent equations

$$P_{ij}(a_{ij} = 1) = \frac{\exp(\beta_0 + \xi_j + \beta_{0i} + X_j \beta_i)}{1 + \exp(\beta_0 + \xi_j + \beta_{0i} + X_j \beta_i)}, \quad \forall i \in \{1, \dots, I\} \quad (8)$$

Each of these equations is a standard single agent choice model. In this model existence or multiplicity of equilibria are not concerns. These factors are also not a concern for unidirectional social interactions, i.e. interaction between a leader and followers. The leader's choice is modeled as in a regular choice model, given she does not care about followers' choices. The leader's choice can be treated as exogenous to followers' choices. Examples of unidirectional social effects are time lagged social effects (Manchanda et al. 2008), and asymmetric social effects between opinion leaders and followers (Nair et al. 2007).

Multiplicity and existence become issues only when agents have mutual

influences. Note that Eq. (7) defines a mapping  $G:[0,1]^I \rightarrow [0,1]^I$  such that  $P_j = G(P_j)$ , where  $P_j \equiv (P_{1j}, \dots, P_{Ij})$ . Since  $G$  is continuous in  $P_j$ , we can show the existence of at least one solution to  $P_j$  by Brouwer's fixed point theorem. It is harder to characterize the number of solutions or equilibria; multiple solutions may exist for this non-linear equation system (Brock and Durlauf 2001).

We illustrate the multiplicity of equilibria in such a system by a simulation. For this illustration we assume all individuals have the same taste parameter  $\beta$  in employer attributes, and  $\gamma$  in endogenous social effects. So we have  $\beta_i = \beta$  and  $\gamma_i = \gamma$ . We further assume that  $\beta_0 = \beta_{0i} = \xi_j = 0$  without loss of generality. Then Eq. (7) reduces to

$$P_{ij}(a_{ij} = 1) = \frac{\exp(\beta X_j + \frac{\gamma}{I-1} \sum_{l \neq i} P_{lj}(a_{lj} = 1))}{1 + \exp(\beta X_j + \frac{\gamma}{I-1} \sum_{l \neq i} P_{lj}(a_{lj} = 1))}, \quad \forall i \quad (9)$$

Further, we assume that the social utility can be approximated by  $\gamma \sum_{l \neq i} P_{lj}(a_{lj} = 1) / (I-1) \approx \gamma \sum_l P_{lj}(a_{lj} = 1) / I$ . This assumption is reasonable when the number of individuals is large. Under the aforementioned assumption that the taste parameters  $\beta$  and  $\gamma$  are the same for all individuals, the variation across individuals' choices comes from the private utility shock  $\epsilon_{ij}$ . Note that  $\sum_l P_{lj}(a_{lj} = 1) / I$  is the aggregate choice probability, which can be interpreted as the percentage of the class that submits resumes to internship  $j$ . By the assumption of identical distribution in Assumption 1, the choice probabilities are identical, i.e.  $P_{ij}(a_{ij} = 1) = P_{i'j}(a_{i'j} = 1)$ , for all  $i$  and  $i'$ . To simplify notation, let  $\sigma \equiv P_{ij}(a_{ij} = 1)$ . Then  $\sigma = P_{ij}(a_{ij} = 1) = \sum_l P_{lj}(a_{lj} = 1) / I$ . So Eq. (7) implies

$$\sigma = \frac{\exp(\beta X_j + \gamma \sigma)}{1 + \exp(\beta X_j + \gamma \sigma)} \quad (10)$$

We define two functions

$$\begin{aligned} f_1(\sigma) &= \sigma \\ f_2(\sigma) &= \frac{\exp(\beta X_j + \gamma \sigma)}{1 + \exp(\beta X_j + \gamma \sigma)} \end{aligned} \tag{11}$$

In Figure 1 (various panels) we examine how the number of solutions to  $\sigma$  changes as a function of marginal social effects  $\gamma$ , holding  $\beta X_j$  constant. That is, we plot the two functions in (11) against the aggregate choice probability  $\sigma$ . The points where the two functions intersect (i.e. fixed points) are solutions to the aggregate choice probability in Eq. (10). In Figure 1, the horizontal axis is the choice probability in Eq. (10). The vertical axis displays in Eq. (11). Note that is the diagonal line from the origin in each graph. The intersections of the two curves and are the fixed points. Notice the number of fixed points increases in the marginal social effect. We observe that when social effects  $\gamma$  are negative ( $\gamma < 0$  in Figure 1.a), or there are no social effects ( $\gamma = 0$  in Figure 1.b), the solution is unique.

**Proposition 1.1:** The solution to Eq. (10) is unique when  $\gamma \leq 0$

Proof: See Appendix A.

Multiple solutions of choice probability appear under positive social effects. As  $\gamma$  increases and passes a threshold ( $\gamma \approx 3$  in Figure 1.f), we observe two solutions. And as  $\gamma$  increases further, we observe three solutions ( $\gamma = 3.5$  in Figure 1.g). If an individual expects her peers to make a particular choice with high probability, her choice probability is also high. If she expects her peers' choice probability to be low, her choice probability is also low. Such coordination makes multiple solutions possible. Brock and Durlauf (2001) provide a proof on the maximum number of solutions (see their Proposition 2).

The multiplicity of equilibria does not occur in the presence of negative social effects. When an individual expects peer choice probability is high, she will lower her

own probability, and vice versa. There is a breakdown of the coordination of choices (to reach the same choice as peers) that is present in positive social effects scenario. Similarly, when there is no social effect, the feedback mechanism necessary for multiple solutions no longer exists. So choice probability is unique for negative or zero social effects.

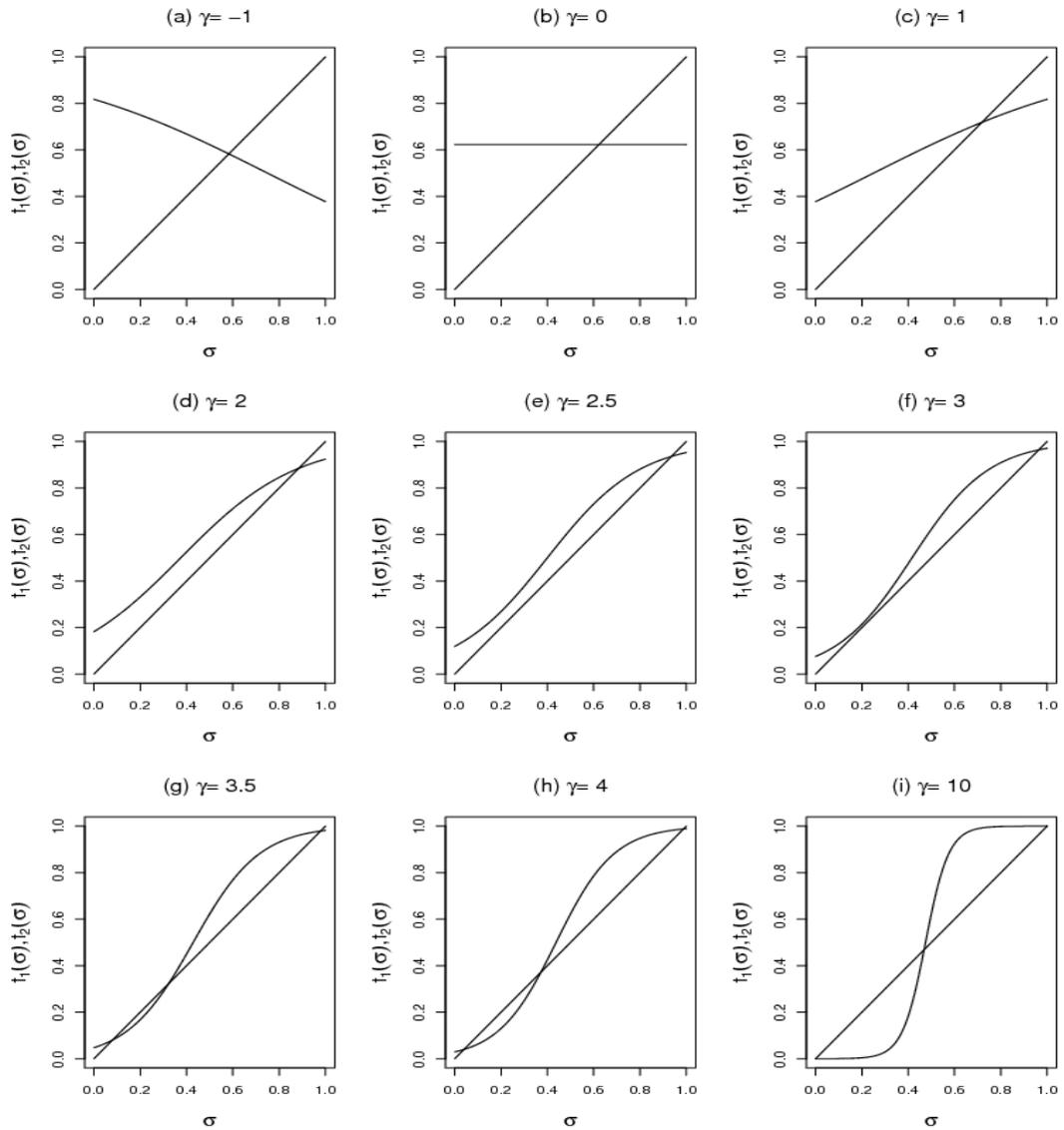


Figure 1 Existence and Multiplicity of Choice Probability as a Function of Marginal Social Effects.

The social multiplier phenomenon is also illustrated in the simulation. Intuitively, the more favorable an internship  $j$ 's attributes  $X_j$ , the more likely individual  $i$  is to apply to this internship. Individual  $i$ 's decision then contributes to the average group choice, and the average choice in turn influences peer decisions. Note that in the simulation we assume that individuals have the same taste parameters  $\beta$  and  $\gamma$ ;  $X_j$  is invariant across these simulations. When there are no social effects (i.e.  $\gamma=0$  in Figure 1.b), the aggregate choice probability is 0.5. When there are positive social effects, for example when  $\gamma=1$  in Figure 1.c, we observe that aggregate choice probability  $\sigma$  is greater than 0.6. As social effects cross the threshold of multiple solutions ( $\gamma \approx 3$  in Figure 1.f), the largest solution of aggregate choice probability continues to grow. However, the other two solutions of aggregate choice probability are lower than 0.5 in Figure 1.g. So the increase in the aggregate choice probability is entirely due to the social effects, thereby illustrating the social multiplier effect.

## **1.4 Identification and Estimation**

### **1.4.1 Identification**

Multiple solutions (equilibria) of choice probability imply that the mapping from parameter space to sample space may not be one-to-one. Existing literature typically assumes a one-to-one relationship exists. For example, Hartmann (2008) assumes that the best (Pareto optimal) equilibrium is played by individuals. In our data, unique equilibrium is likely because of the larger heterogeneity in preferences due to large group size. Brock and Durlauf 2001 (Proposition 2) shows that multiple equilibria appear when the social utility is relatively strong compared to private utility.

In a larger group, group members are likely to have differences in tastes and private utility shocks. So, the social effects tend to be weaker due to stronger private

utilities. The simulation in Bajari et al. (forthcoming) shows that the occurrence of multiple equilibria decreases in the number of group members. In a group with five members, single equilibrium appears ninety three percent of the time. Therefore, in our empirical application with 221 students, it is reasonable to assume unique equilibrium. Furthermore, our empirical results show negative social effects. We have proved in proposition 1.1 that unique equilibrium exists in a negative social effects scenario.

Next, we discuss identifying endogenous effects distinctly from exogenous and correlated effects. The challenge of identifying endogenous effects is documented by Manski (1993) in the context of a linear model. One source of difficulties in identification is the existence of multiple groups in the data. To the best of our knowledge, all existing papers model multiple groups. For example, in Manchanda et al. (2008), each physician identifies their social network, and these vary by physicians. We model the entire MBA class as the relevant social group that could influence choice. Formally, let  $y$  denote characteristics of the group and  $w_i$  denote covariates of individual  $i$  in the group. Then utility is expressed as

$$u_{ij} = \beta_0 + \beta_{0i} + \xi_j + \beta_i X_j + \frac{\gamma_i}{I-1} \sum_{l \neq i} a_{lj} + \delta_i E(w_i | y) + \zeta_i y + \epsilon_{ij} \quad (12)$$

This utility specification is the same as in Eq. (2) with two additional terms -  $\delta_i E(w_i | y)$  and  $\zeta_i y$  - which are used in existing papers to express exogenous effects and correlated effects respectively. For example,  $y$  can be an indicator of a business school. Students attending the same business school face the same institutional environments, such as facing the same travel distance to interviews at corporate sites. So if  $\zeta_i \neq 0$ , then correlated effects exist.  $w_i$  usually describes social and economic status, for example, racial composition in the landmark Colman et al. (1966) study on

school integration. If  $\delta_i \neq 0$ , exogenous effects exist.

**Assumption 3:** The group attributes  $y$  and covariates  $E(w_i | y)$  are invariant over employer  $j$ , for all  $j \in J$ .

Under assumption 3,  $\beta_{0i}$ ,  $\delta_i E(w_i | y)$ , and  $\zeta_i y$  are not separable. Therefore we obtain

$$u_{ij} = \beta_0 + \beta_{0i} + \xi_j + \beta_i X_j + \frac{\gamma_i}{I-1} \sum_{l \neq i} a_{lj} + \epsilon_{ij}$$

In summary, the exogenous and correlated effects collapse into the individual specific parameter  $\beta_{0i}$ . They cannot be separately identified. But they do not affect identification and estimation of the effect in which we are most interested, namely, the endogenous effects.

The other identification assumptions are as follows. First, the global intercept in utility and dummy variables for firm fixed effects cannot all be identified. We normalize the fixed effects of firm one to zero. So the estimate of firm  $j$ 's fixed effects should be interpreted as the difference between firm  $j$  and firm one. Second, since we have an overall intercept and individual specific intercepts, individual specific intercepts cannot all be identified. So we normalize individual specific intercept of individual one. Third, the specification in equation (3) implies that  $Z_j \alpha_o$  and  $\xi_j$  are not both identified. We discuss how to resolve this issue in the estimation section. Fourth, we use scores on GMAT verbal, quantitative, and analytical writing sections, and omit the total score to avoid possible collinearity.

### 1.4.2 Estimation

The proposed binary choice model with social effects results in the simultaneous system of non-linear equations defined in (9). We adopt the two-stage estimation methods used in Manski (1983), Hotz and Miller (1993) and Bajari et al. (forthcoming). Let  $V_{ij}$  denote the deterministic part of the expected utility  $\tilde{u}_{ij}$ , defined

in Eq. (5).

$$V_{ij} \equiv \beta_0 + \beta_{0i} + \xi_j + X_j \beta_i + \frac{\gamma_i}{I-1} \sum_{l \neq i} P_{lj}(a_{lj} = 1 | X_j, \xi_j) \quad (13)$$

The idea is that, in the first stage, we estimate choice probabilities  $P_{ij}(a_{ij} = 1)$  and infer  $V_{ij}$ . Then we estimate the population level parameters in the second stage. We observe each individual make fifty eight binary choices (to apply or not for the internship), one for each internship employer  $j$ . In the first stage we apply non-parametric methods to separately estimate each individual's empirical choice probability given a particular employer. For a given internship employer  $j$ , the empirical choice probability is the conditional mean:

$$P_{ij}(a_{ij} = 1 | \xi_j, X_j) = E(a_{ij} = 1 | \xi_j, X_j) \quad (14)$$

The employer-specific effects  $\xi_j$  in the model represent characteristics that are common knowledge to all group members, but unobservable to researchers. Researchers usually assume that the internship fixed effect  $\xi_j$  is a function of observed covariates (Newey 1994, Bajari et al. forthcoming); we make a similar assumption as follows.

**Assumption 4:**  $\xi_j$  is a non-linear smooth function of  $X_j$ .

Therefore, we have

$$P_{ij}(a_{ij} = 1 | \xi_j, X_j) = P_{ij}(a_{ij} = 1 | X_j) = E(a_{ij} = 1 | X = X_j) \quad (15)$$

The goal is to estimate the choice probability  $P_{ij}$ , which is a non-linear function of covariates  $X_j$ . We use a flexible functional form that can approximate the unknown non-linear function. Let  $E(a_{ij} | X_j) \equiv m_i(X_j)$  denote conditional mean, where  $m_i(\cdot)$  is an unknown non-linear function. This is equivalent to the following stochastic equation:

$$a_{ij}(X_j) = m_i(X_j) + \zeta_{ij} \quad (16)$$

where  $E(\zeta_{ij} | X_j) = 0$ ,  $E(\zeta_{ij}\zeta_{i'j'} | X_j) = 0$  for all  $i' \neq i, j' \neq j$ . To illustrate the estimation, we consider our empirical setting.  $X_j \equiv (X_{j1}, X_{j2}, X_{j3})$  where  $X_{j1}$  is a binary variable indicating whether the employer is in finance industry,  $X_{j2}$  and  $X_{j3}$  are continuous variables measuring salary and firm size respectively. Then equation (16) is equivalent to

$$a_{ij}(X_{j1}, X_{j2}, X_{j3}) = m_{i1}(X_{j2}, X_{j3})I\{X_{j1} = 0\} + m_{i2}(X_{j2}, X_{j3})I\{X_{j1} = 1\} + \zeta_{ij} \quad (17)$$

where  $m_{i1}(\cdot)$  and  $m_{i2}(\cdot)$  are two nonlinear functions. To enable estimation, we will approximate them with basis functions. Consider  $m_{i1}(X_{j2}, X_{j3})$ . We can construct basis functions of a single variable, for example,  $\{h_1(X_{j2}), \dots, h_{k_1}(X_{j2})\}$  and  $\{l_1(X_{j3}), \dots, l_{k_2}(X_{j3})\}$  where  $\{h_i\}$  and  $\{l_i\}$  are B-spline basis functions of single variables (Eilers and Marx 1996). Then the tensor product spline is defined as  $h_i(X_{j2})l_m(X_{j3})$ . The approximation for  $m_{i1}(X_{j2}, X_{j3})$  is

$$m_{i1}(X_{j2}, X_{j3}) = \sum_{i=1}^{k_1} \sum_{m=1}^{k_2} \theta_{im} h_i(X_{j2}) l_m(X_{j3}) \quad (18)$$

$m_{i2}(\cdot)$  can be similarly approximated by tensor product splines (see Wood 2006 for details). Hence equation (16) is reduced to a linear regression..

Given the estimated choice probability  $\hat{P}_{ij}$ , Hotz and Miller (1993) show that  $\hat{V}_{ij}$  can be computed from  $\hat{P}_{ij}$ . Equation (7) implies

$$\hat{V}_{ij} = \log(\hat{P}_{ij}(a_{ij} = 1 | X_j, \Theta)) - \log(1 - \hat{P}_{ij}(a_{ij} = 1 | X_j, \Theta)) \quad (19)$$

The estimates  $\hat{P}_{ij}$  and  $\hat{V}_{ij}$  reduce the random utility model to an easy-to-estimate linear model.

Next we estimate taste parameters. Let  $\tau_{ij}$  denote sampling error for  $\hat{V}_{ij}$ .

$$\hat{V}_{ij} = \beta_0 + \xi_j + \beta_{0i} + X_j \beta_i + \frac{\gamma_i}{N-1} \sum_{l \neq i} \hat{P}_{lj} + \tau_{ij} \quad (20)$$

where  $\tau_{ij}$ s have iid normal distribution,  $\tau_{ij} \sim N(0, \Omega_\tau)$ ,  $E(\tau_{ij} \tau_{ij}) = \sigma_{ij}$ , and  $E(\tau_{ij} \tau_{i'j'}) = 0$  if  $i' \neq i$  or  $j' \neq j$ . Let  $\tilde{X}_{ij} \equiv (X_j, \sum_{l \neq i} \hat{P}_{lj} / (N-1))$ .

