

©Copyright 2012

Lin Hao

**The New Era of Electronic Commerce: Mobile Commerce, Electronic Book Market and  
Novel Online Retail Strategies**

Lin Hao

A dissertation  
submitted in partial fulfillment of the  
requirements for the degree of

Doctor of Philosophy

University of Washington

2012

Reading Committee:

Yong Tan, Chair

Ming Fan

Yong-Pin Zhou

Program Authorized to Offer Degree:

Business Administration

University of Washington

**Abstract**

The New Era of Electronic Commerce: Mobile Commerce, Electronic Book Market and  
Novel Online Retail Strategies

Lin Hao

Chair of the Supervisory Committee:

Professor Yong Tan

Information Systems and Operations Management

This dissertation examines three subareas of modern electronic commerce—the mobile app market, the electronic book market, and the online retail market—each of which is presented in a separate essay. Essay 1 focuses on how consumer rating behavior, i.e., how consumers give and evaluate ratings, influences the mobile app developers’ pricing and quality decisions as well as the platform owner’s decision on revenue sharing policy. Essay 2 analyzes two popular pricing models in the electronic book market, the wholesale pricing model and the agency pricing model. This effort characterizes the dynamics of e-book and e-reader prices under both models’ equilibriums and then compares them. Essay 3 uses a data set from a popular online marketplace to study what factors affect the conversion rate dynamics of online retail. The results provide important guidance for online sellers who want to improve their conversion rates based on the specific statuses of their stores.

## TABLE OF CONTENTS

	Page
List of Figures .....	iii
List of Tables .....	iv
Acknowledgements.....	v
Dedication .....	vi
Preface .....	vii
Essay 1: Economic Value of Ratings in Mobile App Market .....	vii
Essay 2: Pricing Models in the Electronic Book Market.....	vii
Essay 3: Conversion Rate Dynamics in Online Retail .....	viii
Essay 1 Economic Value of Ratings in Mobile App Market .....	1
1.1 Introduction .....	1
1.2 Literature Review .....	4
1.3 The Model .....	7
1.4 Asymmetric Rating-Utility Conversion Rates.....	17
1.4.1 The Platform Owner and Developers' Optimal Choices .....	17
1.4.2 Analysis of Social Welfare.....	21
1.5 Unobserved Cost Rate .....	22
1.5.1 Unobserved $h$ .....	23
1.5.2 Self-selection.....	23
1.6 Conclusions .....	24
Essay 2 Pricing Models in the Electronic Book Market .....	26
2.1 Introduction .....	26
2.2 Literature Review .....	28
2.3 The Model .....	29
2.3.1 Model Setup .....	29
2.3.2 The Wholesale Model .....	31
2.3.3 The Agency Model.....	34
2.4 Comparison and Social Welfare Analysis .....	39
2.5 Discussion.....	44
2.6 Conclusions .....	45
Essay 3 Conversion Rate Dynamics in Online Retail.....	48
3.1 Introduction .....	48
3.2 Literature Review .....	48

3.3 Theory Development .....	49
3.4 Empirical Model .....	54
3.4.1 State-Dependent Conversion Rates .....	54
3.4.2 State-Transition Probabilities .....	56
3.4.3 Likelihood Function .....	57
3.5 Data and Variable Description .....	58
3.6 Estimation Results and Discussion .....	60
3.7 Conclusions .....	64
References .....	65
Appendices .....	69
Appendix A: Additional propositions and proof of Essay 1 .....	69
Proof of Proposition 1.1 .....	70
Proof of Proposition 1.2 .....	73
Proof of Corollary 1.3.1 and Corollary 1.3.2 .....	74
Proof of Corollary 1.3.4 .....	75
Proof of Corollary 1.3.5 .....	77
Proof of Corollary 1.4.1 and Corollary 1.4.2 .....	81
Proof of Corollary 1.4.3 .....	83
Proof of Proposition 1.3 .....	84
Proof of Proposition 1.4 .....	85
Appendix B: Proof of propositions in Essay 2 .....	88
Proof of Proposition 2.1 .....	88
Proof of Proposition 2.4 .....	92
Proof of Proposition 2.5 .....	93
Proof of Proposition 2.6 and Proposition 2.7 .....	94
Proof of Proposition 2.8 .....	94
Appendix C: Additional tables in Essay 3 .....	99
Vita .....	100

## LIST OF FIGURES

Figure Number	Page
1.1 Developers' optimal choice versus $h$ .....	11
1.2 Developers' optimal choice versus $s$ .....	11
1.3 Region distributions and developers' optimal choice across regions .....	13
1.4 Platform owner's optimal choice of $s$ .....	15
1.5 Region distribution versus $k_R$ ( $r_R \geq \hat{r}_R$ ) .....	20
1.6 Region distribution versus $k_R$ ( $r_R < \hat{r}_R$ ) .....	20
2.1 E-book retail price versus retailer's revenue sharing percentage .....	37
2.2 E-reader price versus e-book retail price .....	38
2.3 E-reader and e-book retail prices under two pricing models .....	40
2.4 Social welfare under two pricing models .....	42
2.5 Business profit under two pricing models .....	43
3.1 Single-regime model .....	51
3.2 Regime-switching model .....	51
3.3 Hidden Markov model of seller's conversion rate .....	54

## LIST OF TABLES

Table Number	Page
1.1 Model parameters and decision variables.....	9
1.2 Platform owner's (P.O.'s) interest versus social welfare (S.W.).....	22
2.1 Comparison between two pricing models .....	42
3.1 Variables and descriptive statistics.....	58
3.2 Estimation results .....	61
3.A1 Correlation table .....	99

## **ACKNOWLEDGEMENTS**

The author wishes to express his gratitude to Department of Information Systems and Operations Management for the long-term support to his doctoral study. He also wants to thank every member in his doctoral committee for always being supportive during the last two years. He especially thanks the committee chair, his academic advisor, Professor Yong Tan, without whom none of this would have been possible. The author is greatly grateful to Professor Ming Fan for his help through the years and really enjoys working with him. The author expresses sincere appreciation to Professor Arvind Tripathi for his always timely advice and support in critical moments. The author is also thankful to receive insightful comments and contributions from Professor Yong-Pin Zhou.



## **DEDICATION**

To my wife *Dawn* and my parents *Xingbi Lin* and *Yungui Hao*

## PREFACE

This dissertation consists of three essays which study issues in the mobile app market, the electronic book market, and online retail. In the following paragraphs, I briefly introduce the topics of each essay and provide an overview of the findings and contributions.

### **Essay 1: Economic Value of Ratings in Mobile App Market**

This essay investigates the influence of ratings on an emerging third-party software application market, mobile app market. Given the nature of software application as experience good and the scarcity of other quality signals such as advertising or branding for majority of app developers, consumers' *ex ante* belief on an app's utility relies on the app's rating which is derived from the *ex post* utility received by previous peer consumers. An analytical framework is developed to explicitly characterize this bidirectional rating-utility conversion process with a newly introduced concept "reservation rating." After integrating this conversion process into utility functions, we derive the market equilibrium and reveal how the changes in consumer rating behavior affect mobile app developers' pricing and quality decisions as well as the platform owner's decision on revenue sharing policy. We also investigate how consumer rating attitude affects the social welfare, suggesting consumers' responsibility to the overall goodness of app community in terms of their proper rating behavior. The rating-dependent utility function enables us to derive a self-selection mechanism to achieve a separating equilibrium in which high and low cost rate developers choose differentiated revenue sharing percentages.

### **Essay 2: Pricing Models in the Electronic Book Market**

We recently observed a trend of decreasing e-reader prices and increasing e-book retail prices. In this essay we develop a game theoretic model to study the underlying reasons for this price trend. We find one potential cause to be a change in pricing models, from the wholesale pricing model to the agency pricing model. In the wholesale pricing model, the retailer decides both the e-reader and the e-book's retail prices while in the agency pricing model the retailer decides only the e-reader price and lets the publisher decide

the e-book retail price. We find that the optimal pricing strategy for the retailer in the wholesale pricing model is to use a two-part tariff pricing structure, i.e., pricing e-books at the wholesale price set by the publisher and marking up e-readers for profit. Under the agency pricing model, however, the retailer's optimal strategy changes. The retailer prefers a lower e-reader price and makes a profit through the revenue sharing on e-book sales. We also show numerically that although the publisher gains the e-book retail price control in the agency pricing model, in equilibrium the publisher is potentially worse off in terms of overall profit.

### **Essay 3: Conversion Rate Dynamics in Online Retail**

In this study, we use a proprietary data set from an online marketplace to study the conversion rate dynamics in online retail. We examine how seller-level covariates, such as online sellers' pricing and product strategies, marketing efforts, service responsiveness, reputation scores, product quality ratings, and other attributes, affect conversion rates. Specifically, we address the following research questions: (i) How do sellers' covariates affect their conversion rates? (ii) Is the relationship between the conversion rate and sellers' covariates state dependent? (iii) If the relationship is state dependent, what are the factors that determine the states and the state transitions? A hidden Markov model is adopted in the effort to answer these questions. The estimation results indicate that there are two states that affect conversion rate dynamics. The relationship between the conversion rate and the sellers' covariates is state dependent, that is, given the different states of sellers, the effects of the sellers' covariates on their conversion rates are different. We also estimate the thresholds between states. The results provide important guidance for sellers regarding what attributes they need to improve, to increase their conversion rates.

## **Essay 1 Economic Value of Ratings in Mobile App Market**

### **1.1 Introduction**

The traditional third-party application market is now heading towards its next phase. It has been witnessed that a novel third-party application market, mobile app market, is currently experiencing its explosive growth despite the recent economic downturn. Apple App Store, the world's leading mobile app market launched in July 2008, reached total number of 10 billion downloads in January 2011 while it was just 3 billion in January 2010 (Spain 2011). Kaufman Brothers LP estimated that over \$1 billion revenue was generated from more than 350,000 applications inside the Apple App Store in 2010. Apple Inc. reported that app developers have been paid more than \$4 billion in total since July 2008 till March 2012 (Zeman 2012). It implies that the life-time gross revenue of Apple App Store is more than \$5.7 billion. Following Apple's move, Google, Microsoft, Research in Motion (RIM) and Amazon opened their own mobile app stores. Gartner Inc. reported that the total revenue of entire mobile app market hit \$5.2 billion in 2010. It will further increase to \$35 billion in 2014 according to International Data Corp. (IDC)'s projection. Meanwhile, the tremendous success in this novel market is initiating a new trend which may change the whole climate of future third-party application market. Enlightened by the success of Apple App Store, Apple extended the same business model to desktop and laptop applications. On January 6, 2011, Mac App Store was opened with more than 1,000 computer apps out of the gate. In just 20 days, Pixelmator, a small software company, achieved 1 million dollar sale from its one single image processing application sold at unit price \$29.99. Traditional giant software companies such as Autodesk also joined the Mac App Store. The market is so enticing that even Microsoft, Apple's major competitor, is considering bringing its Microsoft Office's Mac version onto Mac App Store. "It's something we are looking at," said Amanda Lefebvre, Microsoft's senior marketing manager (Bradley 2011).

Concerning this emerging economy, adverse selection caused by information asymmetry is a serious issue which could potentially dismantle the market (Akerlof 1970). Adverse selection exists in this market due to the following reasons. First, software products belong to experience goods of which consumers can hardly observe the true quality *ex ante* (Shapiro 1985). Second, even if the "objective" true quality is

observed, it could imply different utilities for consumers depending on their subjective valuation of quality. The discounted utility is not necessarily known to consumers *ex ante* (Chen and Xie 2008). Third, a characteristic of app market, that is, the mixture of individual and organizational app developers, generates greater belief dispersion on quality than in the situation where vendors are relatively more homogenous. Thus, without the signals to distinguish the quality of apps, consumers are almost clueless about their *ex post* utility so that they can hardly figure out their corresponding willingness to pay.

To counter adverse selection, rating, which results from consumers' *ex post* experience, is considered a good instrument helping consumers formulate the correct *ex ante* beliefs on *ex post* net utility (Li and Hitt 2010, Sun 2012). Though it is not the only instrument, it is much more influential than other ones in app market. First, app platform owner often aggregates all the ratings and make them available at one single site, for instance, iTunes for iPhone apps. Hence they are easy to be found. Second, consumers are allowed to rate only after they purchase an app. This helps, to a large extent, avoid shilling behavior so that maintain the credibility of ratings. Third, other signaling devices such as advertising and branding are usually weak due to individual developers' budget constraint or lack of marketing capability. Therefore, rating is the main criterion that consumers rely on in app market.

Besides consumers, developers and the platform owner also pay significant attentions on ratings. The evidence from real business practices shows that rating is one of the hottest topics and the biggest concerns among developers. Tens of thousands of posts on iPhone dev sdk.com, one of the most popular online iPhone developer communities, are related to how to improve ratings and how ratings affect app sales revenue. All these facts indicate that in app market ratings play a very significant role in determining both the success of an app and the prosperity of the market.

Given its great importance, this study focuses on the impact of rating on app market in terms of consumers' purchase decisions, developers' choices of app price and quality level, the platform owner's choice of revenue sharing policy and the social welfare. We first parameterize consumer rating behavior and integrate it into consumers' utility function. We then derive developers' optimal choice of app price and quality level based on the integrated rating-dependent utility function. Because of the revenue sharing

contract between the developers and the platform owner, the platform owner's decision on revenue sharing policy will also be associated with the parameters of consumer rating behavior.

We assume that consumer rating behavior consists of two processes. One is to give *ex post* ratings based on *ex post* user experience which is quantified by consumers' received net utility after purchase (Kuksov and Xie 2010). The other is to evaluate *ex post* ratings and translate them into *ex ante* perceived net utility. Combining these two processes we construct a bidirectional rating-utility framework in which we call the former utility-to-rating process and the latter rating-to-utility process.

We account for one essential aspect of rating behavior, the subjectivity. Subjectivity is also called "systematic rater error" as one type of rating errors discussed in the vast management literature regarding performance rating (Kane 1994, Kane et al. 1995, Yun et al. 2005, Borman 1977, Saal et al. 1980). In the utility-to-rating process, it is the consumer's subjective choice to rate 4 or 6 in a typical 10-point rating system when her received net utility is zero. Similarly, in the rating-to-utility process, which rating in the consumer's mind corresponds to her perceived *ex ante* zero net utility is also subjective. Having random rating errors removed, such subjectivity in both processes has been demonstrated to be systematic (Kane 1994) if no effort is made to control for its systematic sources. Obviously, this is the case for app market since consumers are under no control for their characteristics such as maturity, conscientiousness and degree of sophistication in using apps, all of which could serve as such systematic sources. To model the subjectivity in the utility-to-rating process, we assume a linear function between *ex post* rating and received net utility. We call this linear function the rating function. For the rating-to-utility process, we introduce a new concept: reservation rating. Reservation rating is defined as the *ex post* rating which signifies zero received net utility in consumers' *ex ante* perception. Based on reservation rating, a linear function between *ex post* rating and *ex ante* perceived net utility is established with potentially different slope and intercept than the rating function.

This work contributes to the existing literature in the following two aspects. The first is the bidirectional rating-utility framework we construct. While we take app market as our research context, this framework can be applied in any market where consumer rating plays an important role in purchase

decisions. The concept of reservation rating introduced for rating-to-utility process is an attempt to systematically model how consumers interpret ratings into net utility. Second, our findings provide the guidance for developers and the platform owner on how to optimally react to the changes in consumer rating behavior by adjusting their choices of quality, price, and revenue sharing percentage. We discover that rating leniency is detrimental to app quality. The analysis on social welfare demonstrates the potential misalignment between the interest of platform owner and the social welfare when consumers' rating behavior changes. It is further shown that the platform owner can discriminate developers through a self-selection mechanism. We prove that there exists a separating equilibrium in which high and low cost rate developers choose different revenue sharing percentages.

The rest of this essay is organized as follows. In Section 1.2 we present a review of the related literature. In Section 1.3 we propose the base model. Section 1.4 extends the base model by imposing asymmetric rating-utility conversion rates and analyzes the social welfare under asymmetric rating-utility conversion rates. In Section 1.5 we study the platform owner's optimal revenue sharing policy when developers' cost rates are unobservable. We derive a self-selection mechanism to achieve separating equilibrium between developers with high and low cost rates. We conclude this essay in Section 1.6. Some propositions and the proof of propositions are included in Appendix A.

## **1.2 Literature Review**

The literature on the effects of online word-of-mouth (WOM) has been proliferating rapidly during the last ten years. Most of them are empirical work in the context of movie and book industry, focusing on the effects of online WOM on predicting or influencing sales revenue (Dellarocas et al. 2004, Chevalier and Mayzlin 2006, Liu 2006). Dellarocas et al. (2004) demonstrates that online rating is a useful proxy for WOM in movie industry and it serves as one of the predictors for a movie's total revenue. Dellarocas and Narayan (2006) identify three metrics of online word-of-mouth: valence, variance and volume. Valence is usually denoted by the average numeric average rating. Variance is usually measured by its statistical variance or entropy (Godes and Mayzlin 2004). Volume is counted as the number of ratings. Liu (2006) shows that online WOM has significant explanatory power for box office revenue and the volume is a

stronger predictor than valence. Chevalier and Mayzlin (2006) show that improvement of online WOM valence increases book sales based on the data from Amazon.com and BarnesandNoble.com. Forman (2008) demonstrates that raters give more positive rating to the review with identity information and the disclosure of identity information increases the sales. Researchers have also investigated the connection between consumer ratings and sales in the context of other markets such as beer, DVD and video games (Clemons et al. 2006, Hu et al. 2008, Zhu et al. 2010). However, little attention has been paid to software market. One reason might be that online software selling has not been widely established until app market appears. Zhou and Duan (2010) find that, from CNET download.com, the increase in product variety strengthens the impact of positive consumer reviews but weakens the impact of negative ones. They also show that positive expert reviews lead to more software downloads.