Combining equations (3) and (20), we obtain

$$\hat{V}_{ij} = \beta_0 + \tilde{\xi}_j + \beta_{0i} + \tilde{X}_j \bar{Z}_i \alpha + \tilde{X}_j \psi_i + \tau_{ij} \quad (21)$$

where  $\tilde{\xi}_j \equiv \xi_j + \tilde{X}_j \alpha_0$ . Equation (21) can be estimated by GLS detailed in Hsiao (2003). Finally we discuss how to identify  $\xi_j$  and  $\alpha_0$ . We take the approach in Nevo (2000) and assume that  $E(\xi_j | \tilde{X}_j) = 0$ . Let  $\hat{\tilde{\xi}}_j$  denote the estimated  $\tilde{\xi}_j$ , and  $V$  its variance and covariance matrix. Then

$$\hat{\alpha}_0 = (\tilde{X} V^{-1} \tilde{X})^{-1} \tilde{X} V^{-1} \hat{\tilde{\xi}} \quad (22)$$

with  $\text{Var}(\hat{\alpha}_0) = (\tilde{X} V^{-1} \tilde{X})^{-1}$ . And

$$\hat{\xi}_j = \hat{\tilde{\xi}}_j - \tilde{X} \hat{\alpha}_0 \quad (23)$$

## 1.5 Empirical Results

### 1.5.1 Model Selection

First we compare our model with benchmark specifications to show that it is important to include fixed internship effects and social effects. In-sample and out-of-sample model fit statistics are summarized in

Table 2.. Considering first the in-sample results, the baseline homogenous logit model assumes that tastes for internship attributes are homogeneous across students and yields an Akaike Information Criterion (AIC) statistic of 73,446.24. Expectedly, the heterogenous logit model fits better with AIC of 67,943.77. Also, both the

heterogeneous logit model with fixed internship effects (but without social effects), and the heterogeneous logit with social effects, fit better than the heterogeneous logit model, as expected. Finally, the proposed heterogeneous logit model with both fixed internship effects and social effects has a better fit than all other specifications.

We also compare three models in terms of out-of-sample fits: homogeneous logit, heterogeneous logit, and heterogeneous logit with social effects. While the estimation sample consists of 221 MBA students and 40 internships, the hold-out sample consists of the same group of students but 18 other internships. This is because we need to estimate social effects as a function of all other students' choices for any job, and therefore the hold-out sample cannot consist of students not included in the estimation sample. The root mean square error shows that the heterogeneous logit with social effects predicts better than the heterogeneous logit model, and the heterogeneous logit performs better than the homogeneous logit. Since the set of internship employers is different between the estimation and hold-out samples, estimates on internship employer fixed effects are unavailable for the hold-out sample. Therefore, out-of-sample fits cannot be computed for model specifications with fixed internship effects.

Table 2 Model Fit

Model	# parameters (p)	In-sample AIC*	Out-of-sample RMSE**
Homogeneous logit	4	73446.24	21.10
Heterogeneous logit	237	67943.77	17.85
Heterogeneous logit with social effects	244	67677.96	17.13
Heterogeneous logit with fixed effects	276	67314.42	N/A
Heterogeneous logit with fixed effects and social effects	283	67083.81	N/A

### 1.5.2 Effects of Internship Employer Attributes

We discuss here the influence of internship employer attributes, namely industry, salary, and firm size, on students' decisions to apply for an internship. We model students' tastes for employer attributes as heterogeneous parameters, and project these parameters on demographic and other characteristics of students. We first report the interaction of student characteristics and employer attributes. The population-level estimates in Table 3 inform us how characteristics of students predict the marginal employer attribute effects. We find that male and domestic students, relative to female and international students respectively, have larger marginal effects if an employer is from the finance industry. In other words, internships in the finance industry attract more male and domestic student applicants. Male, domestic, and high GPA students have larger marginal effects of firm salary. Similarly, male, domestic, higher GMAT quantitative scores, and higher GPA scores imply stronger effects of firm size.

Next, we show the main effects of internship employer attributes. For this purpose, we consider a representative student or individual denoted by  $i^*$ . The representative individual  $i^*$  has the mean individual characteristics reported in Table 1. We define the main effects as the change in expected log odds of applying to an internship employer versus not applying, when we change one of the employer attributes by one unit, and keep the others constant. In other words, main effects are defined as

$$\Delta \log \frac{P_{i^*}}{1 - P_{i^*}} \equiv E(\beta_{i^*k}) = \alpha_{ok} + Z_{i^*} \alpha_k \quad (24)$$

where  $\Delta \log(P_{i^*} / (1 - P_{i^*}))$  denotes the change in expected log odds,  $\beta_{i^*k}$  is the taste parameter of  $k$ th employer attribute for  $i^*$ ,  $Z_{i^*}$  is the characteristic vector of the

representative individual as indicated in Table 1,  $\alpha_{ok}$  is defined in Eq. (3), and  $\alpha_k \equiv (\alpha_{1k}, \dots, \alpha_{Mk})$  where  $(\alpha_{1k}, \dots, \alpha_{Mk})$  are defined in Eq. (3). Since  $\beta_{i^*}$  has a random component as shown in Eq. (3), we take expectation of  $\beta_{i^*}$ . We find that favorable employer characteristics - higher salary, and larger firm size, increase the expected log odds (Table 4). Furthermore, an employer in the finance sector has higher odds of getting an internship application.

Table 3 Interaction Effects of Student Demographics with Internship Employer Attributes

Parameter	Industry			Salary			Size		
	Mean	Std. Error	Pr(> t )	Mean	Std. Error	Pr(> t )	Mean	Std. Error	Pr(> t )
Intercept	-4.867	1.930	**	-0.303	0.264		-0.165	0.045	***
Gender	4.036	0.937	***	0.122	0.041	***	0.018	0.007	**
International	2.808	1.013	***	-0.225	0.045	***	0.020	0.008	**
experience	0.099	0.157		0.007	0.007		0.001	0.001	
GMAT V	-0.082	0.094		0.005	0.004		0.000	0.001	
GMAT Q	0.058	0.103		-0.006	0.005		0.0014	0.0008	*
GMAT AWA	0.548	0.594		-0.017	0.026		-0.008	0.005	***
GPA	1.532	1.411		0.184	0.062	***	0.046	0.011	***

Table 4 Main Effects of Internship Employer Attributes on Probability of Choice

Main Effect	Change in Expected Log Odds of Applying
Finance relative to non-Finance Sector	0.036
Increase in Annual Salary by \$1000	0.546
Increase in Firm Size by \$1 Million	0.041

### 1.5.3 Negative Social Effects

Next we discuss social effects. We project heterogeneous social effects onto student characteristics. We report the interaction of individual demographics and social influences in Table 5. The main social effects for the representative student  $i^*$  can be computed by Eq. 24 and the estimates in

Table 5. We find that the average marginal social effects are negative, in particular,  $\gamma = -0.403$ . Our results differ from the typical positive social effects findings in the marketing literature (Yang and Allenby 2003, Yang et al. 2006, Hartmann forthcoming). The endogenous social effects found in these studies describe conformity behavior, i.e. a consumer chooses goods that his neighbor, spouse, or golf partner also chooses. Our empirical results provide a case for congestion. Although the Career Management Center encourages all students to apply to each internship, students recognize that only a small number of students will obtain any given internship.

Looking at the interactions of student characteristics with social effects, we find there is indeed heterogeneity in these effects. Our investigation of the heterogeneity in social effects adds to the growing literature in sociology that studies social effects of specific segments of population. For example, sociologists have used social network of men versus women to explain gender segregation of jobs (Fernandez and Sosa 2005) and gender disparities in salary compensation (Belliveau 2005). Specifically, we find that male students with more experience and students with higher GMAT verbal scores are more averse to congestion. This seems paradoxical given these are stronger students on some dimensions (work experience and GMAT scores), and their job search prospects should be less hurt by congestion. One interpretation of this finding is that these stronger students are willing to apply to less popular (and

perhaps non-traditional) internships, relative to the students who have less work experience and lower scores.

Table 5 Interaction Effects of Student Demographics with Social Influence

Variable	Mean	Std. Error	Pr( $t >  t $ )
Intercept	0.256	0.175	
Gender	-0.110	0.040	***
International	-0.010	0.043	
Experience	-0.012	0.007	*
GMAT V	-0.007	0.004	*
GMAT Q	-0.003	0.004	
GMAT AWA	0.018	0.025	
GPA	0.051	0.060	

#### 1.5.4 Counterfactual Analysis and Policy Implication

We discuss next the policy implications of our results for recruiting employers and the career officers in a business school. The main policy instrument for an employer is salary. So we focus on the impact of salary in the presence of social effects. For this analysis we repeat the simulation in section 1.3.2. To simplify our illustration, we consider a group of homogeneous students, each of whom is identical to representative student  $i^*$ . Further, we focus on only one of the internship attributes – salary; similar insights hold for other internship attributes viewed positively by students. We compute the choice probability of the representative student as a function of salary while fixing the social effects parameter  $\gamma$  at one of three values: negative (-0.403), zero, or positive (0.403). Note that the estimated value of  $\gamma$  in our data for

$i^*$  is -0.403. Also, the positive value of 0.403 leads to a unique equilibrium choice probability (larger positive values may lead to multiple equilibria; we do not consider those cases here). In

Figure 2 we show the three choice probability functions corresponding to the three values of the social effects parameter.

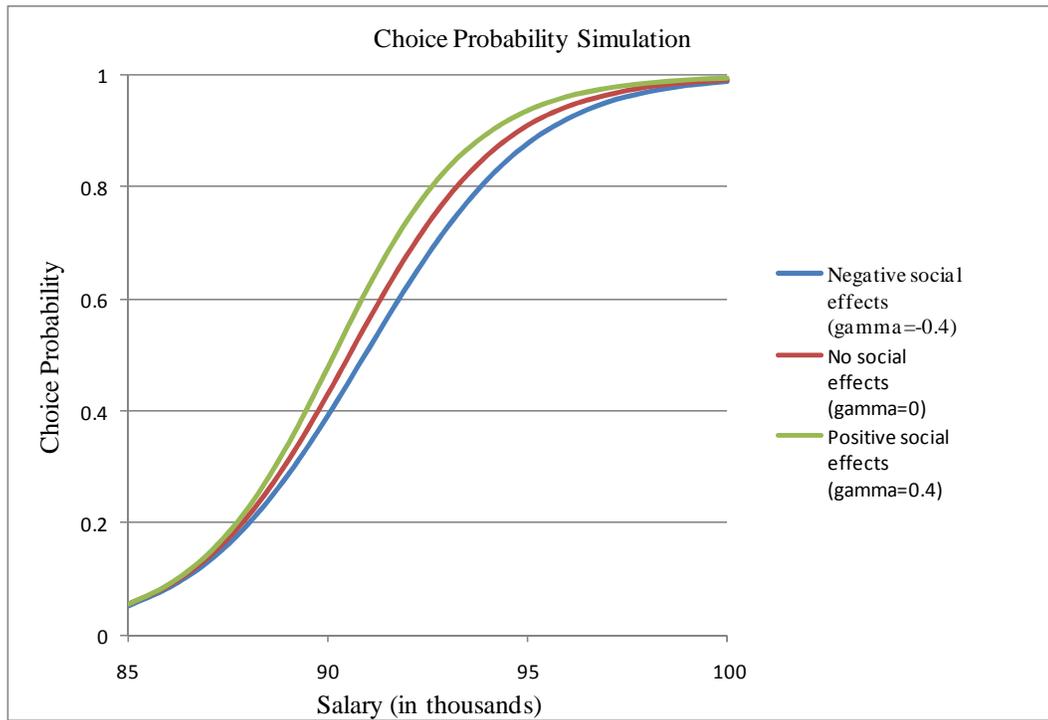


Figure 2 Impact of Salary Under Negative, Zero, and Positive Social Effects

We make the following observations from Figure 2. First, for all three curves, the choice probability is (strictly) increasing in salary. This implies that attractive internship characteristics increase choice probability regardless of the existence or sign of social effects. Second, for each level of salary, the choice probability is strictly increasing in social effects. Thus, negative social effects reduce the choice probability. However, when the salary is low, the internship is unattractive. Therefore, there are few applicants, leading to low congestion. Consequently, at low salaries the choice

probability with negative social effects is close to that without social effects. The largest differences in choice probabilities between the three curves appear in the middle range of the salary. At high salaries, the choice probabilities become close again. This implies that the salary effects will eventually dominate social effects because the social utility component  $\gamma P_{i^*}$  is bounded above by  $\gamma$ .

A third insight from figure 2 is that to achieve a given probability of being chosen, the salary that a prospective employer needs to offer to the representative student  $i^*$  is lower in the presence of positive social effects when compared to a situation of no social effects, and higher with negative social effects. As expected, positive social effects help an employer to recruit. Fourth, we observe that the marginal effect of salary on the choice probability (the slopes of the curves) varies between the three cases. Specifically, when salary is low (high), the marginal effects of salary with positive social effects are higher (lower) than those with negative social effects. This knowledge is useful for employers to predict the impact of a change in salary offered on the likelihood of their receiving an internship application. Finally, for a fixed choice probability, the marginal effect of salary is increasing in social effects.

These observations suggest the following implications for employers. It is obvious that an employer can positively influence the number of internship applications by making attributes such as salary more attractive. An alternative approach is to reduce student concerns about being crowded out, e.g. by directly communicating with the students most prone to negative social effects (in our case students who are male, and have higher work experience and higher verbal GMAT scores).

The next proposition formalizes these observations.

**Proposition 1.2:** Let  $s \in (-\infty, \infty)$  denote salary, and  $\sigma(s | \gamma)$  denote the choice

probability as a function of the salary for  $\gamma < \gamma^*$ , where  $\gamma^*$  is the threshold for multiple equilibria. The equilibrium choice probability is implicitly defined by

$$\sigma(s | \gamma) = \frac{\exp(\beta s + \gamma \sigma(s | \gamma))}{1 + \exp(\beta s + \gamma \sigma(s | \gamma))} \quad (25)$$

Then

- a)  $\sigma(s | \gamma)$  is a cumulative distribution function of  $s$  given parameters  $\gamma$  and  $\beta$
- b) If  $s_1 < s_2$ , then  $\sigma(s_1 | \gamma) < \sigma(s_2 | \gamma)$ , for all  $\gamma < \gamma^*$
- c) If  $\gamma_1 < \gamma_2 < \gamma^*$ , then  $\sigma(s | \gamma_1) < \sigma(s | \gamma_2)$ , for all  $s$ , i.e.  $\sigma(s | \gamma_1)$  first-order stochastic dominates  $\sigma(s | \gamma_2)$
- d) If  $\gamma_1 < \gamma_2 < \gamma^*$ ,  $\int s d\sigma(s | \gamma_1) < \int s d\sigma(s | \gamma_2)$
- e) Suppose  $\gamma_1 < \gamma_2 < \gamma^*$ . There exist open sets  $B_1$  and  $B_2$  such that,  $\sup B_1 < \inf B_2$ , and  $d\sigma(s | \gamma_1) / ds < d\sigma(s | \gamma_2) / ds$  for  $s \in B_1$ , and  $d\sigma(s | \gamma_1) / ds > d\sigma(s | \gamma_2) / ds$  for  $s \in B_2$ .
- f) Suppose  $\gamma_1 < \gamma_2 < \gamma^*$ . For a fixed  $\sigma(s_1 | \gamma_1) = \sigma(s_2 | \gamma_2)$ ,  $d\sigma(s_1 | \gamma_1) / ds < d\sigma(s_2 | \gamma_2) / ds$

See Appendix A for a proof.

## 1.6 Conclusion

We have modeled choice decisions in the presence of endogenous positive or negative social effects. We estimate that negative social effects exist in MBA students' internship choices. The model we propose is general enough to be applicable to other decision contexts in a large group where members interact with all other members and

where members are likely to have private information.

As we have shown, there is a unique equilibrium for the negative social effects found in our case. When positive social effects exist and are sufficiently large, researchers may need to address identification issues caused by multiple equilibria. An important challenge that we do not address is the development of methods to test multiplicity of equilibria, similar to the one proposed in Echenique and Komunjer (forthcoming). Another question to be addressed in future research is that of modeling endogenous group formation. Given our context we were able to assume that all students in a relatively small MBA cohort are members of the group. One approach is to model a group member's decision as a two-stage process (Zanella 2007). In the first stage, an individual decides whether to join a group. In the second stage, she makes a decision given her group membership. Zanella (2007) shows the two-stage process can potentially be represented by a nested logit model. Our model can be seen as the second stage of such a two-stage model, with the group membership taken as given.

## REFERENCES

- Akerlof, George A. (1997), "Social Distance and Social Decisions," *Econometrica*, 65,1005-28.
- Amaldoss, Wilfred and Sanjay Jain (2005), "Conspicuous Consumption and Sophisticated Thinking," *Management Science*, 51,1449-66.
- Bajari, Patrick, Han Hong, John Krainer and Denis Nekipelov (forthcoming), "Estimating Static Models of Strategic Interaction," *Journal of Business and Economic Statistics*.
- Belliveau, Maura A. (2005), "Blind Ambition? The Effects of Social Networks and Institutional Sex Composition on the Job Search Outcomes of Elite Coeducational and Women's College Graduates," *Organization Science*, 16 (March 1), 134-50.
- Blume, Lawrence E. (1993), "The Statistical Mechanics of Strategic Interaction," *Games and Economic Behavior*, 5, 387-424.
- Brock, William A. and Steven N. Durlauf (2001), "Discrete Choice with Social Interactions," *Review of Economic Studies*, 68 (April), 235-60.
- Brock, William and Steve Durlauf (2002), "A Multinomial-Choice Model of Neighborhood Effects," *American Economic Review*, 92, 298-303.
- Coleman, James S., United States. Office of Education, and National Center for Education Statistics (1966), *Equality of Educational Opportunity*. Washington: U.S. Dept. of Health, Education, and Welfare, Office of Education.
- De Paula, Aureo (2009), "Inference in a Synchronization Game with Social Interactions," *Journal of Econometrics*, 148 (1), 56-71
- Echenique, Federico and Ivana Komunjer (forthcoming), "Testing Models w/ Multiple Equilibria by Quantile Methods," *Econometrica*.