Given the association between online WOM and product sales, it is the natural next step for researchers to conceive firms' optimal strategy to leverage such association. A growing body of literature has been devoted to this topic. Chen and Xie (2005) show that a firm should choose advertising instead of price adjustment to improve consumer review when sufficient consumers value the product's horizontal features. Dellarocas (2006) demonstrates how firms' shilling behavior, i.e., post anonymous messages which exalt their products on the purpose of influencing the subsequent consumers' perception, will influence firms' profits and consumers' surplus. Chen and Xie (2008) reveal when and how sellers should adjust their marketing communication strategy to improve consumer reviews. Kuksov and Xie (2010) study whether the firm should give an unexpected frill to early customers to boost their product experience. Li and Hitt (2010) show, analytically and empirically, that unidimensional ratings are more associated with the net value, rather than quality, of the product. Firms need to account for price effects and can better serve the consumers by setting up review systems which explicitly separates the net value and the quality. Jiang and Chen (2007) examine the economic effect of both consumer review and consumer ratings and find that firms may have the incentive to under-charge in the early period. They also investigate how the rating and consumer review will affect the market competition.



Most of the firms' strategies mentioned above are implicitly based on an underlying assumption that ratings can signal the underlying true quality (or true value) to novice consumers at least in an expectation sense. However, this assumption does not always hold. The topic concerning how precise the ratings reflect the underlying true quality, which is an issue involved with consumer's rating behavior, becomes increasingly popular in the rating-related research area. Hu et al. (2006) empirically show that in the presence of under-reporting, i.e., only extremely satisfied or extremely unsatisfied consumers rate, consumers might not extract the true quality from the ratings if only valence (mean value) is known. Sun (2012) analytically demonstrates that novice consumers can figure out product's true quality from the distribution of the ratings given by earlier consumers. Li and Hitt (2008) show that later consumers' attempts to recover the true quality information from the ratings may fail because of the biased ratings left by earlier buyers whose preferences on quality are different from later ones'. Hu et al. (2009) summarizes two types of self-selection biases, purchasing bias mentioned in Li and Hitt (2008) and under-reporting bias mentioned in Hu et al. (2006), as two reasons which potentially lead to subsequent consumers' biased perception on true quality. Moe and Trusov (2010) and Moe and Schweidel (2011) empirically show that consumers' rating behavior is significantly affected by previously posted ratings. Lee et al. (2009) demonstrates that social imitation and learning affect can influence user rating generation.

Another important aspect of rating behavior which influences appraisal accuracy is the systematic rater error. While having not been documented in the context of online WOM, it has been studied in the management literature for decades. Kane (1994) summarizes rating errors into multiple categories and points out that leniency, severity and non-differentiation are the three major ones which could be potentially systematic. Kane et al. (1995) demonstrate that rating leniency is the most troublesome rating error and find that it has a relatively stable response tendency. Spence and Keeping (2010) suggest that when managers give performance ratings to their employees, more experienced managers are associated with lower ratings. Berger et al. (2010) empirically show that under the situation where employees' bonus payments are associated with ratings, putting a forced differentiated distribution requirement on ratings actually leads to higher productivity. The systematic rater error especially applies to our study since by

analogy consumers in our context are more or less like managers who rate the developers' "performance," developers' apps.

### 1.3 The Model

Suppose that the platform owner, app developers and app consumers are the three players in the app market. A three-stage dynamic game theoretic model is constructed to study the equilibrium of the market. In the first stage, the platform owner determines and publicizes developers' revenue sharing percentage. By observing this percentage at the beginning of the second stage, developers choose either not to adopt and then exit the market, or to adopt and then determine the optimal quality level and optimal app price. If developers choose not to adopt, the game ends. Otherwise, in the third stage, consumers decide whether to purchase. If a purchase occurs, consumers rate the app based on the received net utility. The third stage is repeated until after a sufficient time period the rating becomes steady. The goals of the platform owner and developers are to maximize their own benefits in this steady-state. Developers would choose to participate when the profit is greater than or equal to zero. Likewise, consumer would make the purchase when the expected net utility is greater than or equal to zero. In the following, we start with several essential preliminaries which serve as the foundation of our model.

**Rating Function.** We assume a linear function between the received net utility and the *ex post* rating. The rating  $r$  is between 0 and 1 after normalization. We propose

$$r = \min \{1, \max \{0, ku + r_0\}\}.$$

In this expression,  $u$  is the received net utility.  $r_0$  is the *ex post* zero net utility rating which is an important concept in terms of measuring the degree of severity. A low  $r_0$  indicates that consumers are severe in giving ratings.  $k$  is the rating-utility conversion rate. It represents sensitivity of the *ex post* rating on the difference in received net utility.

**Received Net Utility.** Following Chen and Xie (2008), we partition consumers into two groups. One consists of all high valuation consumers who consider the app match their taste so that they appreciate the quality. The other consists of all low valuation consumers who find the app a mismatch of their taste. The

received net utility, which is associated with the app's "objective" true quality level  $q$ , its price  $p$  and consumers' valuation for quality, is

$$u_{\text{matched}} = q - p, u_{\text{unmatched}} = -p.$$

We denote the fraction of consumers who belong to high valuation group by  $b$ . It is worth noting that  $b$  is an indicator of developers' marketing performance. A large  $b$  represents a high level of marketing performance since most of consumers are "matched." We assume that  $b \in (0,1)$ .

**Reservation Rating.** Following the definition of "reservation rating" in the introduction section, we further explain how it fits into economic utility theory. As we mentioned, reservation rating is intuitively a bar which an app needs to pass to be considered for purchase. It also fundamentally affects consumers' willingness to pay as follows. It works as a "ruler origin" to measure the *ex ante* perceived net utility of an app. For instance, with regard to a 5-star rating system, when a consumer with reservation rating 3.5 stars observes an app with a rating of 4.5 stars, she would expect some positive net utility from the app. In other words, combining reservation rating and *ex post* rating, consumers will figure out their *ex ante* perceived net utility. In the base model, reservation rating  $r_R$  is assumed to be homogenous among all consumers. Suppose  $r$  is the *ex post* rating given by high valuation consumers. The *ex ante* net utility perceived by high valuation consumers is

$$u_e = \frac{r - r_R}{k}.$$

In order to distinguish the degree of criticism between utility-to-rating and rating-to-utility processes, we define that consumers with low  $r_0$  are "severe" and consumers with high  $r_R$  are "critical." The values of  $r_R$  and  $r_0$  are between 0 and 1. Since consumers are generally more critical in evaluating ratings than giving ratings, we assume that  $r_R > r_0$ .

Now, we derive consumers' expected net utility  $E(U)$ . We assume that both true quality level and type of valuation are unknown to consumers before they experience the app. However, by examining the

distribution of app's *ex post* rating they would discover that the probability of being in the high valuation group is  $b$  and that in the low valuation group is  $1-b$ . Hence, the expected net utility is

$$E(U) = (1-b) \cdot (-p) + b \cdot u_e.$$

If  $E(U) \geq 0$ , consumers would make the purchase.

The developers' profit is

$$u_D = ps - hq^2,$$

where  $h$  is the cost rate on quality and  $h > 0$ . The quadratic form of cost represents the diminishing return of investment on quality (Choudhary 2007). The condition for developers to participate is  $u_D \geq 0$ .

Developers maximize their profits by choosing the optimal  $p$  and  $q$  under the constraint  $E(U) \geq 0$ .

Table 1.1 Model parameters and decision variables

<b>Parameters</b>	
$k$	Rating-utility conversion rate
$r_0$	Zero net utility rating
$r_R$	Reservation rating
$r$	<i>ex post</i> rating from high valuation group consumers
$b$	Fraction of consumers in the high valuation group
$h$	Developers' cost rate on quality
<b>Decision Variables</b>	
$q$	Quality level of the app
$p$	Price of the app
$s$	Developers' revenue sharing percentage

The platform owner's revenue is

$$u_P = p(1-s),$$

where  $s$  is bounded between 0 and 1. The platform owner's goal is to find the optimal  $s^*$  which maximizes  $u_P$ . In our model the platform owner's cost is neglected. Regarding the variable costs such as app hosting cost, we assume that they are already covered by a fixed annual fee paid by developers. For example, Apple iOS platform charges \$99 for annual membership to allow a developer upload her apps

on the Apple App Store. We speculate that this membership fee may cover the cost, but not be the major revenue source. The evidence is that nowadays roughly 350,000 apps are available on the Apple App Store. Even under the most extreme case that they correspond to 350,000 developers, the annual revenue from membership is 35 million at most, which account for less than 4% of the total platform owner's app sales revenue. Once a developer decides to participate, the membership fee is a sunk cost which does not affect her choice of app quality level and price.

Table 1.1 summarizes the notations of the model. Based on the above model setup, we have the following proposition.<sup>1</sup>

**Proposition 1.1** *Given the revenue sharing percentage  $s$ , developers' optimal choice of price  $p^*$  and quality level  $q^*$  are:*

i. *Region 1-1 (self-driven): where  $b > b_1$  and  $0 < h \leq h_1 s$ ,*

$$p_1^* = \frac{b(1-r_R)}{k(1-b)}, q_1^* = \frac{br_0 - r_R b - r_0 + 1}{k(1-b)};$$

ii. *Region 1-2 (platform owner-driven): where  $b > b_1$  and  $h_1 s < h \leq h_2 s$ ,*

$$p_2^* = \frac{1}{2} \frac{sb^2}{h} + \frac{b(-r_R + r_0)}{k}, q_2^* = \frac{1}{2} \frac{sb}{h};$$

iii. *Region 1-3 (poor marketing): where  $b \leq b_1$  and  $0 < h \leq h_3 s$ ,*

$$p_3^* = \frac{b(1-r_R)}{k(1-b)}, q_3^* = \frac{br_0 - r_R b - r_0 + 1}{k(1-b)}.$$

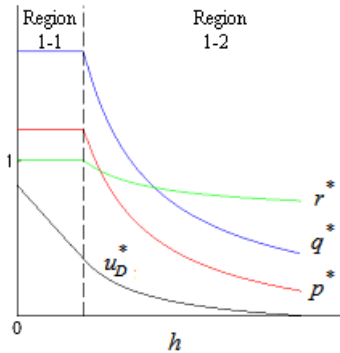
*In all other regions of  $(b, h)$ , developers cannot make non-negative profit. The above thresholds are:*

$$b_1 = 1 - \frac{1-r_R}{r_R-r_0}, h_1 = \frac{1}{2} \frac{bk(1-b)}{br_0 - r_R b - r_0 + 1}, h_2 = \frac{1}{4} \frac{bk}{r_R - r_0}, \text{ and } h_3 = \frac{kb(1-b)(1-r_R)}{(br_0 - r_R b - r_0 + 1)^2}.$$

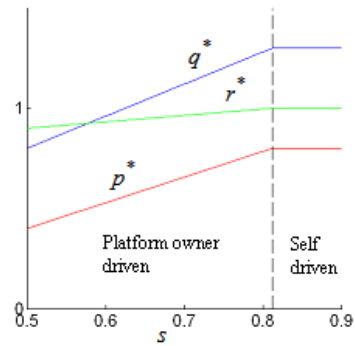
Proposition 1.1 shows that developers' optimal choice of price and quality level is divided into three regions, depending upon their marketing performance  $b$  and cost rate  $h$ . When developers increase the quality level  $q$ , higher quality level will lead to higher *ex post* ratings from high valuation consumers.

<sup>1</sup> Proposition 1.1 can be shown by solving the corresponding Lagrangian in different regions defined by the model parameters. More detailed proof of propositions and corollaries can be found in Appendix A.

Therefore the overall *ex post* ratings get elevated, which increases consumers' *ex ante* perceived net utility as well as their willingness to pay. Thus, developers can gain by charging a higher app price  $p$  to exploit the additional willingness to pay. For developers in Region 1-1, because of low cost rate ( $0 < h \leq h_1s$ ), the marginal gain is always greater than marginal cost of quality till the rating  $r$  reaches its maximum. Therefore, the optimal quality level  $q_1^*$  is the one at which high valuation consumers will give the maximum rating (note that  $r = k(q_1^* - p_1^*) + r_0 = 1$ ). In Region 1-2, when marketing performance is good but the cost rate is high ( $h_2s \geq h > h_1s$ ), the marginal cost of quality increases faster than that in Region 1-1 and will surpass the marginal gain from additional consumers' willingness to pay before the maximum  $r=1$  is realized. Thus, the optimal  $r$  in Region 1-2 will be between  $r_R$  and 1. In Region 1-3, optimal price  $p_3^*$  and quality level  $q_3^*$  are of the same analytical expressions as  $p_1^*$  and  $q_1^*$  in Region 1-1. This indicates when their marketing performance is poor but the cost rate is relatively low ( $0 < h \leq h_3s$ ), developers' optimal choice is to produce at the quality level which yields  $r=1$ . The intuition is that since only a small portion of consumers belong to high valuation group who appreciate the app quality, developers need to provide sufficient satisfaction to them in order to make them rate as high as possible. Otherwise the subsequent consumers' *ex ante* perceived net utility as well as their willingness to pay would be too low because the overall rating is too low. Figure 1.1 depicts developers'  $p^*, q^*, r^*, u_D^*$  in Regions 1-1 and 1-2. All of them are non-increasing function of developers' cost rate  $h$ .



**Figure 1.1** Developers' optimal choice versus  $h$



**Figure 1.2** Developers' optimal choice versus  $s$

**Corollary 1.1.1**  $\partial p_2^* / \partial s > 0$  and  $\partial q_2^* / \partial s > 0$ .

Corollary 1.1.1 shows that for developers in Region 1-2, their optimal choices of quality level and price are driven by the revenue percentage they obtain. By increasing  $s$ , the platform owner can provide more incentive for developers to improve the quality of the app. We hence name Region 1-2 as “platform owner-driven,” and refer the developers in Region 1-2 as “platform owner-driven developers.” In Region 1-1, the optimal quality level is not affected by  $s$ , and therefore the developers are “self-driven.” We further name Region 1-3 as “poor marketing region.” Since the optimality in the poor marketing region shares many similarities with the self-driven region plus it is unlikely to be the main component of the market (for example, assuming  $r_R = 0.7$  and  $r_0 = 0.5$ , we have  $b_1 = -0.5 < 0$ ), we focus our attention on the self-driven and platform owner-driven regions.

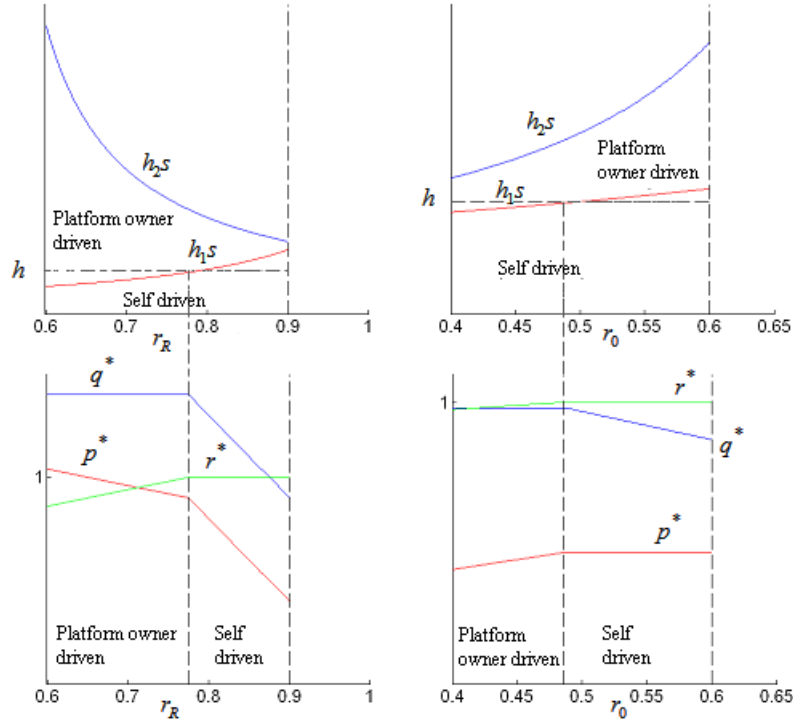
Notice that the boundaries of the regions ( $h_1 s, h_2 s$ ) are proportional to revenue sharing percentage  $s$  set by the platform owner at the first stage of the game. Therefore, given developers’ cost rate  $h$ , whether they choose to be in the self-driven or platform owner region, depends on not only on consumers’ rating behavior ( $k, r_0, r_R$ ) but also the platform owner’s choice of  $s$ . Figure 1.2 illustrates that as  $s$  increases, platform owner-driven developers are incentivized to increase their  $q^*$  and  $p^*$  until the rating  $r$  hits the maximum. By then they become self-driven developers.

**Corollary 1.1.2**  $\partial q_1^* / \partial r_R < 0$ ;  $\partial q_2^* / \partial r_R = 0$ . In all regions,  $\partial p^* / \partial r_R < 0$ .

Corollary 1.1.2 indicates that when consumers become more critical in evaluating ratings, developers need to charge a lower price. Platform owner-driven developers keep the same quality level while self-driven developers choose a lower quality level. This is because in the self-driven region, as long as the maximum rating is retained, further quality improvement will not promote consumers’ perception on *ex ante* net utility. Since the price drops, the quality level required to achieve the maximum rating can be set lower.

**Corollary 1.1.3**  $\partial q_1^* / \partial r_0 < 0$ ;  $\partial p_1^* / \partial r_0 = 0$ .  $\partial p_2^* / \partial r_0 > 0$ ;  $\partial q_2^* / \partial r_0 = 0$ .

When consumers become more severe in giving ratings, which means a lower  $r_0$ , self-driven developers optimally respond by setting a higher quality level but keep the price unchanged. This is because when  $r_0$  decreases, the rating  $r$  will no longer be at the maximum. Thus, self-driven developers regain the incentive to choose a higher quality to push it back to the maximum since consumers would be able to observe and appreciate such improvement. On the other hand, for platform owner-driven developers, the best response is to decrease the price.



**Figure 1.3** Region distributions and developers' optimal choice across regions

Notice that the boundaries of the regions  $(h_1s, h_2s)$  are also affected by  $r_0$  and  $r_R$ . Figure 1.3 shows that when  $r_R$  increases, the platform owner-driven region shrinks but the self-driven region expands. However as  $r_0$  increases, both regions expand. Platform owner-driven developers may switch to be self-driven as  $r_R$  or  $r_0$  increases. Figure 1.3 also depicts how developers' choice of optimal quality and price changes across different regions.

**Proposition 1.2** *The platform owner's optimal choice of developers' revenue sharing percentage  $s^*$  is:*



i. *Region 2-1 (squeezing): when  $b > b_1$  and  $0 < h \leq h_{1\alpha}$ ,*

$$s_1^* = \frac{2h(1 + br_0 - br_R - r_0)}{bk(1 - b)};$$

ii. *Region 2-2- $\alpha$  (encouragement): when  $b > b_1$  and  $h_{1\alpha} < h \leq 2h_2/3$ ,*

$$s_{2\alpha}^* = \frac{1}{2} + \frac{h(r_R - r_0)}{bk};$$

iii. *Region 2-2- $\beta$  (retention): when  $b > b_1$  and  $2h_2/3 < h \leq h_2$ ,*

$$s_{2\beta}^* = \frac{4h(r_R - r_0)}{bk};$$

iv. *Region 2-3: when  $b \leq b_1$  and  $0 < h \leq h_3$ ,*

$$s_3^* = \frac{h(1 + br_0 - br_R - r_0)^2}{b(1 - r_R)k(1 - b)}.$$

*The threshold is:*

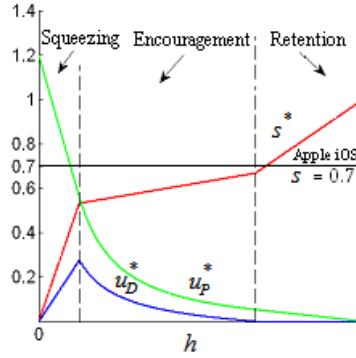
$$h_{1\alpha} = \frac{1}{2} \frac{bk(1 - b)}{2 + br_0 - br_R - r_0 - r_R}.$$

Proposition 1.2 characterizes the subgame perfect Nash equilibrium (SPNE). It shows how the platform owner in the first stage should offer different revenue sharing percentage  $s$  based on developers' cost rate  $h$  and marketing performance  $b$ . The SPNE endogenizes developers' best response in the second stage. For example, if the platform owner knows that the developers in Region 2-1 will maximize  $u_p$  by choosing the self-driven region in the second stage, the platform owner will offer  $s_1^*$  in the first stage to incentivize them to do so. In this sense, Region 2-1 corresponds to the self-driven region in the second stage. Both Regions 2-2- $\alpha$  and 2-2- $\beta$  correspond to the platform owner-driven region in the second stage.

**Corollary 1.2.1**  $\partial s^* / \partial h > 0$  where  $s^* = s_1^*, s_{2\alpha}^*, s_{2\beta}^*$ , or  $s_3^*$ .

Figure 1.4 depicts the optimal revenue sharing percentage  $s$  versus cost rate  $h$ . Figure 1.1 shows that when platform owner-driven developers' cost rate increases, the optimal quality level  $q_2^*$  decreases. Corollary 1.2.1 suggests that in this situation the platform owner should offer a higher revenue sharing

percentage to encourage developers to produce at a higher quality level. The loss due to a lower sharing percentage for the platform owner will be more than compensated by the additional revenue gained from a higher quality level.



**Figure 1.4** Platform owner's optimal choice of  $s$

Region 2-1 corresponds to the self-driven region where the optimal quality  $q_1^*$  is not a function of revenue sharing percentage. In this region, the platform owner can squeeze developers by giving a lower sharing percentage. Due to their low cost rate, developers can still obtain non-negative utility by choosing the quality level  $q_1^*$  at which the rating  $r=1$  and charging the price  $p_1^*$ . Squeezing developers may be dangerous since developers may leave. One reason why low cost rate developers do not leave this platform even when receiving low revenue sharing percentage is we assume that the platform owner is a monopoly or has the monopoly power on the market. For example, Apple still grasps roughly 70% of the entire mobile app market and this number is 99.4% in 2009. Developers are allured by the large consumer base. Another reason is that consumers using different platforms are usually separated, for example, there is little chance that one consumer uses both iPhone and Andriod phone. Since cannibalization is a mild issue, developers are willing to release their app on another platform as long as profitable.

Based on the above discussions, we give more intuitive names to the regions in Proposition 1.2. Region 2-1 is hereafter referred to as “squeezing region.” Developers in the squeezing region will be incentivized to choose the self-driven region in the second stage. Region 2-2- $\alpha$  is named as “encouragement region.” Region 2-2- $\beta$  is referred to as “retention region” where developers make zero profit and are hence on the verge of exiting the market. It can be observed in Figure 1.4 that developers

whose cost rate is at the intersection of the squeezing and encouragement regions can obtain the highest profit. Lower cost rate developers' profit will be squeezed. Higher cost rate developers' won't be able to reach that highest point. The extreme case is when the cost rate is close to zero the platform owner can take almost all the revenue but leave very little share to developers. Developers' participation can still be justified because developing app costs nearly nothing for them.

**Corollary 1.2.2**  $s_2^* > 1/2$  where  $s_2^* = s_{2\alpha}^*$  or  $s_{2\beta}^*$ .

Corollary 1.2.2 coincides with our observation from real business practice, for instance, Apple iOS platform offers a revenue sharing percentage of 70%. In app market, given the fact that a large portion of developers are individuals or development teams made up of several individuals whose cost rates are high, the platform owner may consider them more likely to be in the encouragement or retention, rather than the squeezing region. Therefore, it is optimal for the platform owner to offer a revenue sharing percentage greater than one half.

**Corollary 1.2.3**  $\partial s_1^* / \partial r_R < 0$ ,  $\partial s_2^* / \partial r_R > 0$  where  $s_2^* = s_{2\alpha}^*$  or  $s_{2\beta}^*$ .

According to Proposition 1.1, when consumers become more critical in evaluating ratings, developers in the encouragement and retention regions would decrease the price. Such price drop reduces the total revenue. A raise in sharing percentage would encourage developers to choose a higher quality level and increase the price, which eventually benefit the platform owner. However, the strategy in the squeezing region appears to be a little counterintuitive. It states that in the squeezing region when consumers become more critical in evaluating ratings, the platform owner prefers squeezing developers more, rather than incentivizing them by offering higher revenue sharing. The underlying logic is the following. Notice that the threshold  $h_{1\alpha}$  increases with reservation rating  $r_R$ . If the developers are in the squeezing region, they would still choose the self-driven region at the second stage when  $r_R$  increases. A higher  $s_1^*$  would not promote the choice of quality level but only subsidizing more revenue sharing to developers. Instead, if the platform owner decreases  $s_1^*$ , as long as the decreased  $s_1^*$  still maintains

developers' choice of price and quality level, i.e., staying in the self-driven region, the total app sales revenue won't change but the platform owner simply cut more share.

**Corollary 1.2.4**  $\partial s_1^* / \partial r_0 < 0$ ,  $\partial s_2^* / \partial r_0 < 0$  where  $s_2^* = s_{2\alpha}^*$  or  $s_{2\beta}^*$ .

When  $r_0$  increases, i.e., consumers become more lenient in giving ratings, developers in the encouragement and retention regions would charge a higher price but maintain the same level of quality. Higher price indicates higher revenue. In this case the platform owner can extract larger portion of the revenue by decreasing the revenue sharing percentage. For developers in the squeezing region, the optimal price does not change when consumers are more lenient in giving ratings. Therefore, the explanation why  $s_1^*$  is decreasing in  $r_0$  is similar to the explanation that why  $r_R$  is decreasing in  $s_1^*$  in the squeezing region in Corollary 1.2.3.

**Corollary 1.2.5** *For the squeezing, encouragement, and retention regions, in SPNE developers' profit satisfies  $\partial u_D^* / \partial r_0 \geq 0$  and  $\partial u_D^* / \partial r_R \leq 0$ . For the platform owner's revenue,  $\partial u_P^* / \partial r_0 > 0$  and  $\partial u_P^* / \partial r_R < 0$ .*

## 1.4 Asymmetric Rating-Utility Conversion Rates

In this section, we extend the base model by setting different conversion rates between rating-to-utility and utility-to-rating processes. We denote the conversion rate in the rating-to-utility process by  $k_R$ , and that in the utility-to-rating process by  $k_U$ . We assume that  $k_R > k_U$ . This assumption suggests that consumers are more sensitive to the change in received net utility when they give ratings than the change in *ex post* ratings when they translate ratings to their perceived net utility as well as the corresponding willingness to pay.

### 1.4.1 The Platform Owner and Developers' Optimal Choices

The problem can be solved in the similar way to the symmetric (same- $k$ ) case. The solution is presented in Proposition 1.A1 in Appendix A. Similarly, we identify three regions for developers: self-driven, platform owner-driven, and poor marketing. Their corresponding optimal price and quality are denoted by

$(p_1^{A*}, q_1^{A*})$ ,  $(p_2^{A*}, q_2^{A*})$ , and  $(p_3^{A*}, q_3^{A*})$ ,<sup>2</sup> respectively. Proposition 1.A2 in Appendix A describes the platform owner's optimal choice of revenue sharing percentage. Similar to those in Proposition 1.2, we have, respectively,  $s_1^{A*}$  for the squeezing region,  $s_{2\alpha}^{A*}$  for the encouragement region,  $s_{2\beta}^{A*}$  for the retention region, and  $s_3^{A*}$ . The following corollaries characterize how asymmetric conversion rates  $k_U$  and  $k_R$  affect the optimal quality  $q^{A*}$ , optimal price  $p^{A*}$ , and revenue sharing  $s^{A*}$  in different regions.

**Corollary 1.3.1** *In all regions,  $\partial q^{A*} / \partial k_R < 0$  and  $\partial p^{A*} / \partial k_R < 0$ .*

**Corollary 1.3.2** *In the self-driven region,  $\partial q_1^{A*} / \partial k_U < 0$  and  $\partial p_1^{A*} / \partial k_U = 0$ . In the platform owner-driven region,  $\partial q_2^{A*} / \partial k_U > 0$  and  $\partial p_2^{A*} / \partial k_U > 0$ .*

Corollary 1.3.2 suggests that when  $k_R$  decreases, i.e., consumers are willing to pay more for a higher rating, developers in any region have the incentive to offer a higher quality level and charge a higher price. Corollary 1.3.2 suggests that when  $k_U$  increases, i.e., consumers are more sensitive to the change in received net utility and willing to give more differentiated ratings, developers in the platform owner-driven region have the incentive to choose a higher quality level as well as a higher price. The reason is that when consumers appreciate high quality apps by giving more differentiated ratings, the benefit of producing high quality is augmented. On the contrary, if  $k_U$  decreases, i.e., consumers don't appreciate high utility apps and give alike ratings for good and bad apps, platform owner-driven developers have less incentive to produce at a higher quality level.

Developers in the self-driven region, however, acts differently when  $k_U$  increases. They will keep the price unchanged but choose a lower quality level. This is because whenever the rating  $r$  reaches its maximum, any further improvement on quality will no longer be reflected by the rating. Consumers won't recognize such effort in the rating-to-utility process. Therefore self-driven developers cannot take advantage of increased  $k_U$  in the same way as platform owner-driven developers. Nonetheless, they can

---

<sup>2</sup> Here a superscript  $A$  is added to denote the asymmetric case.

benefit from higher  $k_U$  through the reduction of the cost on quality since it would be less costly for them to achieve the quality level which realizes the maximum rating.

It should be noted that self-driven developers' quality level  $q_1^{A*}$  is decreasing in  $r_0$  and platform owner-driven developers'  $q_2^{A*}$  is increasing in  $k_U$ . This observation supports the finding in management literature that over-leniency is a significant problem. Berger et al. (2010) discover empirically in corporate environment a forced distribution on performance ratings will lead to higher productivity. We show that when  $r_0$  decreases (less lenient) and  $k_U$  increases (more distributed), the productivity indicator  $q$  increases.

**Corollary 1.3.3** *In the encouragement or retention region,  $\partial s_2^{A*} / \partial k_R > 0$  and  $\partial s_2^{A*} / \partial k_U < 0$ , where  $s_2^{A*} = s_{2\alpha}^{A*}$  or  $s_{2\beta}^{A*}$ .*

When consumers are unwilling to pay differentiated price for high rating app (i.e. decreasing  $k_R$ ), or consumers are unwilling to give differentiated ratings to high utility app (i.e. increasing  $k_U$ ), according to Corollaries 1.3.1 and 1.3.2, developers in the encouragement or retention region choose a lower quality level. Therefore the platform owner should offer higher revenue sharing to developers so as to maintain the quality level and hence the price for the platform owner's overall profit.

**Corollary 1.3.4** *In the squeezing region,  $\partial s_1^{A*} / \partial k_U < 0$ . With respect to  $k_R$ ,*

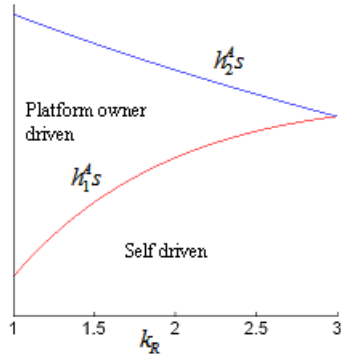
- i. *Case 1: if  $r_R \geq \hat{r}_R = (3 + r_0) / 4$ ,  $\partial s_1^{A*} / \partial k_R < 0$ ;*
- ii. *Case 2: if  $r_R < \hat{r}_R$  and  $k_R > \hat{k}_R = \sqrt{\frac{1-r_R}{1-r_0}} \frac{bk_U}{(1-b)}$ ,  $\partial s_1^{A*} / \partial k_R > 0$ ;*
- iii. *Case 3: if  $r_R < \hat{r}_R$  and  $k_U < k_R \leq \hat{k}_R$ ,  $\partial s_1^{A*} / \partial k_R < 0$ .*

Corollary 1.3.4 shows that when  $r_R$  is low and  $k_R$  is high (Case 2), the platform owner would increase  $s_1^{A*}$  when  $k_R$  increases. As discussed earlier, low cost rate developers target the maximum rating because the revenue benefit from charging a higher price outweighs the required additional cost spent on higher quality. When  $k_R$  increases, such benefit diminishes significantly but the cost on quality remains

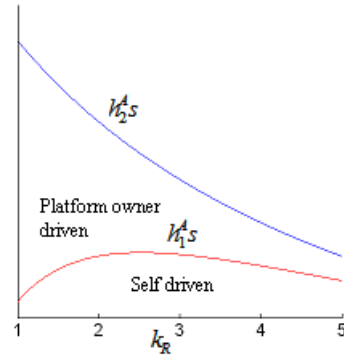
the same because  $k_U$  is unchanged. Therefore, developers will be less willing to choose a high quality level but more inclined to switch to platform owner-driven region. On the other hand, according to Figure 1.3, a high  $r_R$  shrinks the scope of platform owner-driven region. Therefore if  $r_R \geq \hat{r}_R$ , even a high  $k_R$  cannot render developers switch to platform owner-driven region. But if  $r_R$  is low ( $r_R < \hat{r}_R$ ) as in Case 2 in which the scope of platform owner-driven region is large, when  $k_R$  increases, developers would potentially switch to platform owner-driven region, which is detrimental for platform owner's revenue. Hence, it is to the platform owner's best interest to offer a higher  $s_1^{A*}$  to keep developers in the self-driven region.

The above reasoning can be justified by Figures 1.5 and 1.6 drawn based on Proposition 1.A1 in Appendix A. Figure 1.5 presents the situation for Case 1. It shows that when  $r_R \geq \hat{r}_R$ , developers in the self-driven region would always choose the self-driven region when  $k_R$  increases. Figure 1.6 illustrates the situation for Cases 2 and 3 where  $r_R < \hat{r}_R$ . Developers in the self-driven region will stay in the self-driven region if  $k_R$  is in the range indicated by Case 3. However if  $k_R$  is high (Case 2), increase in  $k_R$  may make them switch to platform owner-driven region.

**Corollary 1.3.5** *For squeezing, encouragement, and retention regions, in SPNE the platform owner's revenue  $\partial u_p^* / \partial k_U > 0$ ,  $\partial u_p^* / \partial k_R < 0$ ,  $\partial u_p^* / \partial r_0 > 0$ , and  $\partial u_p^* / \partial r_R < 0$ .*



**Figure 1.5** Region distribution versus  $k_R$  ( $r_R \geq \hat{r}_R$ )



**Figure 1.6** Region distribution versus  $k_R$  ( $r_R < \hat{r}_R$ )

### 1.4.2 Analysis of Social Welfare

In this subsection, we analyze the impact of consumers' rating behavior on social welfare of app market.

The social welfare is defined as:

$$W = bq - hq^2.$$

**Corollary 1.4.1** *In the encouragement or retention region,  $\partial W / \partial r_R > 0$  and  $\partial W / \partial r_0 < 0$ .*

In the encouragement or retention region, if  $r_R$  increases or  $r_0$  decreases, the social welfare increases. In these regions, there exist developers' quality efforts which are not fully pushed out since consumers are lenient towards developers. When consumers become more critical in evaluating ratings or more severe in giving ratings, such efforts will be pushed out to increase the app quality and hence the social welfare increases.