- Eilers, Paul H. C. and Brian D. Marx (1996), "Flexible Smoothing with B-Splines and Penalties," *Statistical Science*, 11, 89-102.
- Fernandez, Roberto M. and Isabel Fernandez-Mateo (2006), "Networks, Race, and Hiring," *American Sociological Review*, 71, 42-71.
- Fernandez, Roberto M. and M. L. Sosa (2005), "Gendering the Job: Networks and Recruitment at a Call Center," *American Journal of Sociology*, 111, 859-904.
- Frank, Robert H. (1985), *Choosing the Right Pond: Human Behavior and the Quest for Status*. Oxford University Press.
- Glaeser, Edward L., Bruce Sacerdote, and Jose A. Scheinkman (1996), "Crime and Social Interactions," *Quarterly Journal of Economics*, 111, 507-48.
- Glaeser, Edward L., Bruce Sacerdote, and José A. Scheinkman (2003), "The Social Multiplier," *Journal of the European Economic Association*, 1, 345-353.
- Hartmann, Wesley R. (forthcoming), "Demand Estimation with Social Interactions and the Implications for Targeted Marketing," *Marketing Science*.
- Hotz, V. J. and Robert A. Miller (1993), "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, 60, 497-529.
- Hsiao, Cheng (2003), "Analysis of Panel Data," *Econometric Society Monographs* (No. 34), Cambridge University Press.
- Jones, Stephen R. G. (1984), *The Economics of Conformism*. Oxford University Press: Basil Blackwell Publisher.
- Krauth, Brian V. (2006), "Simulation-Based Estimation of Peer Effects," *Journal of Econometrics*, 127 (July), 243-71.
- Manchanda, Puneet, Ying Xie, and Nara Youn (2008), "The Role of Targeted Communication and Contagion in New Product Adoption," *Marketing Science*, 27 (6), 961-976.

- Manski, Charles F. and David A. Wise (1983), *College Choice in America*, Cambridge, Mass.: Harvard University Press.
- Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, 531-42.
- Manski, Charles F. (2000), "Economic Analysis of Social Interactions," *The Journal of Economic Perspectives*, 14 (Summer), 115-36.
- Milgrom, Paul and John Roberts (1990), "Rationalizability, Learning, and Equilibrium in Games with Strategic Complementarities," *Econometrica*, 58, 1255-77.
- Nair, Harikesh, Puneet Manchanda, and Tulikaa Bhatia (2006), "Asymmetric Peer Effects in Physician Prescription Behavior: The Role of Opinion Leaders," working paper.
- Nevo, Aviv (2000), "A Practitioner's Guide to Estimation of Random-Coefficients Logit Models of Demand," *Journal of Economics & Management Strategy*, 9, 513-48.
- Newey, Whitney K. (1994), "The Asymptotic Variance of Semiparametric Estimators," *Econometrica*, 62, 1349-82.
- Puhani, Patrick A. (2000), "The Heckman Correction for Sample Selection and its Critique," *Journal of Economic Surveys*, 14, 53-68.
- Stanley, H. E. (1971), *Introduction to Phase Transitions and Critical Phenomena*. Oxford: Oxford University Press.
- Topkis, Donald M. (1979), "Equilibrium Points in Nonzero-Sum n-Person Submodular Games," *SIAM Journal on Control and Optimization*, 17, 773-87.
- Van den Bulte, Christophe and Yogesh V. Joshi (2007), "New Product Diffusion with Influencers and Imitators," *Marketing Science*, 26 (May), 400-21.
- Wood, Simon N. (2006), "Low-Rank Scale-Invariant Tensor Product Smooths for Generalized Additive Mixed Models," *Biometrics*, 62 (12), 1025-36.

- Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40, 282-94.
- Yang, Sha, Vishal Narayan, and Henry Assael (2006), "Estimating the Interdependence of Television Program Viewership between Spouses: A Bayesian Simultaneous Equation Model," *Marketing Science*, 25, 336-49.
- Zanella, Giulio (2007), "Discrete Choice with Social Interactions and Endogenous Memberships," *Journal of the European Economic Association*, 5 (03), 122-53.

## **Chapter 2 An Equilibrium Analysis of Online Social Content-sharing Websites**

### **2.1 Introduction**

User generated content is an emerging online phenomenon. Individuals are no longer audiences of traditional media industry. Hundreds of web sites allow ordinary people to create and share diverse kinds of content including videos (YouTube), photographs (Flickr), research papers (CiteULike), online bookmarks (del.icio.us), presentations (slideshare), documents (Scribd), Citizen journalism (Slashdot), Encyclopedia (Wikipedia), food (All recipes), and fashion (Polyvore). Each website hosts a large ecosystem. Forty four million visitors worldwide used Yahoo Flickr in March 2008, and over twenty six million users have contributed four billion photographs since the site's inception (Graham 2008, Champ 2009).

Firms recognize business potential of aforementioned content. Yahoo paid an estimated 40 million dollars for Flickr in 2005, and Google paid an estimated 1.6 billion dollars for YouTube in 2006. But it is not all that clear how to manage a content system because content differs from traditional goods. Content is free. Content consumers do not pay to consume, nor do content producers make profit in production. So YouTube and Flickr need to adopt business policies different from ones for traditional goods. The starting point for policy analysis is to understand who a content firm is. A content firm plays at least three roles: (1) a Web2.0 firm, (2) a central planner for consumer welfare, and (3) a profit seeker. First, YouTube and Flickr are built upon Web2.0 technology. A primary objective of Web2.0 is to encourage user participation and to transform users from content consumers to producers (Oreilly 2007). Second, content website providers are hosts and guardians

of social activities involving millions of users. Firm's own business ethics, such as Google's "Do no evil" code of conduct, call for consideration of consumer welfare. In addition, Firms like Google and Yahoo face constant scrutiny from lawmakers (Hausman and Sidak 2009). Finally, content firms are profit maximizers. Monetizing content is an ongoing quest. One profit opportunity is to charge content producers. Yahoo Flickr, for example, sells storage space for photo producers.

This paper provides a modeling framework to analyze policy options. The aforementioned objectives all involve individual content users: Web2.0 and profit objectives are concerned with producer, and consumer welfare objective is concerned with consumer. So we first model consumer and producer behaviors. Content is differentiated in quality, and is ranked from low to high by a number. We assume that a consumer is only satisfied by content above a threshold quality. Therefore lower quality is an *imperfect substitute* for high quality (Rosen 1981). Consumers need to search for satisfactory quality. Since the number of consumers is large, each of them has no influence on the whole system and can only take *sampling probability*—probability to find a particular quality as given. A utility maximizing consumer takes sampling probability as input to decide optimal demand. On supply side, producers no longer make profits because content is free. Content systems allow consumers to leave positive feedback on a piece of content if they are satisfied by the quality (we call this *endorsement*). We assume producers are motivated by endorsement. Since the number of producers is large, producers take endorsement as given. Producers take endorsement as input to decide optimal production.

Sampling probability is a key policy instrument for content firms (Bernardo and Wu 2007).<sup>6</sup> But before policy intervention, an understanding of origin of sampling

---

<sup>6</sup> Only sampling probability is considered to be a policy instrument in this paper, while manipulation of endorsement will be considered an internet fraud. Policy makers can use the framework in this paper to analyze consequence of fraudulent manipulation of endorsement.

probability and endorsement is required. They are exogenous variables in the individual models, but to content firms yet to intervene, both variables are created within the system or endogenous. The individual models alone do not explain genesis and persistence of sampling probability and endorsement. Yet both are central to policy issues. The importance of sampling probability can be argued from attention economics (Pieters 1999, Dukas 2004). In a seminar paper (Simon 1971), Herbert Simon makes a general observation that we live in a world in which information is abundant whereas attention is scarce. Content can be considered as one form of information. Sampling probability informs us about which content will get limited attention. Herbert Simon is concerned with organization design in an information rich world, and he argues that an information processing subsystem “can transform (‘filter’) information into an output that demands fewer hours of attention than the input information.” Recently computer scientists start to implement this goal by designing policies to allocate scarce attention to rich information (Bernardo and Wu 2007). We will show a content system without a centralized force overseeing entire system, possesses an ability to filter information. Attention is a consumer objective. We will show sampling probability is also important to achieving content website objectives.

It is also important to study endorsement because interesting pattern emerges on endorsement and resulting production. We collect a sample of about 40 million photographs taken by about 205,000 users from Yahoo Flickr (see Table 6 for a summary statistics). We plot number of users against maximum endorsement per photo in Figure 3. The plot shows that a small number of producers enjoy a large per photo endorsement. Pattern also emerges in content production. We plot number of users against number of photo produced by a user in Figure 4. The plot shows that a small number of producers produce a large number of photos. Perhaps the most

striking fact is that production inequality has been observed across different types of content, and sometimes characterized as 1-9-90 rule, i.e., an empirical observation that heavy producer, light producer, and lurkers take up 1%, 9%, and 90% population respectively (Nonnecke and Preece 2000). Understanding production inequality is relevant to firm objectives. Inequality suggests that a primary objective of Web2.0 firms— democratizing content generation by promoting individual participation and content diversity—may have failed. Furthermore, firms build their business models on the premise that majority of users will produce in an equal fashion. One such company, Splashcast, folded its operation in 2009 after its CEO made the following remark: “Most of us would rather consume than create. This is one of the big ticket findings of the Web 2.0 technology wave” (Schonfeld 2009).

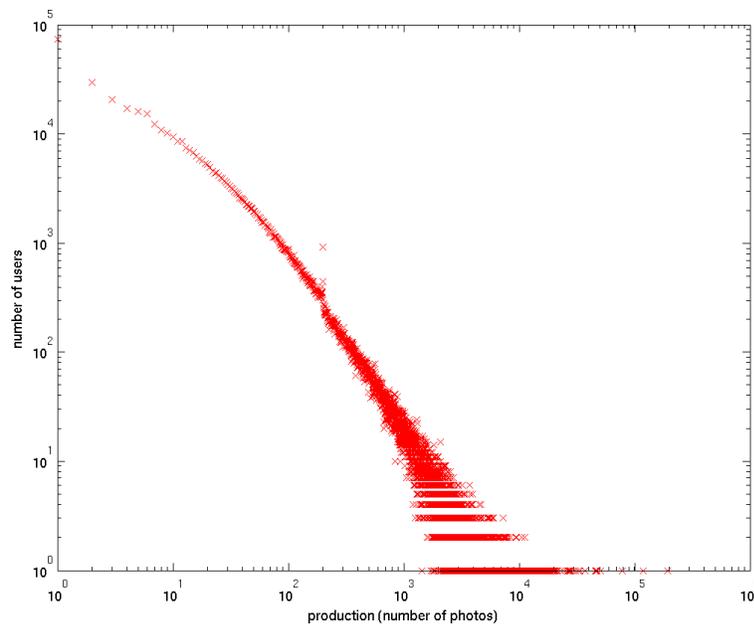


Figure 3 Inequality in Production (X-axis shows number of photos, and y-axis shows number of producers. Both are in log scale).

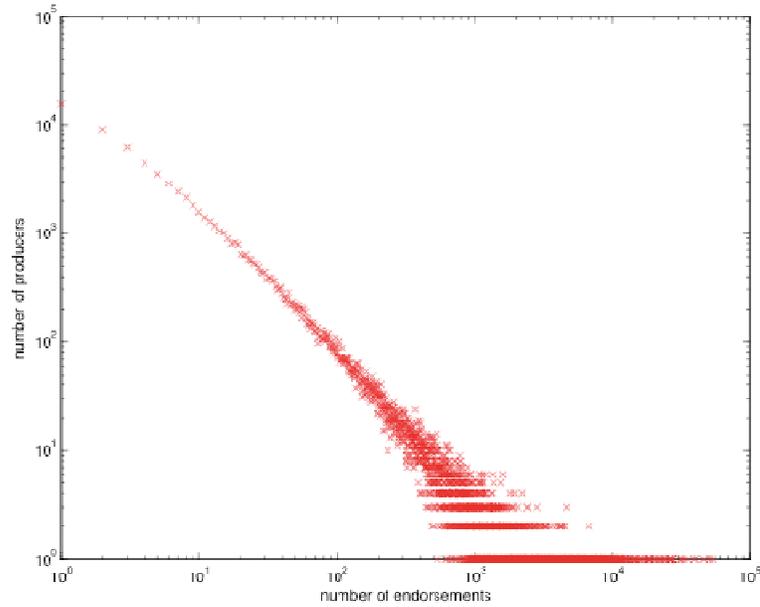


Figure 4 Inequality in Production (X-axis shows number of photos, and y-axis shows number of producers. Both are in log scale).

Table 6 Summary Statistics

variable	mean	range
Number of endorsements per producer	59.07	0 - 53,659
Number of photos per producer	189.8	1-115,770
Number of endorsements per photo	9.24	0 - 6127
Sampling probability in current period	2.46e-8	0 – 2.57e-3
Sampling probability in previous period	2.67e-8	0 – 8.08e-4
Photo quality score	1.08411e-06	4.68252e-10 – 1.06115e-3

The main part of this paper studies how optimal consumer behavior and optimal producer behavior interact to give rise to endogenous sampling probability and endorsement. We frame the managerial questions to a problem of studying

whether a content system can reach equilibrium—a state in which interacting users will not change their behaviors. A useful parallel to guide our study of content system is *market* for traditional goods. The market mechanism is well known: myopic individuals make decentralized decisions, yet the system can reach equilibrium in which no individual wants to change his/her decision. The individuals need not know the decisions of others as long as they know prices of goods. So a fundamental role of price is to enable a decentralized market to function. The celebrated general equilibrium theory shows that equilibrium prices exist (Arrow and Debreu 1954, Debreu 1959). However, two features of user generated content prevent application of the market mechanism: content is free, and content is non-rival, i.e., one individual's consumption of a piece of content does not reduce the amount that another can consume. Since content is free, price as the sufficient statistic for decentralized decision-making is missing. Market equilibrium is defined as a state of balance—market clearing, i.e., demand meets supply. But market clearing is not relevant for content due to its non-rival feature. To fill the missing content mechanism, we develop and characterize a content equilibrium from first principles.

First, we demonstrate that a decentralized content system may reach equilibrium under two conditions: (1) a user makes independent decisions between content consumption and production, and (2) quantity demand varies inversely with search cost. A content equilibrium has the following features. The first feature is the information for decentralized decision making. In a market, consumers and producers use the same information, i.e., price. In a content system, consumers and producers use different information, i.e., search cost and endorsement. We show that search cost and endorsement are related. The second feature is the relationship between demand and supply. In a market, total demand equals total supply. In a content system, the role of market clearing is replaced by *endorsement clearing*, i.e., the total endorsements

given by the consumers equals the total endorsements received by the producers. These two features highlight the significance of endorsement. Endorsement not only provides feedback to individual producers, but also enables a decentralized system to reach equilibrium. The third feature is where regularity appears at equilibrium. At a content equilibrium, regularity lies in the distribution of content, i.e., the relative amounts of different content in a system. It explains empirical regularity across content platforms is observed on distribution such as 1-9-90 rule rather than absolute measures. The fourth feature is equilibrium outcome. The system outcome only depends on the demand side condition. This result is complementary to the one-sided equilibrium in which the market outcome only depends on the supply side, as shown by Paul Samuelson in his Non-substitution Theorem (Samuelson 1951). The fifth feature is optimality. Inequality seems to suggest that a content system is like an oligopoly market, and bad for consumer welfare. We show the contrary. The sixth feature is self-organization. We demonstrate that a content system can steer itself towards the equilibrium without any exogenous forces. Such force is significant because a content system is fundamentally decentralized. In social network literature, researchers have discovered a remarkable small world phenomenon—short paths exist between individuals (Milgram 1967). Even more remarkably, Jon Kleinberg shows why these short paths can be found by individuals with no global knowledge except their immediate neighbors (Kleinberg 2000). This paper established a similar type of collective intelligence—a content system not only has a good state (Pareto optimal equilibrium), but also knows how to find it from many other perhaps sub-optimal states.

Second, we show inequality arises in a content equilibrium. High quality producers will always earn more endorsements and produce more content than low quality producers. The behavioral drive is imperfect substitution. But it is the non-

rival feature of content systems (a result of technology) that puts the behavior to work towards the advantage of high quality content. High quality content can be consumed by consumers with taste for high quality, and earn their endorsements. In the meantime, the same content can also be consumed by consumers with taste for low quality, and therefore compete for a share of endorsements for low quality content. Imperfect substitution prevents the reverse process—low quality producers cannot earn endorsements from high quality taste consumers. The selective nature of the content equilibrium helps achieve the organization design goal in Simon (1971). If the content system fails to deliver a democratic web, it succeeds in providing collaborative filtering – a mechanism that makes high quality content easier to find, and that promotes the production of high quality content. It is also worth noting that producer quality decides inequalities in endorsement, production, and sampling probability only in a qualitative way. Size of the inequalities is independent from quality difference between low and high quality producers. This is a result from the fact that equilibrium is independent from supply side condition. The disconnection between reward and talent has been observed in labor market (Lazearand and Rosen 1981).

We use our basic model to characterize other forces on the inequality. The first force is incomplete endorsement. Our basic model assumes that consumers will endorse when their demand is met (which we call *complete endorsement*). In reality, consumers may not always endorse a piece of content even when it satisfies their demand (which we call *incomplete endorsement*). This is a problem similar to survey non-response. We show that if consumers want their preferred content, they should endorse. The second force is search by endorsement. We show that endorsement may give extra advantage to high quality content over time. In the basic model, high quality content producers have an advantage because they can earn endorsements from consumers who look for low quality content. The basic model assumes that content

has uniform sampling probability—the probability that a piece of content is found by consumers. However, consumers may return to the same producers whom they endorse. Therefore, endorsement can be prior information for locating content in the future. Since the high quality producers have more endorsements, high quality content will be easier to find. So the inequality will increase over time because the sampling probability is changing over time. This is similar to the preferential attachment process (“rich-get-richer”) which can lead to a power law distribution—a highly skewed distribution (Newman 2005). The third force is entry of new producers. We demonstrate that new producers entering the system counter-balance the skew towards high quality content for existing producers in short term. New content producers of different qualities have equal production because they have not received any preferential endorsement. So the addition of their content decreases the search cost of existing low quality content. In long run, however, content system returns to equilibrium state in terms of sampling probability.

Third, we discuss how to extend our model to more general conditions. We extend our model from two qualities to multiple qualities. Multiple qualities account for finer differentiation among producers, and therefore explain why some high quality producers produce more than others. When demand is nonlinear in inverse of search cost, our equilibrium model can be used to construct bounds for existing nonlinear demand specifications in the literature. We also relax the independence assumption between a user’s consumption and production decisions. Using the proposed framework, we provide policy recommendations to content platform providers. Our analysis shows that sampling probability is a key policy instrument because it affects consumer demand. A content site provider can decide which content appears more often than others, and therefore change sampling probability. A website provider can also use sampling probability to influence producers given our insight to

interaction between consumer and producer. Consumers always like content with higher qualities than their threshold ones. So as a central planner, a site provider should promote high quality content. However, Web2.0 objective of promoting individual participation / content diversity implies that sampling probabilities should not be concentrated on a small set of content. A compromise will be to promote existing content according to quality ranking. Content websites make profit from selling services to content producers. Sampling probability influences consumer decision on endorsement, and therefore indirectly influences producer decision. Producers are more likely to purchase services when they produce more. An efficient policy appears to be assigning sampling probability to high quality content because the resulting endorsement is not split between high and low quality producers. However, high quality content is more than low quality one at equilibrium. So in terms of unit endorsement, it makes no difference whether to promote high or low quality content. The optimal policy is to promote existing content in an equal fashion. Therefore, a central planner and a diversity promoter are competing roles. Interestingly, promoting diversity and making profits can be consistent.

Our model provides three additional managerial implications. First, content equilibrium allows us to define *relative economic value of content*. The economic value of goods is captured by price (Debreu 1959). For content without price, endorsement captures the value of content. As the value of goods is determined by interaction between consumers and producers of goods, the value of content is determined by interaction between consumers and producers of content. Second, our analysis provides a first benchmark of content consumer welfare for policy makers who potentially need to regulate content systems. Third, since search cost and endorsement are endogenous, researchers and firms need to consider simultaneity issue when they conduct empirical analysis to estimate content consumer and producer

behaviors.

The rest of the paper is organized as follows. After discussing related literature in Section 2.2, we describe a basic model of content equilibrium in Section 2.3. We use the basic model to study inequality in Section 2.4. We introduce various extensions of the basic model in Section 2.5. We investigate how to use our framework to manage content websites in Section 2.6, and conclude in Section 2.7.

## **2.2 Related Literature**

User generated content is an emerging phenomenon, so we use its features as the starting point of comparison with the extant literature. Content is like public goods in that it exhibits non-rival feature (although unlike content, public goods have prices). The public goods literature (Bergstrom et al. 1986) focuses on the provision issue of public goods. Researchers find that private provision of public goods is inefficient when compared to a competitive market due to the free-riding problem. Free-riding is not an issue here because the cost of giving an endorsement (clicking a mouse button) is negligible, and because the large number of users in a content system prevents any single user from changing the system. We show that a content system can be efficient despite the non-rival feature.

Content is an online phenomenon. The large body of click stream literature (Bucklin et al. 2002) focuses on consumer choice behavior on the internet, while our study focuses on the system behavior of consumers and producers. We make certain abstractions in consumer behavior in order to focus on system dynamics. Further extensions to our model can be made by adding richer consumer behavior found in the click stream literature. Different from goods in the click stream literature, content is free and non-rival.

The two-sided market literature (Rochet and Tirole 2006) includes diverse off-

line examples such as credit card and video consoles. The unifying theme is that the platform provider adopts a pricing strategy that uses one side to subsidize the other. Subsidization takes place in online content-sharing websites. For example, it is free to view a photo on Flickr. However, the goods in the two-sided market literature all have prices, and one side gets a monetary reward. In comparison, user-generated content is free, and the producer does not receive any monetary reward. So before we can study a pricing strategy for content site providers, we need to fill the gap on how the two sides interact on free content. In addition, the literature does not address why inequality arises across online content systems.

There is a growing stream of literature on a type of user generated content—user review, e.g. book and DVD reviews on Amazon (Chevalier and Mazlin 2006, Ghose and Ipeiritis 2008, Archak et al. 2008, and Lee and Bradlow 2009). The type of content in this paper differs from user reviews in that they are the primary subjects to consume whereas user reviews are used to facilitate consumption of other products. Consumers purchase books and DVD's produced by publishing companies. User review serves as a medium for users to exchanging product and consumption information. Photos on Flickr and videos on YouTube are end consumption subjects like books and DVD's rather than their reviews. But unlike books and DVD's, photos and videos are free. Market is the underpinning for studying user reviews. Our paper provides a system mechanism for content as primary consumption subjects. Two papers studying content as primary subjects are Ghose and Han (2009) and Kumar (2009), both of which study dynamic user behavior. Ghose and Han (2009) study learning behavior due to uncertainty in content quality. Our paper differs from their single agent model framework, and studies the interaction of agents in an entire system. As we mention in future research, our model can be extended by including learning behavior. Kumar (2009) and our paper are complementary in terms of the

type of content in question. He studies the content available only to one's friends and peers whereas we study the content available to general public. Kumar studies the competitive interaction among friends and peers assuming the competition will reach Markov Perfect Equilibrium. Our paper assumes only myopic rather than strategic behavior, and proposes a content equilibrium in which the general population interacts.

Inequality has been studied in a variety of contexts by sociologists and computer scientists. The striking feature they find is that the inequality follows a precise form, namely the power law (Simon 1955) and other related distributions. Sociologists and computer scientists have proposed various generative models for the power law (Newman 2005). These models typically lack micro foundation. Rosen (1981) studies a looser form of inequality—convexity. Rosen studies what he terms the *superstar phenomenon*, the observation that a small fraction of participants tend to dominate a field, earning the majority of the money. A few highly-paid athletes dominate a sport, for example. Rosen attributes the superstar phenomenon to two micro foundations. First, consumers prefer high quality, and therefore lower quality is an imperfect substitute for high quality. Second, technology enables a producer to serve a large market share without incurring proportional cost. Imperfect substitution applies to content because content is differentiated in quality. And it was Rosen's extraordinary vision on technology that makes his work particularly relevant to content. Rosen writes: "The practical importance of all this is related to technical changes that have increased the extent of scale economies over time in many activities. Motion pictures, radio, television, photo reproduction equipment, and other changes in communications have decreased the real price of entertainment services, but have also increased the scope of each performer's audience." The Internet pushes Rosen's vision to an extreme in which (1) content is free and (2) a content producer

can now reach almost any Internet user on Earth. But Rosen shows that imperfect substitution and technology lead to a superstar phenomenon through a market process which is lacking in content systems. In this paper, we provide the first analysis to shed light on a content process and how it generates inequality starting from Rosen's micro foundations. Compared to the convexity in Rosen (1981), we study a more general form of inequality in the basic model, i.e., higher quality producer produces more than lower quality producer. Then in an extension we discuss how preferential attachment—a process leading to a power law effect—may arise. An alternative micro foundation for inequality is proposed in Adler (1985). Adler suggests that an opera patron needs to learn about opera in order to better appreciate it. It is easier for the patron to gain knowledge if an opera singer has a large fan base. Inequality will thus arise even if there is no difference in quality depending on the initial random fan base. Our paper uses Rosen's micro foundation but a different system process.

Existence of content equilibrium is a question similar to the existence of competitive equilibrium in the competitive market literature. Our goal is not to achieve the generality of the analysis in Arrow and Debreu (1954) and Debreu (1959) among others, but to provide a simple model to show the *possibility* of content equilibria. In contrast to competitive markets which are enabled solely by price, we show that content consumers and producers use two different sources of information: search cost and endorsement. The structure of content systems makes the two related. We also show that, since content does not clear as goods do, we need the balance of endorsement to unite the two sides of a content system. The specific nature of our model allows us to show that regularity lies in the distribution of production. A static method (fixed-point theorem) is used to establish competitive equilibrium, whereas we construct a novel Liapnov function which shows not only the existence of equilibrium but also how a content system can get there. Our approach is in the same spirit as the

Tâtonnement process used to establish stability of a competitive market (Arrow and Hahn 1971). The significance of the Tâtonnement process is that it demonstrates an optimal state of economy can be reached in a decentralized fashion.

## **2.3 Basic Model of Content Equilibrium**

We start with a discussion of content features and individual behaviors in Section 2.3.1. Then we detail content user behaviors and their interaction without assuming equilibrium in Section 2.3.2. We demonstrate content equilibrium in Section 2.3.3.

### **2.3.1 Content and Individual**

We have mentioned two features of content: free and non-rival. Content is also *differentiated*. Like differentiated goods, each piece of content is unique in dimensions specific to the type of content. For example, nearly every one of the over 4 billion photographs on Flickr is unique in subjects, compositions, exposure settings, etc. We follow Rosen (1981) and assume that the differences can be summarized by one number referred to as quality  $z$  (see Table 7 for a key to notation). Rosen uses a continuous quality  $z \in (0, \infty)$ , while we use a discrete version  $z = \{z_i, i = 1, \dots, I\}$  where  $I$  is potentially enormous, but still finite, and  $z_1 < \dots < z_I$ . This can be motivated by having quality on a continuum  $(z_1, z_I)$ . Since users may not be able to distinguish tiny difference in quality, we can divide the continuum into a finite number of intervals, and  $z_i$  is the average quality in an interval.

Table 7 Notation and Variable Explanation

$i \in \{1, \dots, I\}$	content quality index
$z_i$	Threshold content quality of type $i$ consumer
$z_i^p$	Content quality by type $i$ producer
$x_i(t)$	quantity demanded for quality $z_i$ during period $t$
$\alpha_i$	exogenous preference parameter by consumer with threshold quality $z_i$
$W_i$	budget for consumer with threshold quality $z_i$
$x_o^i(t)$	demand for outside good by consumer with threshold quality $z_i$
$p_o$	price for the outside good
$\bar{x}_o$	total endowment of outside good
$p_i(t)$	search cost for quality $z_i$
$w$	wage rate for all consumers
$s_i(t)$	quantity of content that consumer must examine before finding content of quality $z_i$ or higher
$\tau$	time required to examine one piece of content
$F(z_i)$	cumulative distribution function for quality $z_i$
$m_i(t)$	sampling probability - the probability that content selected at random has quality $z_i$
$u_i$	utility for consumer with threshold quality $z_i$
$\bar{u}_i$	a pre-determined level of utility for consumer with threshold quality $z_i$
$Y_i(t)$	total number of content of quality $z_i$ at the end of period $t$
$y_i(t)$	content production of quality $z_i$ during period $t$
$e_i(t)$	total number of endorsement for quality $z_i$ generated during period $t$
$\Delta e_i(t)$	endorsement increase per existing content of quality $z_i$ during period $t$
$E_i(t)$	endorsement per exiting content of quality $z_i$ during period $t$
$\Psi$	monetary value per endorsement
$l_i$	labor (in terms of time) needed to produce each content $z_i$
$q_i$	time budget to consume content $z_i$ and outside good
$T_i$	Total time budget for production and consumption
$\beta_i$	percentages of the demand that translate into endorsements for consumer with threshold quality $z_i$
$\Delta Y$	number of new content by new producers added at each quality level
$\eta$	shape parameter in Hendel and Dube model
$\theta$	shape parameter for non-linear demand
$c$	fixed cost to adjust sampling probability by content site provider

Table 7 Continued

$\lambda_i$	penalizing/regularizing parameter for central planner
$\gamma_i$	penalizing/regularizing parameter for profit maximizer
$B$	normalizing constant for purchasing probability
$K$	constant defined as $K = \psi \sqrt{2\tau w^2}$
$p_s$	price for the service provided by content site provider
$n$	number of consumer and number of producer in stochastic specification
$a$	probability consumers access content by new producers

We also identify the key properties of content systems. First, in Rosen's analysis, one requirement for the superstar phenomenon is direct contact between consumers and producers. Such direct contact is even stronger in a content system, since most sharing sites allow either party to learn the identity of the other, and even to communicate with them. Second, a large number of content consumers and producers are involved. Third, most content systems have endorsement features, allowing consumers to give feedback to producers about their content. For instance, YouTube viewers give 1- to 5-star ratings to videos, while Flickr allows a consumer to mark a photo as her favorite.

Users in content-sharing websites consume multiple pieces of content, analogous in marketing to a situation in which consumers consume multiple quantities of products in multiple categories. The consumers make decisions about which and how much content to view. In our context, we assume that consumers decide what to view based on quality. Each consumer is endowed with an *exogenous threshold quality* at a given time, assumed to result from one's upbringing, education, values, and other factors. Any content at the threshold quality or above can meet the consumer's demand. As mentioned before, the vast majority of content producers do not make profits on their work. Since the motivation is not financial profit, we assume that producers are motivated by gaining endorsements, in a similar way that scholars are

motivated by gaining recognition of their scientific work (Merton 1957). We make the following assumptions on individuals.

#### **Assumptions (individual behavior)**

1. Consumers prefer higher quality to lower quality. In this context, for a consumer endowed with a certain quality preference, he also accepts content of any higher quality.
2. Consumers have *search costs* when they allocate their time in consuming content.
3. The cost to endorse is negligible.
4. Content producers are motivated by endorsement from consumers.

#### **2.3.2 Individual Behavior and System Dynamic**

We use a simple model to describe the essence of a content equilibrium. For the moment we make three simplifying assumptions: (1) there are only two quality scores,  $z_1 < z_2$ ; (2) there are two consumers  $i = 1, 2$ , those who have threshold quality at  $z_1$  and those who have threshold quality at  $z_2$ ; and (3) there are exactly two producers, one producing content with quality  $z_1^p \in [z_1, z_2)$  and the other  $z_2^p \in [z_2, \infty)$ , where  $z_1^p$  and  $z_2^p$  denote product qualities.<sup>7</sup> Note that type 1 consumer likes content produced by both producers whereas type 2 consumer only prefer content of quality  $z_2^p$ . Each of the two consumers can be considered an aggregation of many identical consumers. Similarly, each of the two producers can be considered an aggregation of many identical producers. It is possible that some of the users are both consumers and producers. For the moment, we assume that consumption and production decisions are independent. We begin our analysis by describing system dynamics, without assuming that it is in equilibrium.

---

<sup>7</sup> We thank an anonymous reviewer for suggesting this flexible specification which admits the following specification:  $z_1^p = z_1$ ,  $z_2^p = z_2$ .

Let  $x_1(t), x_2(t)$  denote the demand by the two consumers during a period of time  $t$ . Let  $\alpha_1, \alpha_2$  denote exogenous preference for content by the two consumers, respectively. Both consumers have a fixed budget  $W_1$  and  $W_2$ . A consumer  $i$  decides between content  $x_i(t)$  and a composite good  $x_o(t)$  (we use  $x_o^1$  and  $x_o^2$  to denote the demand of the two consumers for the composite good). The search cost denoted by  $p_i(t), i=1,2$ , is paid by the consumer in order to find a piece of content that satisfies his or her preferences. We assume that search process is sampling with replacement since the amount of content is large.<sup>8</sup> Of course, the actual search on content-sharing sites is free, but consumer still pays with lost time (with which he or she could be earning money). In particular, we model the consumer's cost as  $p_i(t) = w\tau s_i(t)$ , where  $w$  is the consumer's wage rate (we assume equal wage for simplicity),  $s_i(t)$  is the quantity of content that the consumer must examine before finding a piece of content of quality  $z_i$  or higher, and  $\tau$  is the average time required to examine one piece of content. Let  $F(z_i)$  denote the cumulative probability distribution function, so that  $F(z_i)$  indicates the probability that a given piece of content has a quality less than  $z_i$ . Then for an individual seeking quality  $z_i$  or above, the probability he is satisfied is  $1 - F(z_i)$  because any quality higher than  $z_i$  will be satisfactory. Thus the probability of a consumer being satisfied for the first time with the  $n$ -th photo follows a geometric distribution  $(1 - F(z_i))F(z_i)^{n-1}$ , and the expected number of photos he needs to sample in order to find a photo of quality  $z_i$  or higher is  $s_i = 1/(1 - F(z_i))$ . In our context there are exactly two discrete quality values. Let  $m_i(t), i=1,2$  denote the *sampling probabilities* for the low and high quality content—the probability that content selected at random has quality  $z_i$ . Then the expected number of steps to find a quality at least as good as  $z_i$  is:

---

<sup>8</sup> We thank an anonymous reviewer for this clarification.

$$s_1(t) = \frac{1}{m_1(t) + m_2(t)} = 1, \quad s_2(t) = \frac{1}{m_2(t)}. \quad (26)$$

Note that it takes only one step for a consumer looking for low quality, because any content will suffice. Let  $u_1$  and  $u_2$  denote utilities for the two consumers. Supposing that consumers have a quasilinear utility function, they have the following maximization problem:

$$\begin{aligned} \max_{x_i(t), x_o^i(t)} \quad & \alpha_i \ln x_i(t) + x_o^i(t) \\ \text{s.t.} \quad & p_i(t)x_i(t) + x_o^i(t) = W_i, \quad i = 1, 2. \end{aligned} \quad (27)$$

The optimal values of demand are inversely proportional to the search cost:

$$x_1(t) = \frac{\alpha_1}{p_1(t)} = \frac{\alpha_1}{\tau w}, \quad x_2(t) = \frac{\alpha_2}{p_2(t)} = \frac{\alpha_2}{\tau w} m_2(t). \quad (28)$$

The demand for high quality increases with its sampling probability because higher sampling probability makes it easier to find content. Next we make the link between sampling probability and production. We consider the simplest case in which each piece of content has an equal probability of being sampled. Let  $Y_1(t), Y_2(t)$  denote the total number of pieces of content produced at the end of time period  $t$  in the system having low and high qualities, respectively. The sampling probability at period  $t$  is a function of total production output at the end of previous period  $t-1$ :

$$m_1(t) = \frac{Y_1(t-1)}{Y_1(t-1) + Y_2(t-1)}, \quad m_2(t) = \frac{Y_2(t-1)}{Y_1(t-1) + Y_2(t-1)}. \quad (29)$$

In the basic model, we assume that consumers will endorse whenever their demand is met. This is the link between consumer choices and producer incentives. Let  $e_1(t), e_2(t)$  denote the total number of endorsements attracted by low quality content and high quality content during time period  $t$ :

$$\begin{aligned}
e_1(t) &= x_1(t)m_1(t) = \frac{\alpha_1}{\tau w} m_1(t) = \frac{\alpha_1}{\tau w} \frac{Y_1(t-1)}{Y_1(t-1) + Y_2(t-1)}, \\
e_2(t) &= x_1(t)m_2(t) + x_2(t) = \frac{\alpha_1}{\tau w} m_2(t) + \frac{\alpha_2}{\tau w} m_2(t) = \frac{(\alpha_1 + \alpha_2)}{\tau w} \frac{Y_2(t-1)}{Y_1(t-1) + Y_2(t-1)}.
\end{aligned}
\tag{30}$$

The key point is that the demand for low quality will be shared by high quality because of the imperfect substitution and non-rival properties.

We have so far discussed the relationship between consumer behavior and period endorsement. Sampling and endorsement distribution processes are stochastic. We can derive the endorsement distribution in Eq. (30) from alternative stochastic specifications and law of large numbers. Suppose that each type of consumer has  $n$  members, and let  $x_{1j}$  and  $x_{2j}$  where  $j=1, \dots, n$  denote the demands for low and high quality consumers. Suppose that high quality consumers know about the sampling probability  $m_2$  up to an error, i.e.,  $m_{2j}$ , and  $E(m_{2j}) = m_2$ . Consider an experiment with  $n$  samples, each of which has a pair of low and high consumers. Then the period endorsement from each sample are random variables denoted by  $e_{1j}(t)$  and  $e_{2j}$ :

$$\begin{aligned}
e_{1j}(t) &= \sum_{l=1}^{x_{1j}} d_l(t), \\
e_{2j}(t) &= \sum_{l=1}^{x_{1j}} (1 - d_l(t)) + x_{2j}(t).
\end{aligned}
\tag{31}$$

where  $d_l(t)$  denote an indicator function such that  $d_l(t) = 1$  if a demand of low quality consumer is given to a low quality content, and  $d_l(t) = 0$  otherwise. From Eq. (28), we have  $x_{1j}(t) = \alpha_1 / (\tau w)$ ,  $x_{2j}(t) = \alpha_2 m_{2j}(t) / (\tau w)$ . We assume that sampling process and whether particular demand of a low quality consumer is satisfied by a low quality content are independent. Then the expected period endorsements from each sample are

$$\begin{aligned}
E(e_{1j}(t)) &= \frac{\alpha_1}{\tau w} m_1, \\
E(e_{2j}(t)) &= \frac{\alpha_1}{\tau w} m_1 + \frac{\alpha_2}{\tau w} m_2(t).
\end{aligned} \tag{32}$$

Let  $x_1 = n\alpha_1 / (\tau w)$  and  $x_2 = n\alpha_2 / (\tau w)$ . By Law of Large Number, Eq. (30) is an approximation of period endorsement when  $n$  is large.<sup>9</sup>

Next we examine period endorsement on each piece of content. We can write the *endorsement increase per unit existing content*, denoted by  $\Delta e_1, \Delta e_2$  during time period  $t$  for each quality as:

$$\begin{aligned}
\Delta e_1(t) &= \frac{e_1(t)}{Y_1(t-1)} = \frac{\alpha_1}{\tau w} \frac{1}{Y_1(t-1) + Y_2(t-1)}, \\
\Delta e_2(t) &= \frac{e_2(t)}{Y_2(t-1)} = \frac{(\alpha_1 + \alpha_2)}{\tau w} \frac{1}{Y_1(t-1) + Y_2(t-1)}.
\end{aligned} \tag{33}$$

By taking the ratio of these two expressions and simplifying, we have:

$$\frac{\Delta e_2(t)}{\Delta e_1(t)} = 1 + \frac{\alpha_2}{\alpha_1}. \tag{34}$$

Having described consumer behavior, we next specify producer behavior. In content-sharing websites, producers observe endorsement from consumers. For example, Yahoo Flickr shows the total number of endorsement each photo gets. We assume that content producers are motivated by earning endorsement. Endorsement is accumulated over time, and we define *total per-unit endorsement*  $E_1(t), E_2(t)$  at period  $t$  as

$$E_1(t) \equiv \Delta e_1(t) + E_1(t-1), \quad E_2(t) \equiv \Delta e_2(t) + E_2(t-1). \tag{35}$$

Since producers generate new content at every period, older content receives more

---

<sup>9</sup> We thank an anonymous reviewer for the suggestion of stochastic specification.

per-unit endorsement than newer content. We assume that producers are motivated by the maximum per-unit endorsement. So  $E_1(t)$  and  $E_2(t)$  are defined as cumulative per-unit endorsement on the oldest content. As we will show that the initial total unit endorsement  $E_i(0)$  does not change the equilibrium, we only require  $E_i(0) \geq 0$ .