**Corollary 1.4.2**  *$\partial W / \partial k_R < 0$  in the encouragement region and  $\partial W / \partial k_R = 0$  in the retention region.*

*$\partial W / \partial k_U > 0$  if  $h \leq h_{2\alpha}^A$ , and  $\partial W / \partial k_U < 0$  if  $h > h_{2\alpha}^A$  where:*

$$h_{2\alpha}^A = \frac{1}{2} \frac{k_R b k_U^2 (1-b)}{(b k_U - k_R b + k_R)^2 (r_R - r_0)}.$$

In the encouragement region, the social welfare is decreasing in  $k_R$ . This indicates that consumers need to appreciate ratings, that is, be willing to pay higher price for higher rating, for the sake of welfare of the market. If developers' cost rate is relatively lower ( $h \leq h_{2\alpha}^A$ ), the social welfare increases when consumers give more differentiated ratings for good and bad apps (higher  $k_U$ ). On the other hand, if only high cost developers ( $h > h_{2\alpha}^A$ ) on the market, more differentiated rating behavior is detrimental to the social welfare.

**Corollary 1.4.3** *In the squeezing region,  $\partial W / \partial x < 0$  where  $x = r_R, r_0, k_R, k_U$ .*

In the squeezing region, developers' cost rate is very low. In that situation being severe in giving ratings is beneficial for the social welfare. The intuition comes from the fact that developers have a great potential to produce high quality app. Consumers' severity in giving ratings helps to realize such potential



and improve the overall social welfare. However, paradoxically, being more critical in evaluating ratings reduces the social welfare. This is because it fails to force developers to promote the quality as they are already self-driven. Overly critical behavior actually reduces developers' incentive to produce at high quality level because they cannot charge the high price to justify their investment on quality.

**Table 1.2** Platform owner's (P.O.'s) interest versus social welfare (S.W.)

		$r_R$		$r_0$		$k_U$		$k_R$	
		+	-	+	-	+	-	+	-
<b>Squeezing Region</b>	P.O.	-	+	+	-	+	-	-	+
	S.W.	-	+	-	+	-	+	-	+
<b>Encouragement Region</b>	P.O.	-	+	+	-	+	-	-	+
	S.W.	+	-	-	+	+	-	-	+
<b>Retention Region</b>	P.O.	-	+	+	-	+	-	-	+
	S.W.	+	-	-	+	-	+	N	N

Table 1.2 compares the directions of change for platform owner's  $u_p^*$  and the social welfare  $W$  in response to the change of rating parameters  $r_R$ ,  $r_0$ ,  $k_U$  and  $k_R$  in different regions. Positive sign “+” represents “increase” while negative sign “-” represents “decrease.” “N” means “No effect.” We highlight the situations where the interest of platform owner and the social welfare are always misaligned. We can observe that the adjustment of  $k_R$  is always aligned, but that of  $r_0$  is always misaligned. This shows that rating leniency is always detrimental to the social welfare but always favored by the platform owner.

### 1.5 Unobserved Cost Rate

According to Proposition 1.2 in the base model, the platform owner's optimal choice of revenue sharing percentage is a function of developers' cost rate  $h$ . However,  $h$  may be unobservable by the platform owner. In this section, we present the optimal choice of sharing percentage when  $h$  is unobserved. We also demonstrate that the platform owner can offer developers a menu which lists two combinations of *ex post* ratings and revenue sharing percentage to engage them into a self-selection process.

### 1.5.1 Unobserved $h$

Suppose that the platform owner has no information about  $h$ . One approach to model the uncertainty with no prior information is to assume uniform distribution. Assume that  $h$  is uniformly distributed in  $[0, h_U]$  where  $h_U > h_S$ . It represents the situation where developers' cost rate space is not fully covered. It is reasonable since there are always developers whose cost rates are so high that they cannot earn non-negative profit even when all revenue shares go into them.

**Proposition 1.3** *When  $h$  is uniformly distributed, the expected optimal developers' revenue sharing percentage  $s^{N*} = 1/2$ .*

From the base model we can see that the optimal sharing percentage is a linear function of  $h$ . Intuitively speaking, when  $h$  is on the lower side of the distribution, the optimal sharing percentage is less than one half; whereas when  $h$  is on the higher side of the distribution, the optimal sharing percentage is greater than one half. Proposition 1.3 tells the expectation is exact one half.

### 1.5.2 Self-selection

In this subsection, we show that the platform owner can design a menu of  $(s, r)$  to engage developers to a self-selection process. Note that developers with different cost rate have different objectives for the rating. Figure 1.1 shows that self-driven developers aim at  $r = 1$  while platform owner-driven developers' goal is between  $r_R$  and 1. Therefore although cost rate  $h$  is not observable, the rating  $r$  at which developers would like to stay is an observable factor signaling developers' cost rate  $h$ . Hence, if the platform owner can offer different revenue sharing percentages bundled with different realization of  $r$ , developers will self-select the combination which is most profitable to them. This turns the information structure of our three-party game into a screening game in conjunction with a signaling game. The screening game is between the platform owner and developers, and the platform owner is the uninformed party. The signaling game is between developers and consumers, and developers are uninformed. In both games, developers are the informed party who would like to signal app quality to consumers (signaling game) and also engage into a self-selection generated by the platform owner (screening game).

We consider a separating equilibrium which involves self-driven and platform owner-driven developers. Therefore, suppose that there exists a representative self-driven developer with cost rate  $h_L$  in the squeezing region and a representative platform owner-driven developer with cost rate  $h_H$  in the encouragement region (Region 2-2). We derive the following result.

**Proposition 1.4** *There exists a separating equilibrium in which the representative developers  $h_L$  and  $h_H$  would self-select to  $s_L$  and  $s_H$  respectively if the following menu is offered:*

$$s^N = \begin{cases} s_H, & \text{if } r \leq \hat{r}; \\ s_L, & \text{if } r > \hat{r}, \end{cases}$$

where,

$$\hat{r} = r_0 - \frac{(r_R - r_0)(1+b)}{2} - \frac{1}{4} \frac{bk(1-b)}{h_H}, \quad s_L = \frac{2h_L(1+br_0-br_R-r_0)}{bk(1-b)}, \quad \text{and} \quad s_H = \frac{1}{2} + \frac{h_H(r_R - r_0)}{bk}.$$

Proposition 1.4 characterizes a separating equilibrium for the screening game. By offering this menu selection to developers, without information on developers' cost rate  $h$ , the platform owner can incentivize  $h_H$ -type developers to self-select the corresponding platform owner's profit-maximizing  $s_H$  and  $h_L$ -type developers to the corresponding platform owner's profit-maximizing  $s_L$ .

It is interesting to see from Proposition 1.4 that when consumers become more severe in giving ratings, the equilibrium threshold rating  $\hat{r}$  decreases. However, if consumers become more critical in evaluating ratings,  $\hat{r}$  also decreases.

## 1.6 Conclusions

In this essay, we parameterize consumer rating behavior into four parameters  $(k_U, k_R, r_0, r_R)$  and construct a bidirectional rating-utility framework which integrates these parameters into consumers' utility functions. In the equilibrium analysis, we identify three types of developers: self-driven, platform owner-driven and poor marketing. We investigate how the changes in consumer rating behavior  $(k_U, k_R, r_0, r_R)$  affect their optimal choices of quality level and app price, as well as the platform owner's optimal choice

of revenue sharing policy and the social welfare. The over-leniency issue, a well-known problem which has been empirically identified in many behavioral management studies, is analytically observed in our rating-utility economic model. Our analysis on social welfare reveals that some types of changes in consumers rating behavior, though increasing the platform owner's profit, might be not aligned with the interest of social welfare. We also find that when developers' cost rate is unobservable, the platform owner can design a screening game in which there exist a separating equilibrium for high and low cost rate developers.

## Essay 2 Pricing Models in the Electronic Book Market

### 2.1 Introduction

When Amazon.com launched its e-book reader Kindle in 2007, its founder Jeff Bezos it was time for “the last bastion of analog”, books, to be digitized and sold in electronic format (Levy 2007). In 2011, less than four years after introducing Kindle books, Amazon.com sells more e-books than print books, hard cover and paperback combined (Miller and Bosman 2011). Although readers have become more comfortable with e-books, many complain about their high retail prices. As it costs less to produce and distribute an e-book compared to a physical book, consumers expect lower prices for e-books. Instead, they see rising prices (Stone and Rich 2009).

To discern the cause of rising e-book retail prices, we investigate the pricing models in the e-book industry. Currently, there are two main e-book pricing models: (i) the wholesale pricing model (hereafter called the “wholesale model”) and (ii) the agency pricing model (hereafter called the “agency model”) (Trachtenberg 2011b). In the wholesale model, a publisher such as Random House charges a retailer such as Amazon.com a wholesale price. Then, Amazon.com decides the final retail price for the e-book. In the agency model, the publisher decides the final retail price for an e-book and the e-book retailer receives a fixed percentage of the sales revenue. Amazon.com started with the wholesale model and priced digital versions of *New York Times* best-sellers and new releases at just \$9.99 (Stone and Rich 2009). Publishers were unhappy with the low e-book retail prices and pressured Amazon.com to raise them. In 2010, after negotiating with Amazon.com, several publishers eventually switched to the agency model and took control of their own e-book retail prices (Trachtenberg 2011a).

So what was the impact of this switch in pricing models on e-book retail prices? Could it possibly contribute to the increase in e-book retail prices? To answer these questions, we need to investigate how the retail prices are determined in both models. In the wholesale model, what we observed in the e-book market is quite different from that seen in a traditional retail setting where double marginalization occurs, i.e., both the supplier and the retailer have incentives to mark up a product’s price over its marginal cost (Tirole 1988). Double marginalization often leads to lower demand, a smaller profit, and a higher retail

price for the whole supply chain. In response, the supplier wants to use vertical restraints such as resale-price maintenance (RPM) to set up a price ceiling for the retailer to make sure the retail price is not too high (Rey and Tirole 1986). In the e-book market, however, an e-book retailer such as Amazon.com, instead of marking up, wants to set low retail prices for e-books. Why would the e-book retailer charge low e-book retail prices in the wholesale model? This study is aimed at providing possible explanations for that. In the agency model, on the other hand, the publisher decides the e-book retail price. What are the factors that the publisher considers in choosing the e-book retail price? How is the equilibrium retail price determined? And how do the equilibrium prices compared to the wholesale ones? This study also attempts to answer these questions.

We use a game theoretical approach to analyze both the publisher's and the e-book retailer's pricing schemes in both the wholesale and agency model. We identify one potential reason for the e-book retailer in the wholesale model to set low e-book retail prices—the complementary consumption of e-reader and e-book. If the e-book retailer also sells e-readers and consumers need to buy the e-reader to read the e-books, the e-book retailer's optimal scheme is to use a two-part tariff pricing structure, i.e., pricing e-books at the wholesale price set by the publisher and marking up e-readers. Under the agency model, however, the e-book retailer's optimal scheme changes. It prefers a lower e-reader price and makes a profit through the revenue sharing of e-book sales. We show that given a certain range of the revenue sharing percentage, the equilibrium e-book retail price is higher in the agency model compared to that it is in the wholesale model. We also show numerically that although the publisher gains the e-book's retail price control in the agency model, in equilibrium the publisher could be worse off in terms of overall profit.

The rest of this essay is organized as follows. We review relevant literature in Section 2.2. In Section 2.3, we set up the demand functions and analyze both the wholesale and agency models. We provide the social welfare analysis and a comparison of both pricing models in Section 2.4. Discussion is presented in Section 2.5, followed by our conclusions in Section 2.6.

## 2.2 Literature Review

As mentioned earlier, this study is related to the literature of vertical constraints. The problem of double marginalization happens when both upstream and downstream firms have monopoly power such that they want to mark up the price above marginal cost. The markups raise the retail price, lower the demand, and lower the combined profit for the supplier and the retailer (Spengler 1950). A franchise fee or RPM can solve the double marginalization problem (Rey and Tirole 1986). In the former case, the supplier charges the retailer a wholesale price equal to the supplier's marginal cost plus a franchise fee. Then, the retailer has all the incentive to set the retail price as if the supplier and the retailer were vertically integrated. In RPM, the supplier directly imposes a price ceiling for the retailer to mitigate the unfavorable effects of double marginalization.

This study is also related to multiproduct pricing models. A vast body of literature has been devoted to both linear and nonlinear multiproduct pricing models (Whinston 1990, Armstrong 1996, Rochet and Chone 1998). Church and Gandal (1996) assessed the effect of hardware control on software provision in the markets where the consumption benefit of hardware is a function of the variety of available software. Mulhern et al. (1991) studied the optimal pricing and promotion policy when retailers sell multi products for which demand is interdependent. It was shown that retailers can exploit the substitute or complement relationship among products using implicit price bundling.

In the Information Systems (IS) area, there is a rich literature on information goods pricing and distribution strategies. Sundararajan (2004) analyzed two nonlinear pricing models, unlimited-usage (fixed-fee pricing) model and the usage-based model—for their applications in pricing information goods. The results suggested that using fixed-fee pricing in addition to nonlinear usage-based pricing always improves the profit when the transaction cost is nonzero. Lang and Vragov (2005) examined a pricing scheme for distributing digital content over centralized and decentralized networks. Chellappa and Kumar (2005) studied how “free” product-augmenting services affect online sellers’ pricing and customer retention strategies. Fan et al. (2007) developed a model to examine optimal strategies for media providers to utilize online channels to distribute digital media. Mantena et al. (2010) studied the exclusive

contract between vendors of platforms such as video game consoles and vendors of complementary goods. Feng et al. (2009) and Li (2010) examined optimal channel structures and corresponding pricing strategies in distributing digital content. Choudhary (2010) studied vendors' choices of pricing schemes under market competition and showed that the choice of pricing scheme affects buyers' usage levels as well as revenue distribution over different segments of buyers. Thus, vendors could differentiate themselves by choosing different pricing schemes, for example, per user pricing and site licensing. Yu et al. (2011) studied optimal pricing schemes for both digital devices and contents when they are tied.

Our study contributes to the literature in the following ways. One unique aspect of our study is that instead of setting up an aggregated market demand function, we construct the market demand from several consumption patterns that are specifically attached to the e-book market. For example, we model the complementary consumption of e-readers and e-books by assuming that one individual consumer purchases one e-reader with multiple e-books. Moreover, consumers' partiality for electronic reading is captured in our horizontal differentiation setup between e-book and physical book. Therefore, in our model we characterize the price dynamics between e-reader and e-book in the presence of a substitution effect between e-book and physical book. We identify two different business models for the e-book retailer—making profit via e-reader versus making profit via e-book. For the publisher we show how to adjust the e-book price to balance its e-book market and physical book market to obtain the overall maximum profit. We contrast two relevant pricing models in the e-book industry, the wholesale model and the agency model, and reveal their social welfare implications. To the best of our knowledge, this is the first work to compare two popular pricing models in e-book industry, and our results have important implications for both the publishing and e-book retailing industries.

## **2.3 The Model**

### **2.3.1 Model Setup**

Suppose there is one representative publisher and one e-book retailer (hereafter called “retailer”) in the marketplace. The publisher supplies the content of e-books to the retailer. The publisher also sells physical books with the same content. The retailer acquires the content of e-books and makes it available



to the marketplace in the format of e-books along with its e-book reader. Denote the e-book retail price by  $p_E$  and the e-reader price by  $p_D$ . Denote the physical book retail price by  $p_F$ .

We start our analysis by characterizing the individual consumer's purchase quantities. As we mentioned earlier, a consumer decides to buy an e-reader because the consumer plans to use the device repeatedly, i.e., to read multiple e-books (Yu et al. 2011). Therefore, in the spirit of Laffont et al. (1998), we derive one consumer's purchase quantity of e-books as

$$q_E = (b - p_E) / m.$$

As the e-book retail price  $p_E$  drops, one consumer will buy more e-books. One consumer's purchase quantity of physical books follows a similar function:  $q_F = (b - p_F) / m$ . We assume that the e-book and the physical book have same parameter  $(b, m)$ , because the content they deliver is identical. Without loss of generality, we restrict that  $b > p_F$  and  $b > p_E$ .

Based on the above individual demand functions, we derive one consumer's total surplus of purchasing e-reader and e-books as  $CS_E = (b - p_E)^2 / (2m) - p_D$  (Laffont et al. 1998). The consumer's total surplus of purchasing physical books is  $CS_F = (b - p_F)^2 / (2m)$ . Consumers are heterogeneous in terms of their preference toward electronic reading. Following Hotelling's location model (Tirole 1988) and Laffont et al. (1998), we assume consumers are uniformly distributed in  $a \in [0, 1]$ . At  $\alpha = 0$ , consumers mostly favor e-book reading, and at  $\alpha = 1$ , consumers mostly favor physical book reading. As  $\alpha$  increases, consumers' preference for e-book reading decreases and preference for physical book reading increases.

Thus, the net utility gained by one consumer who chooses e-books is  $CS_E - t\alpha$  (this is essentially choosing the combination of an e-reader and multiple e-books; but for simplicity and contrast to physical books, we call it "e-books"). For the consumer who chooses physical books it is  $CS_F - t(1 - \alpha)$ . Suppose

the consumer with the preference parameter  $\alpha^*$  is indifferent about choosing physical books versus e-books. Assume the market is fully covered. Thus, we have

$$CS_E - t\alpha^* = CS_F - t(1 - \alpha^*).$$

Then, we derive the number of consumers who choose the e-book, that is, the size of the e-book market:

$$s_E = \alpha^* = (t - p_D) / (2t) + (p_F - p_E)(2b - p_E - p_F) / (4tm).$$

The size of the physical book market is therefore  $s_F = 1 - s_E$ .

It is worth mentioning that our analysis focuses on one category of books. While admittedly consumers may make different choices between e-book versus physical book for different categories of books, we believe that within one category, the choice is relatively stable. Take textbooks as an example. When consumers consider whether to choose e-books or physical books for textbooks, they first compare the e-book prices and physical book prices, and then calculate if the price differences are sufficiently large to make it worth buying an e-reader. Once they have chosen the e-book, they will generally stick to it for all textbooks as long as the e-book is less expensive than the physical book. Our model approximates this consumer rationale. It is also consistent with prior studies in multichannel research (Balasubramanian 1998, Chiang et al. 2003).

After establishing the demand functions, we analyze the wholesale model and the agency model equilibriums.