We consider a simple model of producer behavior. Let  $y_1(t), y_2(t)$  denote the production for low and high quality content during period  $t$ . The current period's production contributes to the total content available in the next period:

$$Y_i(t) = y_i(t) + Y_i(t-1), \quad i = 1, 2. \quad (36)$$

Producing a piece of content does have a real monetary cost, because making content involves expending labor. We assume that a producer is able to compare the endorsement with his unit labor cost—wage rate  $w$ . Let  $\psi$  denote this monetary value per unit of endorsement, and let  $l_i$  denote labor needed to produce content. To simplify the presentation, assume that the production technology is  $y_i = l_i^{1/2}$  (we will show that our results do not depend on a particular specification of technology). Then the producer solves the following optimization problem:

$$\begin{aligned} \max_{y_i(t)} \quad & \psi E_i(t) y_i(t) - w l_i(t) \\ \text{s.t.} \quad & y_i(t) = l_i(t)^{1/2}. \end{aligned} \quad (37)$$

We call the difference between the monetary value of endorsement and labor cost *quasi-profit* because motivation for producers is endorsement. It is worth noting that current period endorsement  $E_i(t)$  is a function of previous period production  $y_i(t)$ . But since producers take endorsement as given, they do not perceive endorsement as a

function of production.<sup>10</sup> An implicit assumption is that, producers use current period endorsement  $E_i(t)$  to predict next period endorsement  $E_i(t+1)$ . The consistency needed to justify this behavior is guaranteed in equilibrium as shown in Section 2.3.3. The optimal production to maximize quasi-profit is

$$y_i(t) = \frac{\psi E_i(t)}{2w}. \quad (38)$$

We have specified the behavior for both demand and supply sides. Though we take a disequilibrium approach, we have introduced interaction between consumption and production. Unlike in an equilibrium where individual no longer change behaviors, time matters in a disequilibrium environment, and we can order content system dynamics in the following sequence illustrated in Figure 5:

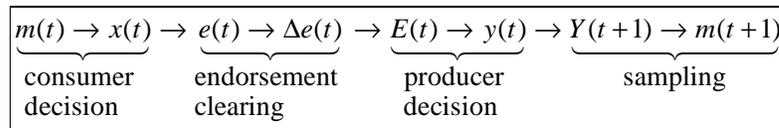


Figure 5 Content System Dynamics

Sampling probability, demand, endorsement, and supply are all state variables in this chain of events, all of which are obviously related.

### 2.3.3 Content Equilibrium

We have described consumer and producer behaviors without assuming the system is in equilibrium. Decisions are made based on exogenous data including

---

<sup>10</sup> Even if producers are forward looking, they cannot change endorsement due to large number of producers in content system. Assuming myopic or forward-looking for producer behavior does not make a difference.

tastes, wealth, technology, and wage rate. In addition, behaviors are hinged upon two endogenous variables—sampling probability and endorsement which are at center stage of policy issues. The individual models with interaction in Section 2.3.2 have given their genesis: consumers hand out endorsement; and producers indirectly decide sampling probability through production. However, the individual models are silent about their persistence as manifested in inequality phenomenon across content platforms. The tool for studying persistence in social science is equilibrium analysis. In this section, we develop a content equilibrium from first principles.

First, we summarize the content system described in Section 2.3.2.

**Definition (User Generated Content system):** A user generated content system (we use **content system** hereinafter) is a collection of individuals consuming and producing free and non-rival content through content-sharing websites. Consumers take sampling probabilities as given and maximize utility. Producers take endorsement as given and maximize quasi profits. Endorsement needs to clear, i.e., total endorsements received by producers equal total endorsement given by consumers.

Dictionary definition of equilibrium is “a state of balance between opposing forces or actions that is either static...or dynamic...” (Equilibrium 2010). Market equilibrium is defined as a state in which goods consumers and producers have no incentive to change their behavior. We can define a similar content equilibrium.

**Definition (Content equilibrium 1):** A content system is in **equilibrium** if content consumers and producers do not change their period decisions over time, i.e.,  $x_i(t) = x_i(t-1)$  and  $y_i(t) = y_i(t-1)$  for  $i = 1, 2$ .

This definition does not immediately apply to our content system. Content producers are motivated by *accumulated* per-unit endorsement  $E_1(t)$  and  $E_2(t)$ . By definition in Eq. (35),  $E_1(t)$  and  $E_2(t)$  are non-decreasing over time. Therefore, it is not clear

whether producers will reach steady decisions. This is where content producers differ from traditional goods producers in addition to non-monetary nature of reward. Goods producers are motivated by reward in a period, whereas content producers are motivated by cumulative reward over multiple periods. This is caused how content websites present endorsement. Content websites show cumulative endorsement rather than period endorsement, and therefore make the former as the sought-after reward for producers. So we next define content equilibrium as a state in which consumers do not change their behaviors. This definition admits the first definition, but not vice versa. We will show later that when time is large, producers will not change their behaviors in equilibrium.

**Definition (Content equilibrium 2):** A content system is in **equilibrium** if content consumers do not change their demand over time, i.e.,  $x_i(t) = x_i(t-1)$  for  $i = 1, 2$ .

Content consumers make demand decisions based on exogenous tastes, income, time to examine one photo, and wage rate, i.e., a exogenous vector of  $(\alpha_i, W_i, \tau, w)$  for  $i = 1, 2$ . Demands are also functions of endogenous sampling probabilities  $m_i$  which in turn, are functions of total production  $Y_i$  for  $i = 1, 2$ . If  $Y_1$  and  $Y_2$  no longer change, then demand will be stable. Therefore, we can also define content equilibrium in terms of total production. But first, we observe the fact that demands are independent from arbitrary scaling of total production  $Y_1$  and  $Y_2$ .

**Proposition 2.1 (Demand is homogeneous of degree zero in production):** Demand depends on *relative* value of total production. In another word,  $x_i(Y_1, Y_2) = x_i(\alpha Y_1, \alpha Y_2)$  for any scalar  $\alpha > 0$ .

See Appendix B for a proof.

This results implies that demands do not change if we normalize decision inputs  $Y_1$  and  $Y_2$  by  $Y_1$  to 1 and  $Y_2 / Y_1$ . The intuition is that sampling probability is a relative value, so scaling total production does not change sampling probability. Note that in market

environment, demand is independent from scaling of price. Using this fact, we give a second content equilibrium definition in terms of normalized total production.

**Definition (Content equilibrium 3):** A user generated content system is in **equilibrium** if relative total production is constant over time, i.e.,

$$\frac{Y_2(t)}{Y_1(t)} = \frac{Y_2(t-1)}{Y_1(t-1)}. \quad (39)$$

In Section 2.3.2, we have shown that  $E_1(t)$  and  $E_2(t)$  are functions of  $Y_1(t-1)$  and  $Y_2(t-1)$ . So a content equilibrium defined by total production can also be defined by endorsement.

**Definition (Content equilibrium 4):** A user generated content system is in **equilibrium** if relative endorsement is constant over time, i.e.:

$$\frac{E_2(t)}{E_1(t)} = \frac{E_2(t-1)}{E_1(t-1)}. \quad (40)$$

Next proposition states that the three content equilibrium definitions are equivalent.

**Proposition 2.2 (Definition equivalence):** Definition 2, 3, and 4 are equivalent.

See Appendix B for a proof.

This result allows us to define content equilibrium from the first principles, i.e., as a state in which consumers have no incentive to change behaviors, and to prove existence of content equilibrium from one of its equivalences—that normalized endorsements no longer change.

Content users care for nothing but their own personal interests in a way similar to consumers and producers in a market behave. Adam Smith (Smith 1776) sees an *invisible hand* guiding an economic system despite seemingly chaotic actions by selfish individuals. A similar question arises for content system. Can a decentralized

content system settle to a harmony? Does equilibrium have some socially desirable features? And is equilibrium fragile? These are the questions addressed in the rest of this section.

**Proposition 2.3 (Existence):** In a content system, a content equilibrium by definition 2 exists, with the property:

$$\frac{E_2(t)}{E_1(t)} = 1 + \frac{\alpha_2}{\alpha_1}, \text{ for all } t.$$

See Appendix B for a proof.

This result implies that, it is possible for a content system to have a state in which consumers have no incentives to change their behaviors. We have argued that producers may still change their behaviors at equilibrium. Nevertheless, consumers and producers, despite knowing nothing about others' decisions, can be compatible in the sense that endorsements given by consumers and endorsements received by producers are equal. Note that compatibility can be easily achieved in an out-of-equilibrium world in which individual can adjust their behaviors when environment changes. But it is much harder when individuals do not change their behaviors. This is the crust of price equilibrium (Arrow and Debreu 1954). This result shows compatibility when consumers do not change behavior while producers still do.

The content consumer specification that quantity is inverse to search cost is important to equilibrium result. This specification captures the opposing effects of sampling probabilities on demand. On the one hand, higher total supply of one quality type relative to the other quality type (and hence higher sampling probability) increases the demand by lowering search cost. On the other hand, relatively high total supply reduces the unit endorsement change. We discuss relaxing this assumption in Section 2.5.2.

This equilibrium is robust to production technology. We use a simple specification for content production technology, i.e.,  $y = l^{1/2}$ . However, it can be easily verified that choice of production technology does not affect the equilibrium results. Thus our model admits a wide range of production technology. The important supply side condition is that producers are motivated by endorsement.

The characterization of content equilibrium is puzzling. We have demonstrated that interaction of demand and supply gives rise to the content equilibrium. But the quantitative feature of content equilibrium only depends on the demand condition—preference  $\alpha_1$  and  $\alpha_2$ . This explains why the specification of production technology does not affect equilibrium endorsement. One-sided equilibrium first appears in Paul Samuelson’s Non substitution Theorem (Samuelson 1951). Samuelson shows that an equilibrium price system depends only on the supply condition when technology is constant-return-to-scale. Our model provides a complementary case where demand is inversely related to the demand condition—search cost. We will use this result to show that the size of endorsement received by a high quality producer may be larger than what his quality warrants in Proposition 2.7.

We have shown existence by the second definition of content equilibrium in which consumers have no incentive to change behaviors. We can show that when time is sufficiently large, producers in a content system will have no incentive to change their behaviors.

**Proposition 2.4 (Production in asymptotic equilibrium):** In content equilibrium when time is large, producers do not change their productions, i.e.,  $y_i(t)$  is a constant for  $i = 1, 2$  when  $t \rightarrow \infty$ . In another word, a content equilibrium by definition 1 exists when time is large.

See Appendix B for a proof.

This result suggests that period production increases in short-run, and hit a ceiling in

long-run. Therefore producers will not change their behaviors in the equilibrium defined by stability of consumer behaviors. It implies that the equilibrium by Definition 1 exists when time is large. Therefore it is possible for consumer and producer to be compatible when neither of them changes behavior.

We have shown that selfish individuals in a content system can reach a state of balance. Next we ask whether such a state is socially desirable. This is a relevant question to potential regulators of content platforms. Policy makers are interested in improving consumer welfare. So we examine whether the content equilibrium can be used as a benchmark for consumer welfare. First, we need to define a set of allocation  $(x_1, x_2)$  that the content equilibrium allocation is compared against. In market context, optimality of competitive equilibrium is established by comparison with all allocations that clear market. Content does not clear, so market clearing allocation no long applies to content. A natural candidate in a content system is the set of allocations that clears endorsement.

**Definition (Endorsement clearing allocation):** An endorsement clearing allocation

$$(x_1, x_2) \text{ satisfies } \frac{x_1(t)}{Y_1(t-1) + Y_2(t-1)} + \frac{x_2(t)}{Y_2(t-1)} = \Delta e_2(t).$$

This definition states the fact that unit endorsement for high quality content comes from the high quality consumer as well as the low quality consumer.

**Proposition 2.5 (Optimality):** When the outside goods supply is fixed, i.e.,  $x_o^1 + x_o^2 = \bar{x}_o$ , the content equilibrium is Pareto optimal among endorsement clearing allocation for any fixed  $\Delta e_2(t)$ .

See Appendix B for a proof.

This result suggests that content systems, despite that individuals care for only their own interests, are socially desirable in terms of consumer welfare. This result is

surprising in light of the inequality observed in Figure 3 and 4 and to discuss in details in Section 2.4. Inequality may alarm lawmakers because only in an oligopoly market, a smaller number of producers produce a lion share. It is well known that oligopoly market is not optimal to consumer welfare. This result eases such concern.

We have established that content equilibrium is a desirable state. But as implied by system dynamic in Section 2.3.2, there exist numerous configurations of a content system. The content equilibrium is only one of infinitely many states a content system can be in. So an important question is how a system reaches equilibrium, and whether it will stay there. This question is relevant to empirical research. Researchers have been increasingly applying equilibrium as identification conditions in empirical analysis. For example, Nash equilibrium is assumed for modeling competition in BLP (1995). If we can find a force that steer a system towards equilibrium, it is more credible to apply equilibrium concept in empirical work. Next we show such a force exists in a content system.

**Definition (Self-organization):** An ability of a content system to steer itself towards equilibrium.

**Definition (Stability):** A state of content system is *asymptotically stable* if the system converges to the state when elapsed time is large.

**Proposition 2.6 (Stability):** Content equilibrium has self-organization feature and is asymptotically stable.

See Appendix B for a proof.

In related social network field, researchers have established a distinguishing feature of social network—small world phenomenon (Milgram 1967). Jon Kleinberg shows that a decentralized social network not only has short path existing between any

two individuals, but also can find it when the social network satisfies certain structural property (Kleinberg 2000). Our result can be interpreted in a similar way—a content system not only has an optimal state, but also can find it in a decentralized manner when the system has a simple behavioral property—demand is inverse to search cost. This specification generates a “gradient” feature for the content system. A second related model is in market theory. The competitive market equilibrium was first established by a fixed-point argument (Kakutani 1941). But it is a static framework, and questions remained as to how a market arrives at its equilibrium. So researchers built an elaborate dynamic structure called the *Tâtonnement* process to show how price system settles to its equilibrium (Arrow and Hahn 1971).

It is not a coincidence that three models: price adjustment model, Kleinberg’s small world search model, and content equilibrium model, despite different specifications and applications (namely, market, social network, and content system), all fall into a class of mathematical models.<sup>11</sup> It is established by Liapunov theory that if a dynamic system satisfies certain conditions, it can reach stability. Upon examination, our model provides another case analogous to the list of stable dynamic systems found in the physics and biological literature. To show that the system is stable, we begin by defining a function  $f(\cdot): R^2 \rightarrow R^2$  that models the dynamics of endorsement, taking  $E_1(t)$  and  $E_2(t)$  and producing  $E_1(t+1)$  and  $E_2(t+1)$ :

$$\begin{aligned} E_1(t+1) &\equiv f(E_1(t), E_2(t)) = \Delta e_1(t+1) + E_1(t), \\ E_2(t+1) &\equiv f(E_1(t), E_2(t)) = \Delta e_2(t+1) + E_2(t). \end{aligned} \tag{41}$$

where  $\frac{\Delta e_2(t+1)}{\Delta e_1(t+1)} = 1 + \frac{\alpha_2}{\alpha_1}$ . We define  $V^*(\cdot): R^2 \rightarrow R$  as a distance measure.

---

<sup>11</sup> Kleinberg (2000) does not make the connection between the search model and Liapunov function.

$$V^*(E_1, E_2) \equiv \left( \frac{E_2}{E_1} - 1 - \frac{\alpha_2}{\alpha_1} \right)^2. \quad (42)$$

The key idea in showing stability is to find a function that summarizes the stable variable. Recall the definition of the discrete version of a Liapunov function (Luenberger 1979):

**Definition (discrete Liapunov function):** Let  $\Omega$  denote state space, and let  $X \in \Omega$  denote the state variable. The state variable evolves by  $X(t+1) = f(X(t))$ . A Liapunov function  $V(X)$  satisfies the following properties: (a)  $V(X)$  is continuous; (b)  $V(\bar{X})$  has a unique minimum at  $\bar{X} \in \Omega$ ; (c)  $V(f(X)) - V(X) \leq 0$ .

We can verify that  $V^*(\cdot)$  is a Liapunov function. Therefore, the content system is stable.

## 2.4 Endogenous Endorsement and Sampling Probability

The content equilibrium outlined in the last section allows us to discuss persistent patterns on endogenous variables in a content system. We characterize distribution of endorsement and sampling probability arising from the content equilibrium in Section 2.4.1, and show that a content system is selective. Other forces not discussed in the previous section may also affect extent of the selectiveness. We consider three of them: incomplete endorsement, search by endorsement, and entry of new producer in Section 2.4.2.

### 2.4.1 Baseline Results

Content equilibrium in Proposition 2.3 immediately gives rise to the following result.

**Proposition 2.7 (Inequality in Endorsement and Sampling Probability):** At the

content equilibrium, the high quality producer has more unit endorsement than low quality producer regardless of consumer preferences. Consequently, the high quality producer produces more content than the low quality producer. High quality content has high sampling probability, that is, high quality content is easier to find than low quality content, which is a collaborative filtering mechanism. Finally, the size of inequality is independent from the difference in producer quality. Formally,  $\frac{E_2}{E_1} = \frac{Y_2}{Y_1} = \frac{m_2}{m_1} = 1 + \frac{\alpha_2}{\alpha_1}$ , and  $\frac{E_2}{E_1}, \frac{Y_2}{Y_1}, \frac{m_2}{m_1}$  are independent from  $z_1^p, z_2^p$ .

See Appendix B for a proof.

First, this result shows that the ratio of cumulative endorsements, total productions, and sampling probabilities are equivalent. Therefore, result on endorsement applies to the other two. This proposition is motivated by the endorsement inequality in a precise form—power-law distribution observed in Figure 3. Existing models (e.g. Yule 1925) generates power-law distribution via a rich-gets-richer process. However, which party ends up being the richer one is unclear. Almost any producer can become a superstar (Alder 1985). This proposition states that producer quality matters in deciding who gets rich in the first place. Second, Simon (1971) raises the important issue of designing organization to promote high quality information in an information rich world. In the same spirit, the content system in this paper promotes high quality content in a content rich world despite that users are un-organized and selfish. Incomplete substitution behavior drives a content system to select high quality producer with the aid of technology which makes content non-rival and enables producers to reach a large number of consumers. Third, producer quality plays a limited role in determining endorsement distribution. Higher quality producer earns more endorsement. But how much more is independent from quality difference. It is argued in labor literature that compensation for top-level management is often

disproportional larger than those for employees although skill differences may be small (Lazear and Rosen 1981). Our result suggests that higher quality producers may earn disproportional endorsement and produce disproportional content although difference in quality may be small.