### 2.3.2 The Wholesale Model

In the wholesale model, suppose the publisher offers the retailer each e-book at the wholesale price  $w_E$ . Consider a two-stage sequential game. At the first stage, the publisher decides  $w_E$ . At the second stage, given  $w_E$ , the retailer decides retail prices for the e-book and the e-reader  $(p_E, p_D)$ . The publisher pays  $c_A$  for author's royalty on each e-book as well as each physical book. We normalize the marginal costs of storing and distributing an e-book to zero for two reasons. First, those costs are very low (Trachtenberg 2011b). Second, normalizing them to zero does not affect the results of the analysis. We also disregard the

cost of digitalizing books since it is one-time fixed cost. Suppose the publisher's total cost for distributing and selling one physical book is  $c_F$ . Then, the publisher's total profit function is

$$\pi_p = (w_E - c_A)q_E s_E + (p_F - c_F - c_A)q_F s_F.$$

As shown in this equation, the publisher's total profit  $\pi_p$  consists of two parts: profit from e-books and profit from physical books.

The retailer's total profit  $\pi_R$  is

$$\pi_R = ((p_E - w_E)q_E + (p_D - c_D))s_E,$$

where  $c_D$  is the marginal cost for the e-reader. The equation shows that the retailer's total profit also consists of two parts: profit from e-books and profit from e-readers.

By using backward induction, we first solve for the retailer's problem, assuming  $w_E$  is already given.

**Proposition 2.1** *Given the publisher's wholesale price  $w_E$ , the optimal e-book retail price  $p_E^*$  and e-reader price  $p_D^*$  in the wholesale model are*

$$p_E^* = w_E, \text{ and}$$

$$p_D^* = (t + c_D) / 2 + (p_F - w_E)(2b - p_F - w_E) / (4m).$$

Proof of propositions is provided in Appendix B.

Proposition 2.1 suggests that the retailer should set the e-book retail price at the wholesale price received from the publisher. It implies that the retailer's optimal scheme is to have a two-part tariff pricing structure under which the retailer makes profit from the e-reader rather than from the e-book.

It is interesting to see that this retailer's pricing scheme is different from traditional double marginalization case in which the retailer has an incentive to set the retail price above the marginal cost. The major difference here is the existence of the e-reader, which is required to read e-books. This complementary relationship between the e-reader and e-book, plus the fact that retailer's profit comes from both parts, incentivizes the retailer to adjust both prices to attain the overall optimal profit point.

Proposition 2.1 suggests that this overall optimal point can be achieved by pricing the e-book at the wholesale price to boost the consumer surplus, and then marking up the e-reader to extract that surplus.

Examining the e-reader price in Proposition 2.1, we find that  $\partial p_D^* / \partial w_E < 0$ . This suggests that when the publisher increases the e-book wholesale price, the retailer should decrease the e-reader price. The reason is that the increase in the e-book wholesale price will lead to an increase in the e-book retail price so that it decreases consumers' perception of the surplus for choosing e-books. Consequently, the surplus the retailer can extract declines. Thus, the price of the e-reader needs to drop.

We substitute the retailer's optimal prices  $(p_E^*, p_D^*)$  from Proposition 2.1 into the publisher's profit function and solve for the publisher's problem to obtain the subgame perfect Nash-equilibrium (SPNE). Unfortunately, the closed-form expression for equilibrium  $w_E$  is not available. However, when  $t$  is sufficiently large, we can find out the following comparative statics in SPNE (see Appendix B for detailed proof).

**Proposition 2.2** *In the wholesale model, (i) the publisher's optimal wholesale price  $w_E^*$  is decreasing in physical book cost  $c_F$ , i.e.,  $dw_E^* / dc_F < 0$ ; and (ii) the optimal e-reader price  $p_D^*$  is increasing in physical book cost  $c_F$ , i.e.,  $dp_D^* / dc_F > 0$ .*

Proposition 2.2 demonstrates how the physical book cost  $c_F$  affects the publisher's equilibrium choice of the e-book's wholesale price  $w_E^*$ . If the physical book cost  $c_F$  increases, the profit margin for one physical book decreases. Then, the e-book becomes relatively more profitable to the publisher than it used to be. As a result, the publisher has the incentive to encourage more customers to choose e-books by reducing the wholesale price, which leads to a lower e-book retail price  $p_E^*$ .

As a consequence of lower e-book retail price  $p_E^*$ , the consumer surplus on e-book consumption increases. The retailer can extract a larger amount of the consumer surplus from customers by increasing the e-reader price. Therefore, the equilibrium e-reader price  $p_D^*$  is increasing in the physical book cost  $c_F$ .

We further investigate the influence of change in the physical book retail price  $p_F$  on equilibrium e-book wholesale price  $w_E^*$ . Since in SPNE the e-book retail price  $p_E^*$  is set equal to the e-book wholesale price  $w_E^*$ , the influence of such a change in  $p_F$  also applies to  $p_E^*$  in the same way as it does to  $w_E^*$ . We define the following two price thresholds:  $p_{th1} = (b + c_A + c_F) / 2$  and  $p_{th2} = b(2b + 2c_A + c_F) / (3b + c_A)$ .

**Proposition 2.3** *In the whole sale model, (i) when  $p_F \leq p_{th1}$ , the publisher's wholesale price  $w_E^*$  is increasing in physical book retail price  $p_F$ , i.e.,  $dw_E^* / dp_F > 0$ ; (ii) when  $p_F \geq p_{th2}$ , the publisher's wholesale price  $w_E^*$  is decreasing in  $p_F$ , i.e.,  $dw_E^* / dp_F < 0$ .*

Proposition 2.3 shows the dynamics between the physical book retail price  $p_F$  and the equilibrium e-book wholesale price  $w_E^*$ . The first price threshold  $p_{th1}$  is essentially the optimal physical book retail price if the publisher only sells physical books. When the physical book retail price  $p_F$  is lower than the threshold  $p_{th1}$ , an increase in  $p_F$  will increase the amount of profit that the publisher gains from one individual consumer who chooses physical books. So, in equilibrium the publisher increases the e-book wholesale price  $w_E^*$ , which converts some of the e-book buyers to physical book buyers, leading to an overall maximized profit.

When the physical book retail price  $p_F$  is sufficiently high ( $p_F \geq p_{th2}$ ), any further increase in  $p_F$  will lower the profit from one individual consumer who chooses the physical book. Therefore, the publisher should reduce the e-book wholesale price  $w_E^*$  to convert some of the physical book buyers to e-book buyers, which eventually maximizes the overall profit.

### 2.3.3 The Agency Model

In the agency model, the publisher determines the e-book's retail price  $p_E$ . The retailer and the publisher then agree upon a revenue sharing contract to split the revenue of e-book sales. We denote the retailer's percentage by  $r$  and assume  $0 < r < 1$ . Consider a two-stage sequential game. At the first stage, the

publisher determines the e-book retail price  $p_E$ . At the second stage, given the first stage  $p_E$ , the e-book retailer decides the e-reader price  $p_D$ .

In the agency model, the publisher's profit  $\pi_p$  is

$$\pi_p = ((1-r)p_E - c_A)q_E s_E + (p_F - c_A - c_F)q_F s_F,$$

where  $(1-r)p_E$  is the revenue the publisher receives from one e-book sale. Similar to the wholesale model, the publisher's profit has two parts—profit from e-books and profit from physical books.

The retailer's profit  $\pi_R$  is

$$\pi_R = (rp_E q_E + (p_D - c_D))s_E,$$

where  $rp_E$  is the revenue the retailer receives from one e-book sale. The retailer's profit also has two parts—profit from e-books and profit from e-readers.

**Proposition 2.4** *Given the e-book retail price  $p_E$ , the optimal e-reader price  $p_D^*$  in the agency model is*

$$p_D^* = (t + c_D)/2 + (p_F - p_E)(2b - p_E - p_F)/(4m) - rp_E(b - p_E)/(2m).$$

Proposition 2.4 shows the retailer's optimal e-reader price  $p_D^*$ , given the e-book retail price  $p_E$ . By deriving  $\partial p_D^* / \partial p_E$ , we can find how the change in e-book retail price affects the e-reader price. When  $p_E$  is relatively low ( $p_E < (1+r)b/(1+2r)$ ), we have  $\partial p_D^* / \partial p_E < 0$ . This result indicates that the retailer should decrease the e-reader price  $p_D$  when the e-book retail price  $p_E$  increases. Although this result is similar to that in the wholesale model, the retailer's underlying rationale is quite different. In the wholesale model, the retailer makes money from the e-reader only. When the e-book retail price increases, the retailer has to reduce the e-reader price to keep a reasonable e-book market size so the retailer can make a profit through the complementary good, the e-reader. In the agency model, the retailer's profit comes from both the e-book and the e-reader. So when the e-book retail price increases, the retailer lowers the e-reader price not because the retailer has to, but because reducing the e-reader price leads to a

larger e-book market size, which increases the total e-book sales revenue. As a consequence of the revenue sharing contract in the agency model, the retailer's profit from e-books increases. Whenever the profit gain from e-books is sufficient to compensate for the revenue loss caused by the e-reader's price drawdown, it stands to reason for the retailer to decrease the e-reader price. In essence, instead of focusing on profits from e-readers, as retailers do in the wholesale model, in the agency model the retailer focuses on profits from e-books.

However, when  $p_E$  is on the high end ( $p_E > (1+r)b/(1+2r)$ ), we have  $\partial p_D^* / \partial p_E > 0$ . When  $p_E$  is too high, it leads to a small e-book market size and low e-book revenue. In this situation, the marginal gain in the retailer's e-book revenue does not compensate for the loss of revenue on e-readers if the e-reader price decreases. Therefore, the retailer increases the e-reader price to maximize total profit.

In the following, we derive the SPNE properties of the agency model.

**Proposition 2.5** *When the retailer's revenue sharing percentage for e-books is less than that for physical books, i.e.,  $r < (c_F / p_F)$ , the equilibrium e-book retail price  $p_E^*$  in the agency model satisfies  $p_E^* < \bar{p}_E$ , where  $\bar{p}_E = (b + c_A / (1-r)) / 2$ .*

The condition in Proposition 2.5,  $r < (c_F / p_F)$ , is quite mild. It means that the retailer's revenue sharing percentage for the e-book channel is less than that for the physical book channel. As shown in Trachtenberg (2011b), this condition is easily satisfied in the real business environment.

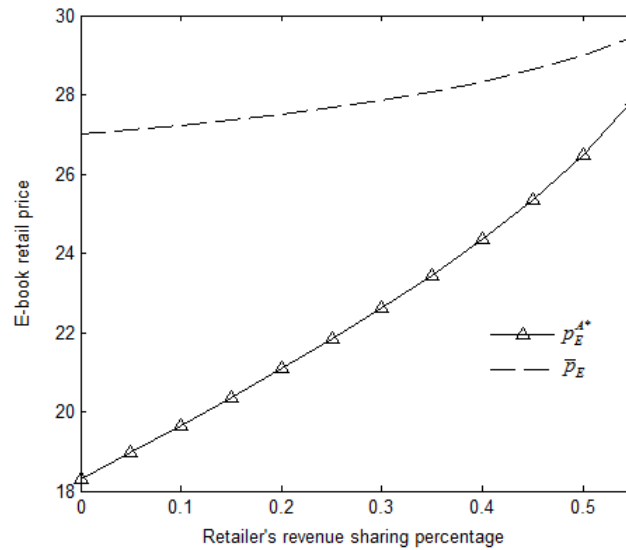
The upper bound for the e-book retail price,  $\bar{p}_E$ , is essentially the optimal e-book retail price as if the publisher sells only e-books. Proposition 2.5 suggests that the substitution effect between the physical book and e-book lowers the equilibrium e-book retail price. The publisher's optimal e-book retail price decreases if that e-book's physical version is also introduced.

In addition, through deriving  $\partial \bar{p}_E / \partial r$ , we find that as the retailer's revenue sharing percentage  $r$  increases, the upper bound for the e-book's price  $\bar{p}_E$  also increases. This is because the increase in the

retailer's revenue sharing percentage reduces the publisher's profit. As a result, the publisher wants to raise the e-book retail price to make up for the loss.

We illustrate these results in Figure 2.1. The parameter values used for drawing the graph basically follow Trachtenberg (2011b) with the following values:  $p_F = 26$ ,  $c_F = 15.6$ ,  $c_A = 4$ ,  $c_D = 70$ ,  $b = 50$ ,  $m = 1$ , and  $t = 400$ .

Proposition 2.4 shows that the equilibrium e-reader price could either be decreasing or increasing in the e-book retail price, depending on whether the e-book retail price is lower or higher than the threshold  $(1+r)b/(1+2r)$ . Based on the result of Proposition 2.5, we want to point out that the latter case, i.e., the e-reader price is decreasing in the e-book retail price, is unlikely to happen. That is because the upper bound of the e-book's price  $\bar{p}_E$  is less than  $(1+r)b/(1+2r)$  as long as  $r \leq 0.5$  and  $c_A < b/4$ . In practice, both conditions are apparently satisfied since the retailer's revenue sharing percentage is around 30% and the author's royalty fee is usually less than 20% of the physical book price (Trachtenberg 2011b). Therefore, we conclude that in the real business environment the equilibrium e-reader price is decreasing in the e-book retail price in the agency model. This result is illustrated in Figure 2.2, which is drawn based on the same parameter values as the ones used in Figure 2.1.

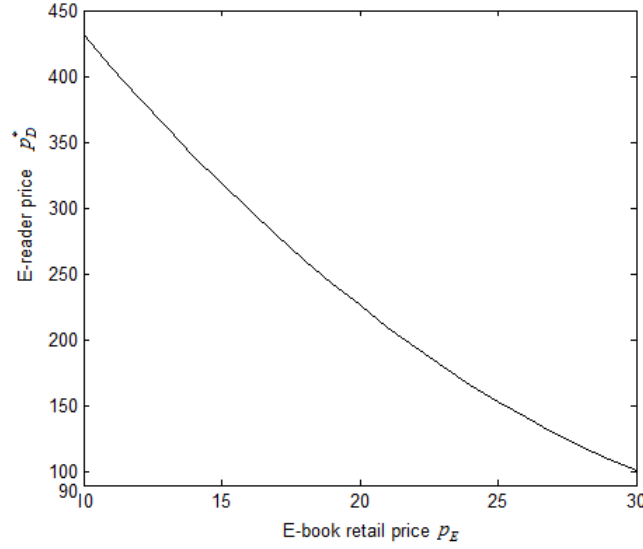


**Figure 2.1** E-book retail price versus retailer's revenue sharing percentage



We further investigate how changes in the physical book retail price and physical book cost affect the equilibrium e-book retail price in the agency model.

**Proposition 2.6** *The equilibrium e-book retail price  $p_E^*$  in the agency model is decreasing in the physical book cost  $c_F$ , i.e.,  $\partial p_E^* / \partial c_F < 0$ .*



**Figure 2.2** E-reader price versus e-book retail price

Define  $\tilde{p}_{th2} = b(c_A + (1-r)(2b + c_A + c_F)) / (c_A + 3(1-r)b)$ . Similar to Proposition 2.2, when  $t$  is sufficiently large (see Appendix B for more detail), we have the following results.

**Proposition 2.7** *In the agency model, (i) when  $p_F \leq p_{th1}$ , the equilibrium e-book retail price  $p_E^*$  is increasing in physical book retail price  $p_F$ , i.e.,  $\partial p_E^* / \partial p_F > 0$ ; and (ii) when  $p_F \geq \tilde{p}_{th2}$ , the equilibrium e-book retail price  $p_E^*$  is decreasing in physical book retail price  $p_F$ , i.e.,  $\partial p_E^* / \partial p_F < 0$ .*

Results from Propositions 2.6 and 2.7 are similar to those in the wholesale model. This is because for all the dynamics between the physical book and e-book retail prices, the underlying principles of the two pricing models are similar: The e-book and physical book are market substitutes. Both Propositions 2.2 and 2.6 show that an increase in the physical book cost will reduce the physical book profit margin such that it gives the publisher the incentive to motivate some consumers to choose e-books over physical books by reducing the e-book retail price. Both Propositions 2.3 and 2.7 show that the dynamic

relationship between the e-book retail price and the physical book retail price depends on the level of the physical book retail price. For mass-market books, the condition  $p_F \leq p_{th1}$  is likely to be satisfied. Therefore, the e-book and physical book retail prices should move together to maximize the total profit.

## 2.4 Comparison and Social Welfare Analysis

In this section, we first compare the equilibrium prices of the wholesale and agency models. Then, we demonstrate the welfare implications.