### 2.4.2 Other Forces

We discuss three forces which are not in the basic model, but influence endorsement and sampling probabilities.

#### Incomplete Endorsement

Endorsement is generated from demand. In the basic model, we assume that an endorsement is given to a piece of content as long as a consumer's quality demand is met. This simple assumption is possible from a cost point of view because the cost to endorse is so low. But consumers may choose not to endorse for a variety of reasons—they may not be familiar with the endorsement mechanism, or they may forget, for example. Such incomplete endorsements are analogous to non-responses in a survey.

We examine how incomplete endorsement rate impacts the inequality. Suppose  $\beta_1, \beta_2$  are the percentages of the demand that translate into endorsements for each type of consumers. Then the total endorsements during time  $t$  are:

$$\begin{aligned} e_1'(t) &= \beta_1 x_1(t) m_1(t) = \frac{\beta_1 \alpha_1}{w} \frac{Y_1(t-1)}{Y_1(t-1) + Y_2(t-1)}, \\ e_2'(t) &= \beta_1 x_1(t) m_2(t) + \beta_2 x_2(t) = \left( \frac{\beta_1 \alpha_1}{w} + \frac{\beta_2 \alpha_2}{w} \right) \frac{Y_2(t-1)}{Y_1(t-1) + Y_2(t-1)}, \end{aligned} \tag{43}$$

and the new equilibrium is:

$$\frac{E_2'(t)}{E_1'(t)} = 1 + \frac{\alpha_2 \beta_2}{\alpha_1 \beta_1}. \quad (44)$$

**Proposition 2.8 (Incomplete endorsement):** The outcome of the equilibrium favors the more highly-endorsed quality.

See Appendix B for a proof.

This result suggests that if consumers want content with a certain level of quality, they should endorse in order to encourage greater production of such content. Therefore, both high and low quality producers can gain advantage depending on whether high and low consumers endorse more. In contrast, the next force we discuss only favors high quality producer.

### Endorsement in searching

In the basic model, each piece of content has equal sampling probability. However, if high quality content is easier to find (and hence has lower search cost) than low quality content, the inequality can be further skewed to high quality. This is possible because endorsement carries information on where to locate high quality content. Consumers can return to the producers they endorse in the previous period to find satisfactory content. In this sense, endorsement has important carry-over effects over time. It rewards high quality producers not only in current period, but also in future periods.

Suppose consumers decide to return to the producers they endorsed in the last period. Then the sampling probability should be weighted by unit endorsement:

$$m_1(t) = \frac{E_1(t-1)Y_1(t-1)}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1)}, m_2(t) = \frac{E_2(t-1)Y_2(t-1)}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1)}.$$

(45)

The total endorsements are

$$\begin{aligned}
 e_1(t) &= x_1(t)m_1(t) = \frac{\alpha_1}{\tau w} \frac{E_1(t-1)Y_1(t-1)}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1)}, \\
 e_2(t) &= x_1(t)m_2(t) + x_2(t) = \frac{(\alpha_1 + \alpha_2)}{\tau w} \frac{E_2(t-1)Y_2(t-1)}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1)}.
 \end{aligned} \tag{46}$$

**Proposition 2.9 (Search by endorsement):** When endorsement is used in search, endorsement is further skewed towards high quality producers in the next period.

See Appendix B for a proof.

This result implies that when endorsement is used for search, the unit endorsement will increase even when the equilibrium point in the basic model is surpassed, i.e.,  $E_2(t)/E_1(t) > 1 + \alpha_2/\alpha_1$ . This process can repeat over time, causing the inequality to become further skewed towards high quality content. This is similar to a preferential attachment (“rich-get-richer”) process which may lead to a power-law distribution. The basic model in Section 2.3 creates an initial inequality based on producer quality. Search by endorsement amplifies the quality induced inequality.

### Entry of new producers

Competing forces also exist that counterbalance the system to skew towards higher quality content for existing producers. In particular, new producers are not subject to the system dynamics. They have not received unequal endorsement. Therefore their entry may bring equal production, which increases the relative amount of low quality content for existing producers.

Suppose that there are equal numbers of new producers for low and high quality content. Since they are new producers, they have not yet received endorsements. Thus no one of them is motivated to produce more content than any

other, so the new producers are likely to contribute equal amounts of content. Suppose that a content system is in equilibrium, and  $\Delta Y$  new content are added at each quality level at the end of period of  $t - 1$ . Then the sampling probabilities are:

$$\begin{aligned} m_1(t) &= \frac{Y_1(t-1) + \Delta Y}{Y_1(t-1) + Y_2(t-1) + 2\Delta Y}, \\ m_2(t) &= \frac{Y_2(t-1) + \Delta Y}{Y_1(t-1) + Y_2(t-1) + 2\Delta Y}. \end{aligned} \quad (47)$$

We have the following result on content sampling probability in comparison to the basic model.

**Proposition 2.10 (Entry of new producers):** Entry of new producers reduces the skew towards higher quality content in terms of sampling probability in short term. In long run, however, the system returns to the equilibrium described in Proposition 7.

See Appendix B for a proof.

Firstly, this result implies that entry of new producer change inequality in sampling probability favoring the low quality content temporarily. In long term, the inequality between low and quality content will return to its equilibrium state. Secondly, there is an inequality between new and existing producers of the same quality because new producers have not been subjected to selection of content system. But in long term, both will have the same sampling probability. So an insight is that inequality in the basic model is independent from time sequence producers enter into a content system.

Timing of entry may matter when inequality depends on particular features of producers which are time dependent. Search by endorsement is one example. Suppose that in period  $t - 1$ , a content system is in equilibrium, and consumers start searching by endorsement. In addition,  $\Delta Y$  new content are added at each quality

level. We assume that consumers sample content of new producers by probability  $a$ .

Then the sampling probabilities are:

$$\begin{aligned} m_1(t) &= \frac{E_1(t-1)Y_1(t-1) + a\Delta Y}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1) + 2a\Delta Y}, \\ m_2(t) &= \frac{E_2(t-1)Y_2(t-1) + a\Delta Y}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1) + 2a\Delta Y}, \end{aligned} \quad (48)$$

while the total endorsements are:

$$\begin{aligned} e_1(t) &= x_1 m_1 = \frac{\alpha_1}{\tau w} \frac{E_1(t-1)Y_1(t-1) + a\Delta Y}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1) + 2a\Delta Y}, \\ e_2(t) &= x_1 m_2 + x_2 = \frac{(\alpha_1 + \alpha_2)}{\tau w} \frac{E_2(t-1)Y_2(t-1) + a\Delta Y}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1) + 2a\Delta Y}. \end{aligned} \quad (49)$$

**Proposition 2.11 (Entry of new producers with search by endorsement):** Entry of new producers reduces overall skew towards higher quality content for existing producers. Existing high quality producer enjoys more inequality than new high quality producers.

See Appendix B for a proof.

When consumers do not use endorsement in search, new producer will have the same inequality as existing producers of the same quality. But when consumers search by endorsement, the initial gap in endorsement between new and existing producers will persist in long run. This result highlights the fact that a process that depends on producer characteristics, e.g. endorsement, may give advantage to established producers. Searching by endorsement or not, both Proposition 2.10 and 2.11 show that entry of new producers help reduce skew to high quality producers in sampling probability. As we will show in Section 2.6, reducing inequality is consistent with policy to implement profit goal for content website providers. Therefore, content

website providers should invest to attract new producers.

## 2.5 Extending Basic Model

We first show how to extend the basic model from two qualities to any finite number of qualities in Section 2.5.1. The basic model assumes that demand is inverse to search cost, and demand decision is independent of production decision. We discuss relaxing these conditions in Section 2.5.2 and Section 2.5.3 respectively.

### 2.5.1 Multiple Qualities

We only discuss two qualities, consumers, and producers for simplicity in the basic model. But the inequality in Figure 3 and 4 suggests that inequality spans across multiple producers which cannot be explained by the basic model. We show how to extend the equilibrium model from two qualities to the case of any arbitrary finite number  $I$  qualities  $z_1 < \dots < z_I$ . The preference parameters are  $\alpha_1, \dots, \alpha_I$ . It can be shown that the average number of steps required to find a piece of content for threshold quality  $z_i$  is

$$s_i = \frac{1}{m_i + \dots + m_I}. \quad (50)$$

The demand of the consumer with threshold quality  $z_i$  is

$$x_i(t) = \frac{\alpha_i}{\tau w s_i} = \frac{\alpha_i}{\tau w} (m_i + \dots + m_I). \quad (51)$$

The endorsement during period  $t$  for producers of quality  $z_i$  is

$$e_i(t) = \frac{(\alpha_1 + \dots + \alpha_i) m_i(t)}{\tau w}. \quad (52)$$

The unit endorsement increase during period  $t$  is

$$\Delta e_i(t) = \frac{e_i(t)}{Y_i(t-1)} = \frac{(\alpha_1 + \dots + \alpha_i)m_i(t)}{\tau w Y_i(t-1)} = \frac{(\alpha_1 + \dots + \alpha_i)}{\tau w \sum_i Y_i(t-1)}. \quad (53)$$

Therefore, the ratio between the unit endorsement increase for quality  $z_i$  and quality  $z_1$  is

$$\frac{\Delta e_i(t)}{\Delta e_1(t)} = 1 + \frac{\alpha_2 + \dots + \alpha_i}{\alpha_1}, \quad (54)$$

where  $2 \leq i \leq I$ . Therefore the equilibrium in the two quality model can be extended to the case of  $I$  qualities. The regularity again appears in the distribution of production in the equilibrium, that is:

$$\frac{y_i(t)}{y_1(t)} = 1 + \frac{\alpha_2 + \dots + \alpha_i}{\alpha_1}, \quad (55)$$

and

$$\frac{Y_i(t)}{Y_1(t)} = 1 + \frac{\alpha_2 + \dots + \alpha_i}{\alpha_1}. \quad (56)$$

This result and search by endorsement in Section 2.4 provide a rich-gets-richer process on multiple qualities that may lead to power-law distribution in Figure 3 and 4. This result implies that among high quality producers, some will produce more than others due to finer differentiation in quality.

### 2.5.2 Bounds for Nonlinear Demand

The basic model assumes that demand is linear to the inverse search cost. More generally, demand can decrease with price in a nonlinear way; for example, an alternative demand specification (Hendel 1999 and Dube 2004) has the following form:

$$x_i \propto (1/p_i)^{1/(1-\eta)}. \quad (57)$$

In this definition, the parameter  $\eta \in (0,1)$  controls the shape of the function, with  $\eta = 0$  making it linear. Note that Hendel and Dube model joint decisions of product and quantity, while quality preferences are given in our model. Interestingly, Dube (2004) estimates that  $\eta = 0.009$  for soft drinks, which makes quantity almost linear in  $1/p_i$ . In the case where linearity is not a good approximation, we can still use the content equilibrium in the basic model to establish bounds on the content system behavior.

We have shown that the content system evolves around a constant—the ratio of unit endorsement increase  $\Delta e_2(t)/\Delta e_1(t)$ , where in particular the state variable  $E_2(t)/E_1(t)$  always moves towards  $\Delta e_2(t)/\Delta e_1(t)$ . When the demand is not linear to the inverse of search cost,  $\Delta e_2(t)/\Delta e_1(t)$  is no longer a constant, but we can still characterize the system. The idea is that if we can bound  $\Delta e_2(t)/\Delta e_1(t)$ , we can learn the range of system dynamics. We use Hendel and Dube’s specification as an example. Under Hendel and Dube’s specification, quantity demanded is proportional to a function of sampling probability  $m^\theta$ , where  $\theta > 1$ . Without loss of generality, let the new demand in the two-quality case be

$$x_1(t) = \frac{\alpha_1}{\tau w}, x_2(t) = \frac{\alpha_2}{\tau w} m_2(t)^\theta. \quad (58)$$

**Proposition 2.12 (Bounds for non-linear demand):** There exist bounds for

$$\text{equilibrium endorsement ratio, i.e., } 1 + \frac{\alpha_2}{\alpha_1} \frac{1}{2^{\theta-1}} < \frac{\Delta e_2(t)}{\Delta e_1(t)} < 1 + \frac{\alpha_2}{\alpha_1}.$$

See Appendix B for a proof.

This result provides a different approach for researchers to characterize content system when the condition for a sharp result in Section 2.3 is not available. We show that the bounds are informative on the distinguishing feature outlined in Section 2.4—

a content system is selective. Using Hendel and Dube’s specification, inequality still exists because the lower bound for equilibrium unit endorsement  $E_2(t)/E_1(t) > 1$  for all  $\theta \in (0,1)$  although it is less skewed when compared to the linear case because now the linear case equilibrium  $1 + \alpha_2 / \alpha_1$  is the upper bound. This result together with the result in Section 2.3 raise interesting observations on the relationship between individual behavior and macro outcome. Under a simple assumption—demand is inverse to search cost, we arrive at an equilibrium with desirable features: optimality and stability. Then we assume consumers are more sophisticated and they can do more complex thinking as implied by Eq. 32. But even under the more general assumption, a content system still works towards desirable states: high quality content is filtered and high quality producers are rewarded. Though we do not study origin of individual behavior in this paper, our results provide cases to study how individual behavior is developed in the first place from sociobiology and evolutionary psychology point of views (Wilson 1975 and Buss 1995).

### **2.5.3 Dependence between Consumption and Production**

An important feature of a content system is that content users both consume and produce content. In the basic model, the decisions to consume and produce contents were assumed to be independent. In this section, we relax this assumption and incorporate a dependency caused by scarcity of time in our model. A user spends time in search as a consumer and in production as a producer. The two activities compete for limited time. The quasilinear utility in the basic model implies that consumption of content is not sensitive to total wealth. To study dependency case, we assume consumers have time budget  $q_1$  and  $q_2$  for consuming content and an alternative goods. Recall the time  $l_1$  and  $l_2$  used in producing content. The total time budgets for consumption and production are  $T_1 = q_1 + l_1$  and  $T_2 = q_2 + l_2$ . In order to have the

consumption as a function of budget, we use a Cobb-Douglas utility function instead of quasilinear utility function:

$$\begin{aligned} \max_{x_i(t), x_o^i(t)} \quad & \alpha_i \ln x_i(t) + (1 - \alpha_i) \ln x_o^i(t) \\ \text{s.t.} \quad & p_i(t)x_i(t) + p_o(t)x_o^i(t) = W_i, \quad i = 1, 2, \end{aligned} \quad (59)$$

where  $p_o(t)$  is the price of the outside goods.

Note that the linearity of demand in the inverse of search cost is no longer sufficient for equilibrium because the unit endorsement change is no longer a constant:

$$\frac{\Delta e_2}{\Delta e_1} = 1 + \frac{\alpha_2 q_2(t)}{\alpha_1 q_1(t)}. \quad (60)$$

But we next show that a content system may still reach equilibrium due to the time constraint. A **content equilibrium with dependent consumption and production** solves the following equation system:

$$\begin{aligned} \frac{\Delta e_2}{\Delta e_1} &= 1 + \frac{\alpha_2 q_2(t)}{\alpha_1 q_1(t)} && \text{(optimal demand),} \\ y_i &= l_i^{1/2}(t), \quad i = 1, 2 && \text{(technology),} \\ y(t) &= \frac{\psi E(t)}{2w} && \text{(optimal supply),} \\ q_i(t) + l_i(t) &= T_i, \quad i = 1, 2 && \text{(time constraint),} \\ \frac{E_2(t)}{E_1(t)} &= \frac{E_2(t-1)}{E_1(t-1)} && \text{(equilibrium condition).} \end{aligned} \quad (61)$$

**Proposition 2.13 (Existence of content equilibrium when consumption and production are dependent for a user):** A sufficient condition for a content equilibrium to exist is that the high quality user commits more overall time than the low quality user, i.e.,  $T_2 > T_1$ . There are multiple (an infinite number of) equilibria. Inequality in endorsement and production still exists in the equilibria.

See Appendix B for a proof.

We have shown that, when consumption and production are independent, content equilibrium exists when demand is inverse to search cost. This result shows that when consumption and production are dependent, under the condition that high quality users spend more time in content activities, content equilibrium also exists. With independent consumption and production, there is a single equilibrium which implies that only one set of consumer and producer behaviors will be compatible. This result shows multiple equilibria which imply that multiple sets of consumer and producer behaviors will be compatible for dependent consumption and production. It is not surprising that inequality still holds at equilibria because imperfect substitution and technology are still the micro foundations in this model. The essence of the basic model in Section 2.3 is interaction—production depends on endorsement from consumer. The insight extends to this case, i.e., one type of users' allocations of their time between consumption and production will influence allocations of the other type. So we expect that interaction pattern emerges across equilibria. First we check whether consumption times  $q_1$  and  $q_2$  are related at equilibria between high and low quality users. We examine the relationship between the two quantities by using a simulation. For simplicity, we set preference and total time as  $\alpha_1 = \alpha_2 = 1$ ,  $T_1 = 1$ , and  $T_2 = 2$ . We find that  $q_2$  is increasing in  $q_1$  as shown in Figure 6. An explanation is that when the low quality user increases consumption time, their production time decreases as does their production. Therefore, it takes fewer pieces of high quality content for the high quality user to earn the same quantity of endorsements, freeing him to spend more time consuming high quality content. This result raises an interesting question on inequality. When high quality users increase consumption, dependency implies high quality production will drop. Does an increase in consumption time by low quality user lead to reduction in inequality? So next we

check the behavior of the equilibrium endorsement ratio  $E_2(t) / E_1(t)$  as a function of the low quality user's consumption time  $q_1$  using the same simulation setup. We find the opposite result holds: the equilibrium inequality increases in  $q_1$  as shown in Figure 7. Therefore, inequality is further skewed towards high quality user. An intuition is that, both types of users reduce production. So reduction of production by high quality users is offset by reduction of production by low quality users. In addition, increase in consumption by low quality user implies increase in endorsement for both low and high quality users. But increase in consumption by high quality user implies increase in endorsement for only high quality users.

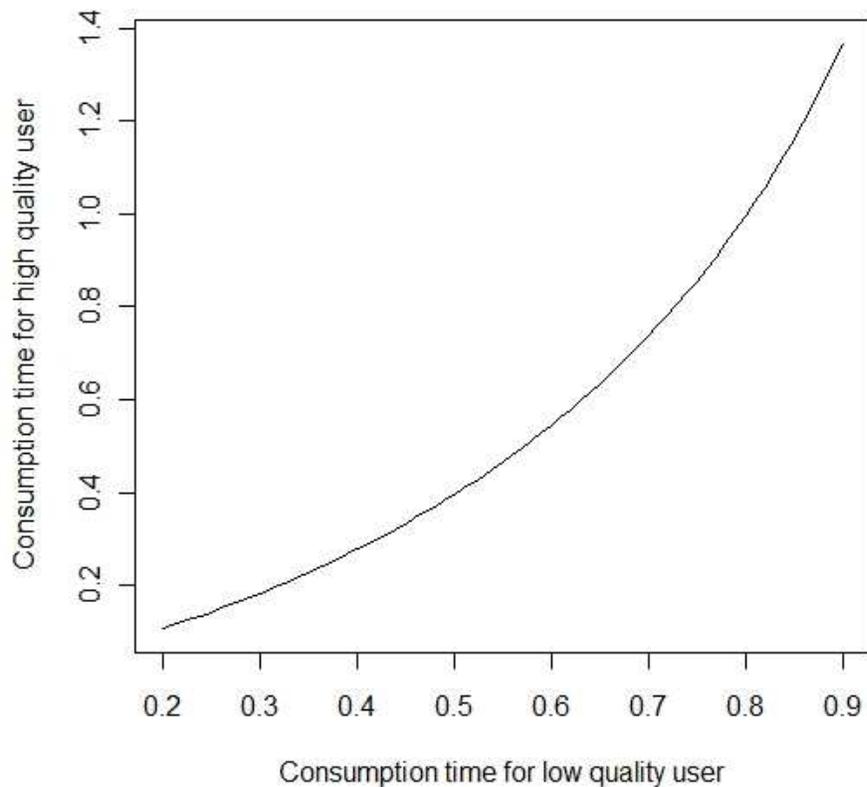


Figure 6 Consumption Time by High Quality User as a Function of Consumption Time by Low Quality User.

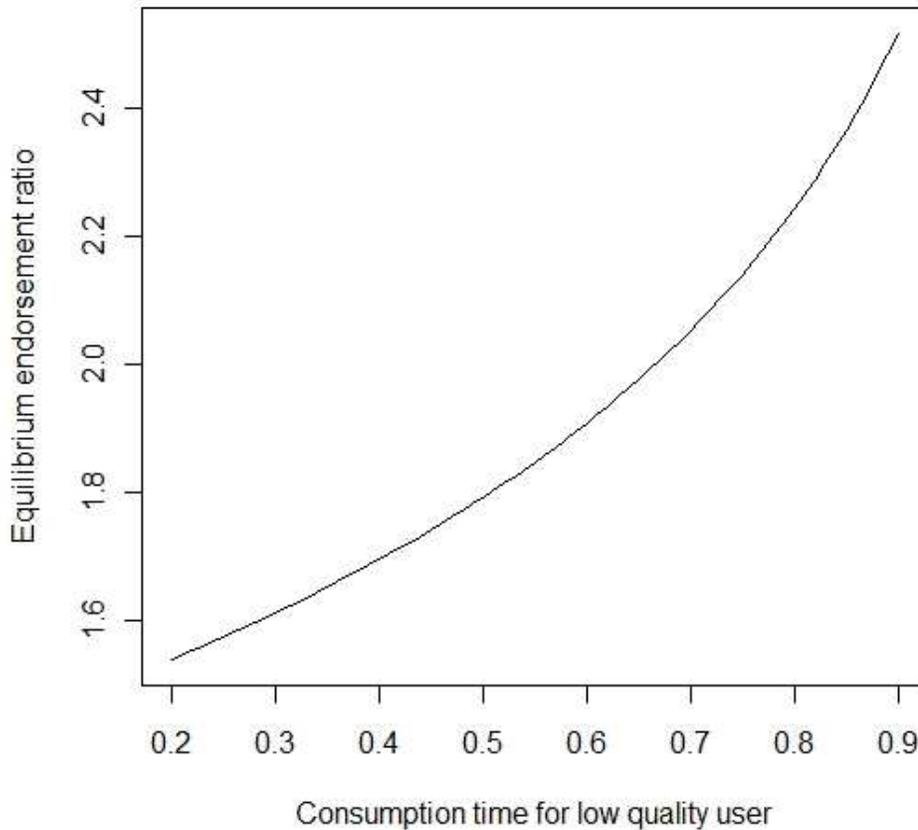


Figure 7 Endorsement Ratio as a Function of Consumption Time by Low Quality User

#### 2.5.4 Content Quality Ranking<sup>12</sup>

In the basic model in Section 2.3, we assume that there exists a universal ranking of content quality for all users. But in reality there may be multiple rankings of content quality. Multiple rankings raise interesting dynamics of the model. We use the basic model for a heuristic discussion. Rather than having one population of

<sup>12</sup> We thank an anonymous reviewer for motivating this discussion.

consumers agreeing upon one ranking, there exists two populations holding two exact opposite ranking orders, i.e.,  $z_2 > z_1$  in ranking order 1, and  $z_2 < z_1$  in ranking order 2. In first case, we assume equal sizes for the two populations. There won't be inequality in endorsement, production, and sampling probability. In second case, suppose that the population with ranking order 1 has a larger size, then inequality will arise in favor of population 1. Another interesting case emerges if we introduce learning into consumer behavior. For example, consumers may modify their quality ranking based on the overall ranking presented by the system. We leave this to future research.

## **2.6 Managing Content System**

In this section, we illustrate managerial implications of our analysis. The site provider wields significant influence over the design and operation of a content-sharing website. In particular, the site providers have influence on the mechanisms by which consumers search new content—the order in which content is ranked in response to a keyword search, for example, or the choice of which content is “featured” on the main page. The importance of these decisions is borne out by our analysis, which shows that the sampling probability is a key factor influencing consumer behavior. How the site providers engineer the sampling probability depends on their objectives. They can simultaneously be a central planner that considers the welfare of consumers, and a profit maximizer that considers its own well-being. In addition, the providers may also have a goal of engineering diversity since part of the appeal of Web 2.0 lies in its democratic nature. We want to highlight the tension among these objectives. We discuss the role of central planner in Section 2.6.1, and the role of profit maximizer in Section 2.6.2. And the role of diversity promoters is discussed in both sections.

### 2.6.1 Central planner

A strategy for improving consumer welfare is to increase the average quality of the content on the site—to maximize the expected quality,  $\sum_i m_i z_i - c$ , where  $c$  is the cost of adjusting the sampling probability. The solution is then  $m_i = 1$ , if  $i = I$ , and  $m_i = 0$ , otherwise, where  $I$  denotes the maximum possible quality index. The downside of this objective is that it leads to a concentration of sampling probability on the best quality content. An alternative objective would be

$$\max_{m_i} \sum_i m_i z_i - \sum_i \lambda_i m_i^2 - c. \quad (62)$$

where the second term penalizes the concentration of sampling probability on a particular quality. This is a similar idea to regularization methods in the statistics literature. The added penalization term ensures that the sampling probabilities are continuously distributed among content of different quality, and therefore is consistent with Web 2.0’s democratic objective. The optimal solution then is

$$m_i = \lambda_i z_i / 2. \quad (63)$$

### 2.6.2 Profit maximization

Content website providers may profit from selling services to producers. An example is Flickr’s “Pro account” feature, which removes storage and bandwidth limits in exchange for a yearly membership fee. A natural way to increase the sales of Pro accounts would be to motivate producers to make and upload more content. From the two quality case discussed in Section 2.3, it is clear that the unit endorsement change is an increasing function of sampling probability. We provide a heuristic argument on how a content site provider may adjust sampling probability to increase the amount of production. Recall that the unit endorsement increase for quality  $z_i$

from Section 2.5.1 is

$$\Delta e_i(t) = \frac{(\alpha_1 + \dots + \alpha_i)m_i(t)}{\tau w Y_i(t-1)}.$$

Let  $\Delta y_i(t) = y_i(t) - y_i(t-1)$  denote the change of period production:

$$\Delta y_i(t) = \frac{\psi \Delta e_i(t)}{2w} = \frac{\psi}{2\tau w^2} \frac{(\alpha_1 + \dots + \alpha_i)m_i(t)}{Y_i(t-1)} = K \frac{(\alpha_1 + \dots + \alpha_i)m_i(t)}{Y_i(t-1)}, \quad (64)$$

where  $K = \psi / (2\tau w^2)$ . A content site provider may wish to encourage production, because content suppliers are more likely to purchase additional services (such as extra storage or bandwidth) when they produce more content. For simplicity, we assume that the purchase probability is proportional to the total content output. So in period  $t$ , for producer  $i$ , the purchase probability is  $[\Delta y_i(t) + y_i(t-1) + Y_i(t-1)] / B$ , where  $B$  is a normalizing constant for probability definition. Note that for the service that removes storage/bandwidth limit, the producers of quality  $z_i \geq z_i^*$  already have the service because their total production  $Y_i$  has surpassed the storage/bandwidth limit. In this case, we only consider the producers of quality  $z_i < z_i^*$ . Suppose that the price for the service is  $p_s$ . Then expected revenue for producer  $i$  in period  $t$  is  $p_s \frac{\Delta y_i(t) + y_i(t-1) + Y_i(t-1)}{B}$ . Then a profit maximization problem is

$$\begin{aligned} \max_{m_i(t)} \sum_i p_s \frac{\Delta y_i(t) + y_i(t-1) + Y_i(t-1)}{B} - c - \sum_i \gamma_i m_i(t)^2 \\ \text{s.t. } \Delta y_i(t) = K \frac{(\alpha_1 + \dots + \alpha_i)m_i(t)}{Y_i(t-1)}. \end{aligned} \quad (65)$$

It can be rearranged to

$$\max_{m_i(t)} \sum_i \frac{p_s K}{B} \frac{(\alpha_1 + \dots + \alpha_i)m_i(t)}{Y_i(t-1)} - \sum_i \gamma_i m_i(t)^2 + \sum_i p_s \frac{y_i(t-1) + Y_i(t-1)}{B} - c. \quad (66)$$

The optimal sampling probability is

$$m_i(t) = \frac{p_s K (\alpha_1 + \dots + \alpha_i)}{B \gamma_i Y_i(t-1)}. \quad (67)$$

Suppose that the system is in equilibrium during the previous period  $t-1$ , then recall from Section 2.5.1 that  $Y_i(t-1) = (\alpha_1 + \dots + \alpha_i)Y_1(t-1) / \alpha_i$ . Then the optimal sampling probability becomes

$$m_i(t) = \frac{p_s K}{B \gamma_i} \frac{\alpha_i}{Y_1(t-1)}. \quad (68)$$

If the penalizing weights  $\gamma_i$ 's are equal, then the optimal sampling probabilities are also equal for all qualities. The intuition for this result is as follows. High quality content receives more endorsements than low quality content due to imperfect substitution and technology. But there is also more high quality content in the system than low quality content. So from a unit endorsement standpoint, it makes no difference for the content site provider to promote one quality of content over the other. Note that if the inequality is further skewed than in the basic model, we will have optimal sampling probability decreasing in the quality. This result implies the tension between profit maximization and consumer welfare. Interestingly, profit maximization requires promoting content diversity in production.

## 2.7 Conclusion

Our main contribution is to provide an analysis of system behavior of content systems and individuals. Due to the large amount of content available, the consumers face a search problem. Our empirical analysis shows the producers are motivated by endorsement. We construct a simple model to show the system can reach equilibrium based on these myopic behaviors. The content equilibrium has good properties such as optimality and stability. Our results show that endorsement is not only important for encouraging individual producers, but also crucial to enable the system to work

together.

We address an important macro phenomenon. We give empirical evidence of inequality in endorsement and production, showing that they follow power law distributions. The micro-foundation needed is Rosen's imperfect substitution and technology. The content mechanism we describe fills the link from the micro foundation to the macro phenomenon. We also show that incomplete endorsement, search by endorsement, and entry of new producers may impact the inequality.

We provide managerial insights to content site providers. Site providers are both private businesses and hosts of a public platform for social activities; therefore their objectives are both related and competing.

Our analysis has a number of practical implications. We provide a theoretical motivation for consumer behavior researchers to consider simultaneity and endogeneity issue. We provide a framework for the design and management purpose of content system providers. We provide a benchmark of consumer welfare for policy makers who regulate content platforms.

We have kept our model simple to show the essence to a content equilibrium. We also have extended the basic model to more general conditions, but even further extensions can be made. For example, the basic model assumes the same production technology for high and low quality producers. Heterogeneous production technology can be explored, which may give further advantage to high quality producers. Another direction is that consumers can find search information through some form of learning. If the learning involves their friends, social networks may also impact the equilibrium.

It is well-known that a market is the benchmark mechanism with many good properties for a decentralized system with myopic decision-makers. So the larger question for this research is, does a similar mechanism exist in a different environment where price is missing and non-rivalry is present? This paper suggests the existence

of such a mechanism. This mechanism is selective, and therefore inequality is bound to arise. On the upside, the system is a collaborative filter for high quality content.

## REFERENCES

- Adler, M. (1985), "Stardom and Talent," *The American Economic Review*, 75 (1), 208-12.
- Alexa: about Flickr.com (2009), "*Alexa: The Web Information Company*," 1 November 2009, web, accessed 1 November 2009.
- Archak, N., A. Ghose, and P. Ipeirotis (2008), "Deriving the Pricing Power of Product Features by Mining Consumer Reviews," working paper, SSRN.
- Arrow, K. J. and G. Debreu (1954), "Existence of an Equilibrium for a Competitive Economy," *Econometrica*, 22 (3), 265–290.
- Arrow, K. J. and F. H. Hahn (1971), *General Competitive Analysis*. San Francisco: Holden-Day.
- Backstrom, L., D. Huttenlocher, and J. Kleinberg (2006), "Group Formation in Large Social Networks: Membership, Growth, and Evolution," paper presented at KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge Discovery and Data-mining.
- Bergstrom, T., L. Blume, and H. Varian (1986), "On the Private Provision of Public Goods," *Journal of Public Economics*, 29 (1), 25-49.
- Huberman, B. A. and F. Wu (2007), "The economics of attention: maximizing user value in information-rich environments," *Proceedings of the 1st international Workshop on Data Mining and Audience intelligence For Advertising (San Jose, California, August 12 - 12, 2007)*, ADKDD '07, ACM, New York, NY, 16-20.
- Berry, Steven, James Levinsohn & Ariel Pakes (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63(4), 841-890.

- Bucklin, R. E., J. M. Lattin, A. Ansari, D. Bell, E. Coupey, S. Gupta, J. D. C. Little, C. Mela, A. Montgomery, and J. Steckel (2002), "Choice and the Internet: From Clickstream to Research Stream," *Marketing Letters*, 245-58.
- Buss, D. M. (1995), "Evolutionary psychology: A new paradigm for psychological science," *Psychological Inquiry*, 6, 1-30.
- Champ, H. (2009), "4,000,000,000," *Flickr Blog*, 12 October 2009, web, accessed 1 November 2009.
- Chevalier, J. and D. Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-354.
- Crandall, D. J., L. Backstrom, D. P. Huttenlocher, and J. M. Kleinberg (2009), "Mapping the World's Photos," paper presented at WWW '09: Proceedings of the International Conference on the World Wide Web.
- Debreu, G. (1959), *Theory of value*. New York: Wiley.
- Dube, J. (2004), "Multiple Discreteness and Product Differentiation: Demand for Carbonated Soft Drinks," *Marketing Science*, 23 (1), 66-81.
- Equilibrium (2010), *Merriam-Webster Online Dictionary*, retrieved July 22, 2010, from <http://www.merriam-webster.com/dictionary/equilibrium>.
- Huberman, B. A. and F. Wu (2007), "The economics of attention: maximizing user value in information-rich environments," *Proceedings of the 1st international Workshop on Data Mining and Audience intelligence For Advertising (San Jose, California, August 12 - 12, 2007)*, ADKDD '07, ACM, New York, NY, 16-20.
- Ghose, A. and S. P. Han (2009), "A Dynamic Structural Model of User Learning in Mobile Media Content," working paper, SSRN.
- Ghose, A. and P. Ipeirotis (2008), "Estimating the Socio-Economic Effect of Reviews: Mining Text and Reviewer Characteristics," working paper, SSRN.
- Graham, J. (2008), "Flickr Rules in Photo Sharing, as Video Tiptoes in," *USA Today*.

- Hendel, I. (1999), "Estimating Multiple-discrete Choice Models: An Application to Computerization Returns," *Review of Economic Studies*, 66 (2), 423-46.
- Kakutani, S. (1941), "A Generalization of Brouwer's Fixed Point Theorem," *Duke Math. J.*, 8 (3), 457-459.
- Kumar, V. (2009), "Why Do Consumers Contribute to Connected Goods? A Dynamic Game of Competition and Cooperation in Social Networks," working paper, Carnegie Mellon University.
- Kleinberg, J. M. (2000), "Navigation in a small world", *Nature*, 406, 845.
- Lazearand, E. and S. Rosen (1981), "Sherwin Rosen Rank-Order Tournaments as Optimum Labor Contracts," *Journal of Political Economy*, 89 (5), 841-64.
- Lee, T. and E. Bradlow (2009), "Automatic Construction of Conjoint Attributes and Levels from Online Customer Reviews," working paper, Wharton School.
- Luenberger, D. G. (1979), *Introduction to dynamic systems: Theory, models, and applications*. New York: John Wiley & Sons.
- McKenzie, L. (1954), "On Equilibrium in Graham's Model of World Trade and Other Competitive Systems," *Econometrica*, 22 (2), 147-161.
- Merton, R. K. (1957), "Priorities in Scientific Discovery: A Chapter in the Sociology of Science," *American Sociological Review*, 22 (6), 635-59.
- Milgram, S. (1967), "The small-world problem," *Psychology Today*, 2, 60-67.
- Newman, M. E. J. (2005), "Power Laws, Pareto Distributions and Zipf's law," *Contemporary Physics*, 46 (5), 323-51.
- Nonnecke, B. and Preece, J. (2000), "Lurker demographics: counting the silent," *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (The Hague, The Netherlands, April 01 - 06, 2000)*, CHI '00, ACM, New York, NY, 73-80.

- Oreilly, T. (2007), "What is Web 2.0: Design Patterns and Business Models for the Next Generation of Software," *Communications & Strategies*, 1, 17-37, available at SSRN: <http://ssrn.com/abstract=1008839>.
- Rochet, J. C. and J. Tirole (2006), "Two-Sided Markets: A Progress Report," *RAND Journal of Economics*, 35, 645-667.
- Rosen, S. (1981), "The Economics of Superstars," *The American Economic Review*, 71 (5), 845-58.
- Samuelson, P. (1951), "Abstract of a Theorem Concerning Substitutability in an Open Leontief Model," in T. Koopmans (Ed.), *Activity Analysis of Production and Consumption*, New York: Wiley.
- Schonfeld, E. (2009), "SplashCast Throws in the Towel on User-Generated Content; Looking For A Buyer," *TechCrunch*, June 4, 2009.
- Simon, H. A. (1955), "On a Class of Skew Distribution Functions," *Biometrika*, 42 (3/4), 425-40.
- Simon, H. A., (1971), "Designing organizations for an information rich world," in M. Greenberger (Ed.), *Computers, Communications and the Public Interest*, Johns Hopkins Press, 38-52.
- Smith, A. (1776), *The Wealth of Nations*, Cannan Ed. (1976) Chicago: University of Chicago Press.
- Yule, G. U. (1925), "A mathematical theory of evolution based on the conclusions of Dr. J. C. Willis," *Philos. Trans. R. Soc. London*, B, 213, 21-87.
- Wilson, E. O. (1975), *Sociobiology: The New Synthesis*, Harvard University Press.

## APPENDIX A

### Proof of Proposition 1.1

Take derivative of  $f_2(\sigma)$  with respect to  $\sigma$ . We have

$$f_2' = \frac{\gamma \exp(\beta X_j + \gamma \sigma)}{(1 + \exp(\beta X_j + \gamma \sigma))^2}$$

Then  $f_2' \leq 0$ . So  $f_2$  and  $f_1$  can only have one intercept.