**Proposition 2.8** *When the retailer's revenue sharing percentage  $r < \min\{c_F / p_F, 3/4\}$  and author's royalty fee  $c_A < p_F / 6$  and physical book cost  $c_F > 4p_F / 9$ , (i) the equilibrium e-book retail price in the agency model is higher than in the wholesale model, i.e.,  $p_E^{A*} > p_E^{W*}$ ; and (ii) the equilibrium e-reader price in the agency model is lower than in the wholesale model, i.e.,  $p_D^{A*} < p_D^{W*}$ .*

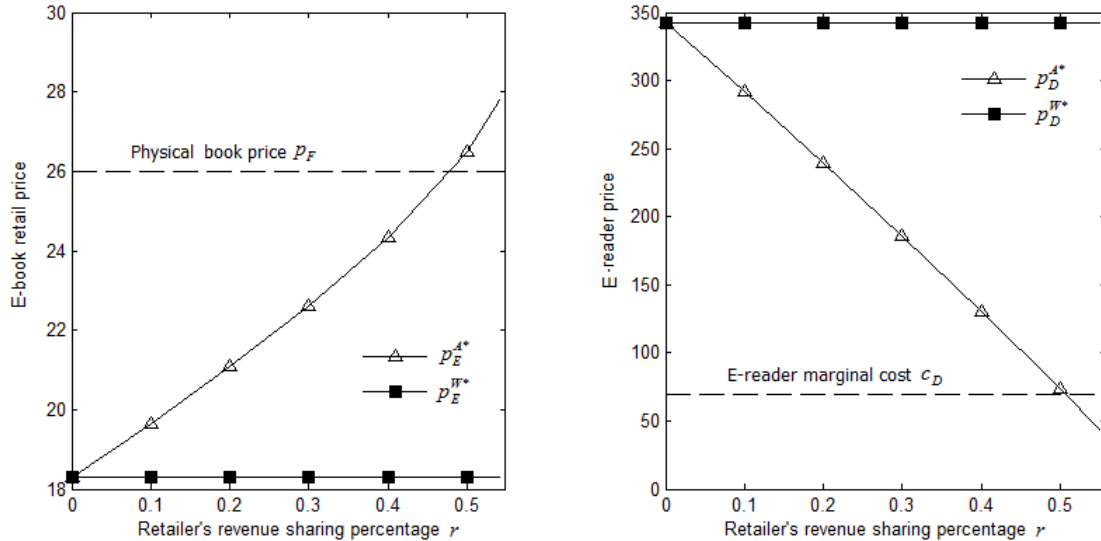
As we mentioned in Proposition 2.5, the condition  $r < \min\{c_F / p_F, 3/4\}$  is quite mild. According to Trachtenberg (2011b), it is safe to believe that in the real business environment the retailer's revenue sharing percentage in the e-book channel is less than its physical channel counterpart. Meanwhile, it is quite common that the retailer's revenue sharing percentage is less than three-fourths. Therefore, in the real business environment the condition  $r < \min\{c_F / p_F, 3/4\}$  should be easily satisfied. The other two conditions  $c_A < p_F / 6$  and  $c_F > 4p_F / 9$  are also satisfied in real business environment according to Trachtenberg (2011b).

Proposition 2.8 indicates that switching from the wholesale model to the agency model leads to an increase in the equilibrium e-book retail price and a decrease in the e-reader price. We plot Figure 2.3 to illustrate such price changes. The parameter values are identical to the ones used in Figures 2.1 and 2.2. We can clearly see that  $p_E^{A*} > p_E^{W*}$  and  $p_D^{A*} < p_D^{W*}$  in Figure 2.3.

As we mentioned earlier, the retailer's business model changes when the pricing model does. In the wholesale model, the retailer keeps the e-book price low and marks up the e-reader. In the agency model, the retailer brings down the e-reader price to enlarge the total e-book revenue and makes a profit through

its e-book revenue share. When the retailer decreases the e-reader price, it pushes the total e-book market size up. So, the publisher loses its previous profit-maximizing division of the e-book and physical book markets, i.e., the e-book market size is overly “large.” Therefore, in response to the retailer’s move, the publisher increases the e-book retail price to restore the profit-maximizing division of the two markets. It is worth mentioning that this *ex post* profit-maximizing division is not the *ex ante* one anymore. Later in Table 2.1, we show the differences.

Figure 2.3 also shows that in the agency model, as the retailer’s revenue sharing percentage increases, the equilibrium e-reader price decreases and the e-book retail price increases. The increase in the retailer’s revenue sharing percentage gives the retailer the incentive to further reduce the e-reader price, because, the more e-book revenue share the retailer takes, the more the retailer wants to focus on selling e-books, and the more the e-reader becomes like a mere vehicle to boost total e-book revenue. When the profit obtained from e-books is high enough, the publisher would even price the e-reader below its cost (as shown in Figure 2.3). Therefore, in response to a decrease in the e-reader price, as we demonstrated earlier, the publisher increases the e-book retail price.



**Figure 2.3** E-reader and e-book retail prices under two pricing models

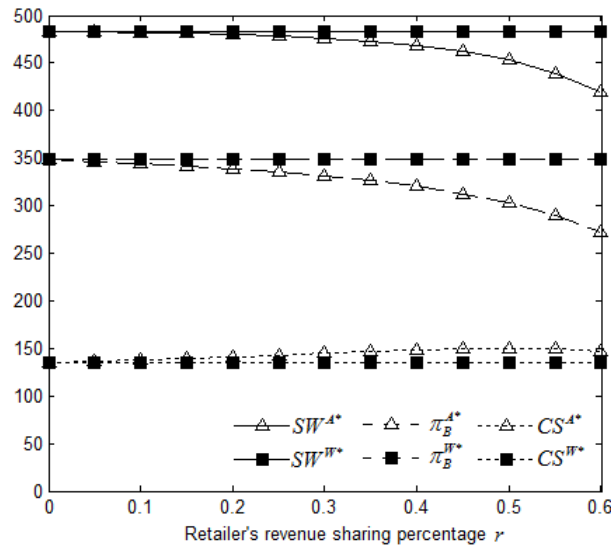
Moreover, when the retailer’s revenue sharing percentage  $r = 0$ , we have  $p_E^{W*} = p_E^{A*}$  and  $p_R^{W*} = p_R^{A*}$ .

This result implies that the wholesale model can be considered a special case of the agency model. In the

agency model, when  $r=0$ , the retailer cannot make any profit from e-books. It can only make profit from e-readers. Thus, the retailer falls back into the same situation as it does with wholesale model, although it may be not by its own volition but driven by an exogenously determined revenue sharing contract. Consequently, equilibrium outcomes turn out to be the same.

Next, we compare the social welfare effects of the two pricing models. Our numerical analysis uses the same parameter values as the ones in Figure 2.1. Denote the social welfare under the agency model by  $SW^{A*}$  and the social welfare under the wholesale model by  $SW^{W*}$ . Denote the total consumer surplus under the agency model by  $CS^{A*}$  and the total consumer surplus under the wholesale model by  $CS^{W*}$ . Define the business profit as the sum of the retailer's profit and the publisher's profit. Denote the total business profit in the agency model by  $\pi_B^{A*}$  and the total business profit in the wholesale model by  $\pi_B^{W*}$ . Figure 2.4 shows that the social welfare, which is the sum of total consumer surplus and total business profit, is lower in the agency model compared to that in the wholesale model, i.e.,  $SW^{A*} < SW^{W*}$ . The figure also shows that the total consumer surplus is higher in the agency model, i.e.,  $CS^{A*} > CS^{W*}$ , and the business profit is lower in the agency model, i.e.,  $\pi_B^{A*} < \pi_B^{W*}$ .

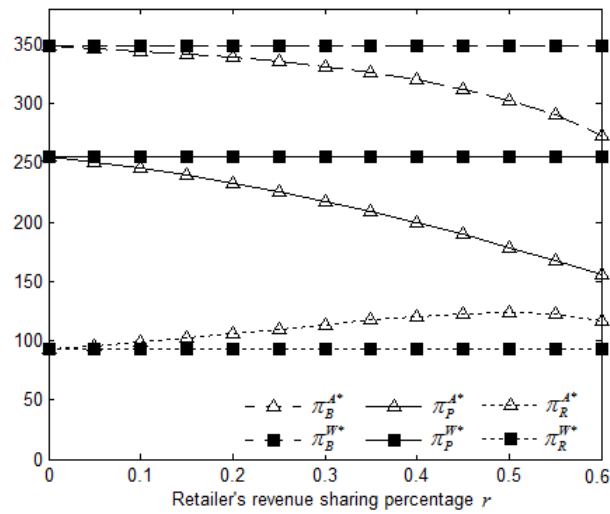
Total consumer surplus is higher in the agency model mainly because the e-reader price is significantly lower (Table 2.1). Although the e-book retail price is higher in the agency model, a significantly lower e-reader price makes consumers better off in terms of overall surplus compared to that in the wholesale model. Moreover, in addition to its influence on total consumer surplus, a lower e-reader price also increases the e-book market size (Table 2.1).



**Figure 2.4** Social welfare under two pricing models

**Table 2.1** Comparison between two pricing models

Variables	Wholesale model	Agency model
E-book retail price $p_E^*$	18.29	22.62
E-reader price $p_D^*$	342.45	185.57
E-book market size $s_E^*$	0.34	0.38
Single customer e-book purchase quantity $q_E^*$	31.71	27.38
Physical book market size $s_F^*$	0.66	0.62
Single customer physical book purchase quantity $q_F^*$	24.00	24.00
Publisher's profit $\pi_P^*$	255.58	217.79
Publisher's physical book profit $\pi_{PF}^*$	101.29	95.74
Publisher's e-book profit $\pi_{PE}^*$	154.30	122.05
Retailer's profit $\pi_R^*$	92.79	113.52
Retailer's e-book profit $\pi_{RE}^*$	0	69.99
Retailer's e-reader profit $\pi_{RD}^*$	92.79	43.53
Total consumer surplus $CS$	134.39	144.76
Total consumer surplus from e-book $CS_E$	31.45	42.95
Total consumer surplus from physical book $CS_F$	102.95	101.81



**Figure 2.5** Business profit under two pricing models

As shown in Figure 2.4, business profit is lower in the agency model. Figure 2.5 gives more details about it by illustrating its two components—retailer's profit and publisher's profit—in two pricing models. While the publisher's profit significantly decreases in the agency model, the retailer's profit increases moderately. It suggests that in equilibrium the publisher actually becomes worse off in the agency model although it has control of the retail price. Part of the reason for this counterintuitive result can be found in Table 2.1. The publisher's profit comes from e-books and physical books. As shown in Table 2.1, profits from both components are lower in the agency model. For physical books, the lower profit is due to the smaller physical book market size, which means that e-books erode the physical book market more seriously in the agency model. For e-books, although it's the market size increases, from 34% to 38%, single customer e-book purchase quantity  $q_E^*$  drops from about 31.71 to 27.38, due to the higher e-book retail price in the agency model. Therefore, for the publisher it is clear that switching to the agency model is not necessarily profitable even if the publisher gains the retail price control in the agency model. For the retailer, switching to the agency model can be profitable because the retailer can benefit from the larger profits from e-books by decreasing the e-reader's price, which eventually makes the overall profit higher than in the wholesale model.

## 2.5 Discussion

Our model generates interesting implications for pricing schemes of e-books and e-readers. It reveals that in the wholesale model, the retailer's optimal pricing scheme is to price e-books at the wholesale price set by the publisher and mark up e-readers. This explains why the retailer wants to charge a low e-book retail price in the wholesale model. The two-part tariff pricing structure essentially solves the traditional double marginalization problem in which the final retail price is too high and the overall demand is too low. The e-reader, which is required to consumer e-books, is the critical component in the wholesale model. It is essentially the vehicle for the retailer to extract consumer surplus generated from purchasing low price e-books.

Comparing the two pricing models, we find the following: (i) the e-book retail price is higher and the e-reader price is lower in the agency model; (ii) social welfare is lower in the agency model; (iii) the total consumer surplus is higher and the business profit is lower in the agency model; and (iv) the retailer's profit is higher while the publisher's profit is lower in the agency model.

Switching from the wholesale model to the agency model has the following implications. First, although the equilibrium e-book retail price is higher in the agency model, it does not lead to a larger physical book market size. On the contrary, the physical book market size is smaller in the equilibrium of the agency model because the equilibrium e-reader price is significantly lower than that in the wholesale model. Such a low equilibrium e-reader price eventually expands the e-book market size. Therefore, if the publisher initially wants to switch to the agency model, gain the e-book retail price control, and then try to protect its physical book market size by adjusting the e-book retail price, we show that in equilibrium, this is not feasible.

Second, based on our model it is clear that a publisher's decision to move from the wholesale model to the agency model is not optimal. Essentially, in the agency model, the publisher has reintroduced the double marginalization problem into the system. As a result, the e-book retail price is set above the supply chain optimal level (notice that the two-part tariff pricing structure in the wholesale model is essentially the supply chain optimal pricing structure and the business profit is lower in the agency pricing model).

Therefore, switching to the agency model and then setting a higher e-book retail price has been proven to be a wrong strategy in equilibrium. It does not help stop the trend toward more people adopting e-books because, in response to the publisher's strategy, the retailer lowers prices for the e-reader. Meanwhile, a higher-than-first-best price reduces single-customer e-book purchase quantity, which eventually leads to a lower publisher's e-book profit as well as a lower total profit.

It would be interesting to explore how a publisher remedies these issues if switch to the agency model is already done and irreversible. Given the low e-reader price that consumers may have already become accustomed to, it is difficult to go back to the wholesale model with a high e-reader price. Then, the publisher faces the problem that the equilibrium e-book retail price in the agency model is too high. However, just reducing the e-book retail price is not sustainable. Although a lower e-book retail price could increase social welfare, it will further lower the publisher's profit. Thus, a more sustainable path is to reduce both the price and the revenue sharing percentage for the retailer. As a high retailer's revenue sharing percentage distorts the e-book retail price, reducing the retailer's revenue sharing percentage gives the publisher incentives to reduce the e-book retail price. In order to cut back the retailer's revenue sharing percentage, which directly reduces the retailer's profit, the publisher has to find ways to compensate the retailer and help develop alternative revenue models. The publisher could first provide additional advanced interactive digital content to the retailer to accompany the original e-books. Then, the publisher encourages the retailer to innovate and develop more advanced, high-margin devices that can take advantage of such interactive digital content. The retailer may indeed have an incentive to do so, since it could use differentiation strategy and target these advanced devices to certain e-book customer segments. The proper e-reader differentiation strategy could eventually increase the retailer's overall profit and improve the publisher's situation.

## **2.6 Conclusions**

In this essay, we study two relevant pricing models that have been used in the marketplace—the wholesale model and the agency model. Our model setup has the following features. First, we consider that consumers buy e-readers for reading more than one e-book. Thus, the e-book retail price not only



affects the e-book market size, but also single-customer e-book purchase quantity. Second, we model the substitution effect between e-books and physical books. Third, we use a game theoretical approach to analyze pricing schemes for the publisher and retailer in both the wholesale and agency models. Our results have interesting and relevant implications for companies that sell e-books and e-readers.

We find that in the wholesale model, the optimal pricing scheme for the retailer is to use a two-part tariff pricing structure, i.e., pricing e-books at the wholesale price set by the publisher and marking up e-readers.

We find that in the SPNE, the physical book costs—for example, printing, distributing, and storing books—affect the e-book retail price under both the wholesale and agency models. In particular, if the physical book costs increase as a result of rising labor and material costs, the profit margin from selling one physical book drops. Then, the publisher is better off shifting the focus from high-cost physical books to e-books by the reducing e-book retail price. This result implies that if we expect the relative cost-benefit of e-books to continue and the profit margin of e-books to be sustainable in the long run, pricing e-books at the correct level to encourage their consumption is strategically important.

Our findings also suggest that the e-reader price is generally decreasing in e-book retail price. A lower e-book retail price in the wholesale model implies a higher e-reader price, while a higher e-book retail price is associated with a lower e-reader price. Our numerical results suggest that in the agency model, it is even possible for the retailer to price the e-reader lower than its marginal cost if it can make enough profit on selling e-books.

We compare the equilibrium prices and social welfare of the wholesale model and the agency model. In the agency model, we find that the equilibrium e-book retail price is higher while the equilibrium e-reader price is lower. Social welfare is lower in the agency model than it is in the wholesale model. Total consumer surplus is moderately higher and overall business profit is lower in the agency model. Our numerical analysis suggests that the agency model—where the equilibrium e-book retail price is higher and the retailer wants to set a significantly lower e-reader price—leads to a smaller physical book market size. This may be against the publisher's initial impetus for switching to the agency model, that is,

preventing the physical book market from being eroded by the e-book through regaining control of the e-book retail price and raising it. According to our analysis, it is indeed infeasible to realize that goal in equilibrium and, moreover, after switching to the agency model, the publisher is actually worse off. We discuss potential ways to improve the publisher's profit if the switch has been made and the agency model is fixed between the publisher and the retailer.

## **Essay 3 Conversion Rate Dynamics in Online Retail**

### **3.1 Introduction**

In the online environment, the purchase conversion rate is an important measure of the effectiveness and success of online stores. There has been increasing interest in studying online conversion rates using click-stream data (e.g., Sismeiro and Bucklin 2004, Moe and Fader 2004). While click-stream data are useful in studying consumer-level behavior for a specific online company, seller-level analysis of the conversion rate is important in understanding how online sellers' strategies and tactics affect the performance of online stores.

In this study, we used a proprietary data set from an online marketplace to conduct seller-level analysis of conversion rates in online retail. The panel data set had a large sample of 2,766 sellers, who we studied over a seven-month period. This unique data set allowed us to examine how sellers' characteristics, such as pricing and product strategies, marketing efforts, service responsiveness, reputation scores, and product quality ratings, affect sellers' conversion rates. Specifically, we addressed the following research questions:

- What are the seller-level factors that affect the conversion rate?
- Is there any hidden state that potentially governs the dynamics of the conversion rate? Is the relationship between the explanatory variables and the conversion rate state dependent?
- If the relationship is state dependent, what are the factors that determine the states?

The organization of the essay is as follows. In Section 3.2, we review prior literature, followed by theory development in Section 3.3. We present our empirical model in Section 3.4, and the descriptive statistics of data and variables in Section 3.5. In Section 3.6, we show the estimation results and discuss their implications. The essay concludes in Section 3.7.

### **3.2 Literature Review**

Prior studies have examined website characteristics and consumer-behavior-related conversion rates. Mandel and Johnson (2002) show that page design can affect purchase decisions. Moe and Fader (2004) develop an individual probability model for visit-to-purchase conversion. Based on previous-visit patterns

and purchases, the model predicts the subsequent visits that are likely to convert to purchases. Sismeiro and Bucklin (2004) develop and estimate a conversion rate model using click-stream data. The model predicts conversion rate by linking the purchase decision to what visitors do and what they are exposed to while visiting the seller's site.

At the seller level, Perdikaki et al. (2012) examine the effect of traffic on sales and conversion rates in brick-and-mortar stores. They break down sales volume into conversion rate and basket value and analyze the impact of traffic. They find that stores' sales volume exhibits diminishing returns to scale with respect to traffic.

Our study uses a panel data set to examine the dynamics of conversion rates across a large number of startup sellers. We believe the study of conversion rates at the seller level provides unique insights into how sellers' product and pricing strategies, marketing efforts, and service responsiveness, together with reputation scores and product quality ratings, affect conversion rates. The results provide important guidance regarding what attributes the seller needs to improve and how much the seller needs to improve them to become more likely to jump to a more favorable status.