### Proof of Proposition 1.2:

a) And we observe that  $\lim_{s \rightarrow -\infty} \sigma(s | \gamma) = 0, \lim_{s \rightarrow \infty} \sigma(s | \gamma) = 1$ , for all  $\gamma < \gamma^*$ . Let the support for  $\sigma(s | \gamma)$  be  $(-\infty, \infty)$ .  $\sigma(s | \gamma)$  is non-decreasing in  $s$ , and continuous. So  $\sigma(s | \gamma)$  is a cumulative distribution function of  $s$ .

b) Rewrite equation (25) to  $\beta s + \gamma \sigma(s | \gamma) = \log \sigma(s | \gamma) - \log(1 - \sigma(s | \gamma))$ . Since  $s$  is continuous in  $\sigma(s | \gamma)$  on  $(0, 1)$ ,  $\sigma(s | \gamma)$  is continuous in  $s$  on  $(-\infty, \infty)$ . Take derivative with respect to  $s$  in the interior of support, we obtain

$$\frac{d\sigma(s | \gamma)}{ds} = \beta \left[ \frac{1}{\sigma(s | \gamma)(1 - \sigma(s | \gamma))} - \gamma \right]^{-1} \quad (69)$$

So  $d\sigma(s | \gamma) / ds > 0$  for  $\gamma < 0$ . Note that  $\inf(1 / [\sigma(s | \gamma)(1 - \sigma(s | \gamma))]) = 4$ . Now consider  $0 < \gamma < \gamma^*$ . So a sufficient condition for  $d\sigma(s | \gamma) / ds > 0$  to hold is  $\gamma < 4$ . However,  $\sigma(s | \gamma)$  is also a function of  $\gamma$ . So if  $\gamma > 4$ ,  $d\sigma(s | \gamma) / ds > 0$  may still hold. Note in our case,  $\gamma < 4$  holds because we observe that  $\gamma^* \leq 3.5$  from Figure 1.g

c) We fix the salary  $s$ . Then Eq. (25) defines an implicit function of  $\sigma(\gamma | s)$ . And  $\sigma(\gamma | s)$  is continuous in  $(-\infty, \gamma^*)$ . Take derivative with respect to  $\gamma$  in  $(-\infty, \gamma^*)$ , we obtain

$$\frac{d\sigma(\gamma|s)}{d\gamma} = \sigma \left[ \frac{1}{\sigma(\gamma|s)(1-\sigma(\gamma|s))} - \gamma \right]^{-1} \quad (70)$$

So  $\frac{d\sigma(\gamma|s)}{d\gamma} > 0$  for  $\gamma < \gamma^* < 4$ .

d) The statement immediately follows from b).

e) Proof by contradiction. If  $B_1$  does not exist, then  $d\sigma(s|\gamma_1)/ds \geq d\sigma(s|\gamma_2)/ds$  in the entire interior of support. It implies  $\sigma(s|\gamma_1) \geq \sigma(s|\gamma_2)$ , contradicting to the first order dominance statement in b). If  $B_2$  does not exist, then  $\lim_{s \rightarrow \infty} \sigma(s|\gamma_1) < \lim_{s \rightarrow \infty} \sigma(s|\gamma_2) = 1$ , contradicting to the fact that  $\lim_{s \rightarrow \infty} \sigma(s|\gamma_1) = 1$ .

f) The statement follows from equation (69).

## APPENDIX B

### Proof of Proposition 2.1

Check Eq. (28).

### Proof of Proposition 2.2

The equivalence between Definition 2 and 3 is evident. To prove the equivalence of Definition 3 and 4, use Equation (36) and (39). We find that  $Y_2(t)/Y_1(t) = Y_2(t-1)/Y_1(t-1)$  holds if and only if  $y_2(t)/y_1(t)$  is constant over time, and if and only if  $\frac{E_2(t)}{E_1(t)}$  is a constant over time by Eq. (38).

### Proof of Proposition 2.3

We plug the unit endorsement of Eq. (35) into Definition 4 of equilibrium and use the ratio of unit endorsement increase of Eq.(34). A solution exists. The content system is in equilibrium if and only if:

$$\frac{E_2(t)}{E_1(t)} = \frac{\Delta e_2(t)}{\Delta e_1(t)} = 1 + \frac{\alpha_2}{\alpha_1}. \quad (71)$$

### Proof of Proposition 2.4

By the second definition of content equilibrium,  $Y_2(t)/Y_1(t)$  is a constant over time. Therefore, period endorsements available  $e_1(t), e_2(t)$  are fixed by Eq. (30). Since period productions  $y_1(t)$  and  $y_2(t)$  are positive,  $Y_1(t)$  and  $Y_2(t)$  are increasing, and unit endorsement changes  $\Delta e_i \rightarrow 0$  for  $i = 1, 2$  as  $t \rightarrow \infty$  by Eq. (33). Therefore accumulated per unit endorsements  $E_1(t), E_2(t)$  will not change, and period productions  $y_1(t), y_2(t)$  become stable.

### Proof of Proposition 2.5

A Pareto efficient allocation is to maximize the utility of one consumer while holding the utility of the other consumer at a given level  $\bar{u}_2$ .

$$\begin{aligned} \max_{x_1(t), x_o^1(t)} \alpha_1 \ln x_1(t) + x_o^1(t) \\ \text{such that } \alpha_2 \ln x_2(t) + x_o^2(t) = \bar{u}_2. \end{aligned} \quad (72)$$

From the endorsement clearing allocation definition, we have

$$x_2(t) = \Delta e_2(t) Y_2(t-1) - \frac{Y_2(t-1)}{Y_1(t-1) + Y_2(t-1)} x_1(t). \quad (73)$$

Substituting  $x_2(t)$  to the objective function, we have the unconstrained maximization problem:

$$\max_{x_1} \alpha_1 \ln x_1(t) + \alpha_2 \ln x_2(t) + \bar{x}_o - \bar{u}_2, \quad (74)$$

which has the first order condition:

$$\frac{x_2(t)}{x_1(t)} = \frac{\alpha_2}{\alpha_1} \frac{Y_2(t-1)}{Y_1(t-1) + Y_2(t-1)}. \quad (75)$$

In the content equilibrium,

$$x_1 = \frac{\alpha_1}{\tau w}, \quad x_2 = \frac{\alpha_2}{\tau w} \frac{Y_2}{Y_1 + Y_2}. \quad (76)$$

So the necessary condition for Pareto optimality is satisfied. So the content equilibrium is optimal in terms of consumer welfare.

### Proof of Proposition 2.6

First, we show that a content system with consumers and producers has a self-organizing feature. This fact can be seen by noting that if the unit endorsement ratio is

greater than the equilibrium value, i.e.,  $\frac{E_2(t-1)}{E_1(t-1)} > 1 + \frac{\alpha_2}{\alpha_1}$ , then the system will reduce

the ratio in the next period, i.e.,  $\frac{E_2(t)}{E_1(t)} < \frac{E_2(t-1)}{E_1(t-1)}$ . If the unit endorsement ratio is

smaller than the equilibrium value, i.e.,  $\frac{E_2(t-1)}{E_1(t-1)} < 1 + \frac{\alpha_2}{\alpha_1}$ , then the system will increase

the ratio in the next period, i.e.,  $\frac{E_2(t)}{E_1(t)} > \frac{E_2(t-1)}{E_1(t-1)}$ . This self-adjusting process will

eventually lead the system to equilibrium regardless of the initial system state.

Formally, we define two sequences. The first sequence is

$$S \equiv \left\{ \frac{E_2(0)}{E_1(0)}, \dots, \frac{E_2(t)}{E_1(t)} \dots \right\} \quad \text{where} \quad \frac{E_2(0)}{E_1(0)} \geq 1 + \frac{\alpha_2}{\alpha_1}; \quad E_i(t) \equiv \Delta e_i(t) + E_i(t-1), \quad i = 1, 2;$$

$$\frac{\Delta e_2(t)}{\Delta e_1(t)} = 1 + \frac{\alpha_2}{\alpha_1}. \quad \text{The second sequence is defined as} \quad S' \equiv \left\{ \frac{E_2'(0)}{E_1'(0)}, \dots, \frac{E_2'(t)}{E_1'(t)} \dots \right\} \quad \text{where}$$

$$\frac{E_2'(0)}{E_1'(0)} \leq 1 + \frac{\alpha_2}{\alpha_1}; \quad E_i'(t) \equiv \Delta e_i'(t) + E_i'(t-1), \quad i = 1, 2; \quad \frac{\Delta e_2'(t)}{\Delta e_1'(t)} = 1 + \frac{\alpha_2}{\alpha_1}. \quad \text{Let us examine}$$

sequence  $S$ . We find  $S$  is monotone decreasing, i.e.,  $\frac{E_2(t)}{E_1(t)} \leq \frac{E_2(t-1)}{E_1(t-1)}$  for all  $t$ .  $S$  is

bounded, i.e.,  $\frac{E_2(t)}{E_1(t)} \geq 1 + \frac{\alpha_2}{\alpha_1}$ . Therefore  $S$  converges. The limit cannot be strictly

greater than  $1 + \frac{\alpha_2}{\alpha_1}$  because a smaller  $\frac{E_2(t)}{E_1(t)}$  can be found, so the limit of the sequence

must be  $1 + \frac{\alpha_2}{\alpha_1}$ . The same argument applies to sequence  $S'$ .

### Proof of Proposition 2.7

The inequality in unit endorsement holds true because  $\frac{E_2(t)}{E_1(t)} = 1 + \frac{\alpha_2}{\alpha_1}$ , even if

preference for low quality may be higher, i.e.,  $\alpha_1 > \alpha_2$ . The inequality in production holds because production per period is linear in unit endorsement. Therefore,  $\frac{Y_2(t)}{Y_1(t)} = 1 + \frac{\alpha_2}{\alpha_1}$  when  $t$  is large. And it is obvious that high quality content has a higher sampling probability, i.e.,  $m_2 > m_1$ .

### Proof of Proposition 2.8

If the high quality consumer has a higher rate of endorsement, i.e.,  $\beta_2 > \beta_1$ , then the high quality producer receives larger share of unit equilibrium when compared to the basic model, i.e.,  $\frac{E_2'(t)}{E_1'(t)} > 1 + \frac{\alpha_2}{\alpha_1}$ . Therefore, the share of higher quality is also higher when compared to the basic model, i.e.,  $\frac{Y_2'(t)}{Y_1'(t)} > 1 + \frac{\alpha_2}{\alpha_1}$ . The results are opposite if the low quality consumer has a higher rate of endorsement, i.e.,  $\beta_2 < \beta_1$ .

### Proof of Proposition 2.9

The unit endorsement increases are:

$$\begin{aligned}\Delta e_1(t) &= \frac{e_1(t)}{Y_1(t-1)} = \frac{\alpha_1}{w} \frac{E_1(t-1)}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1)}, \\ \Delta e_2(t) &= \frac{e_2(t)}{Y_2(t-1)} = \frac{(\alpha_1 + \alpha_2)}{w} \frac{E_2(t-1)}{E_1(t-1)Y_1(t-1) + E_2(t-1)Y_2(t-1)},\end{aligned}\tag{77}$$

while the ratio of unit endorsement increase is:

$$\frac{\Delta e_2(t)}{\Delta e_1(t)} = (1 + \frac{\alpha_2}{\alpha_1}) \frac{E_2(t-1)}{E_1(t-1)} > \frac{E_2(t-1)}{E_1(t-1)},\tag{78}$$

and combining Eq. (35) and Eq. we obtain:

$$\frac{E_2(t)}{E_1(t)} > \frac{E_2(t-1)}{E_1(t-1)}.\tag{79}$$

### Proof of Proposition 2.10

Note that the relative sampling probability is

$$\frac{m_2(t)}{m_1(t)} = \frac{Y_2(t-1) + \Delta Y}{Y_1(t-1) + \Delta Y} < \frac{Y_2(t-1)}{Y_1(t-1)} \quad (80)$$

where  $\frac{Y_2(t-1)}{Y_1(t-1)} = 1 + \frac{\alpha_2}{\alpha_1} > 1$  is the relative sampling probability in the equilibrium.

Total endorsements in the next period are

$$\begin{aligned} e_1(t) &= x_1 m_1 = \frac{\alpha_1}{\tau w} \frac{Y_1(t-1) + \Delta Y}{Y_1(t-1) + Y_2(t-1) + 2\Delta Y}, \\ e_2(t) &= x_1 m_2 + x_2 = \frac{(\alpha_1 + \alpha_2)}{\tau w} \frac{Y_2(t-1) + \Delta Y}{Y_1(t-1) + Y_2(t-1) + 2\Delta Y}. \end{aligned} \quad (81)$$

So the ratio of unit endorsement between low and high quality content for both existing and new producers is still

$$\frac{\Delta e_2}{\Delta e_1} = 1 + \frac{\alpha_2}{\alpha_1}. \quad (82)$$

By the self-organization mechanism, new producers will reach the same inequality specified in Proposition 7 given enough time.

### Proof of Proposition 2.11

In the case of search by endorsement without entry of new producers, relative sampling probability of high quality content with respect to low quality content in

period  $t$  is  $\frac{m_2(t)}{m_1(t)} = \frac{E_2(t-1)Y_2(t-1)}{E_1(t-1)Y_1(t-1)}$ . With entry of new producers, it becomes

$\frac{m_2'(t)}{m_1'(t)} = \frac{E_2(t-1)Y_2(t-1) + a\Delta Y}{E_1(t-1)Y_1(t-1) + a\Delta Y}$ . Therefore,  $\frac{m_2'(t)}{m_1'(t)} < \frac{m_2(t)}{m_1(t)}$ . The ratio of unit

endorsement increase between existing high quality content and low quality content is

$\frac{\Delta e_2}{\Delta e_1} \Big|_{\text{existing producers}} = (1 + \frac{\alpha_2}{\alpha_1}) \frac{E_2(t-1)}{E_1(t-1)}$ , while the ratio between new high quality content

and low quality content is  $\frac{\Delta e_2}{\Delta e_1} \Big|_{\text{new producers}} = 1 + \frac{\alpha_2}{\alpha_1}$ .

$\frac{\Delta e_2}{\Delta e_1} \Big|_{\text{existing producers}} > \frac{\Delta e_2}{\Delta e_1} \Big|_{\text{new producers}}$  since  $\frac{E_2(t-1)}{E_1(t-1)}$  by equilibrium condition.

### Proof of Proposition 2.12

Total endorsements are

$$\begin{aligned} e_1(t) &= x_1(t)m_1(t) = \frac{\alpha_1}{\tau w} m_1(t), \\ e_2(t) &= x_1(t)m_2(t) + x_2(t) = \frac{\alpha_1}{\tau w} m_2(t) + \frac{\alpha_2}{\tau w} m_2(t)^\theta. \end{aligned} \tag{83}$$

Then we have  $m^\theta < m$  since sampling probability  $m_i \in (0,1)$  and  $\theta > 1$ . Then

$$e_2(t) < \frac{\alpha_1 W}{\tau w} m_2(t) + \frac{\alpha_2 W}{\tau w} m_2(t).$$

So

$$\frac{\Delta e_2(t)}{\Delta e_1(t)} < 1 + \frac{\alpha_2}{\alpha_1}. \tag{84}$$

Hence  $1 + \alpha_2 / \alpha_1$  is an upper bound for the non-linear demand system.

Next we search for a lower bound. Note that  $m^{\theta-1}$  is increasing in  $m$  because  $\theta > 1$  and  $m > 0$ . We know that  $m_1 + m_2 = 1$ , and  $m_2 > m_1$  due to inequality in output, so it must be that  $m_2 > 1/2$ . This gives  $m_2^{\theta-1} > 2^{1-\theta}$ , so we obtain  $m_2^\theta > 2^{1-\theta} m_2$  for all  $m_2$ ,

$$e_2(t) > \frac{\alpha}{\tau w} m_2(t) + \frac{(1-\alpha)}{\tau w} 2^{1-\theta} m_2(t).$$

Then

$$\Delta e_1(t) = \frac{\alpha_1}{\tau w} \frac{1}{Y_1(t-1) + Y_2(t-2)}, \quad \Delta e_2(t) > \left( \frac{\alpha_1}{\tau w} + \frac{\alpha_2}{\tau w} 2^{1-\theta} \right) \frac{1}{Y_1(t-1) + Y_2(t-2)}.$$

Therefore

$$\frac{\Delta e_2(t)}{\Delta e_1(t)} > 1 + \frac{\alpha_2}{\alpha_1} \frac{1}{2^{\theta-1}}. \quad (85)$$

So  $1 + \frac{\alpha_2}{\alpha_1} \frac{1}{2^{\theta-1}}$  is a lower bound.

### Proof of Proposition 2.13

The system of equations implies that in the equilibrium  $\frac{E_2}{E_1} = \frac{\Delta e_2}{\Delta e_1}$  (we omit time  $t$  in

the equilibrium). Since  $\frac{E_2}{E_1} = \frac{y_2}{y_1} = \sqrt{\frac{l_2}{l_1}}$ , then we have

$$1 + \frac{\alpha_2 q_2}{\alpha_1 q_1} = \sqrt{\frac{T_2 - q_2}{T_1 - q_1}}. \quad (86)$$

which is  $(1 + \frac{\alpha_2 q_2}{\alpha_1 q_1})^2 = \frac{T_2 - q_2}{T_1 - q_1}$ . It implies  $q_2 < T_2$  when  $q_1 < T_1$ . Let

$A \equiv \frac{\alpha_2}{\alpha_1 q_1}$ ,  $B \equiv \frac{T_2}{T_1 - q_1}$ ,  $C \equiv \frac{-1}{T_1 - q_1}$ . Then we have  $(1 + A q_2)^2 = B + C q_2$ , which is

$A^2 q_2^2 + (2A - C) q_2 + (1 - B) = 0$ . Solving the quadratic equation, we obtain

$$q_2 = \frac{(C - 2A) \pm \sqrt{(2A - C)^2 - 4A^2(1 - B)}}{2A}. \quad (87)$$

It can be shown that, when the consumption time does not exceed the total time, i.e.,

$$q_1 < T_1,$$

$$(2A-C)^2 - 4A^2(1-B) = \frac{4\alpha_2^2}{\alpha_1^2 q_1^2} \frac{T_2}{T_1 - q_1} + \frac{4\alpha_2}{\alpha_1 q_1} \frac{1}{T_1 - q_1} + \left(\frac{1}{T_1 - q_1}\right)^2 > 0. \quad (88)$$

Therefore, a solution exists for  $q_2$ , i.e.,

$$q_2 = \frac{\sqrt{(2A-C)^2 - 4A^2(1-B)} - (2A-C)}{2A} \\ = \frac{\sqrt{\frac{4\alpha_2^2}{\alpha_1^2 q_1^2} \frac{T_2}{T_1 - q_1} + \frac{4\alpha_2}{\alpha_1 q_1} \frac{1}{T_1 - q_1} + \left(\frac{1}{T_1 - q_1}\right)^2} - \sqrt{\frac{4\alpha_2^2}{\alpha_1^2 q_1^2} + \frac{4\alpha_2}{\alpha_1 q_1} \frac{1}{T_1 - q_1} + \left(\frac{1}{T_1 - q_1}\right)^2}}{2\left(\frac{\alpha_2}{\alpha_1 q_1}\right)^2}.$$

A sufficient condition for a non-negative  $q_2$  is  $T_2 > T_1$ . The inequality of endorsement

and production holds because  $\frac{E_2}{E_1} = \frac{y_2}{y_1} = 1 + \frac{\alpha_2 q_2}{\alpha_1 q_1} > 1$ .