### 3.3 Theory Development

In this section, we develop the theoretical foundation of the seller's conversion rate dynamics. Online sellers regularly monitor the status of their stores via a key set of variables, such as reputation scores. These variables are of great concern to sellers, because these variables highly influence consumers' *ex ante* perception of expected purchasing utility, which subsequently affects consumers' purchase decisions. In addition, sellers' characteristics and strategies, such as price, marketing efforts, and service responsiveness, also affect conversion rates. We call these variables—that could potentially affect the conversion rates—explanatory variables.

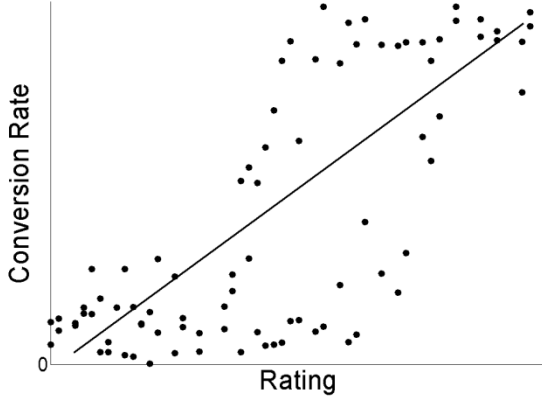
For sellers, it is important to understand how changes in the values of explanatory variables can affect conversion rates. These answers could provide important guidance for sellers to increase their conversion rates. In order to examine the relationship between conversion rates and the explanatory variables, a common method is to approximate it using one linear function, in which the conversion rate is

the dependent variable and the explanatory variables are the independent variables. We take a commonly believed explanatory variable, the seller's product quality rating, for our example. Figure 3.1 illustrates the approximated linear function. The  $x$ -axis of the figure represents the seller's product quality ratings and the  $y$ -axis represents the corresponding conversion rates. Dots represent the observations of "product quality rating/conversion rate" pairs from different sellers. The estimation result, represented by the upward straight line, indicates a positive correlation between the product quality rating and the conversion rate. The results suggest that sellers may be able to improve their conversion rates by improving product quality ratings.

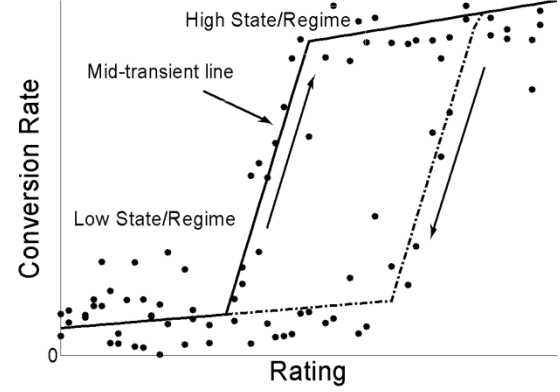
While a simple linear function can reveal the direction of the relationship, it neglects a significant amount of information if the consumer's purchase decision-making process follows a "regime-switching" framework. In a regime-switching model, consumers conceive of single or multiple thresholds when assessing the value of some of the explanatory variables. We call these explanatory variables "state variables."

The thresholds divide the value range of state variables into multiple regimes. In contrast to the single-regime case, where a single linear function governs the relationship between conversion rate and the explanatory variables over the explanatory variables' entire value range, the relationship between the conversion rate and the explanatory variables is different across different regimes in the regime-switching model, that is, the unit change in explanatory variables can have different effects on consumers' purchase decisions across different regimes. We take the product quality rating as an example of a state variable. When assessing the product quality rating, consumers usually conceive a minimum acceptable rating as the threshold for considering a purchase, such as 3 stars in a 5-star rating scale system. Thus, an increase in product quality rating from 1 star to 2 stars might not affect a consumer's purchase decision. However, once the rating crosses the threshold, perhaps 3 stars, the rating will start to have influence on the consumer's purchase decision. The conversion rate may have a sudden, positive jump from around 3 stars to the 3.5 stars area, because consumers now think the rating is "good enough" for them to make purchases. Further, when the rating is above 4.5 stars, the marginal benefit of a rating increase diminishes

because the rating is already in the consumers' comfort range. This stepwise phenomenon is illustrated in Figure 3.2, in which the observations are identical to Figure 3.1. Instead of a straight line, however, a stepwise line is employed for the estimation, which can help us capture a more accurate picture of the purchase decision process in reality. The single-regime model can be considered a special case of the regime-switching model.



**Figure 3.1** Single-regime model



**Figure 3.2** Regime-switching model

The focus of this study is to estimate the underlying regime switching. This is a nontrivial task, especially in the panel context. It is related to the following questions. First, what are the thresholds conceived by consumers? Second, in a given time period, which regime does a seller belong to? Is the seller in the high regime or the lower one? Third, what is the relationship between the conversion rate and the explanatory variables in each regime? We develop a hidden Markov model (HMM) to investigate these issues and estimate the regime-switching model. An HMM can be characterized by a combination of the following three components: (i) the state-dependent outcome probability distribution  $F$ , (ii) the state-transition probability matrix  $G$ , and (iii) the initial distribution  $\pi$ .

To apply the HMM, we first need to define what hidden states exist in our research context. In our study, threshold values are not known. They are random variables. Therefore, given a set of specific values for state variables, which regime a seller belongs to is an unobserved random variable. This unobserved random variable, which represents the seller's unobserved regime location, is the seller's hidden state. In Figure 3.2, for instance, there are potentially two regimes: one is depicted by the upper

right arm of the curve and the other is depicted by the lower left arm. Without knowing where the mid-transient line is located, we cannot conclude with certainty which arm a specific dot (observation) is associated with, that is, we do not know for sure the seller's hidden state. Once the transient line is given, we can consider the seller on the upper right side to be in the high business state and the seller on the lower left side to be in the low business state. The terms "low business state" and "high business state" come intuitively from the fact that the high state has a higher baseline conversion rate than the low state.

We want to mention that we do not model the mid-transient region as a state. Although our model can be flexible in generating more than two states, we hypothesize a two-state setting for simplicity. We assume the slope of the transient line is so steep that the mid-transient region is too narrow to have significant practical implications. We believe the most significant implication of the transient line is its position, not the inside-transient-region dynamics.

As the values of state variables change over time, a seller's regime location can change as well. Therefore, the seller's hidden state can also change over time. A seller may switch from the low state to the high state or vice versa, once the values of the state variables cross the corresponding threshold values. Due to the probabilistic nature of the threshold values, the state switching also operates in a probabilistic manner. The state-transition probabilities are loaded in the state-transition matrix  $G$  of the HMM. Since the switching is apparently caused by variation in the values of the state variables, the state-transition probabilities depend on the values of the state variables.

It is intriguing to postulate that threshold values themselves may be associated with the path taken by state variables. To return to the example of the product quality rating, if a rating of 2 starts in the low state, reaches 4.5 stars in the high state, and then starts to fall, the gain in conversion rate during the rating ascending process may already have been completely "consumed," that is, reverted back to the low state, when the rating drops from 4.5 stars to 3 stars. The economic underpinning of such a phenomenon comes from Prospect theory, which suggests that the individual perception of loss and gain depends not only on the absolute magnitude of the change in utility, but also the reference point of where she started from (Kahneman et al. 1979). This is illustrated in Figure 3.2 by the dashed line at the lower level of the curve.

The higher-level solid line represents the path taken by the rating as it is ascending. The lower-level dashed line represents the path taken by the rating as it is descending. While an increase from 2 stars to 4.5 stars would be considered a fair improvement, falling from 4.5 stars to 3 stars may have already been counted as an “unacceptable” bad signal that damages the conversion rate greatly, that is, puts the seller back in the low state. In the HMM, by modeling the transition probabilities as state dependent, we allow our model to finely capture this potential effect.

After defining the hidden states and demonstrating how they can switch back and forth, we turn our attention to the state-dependent outcome probability distribution  $F$ , which characterizes the relationship between the conversion rate and the explanatory variables given a specific state. It is important to emphasize that in our study, state variables are a subset of explanatory variables. The criterion for an explanatory variable to be incorporated into a state variable set is that it has structural influence—that is, it can effect state switching—on the conversion rate. However, other than its influence on state switching, given a specific state, a change in the value of a state variable may still influence the conversion rate, as it is the nature of an explanatory variable. Product rating is a good example because it can serve as both a state variable and an explanatory variable. Price can serve as an explanatory variable, but not a state variable. Price can influence the conversion rate in a specific given state, but it has no structural impact on the conversion rate.

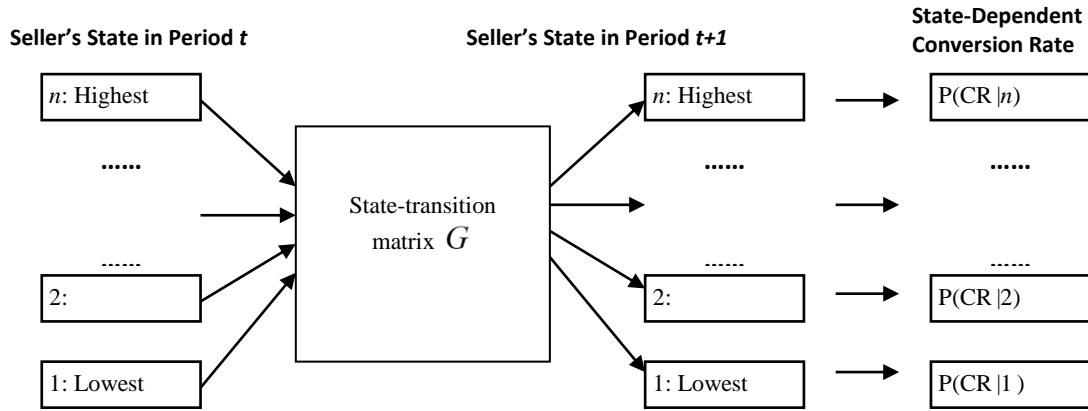
The estimation result of state-dependent outcome distribution  $F$  is able to reveal the answer for a long-time puzzle in the seller’s mind: for some seller, tweaking the explanatory variables, for example reducing the price, effectively boosts her conversion rate, while the same operation done by another seller in the similar context doesn’t have the similar effect, or even quite far from. Some sellers may feel regardless the operations they take they can hardly effectively increase the conversion rate, while looking at other sellers enjoy the high conversion rate seemingly by doing “nothing”. The explanation given by our theory is that those sellers are probably in different hidden states. Furthermore, by knowing which state she is most likely to be seated in at a given time period, the seller can learn from the transition



probability matrix  $G$  that what are the state variables she should improve to make herself switch to a more desirable state.

### 3.4 Empirical Model

In this section, we elaborate our HMM setup. The data sampling time window starts from  $t = 1$  and ends at  $t = \bar{T}$ , in total  $\bar{T}$  periods. All sellers open their online retail stores at time  $t = 1$ . A seller either terminates service at  $t = T_i < \bar{T}$ , or remains open until  $t = \bar{T}$ , in which case  $T_i = \bar{T}$ . We denote the state variables that determine seller  $i$ 's state transition at the end of  $t$  by  $R_{it}$ . We denote explanatory variables that affect seller  $i$ 's conversion rate at time  $t$  by  $O_{it}$ . Seller  $i$ 's unobserved business state at time  $t$  is denoted by  $s_{it}$ . In our model,  $s_{it} = \{1, 2\}$ , where 1 represents the “low business state” and 2 represents the “high business state.” Figure 3.3 illustrates the HMM model in our study.



**Figure 3.3** Hidden Markov model of seller's conversion rate

#### 3.4.1 State-Dependent Conversion Rates

In our model, constructing the state-dependent outcome probability distribution  $F$  is a task of finding the most appropriate specification to characterize the randomness of the conversion rate. Various reports from the mass media suggest conversion rates among online retail stores often possess an extreme under-dispersion property in that for most sellers the conversion rates are lower than 5%. Evidence from our data set supports this statement. These facts indicate that the conversion rate, although a continuous measure, can hardly be assumed to follow either normal or lognormal distribution. As it is arduous to find

a proper continuous distribution for the conversion rate, instead, we characterize it indirectly through the following discrete binomial distribution:

$$\Pr(c_{it} | u_{it}, s_{it}, O_{it}) = \binom{u_{it}}{c_{it}} (p_{it})^{c_{it}} (1 - p_{it})^{u_{it} - c_{it}}.$$

$u_{it}$  is the total number of unique visitors to seller  $i$ 's online store in time  $t$ .  $c_{it}$  is the total number of purchasers among  $u_{it}$ . Both  $u_{it}$  and  $c_{it}$  are observed in our data set. The conversion rate, according to its definition, is strictly generated by  $c_{it} / u_{it}$ . Therefore, the expected value of the conversion rate, which is conditional on the observed  $u_{it}$ ,  $E(c_{it} / u_{it} | u_{it})$ , is equal to  $p_{it}$ . The probability  $p_{it}$  is modeled by a logistic regression as follows:

$$\text{logit}(p_{it}) = \gamma_{s_{it}} O_{it} + \nu_i.$$

$\gamma_{s_{it}}$  is the state-dependent coefficient for  $O_{it}$ . Such state-dependent constructs enable us to estimate a different relationship between the conversion rate and  $O_{it}$  across different states.  $\nu_i$  is the random effect term, capturing the unobserved individual seller's heterogeneity at  $O_{it}$  level.

Our binomial distribution setup essentially results from a hypothetical construct of store visitors' purchase decision-making process. Define

$$U_{ijt} = \gamma_{s_{it}} O_{it} + \nu_i + \varepsilon_{ijt},$$

in which  $U_{ijt}$  represents visitor  $j$ 's net utility of purchasing a representative product in store  $i$  at time  $t$ .

If we assume  $\varepsilon_{ijt}$  to be logistic distribution in  $(0,1)$  and visitors make independent decisions after they observe  $O_{it}$ , then  $p_{it}$  represents the probability that a single visitor makes a purchase in seller  $i$ 's store at time  $t$ . Although this binomial distribution setup is parsimonious due to our aggregated construct of  $U_{ijt}$ , it provides a stronger economic explanation and a more precise statistical approximation than a normal distribution assumption for  $p_{it}$ .

### 3.4.2 State-Transition Probabilities

For the two-state HMM model, the state-transition matrix  $G$  is defined as

$$G = \begin{bmatrix} q_{11} & q_{12} \\ q_{21} & q_{22} \end{bmatrix}$$

where  $q_{j,k} = \Pr(s_{i,t+1} = k | s_{it} = j), 1 \leq t < T_i$ . For each state  $j$ ,  $\sum_k q_{jk} = 1$ . We assume that the state-transition probabilities are governed by an ordered logit model that is, we believe, a suitable choice in our context because (i) the high and low states have the very clear economic interpretation that the high state is more favorable than the low state in terms of conversion rate, and (ii) as stated in the theory development section, the threshold values are our primary interests. Therefore,

$$q_{j,j+1} = 1 - \frac{\exp(\bar{\mu}_j - \beta_j R_{it} - \xi_i)}{1 + \exp(\bar{\mu}_j - \beta_j R_{it} - \xi_i)},$$

$$q_{j,j-1} = \frac{\exp(\underline{\mu}_j - \beta_j R_{it} - \xi_i)}{1 + \exp(\underline{\mu}_j - \beta_j R_{it} - \xi_i)}, \text{ and}$$

$$q_{j,j} = 1 - q_{j,j-1} - q_{j,j+1}.$$

$\bar{\mu}_j$  represents the threshold value of transiting to the next higher state at a given state  $j$ .  $\underline{\mu}_j$  represents the threshold value of transiting to the next lower state given the current state  $j$ . Corresponding to our low/high, two-state context,  $\bar{\mu}_1$  is the threshold value of transiting from the low state to the high state, and  $\underline{\mu}_2$  is the threshold value of transiting from the high state to the low state.

$\beta_j$  denotes the state-dependent coefficients for  $R_{it}$  at a given state  $j$ . As we explained in the theory development section, the reversion of the path from low state to high state is not necessarily the same as the one from high state to low state. The state-dependent setting, however, enables us to estimate two paths separately.

$\xi_i$  is used to capture the seller's unobserved heterogeneity at the state-transition level. We assume that two levels' of unobserved heterogeneity,  $\nu_i$  and  $\xi_i$ , are potentially correlated. To model the

correlation, we assume that they follow a bivariate normal distribution  $N(0, \Sigma)$  of which the covariance structure is

$$\Sigma = \begin{bmatrix} \sigma_{\xi\xi}^{\xi\xi} & \sigma_{\xi v}^{\xi v} \\ \sigma_{v\xi}^{\xi v} & \sigma_{vv}^{\xi v} \end{bmatrix}.$$

It is worth mentioning that the state transition is assumed to happen at the end of a time period. Therefore the state variables at  $t$ ,  $R_{it}$ , together with the state  $s_{it}$  at time  $t$  determine the state probability distribution at time  $t+1$ . The state variables at  $t = T_i$  will not be used.

The last component of the HMM is the initial state distribution  $\pi$ . Since all sellers in our data set are newly born at  $t=1$ , we assume that all of them start from the low state.

### 3.4.3 Likelihood Function

In this subsection, we build the overall likelihood function based on the components we previously specified. Denote the sequence of seller  $i$ 's states during her entire lifespan by  $S_i = \{s_{it} | 1 \leq t \leq T_i\}$ , the sequence of numbers of purchasers for seller  $i$  during her entire lifespan by  $C_i = \{c_{it} | 1 \leq t \leq T_i\}$ , the sequence of numbers of visitors for seller  $i$ 's store by  $U_i = \{u_{it} | 1 \leq t \leq T_i\}$ , the sequence of values of explanatory variables by  $O_i = \{O_{it} | 1 \leq t \leq T_i\}$ , and the sequence of values of state variables by  $R_i = \{R_{it} | 1 \leq t \leq T_i\}$ . First, we derive the likelihood of  $C_i$  as

$$\Pr(C_i | U_i, S_i, O_i) = \prod_{i=1}^{T_i} \Pr(c_{it} | u_{it}, s_{it}, O_{it}).$$

Then, the likelihood of  $S_i$  can be written as

$$\Pr(S_i | R_i) = \Pr(s_{i1}) \prod_{t=2}^{T_i} \Pr(s_{it} | s_{i,t-1}, R_{i,t-1}) = \pi_i \prod_{t=2}^{T_i} q_{s_{i,t-1}, s_{i,t}}.$$

Next, summing up all the possible paths of state evolution that seller  $i$  could take and then taking the integral on the random effects, we derive the marginal distribution of  $C_i$  as

$$\Pr(C_i | U_i, O_i, R_i) = \int_{\xi} \int_{v} \left( \sum_{S_i} \Pr(C_i | U_i, S_i, O_i) \Pr(S_i | R_i) \right) \phi(0; \Sigma) dv d\xi.$$

Finally, by incorporating all of the sellers in the data set, we obtain the following overall likelihood function

$$L(\beta, \gamma, \mu) = \prod_{i=1}^N \Pr(C_i | U_i, O_i, R_i)$$

where  $N$  is the total number of sellers in our data set,  $\beta$  is the set of state variable coefficients at all states,  $\gamma$  is the set of explanatory variable coefficients at all states, and  $\mu$  is the set of threshold values including both upper thresholds  $\bar{\mu}$  and lower thresholds  $\underline{\mu}$  at all states.

### 3.5 Data and Variable Description

The data were collected for a seven-month period from May 2011 to November 2011. All sellers opened their businesses in May 2011. Sellers could be divided into two mutually exclusive groups—ones that sell clothing and ones that sell prepaid refill cards for mobile phones. Some sellers exited in the middle of the seven-month period. The average lifespan of sellers is about 3.7 months. Table 3.1 provides variable descriptions and statistics.

**Table 3.1** Variables and descriptive statistics

Variables	Descriptions	Mean	StDev	Min	Max
<i>buyeripv</i>	Average number of daily product page views per visitor	2.66	5.87	1	344
<i>goodsqualr</i>	Seller's overall product quality rating	1.32	2.18	0	5
<i>reputation</i>	Seller's reputation score	2.15	1.62	1	11
<i>imreplyratio</i>	Response ratio to visitors' inquiries	0.25	0.24	0	1
<i>mkttool</i>	Whether the seller uses promotion tools offered by marketplace owner	0.10	0.30	0	1
<i>productnmb</i>	Seller's total number of products	89.96	161.86	1	3,099
<i>avgprice</i>	Average product price	95.12	163.85	0.80	5,050
<i>ipvuv</i>	Number of unique visitors	149.55	1163.06	1	42,780
<i>puv</i>	Number of purchasers	5.16	39.10	0	1,592

*buyeripv* is an important measure that indicates the overall attractiveness of the sellers' goods portfolio. *buyeripv* does not count the multiple visits to the same product page made by the same visitor. From a seller's perspective, high *buyeripv* implies that the seller has formed a fine portfolio in which a number of products capture consumers' interest.

*goodsqualr* is a very common measure reflecting the degree of satisfaction consumers obtain from the product they purchased. It takes values from 0 to 5, in accord with the widely used 5-star rating scale system. In our study, the marketplace owner does not round the continuous average number to the nearest star or half star.

*reputation* is also a very common and important score widely used in the online marketplace. It indicates the overall credibility and service quality of the seller. The consumer has the right to rate the seller's reputation after making a purchase. The score can be +1, 0, or -1, which represent good, fair, or bad, respectively. *reputation* is accumulated over time. A seller's accumulated total score is categorized into 11 levels in which level 0 is the lowest and level 11 is the highest. All stores start at level 0.

*imreplyratio* is the seller's response ratio to visitors' inquiries. In the online retail environment, consumers usually contact the seller for more details about the goods they are interested in. They often use the in-marketplace instant messaging tool to send inquiries and expect the seller to respond to them in a timely manner. *imreplyratio* is defined as the ratio of the total number of visitors who receive responses from the seller to the total number of visitors who inquire.

*mkttool* is a dummy variable indicating if the seller uses the marketing promotion tools offered by the marketplace owner. These promotion tools are basically designed to bring more visitors to the seller's store. The marketplace owner charges the seller a considerable price for using these promotion tools.

*productnmb* reflects the seller's product variety. The sellers can adjust the number of products listed in the store over time. They may delist some unpopular ones and enlist some popular ones according to the sales outcomes.

*buyeripv*, *goodsqualr*, and *reputation* are selected as state variables  $R_{it}$ . We believe these variables have both long- and short-term effects on the conversion rate. *imreplyratio*, *mkttool*, *productnmb*, and *avgprice* are selected only as  $O_{it}$ .

According to the correlation table shown in Table 3.A1 in Appendix C, multicollinearity is not a serious issue in our study.

### 3.6 Estimation Results and Discussion

Prior to estimating the parameters  $(\beta, \gamma, \mu)$ , we noticed in Table 3.1 that our variables exhibited long-tail properties. One possible reason is that our two groups of sellers, one selling clothes and the other selling prepaid refill phone cards, may have very different value ranges for the explanatory variables. We also suspect that the conversion rate dynamics may be so different across the two groups that solely adding fixed-effect intercepts may not be enough to take care of the heterogeneity. Therefore, we separate the clothing and prepaid card sellers into two groups and generate estimation results separately.

The literature endorses the identification of the general Markov regime switching model when the  $R_{it}$  and  $O_{it}$  variables overlap (Kim 2008). Therefore, as HMM is a special case of general Markov regime switching, any likelihood-based estimation method is suitable for our problem. Given that our likelihood function has no closed-form expression, we chose to estimate our problem using the Bayesian method via Monte Carlo Markov chain (MCMC) sampling. Since we have little prior knowledge about the parameter ranges, we adopt uninformative priors for all of our parameters. We draw multiple sets of initial values for the sampling process and run each chain for 40,000 iterations. The first 20,000 iterations, which are considered a “burn-in period,” are not used for parameter inferences. The estimation results are listed in Table 3.2, which shows that most of the parameters are statistically significant.

As for the state transitions, we can see that higher *buyeripv*, *goodsqualr*, and *reputation* will in general increase the likelihood of switching from a low state to a high state. This result makes intuitive sense, in that the seller needs to incorporate more market-popular products (higher *buyeripv*), improve their product quality ratings (higher *goodsqualr*), and improve their reputation scores (higher *reputation*) to switch from the low state to the high one. For prepaid refill cards, however, the effect of *buyeripv* is not statistically significant in either the low state or high state. One potential reason is that the types of prepaid refill cards available on the market are quite fixed and limited. It is easier and more likely for a seller who sells prepaid refill cards to build a goods portfolio that includes market-popular products than it is for a seller who sells clothing.

**Table 3.2** Estimation results (clothing)

<b>State</b>		<b>Low</b>			<b>High</b>		
$\beta$		<b>Mean</b>		<b>StDev</b>	<b>Mean</b>		<b>StDev</b>
<i>buyeripv</i>	Average number of daily product page views per visitor	12.14	**	4.53	-9.94	**	5.80
<i>goodsqualr</i>	Overall product quality rating	1.57	**	0.57	1.72		1.27
<i>reputation</i>	Reputation score	2.01	*	1.07	1.61		2.40
$\mu$	Thresholds between states	5.12	**	0.40	1.78	**	0.94
$\gamma$							
<i>imreplyratio</i>	Response ratio to visitors' inquiries	-0.41	**	0.07	-5.06	**	0.22
<i>mkttool</i>	Whether the seller uses promotion tools offered by marketplace owner	0.14	**	0.03	-6.25	**	0.17
<i>productnmb</i>	Total number of products	-4.21	**	0.14	4.45	**	0.30
<i>avgprice</i>	Average product price	-7.10	**	0.60	-18.76	**	1.27
<i>buyeripv</i>	Average number of daily product page views per visitor	14.80	**	0.61	-0.65		0.54
<i>reputation</i>	Reputation score	0.89	**	0.08	-1.12	**	0.46
<i>goodsqualr</i>	Overall product quality rating	2.40	**	0.07	3.42	**	0.37
<i>constant</i>	Constant term	-5.40	**	0.06	-1.849	**	0.29

\*  $p < 0.1$ ; \*\*  $p < 0.05$ 

Estimation results (prepaid refill cards)

<b>State</b>		<b>Low</b>			<b>High</b>		
$\beta$		<b>Mean</b>		<b>StDev</b>	<b>Mean</b>		<b>StDev</b>
<i>buyeripv</i>	Average number of daily product page views per visitor	4.78		4.65	2.50		9.16
<i>goodsqualr</i>	Overall product quality rating	1.54	**	0.37	0.08		0.70
<i>reputation</i>	Reputation score	3.57	**	0.64	3.13	**	1.42
$\mu$	Thresholds between states	3.31	**	0.19	0.97	**	0.48
$\gamma$							
<i>imreplyratio</i>	Response ratio to visitors' inquiries	1.27	**	0.11	1.425	**	0.17
<i>mkttool</i>	Whether the seller uses promotion tools offered by marketplace owner	-0.07	*	0.04	1.531	**	0.05
<i>productnmb</i>	Total number of products	-3.05	**	0.13	-2.175	**	0.14
<i>Avgprice</i>	Average product price	-19.03	**	1.49	-8.50	**	1.12
<i>Buyeripv</i>	Average number of daily product page views per visitor	10.88	**	0.63	-0.52		0.42
<i>reputation</i>	Reputation score	3.22	**	0.10	-11.81	**	0.18
<i>goodsqualr</i>	Overall product quality rating	3.01	**	0.10	2.703	**	0.17
<i>Constant</i>	Constant term	-3.79	**	0.06	1.567	**	0.11

\*  $p < 0.1$ ; \*\*  $p < 0.05$



It is interesting to note that when a clothing seller is in the high state, a further increase in *buyeripv* tends to switch the seller back to the low state, that is, to have a structural drawdown on the conversion rate. One potential explanation is the “too many choices” problem. After customers browse a number of products they are interested in but are not vertically differentiated, they may find it quite hard to make a purchase decision. Some of them may even end up making no purchase, because they could not figure out which product to buy. The prepaid refill card sellers do not have the same issue, since prepaid refill cards are often tied to the specific wireless carrier the consumer uses, which makes it much easier for the consumer to make a decision.

The estimation result of  $\bar{\mu}_l$  reveals the threshold which corresponds to the fifty percentage chance transiting from the low state to the high state.

Some common sense is reflected in the estimation results relating to the state-dependent relationship between the conversion rate and the explanatory variables. For example, a lower average price is associated with a higher conversion rate; a higher product quality rating is associated with a higher conversion rate. Moreover, we show that higher *buyeripv* is associated with a higher conversion rate for low-state sellers. This result implies that lower-state sellers may be able to increase their conversion rates by building goods portfolios with more popular products.

The impact of marketing promotion tools on the conversion rate is not only state dependent but also product dependent. For clothing, using the promotion tools will increase the conversion rate for low-state sellers but decrease it for high-state sellers. For prepaid refill cards, the effect is the opposite: it will decrease the conversion rate for low-state sellers but increase it for high-state sellers. Whether the marketing promotion tools increase or decrease the conversion rate, we believe, depends upon the following two aspects: (i) what types of buyers the promotion tools bring to the seller; and (ii) whether the additional buyers brought by the promotion tools have a higher purchasing propensity after visiting the seller’s store, compared to the ones not brought in by the promotion tools. These two aspects explain why high-state clothing sellers could see a decrease in their conversion rates when using promotion tools.

High-state sellers' excellent reputation will attract many casual buyers with much lower intrinsic purchasing propensity than serious buyers to the store once the seller starts using the promotion tools. The overall percentage of serious buyers decreases so that the overall conversion rate decreases. For the low-state clothing sellers, a lack of an excellent reputation will bring fewer casual buyers but more serious buyers. Therefore, the overall conversion rate is likely to increase.

Unlike with clothing, purchasing the prepaid refill cards is often attached to a very specific need for the consumer, sometimes even quite urgent. In addition, the prepaid refill cards often have expiration dates, so consumers rarely shop casually for this product. Therefore, when those consumers with relatively urgent and specific needs are brought by the promotion tools to the store, after confirming the seller's high reputation and ratings, they are likely to make an impulsive purchase, which gives them a higher purchasing propensity than a buyer not brought in by the promotion tools. For low-state sellers, however, using the promotion tools can backfire, and actually decrease the conversion rate. This happens because consumers brought in by the fancy promotion tools may become disappointed after they see that the seller's overall performance record is lower than they expected.

We also found that the effect of *imreplyratio*, the response ratio to visitors' inquiries, is product dependent. Intuition would suggest that a higher *imreplyratio* leads to a higher conversion rate for sellers who sell prepaid refill cards. However, it is rather counterintuitive that a higher *imreplyratio* leads to a lower conversion rate for sellers who sell clothes. One possible reason is that in the online retail environment, clothing products are associated with a greater level of information asymmetry in terms of product characteristics compared to prepaid refill cards. The communication between sellers and customers through the instant messaging tools are often used to reduce such information asymmetry. Because the advertisements for products often highlight their desirable characteristics while hiding their undesirable ones, such dialogue may reveal some undesirable characteristics, drawing down customers' propensity for purchasing. Therefore, being more responsive on instant messaging may not always be good for increasing conversion rates.

There also appears to be an interesting pitfall for sellers in both categories when they are in the high state: a high reputation score can actually depresses the conversion rate. One possible reason is that when a seller's reputation score is very high, the seller's store will be ranked at the top of search results almost every time consumers input a query and ask to see the sellers with the best reputations first. This will bring a seller many casual visitors whose intrinsic purchasing propensities are generally lower than serious buyers'. However, if the seller is ranked at the upper middle of the first page of search results, the consumer who clicks that seller's link is more likely to be a serious buyer with a higher intrinsic purchase propensity.

### **3.7 Conclusions**

In this study, we constructed a hidden Markov model to study the conversion rate dynamics in online retail. We studied how the seller-level covariates affect the conversion rate dynamics and found that conversion rate dynamics are state dependent. We presumed two states for conversion rate dynamics—a low state and a high state—in which the high state represented a more favorable baseline conversion rate. The relationship between the conversion rate and the seller-level covariates differed across the two states. We estimated state-dependent relationships for both the high state and the low state and discussed the implications of the results. We also estimated the state-transition probability, which revealed how the seller-level covariates affected state transition. Our findings provide important guidance for online sellers of all statuses to improve their store conversion rates.

## REFERENCES

- Akerlof, G. A. 1970. The market for “lemons”: Quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, Vol. 84, No. 3, pp. 488-500.
- Armstrong, M. 1996. Multiproduct nonlinear pricing. *Econometrica*, Vol. 64, No. 1, pp. 51-76.
- Balasubramanian, S. 1998. Mail versus mall: A strategic analysis of competition between direct marketers and conventional retailers. *Marketing Science*, Vol. 17, No. 3, pp. 181-195.
- Berger, J., Harbring, C., Sliwka, D. 2010. Performance appraisals and the impact of forced distribution: An experimental investigation. *Management Science*, forthcoming
- Borman, W. C. 1977. Consistency of rating accuracy and rating errors in the judgment of human performance. *Organizational Behavior and Human Performance*, Vol. 20, No. 2, pp. 238-252.
- Bradley, T. 2011. Microsoft considers embracing Mac App Store. *PCWorld* (January 28).
- Chellappa, R. K., Kumar, R. 2005. Examining the role of “free” product-augmenting online services in pricing and customer retention strategies. *Journal of Management Information Systems*, Vol. 22, No. 1, pp. 355–377.
- Chen, Y., Xie, J. 2005. Third-party product review and firm marketing strategy. *Marketing Science*, Vol. 24, No. 2, pp. 218–240.
- Chen, Y., Xie, J. 2008. Online consumer review: Word-of-mouth as a new element of marketing communication mix. *Management Science*, Vol. 54, No. 3, pp. 477–491.
- Chevalier, J. A., Mayzlin, D. 2006. The effect of word of mouth on sales: online book reviews. *Journal of Marketing Research*. Vol. 43, No. 3, pp. 345–354.
- Chiang, W. K., Chhajed, D., Hess, J. D. 2003. Direct marketing, indirect profits: A strategic analysis of dual-channel supply chain design. *Management Science*, Vol. 49, No. 1, pp. 1-20.
- Choudhary, V. 2007. Comparison of software quality under perpetual licensing and software as a service. *Journal of Management Information Systems*, Vol. 24, No. 2, pp. 141–165.
- Choudhary, V. 2010. Use of pricing schemes for differentiating information goods. *Information Systems Research*, Vol. 21, No. 1, pp. 78-92.
- Church, J., Gandal N. 1996. Strategic entry deterrence: Complementary products as installed base. *European Journal of Political Economy*, Vol. 12, No. 2, pp. 331-354.
- Clemons, E. K., Gao, G., Hitt, L. M. 2006. When online reviews meet hyperdifferentiation: A study of the craft beer industry. *Journal of Management Information Systems*, Vol. 23, No. 2, pp. 149–171.
- Dellarocas, C. 2006. Strategic manipulation of internet opinion forums: Implications for consumers and firms. *Management Science*, Vol. 52, No. 10, pp. 1577–1593.
- Dellarocas, C., Farag, N. A., Zhang, X. 2004. Using online reviews as a proxy of word-of-mouth for motion picture revenue forecasting. *Working Paper*.

- Dellarocas, C., Narayan, R. 2006. A statistical measure of a population's propensity to engage in post-purchase online word-of-mouth. *Statistical Science*, Vol. 21, No. 2, pp. 277–285.
- Fan, M., Kumar, S., Whinston, A. B. 2007. Selling or advertising: Strategies for providing digital media online. *Journal of Management Information Systems*, Vol. 24, No. 3, pp. 143–166.
- Feng, Y., Guo, Z., Chiang, W. K. 2009. Optimal digital content distribution strategy in the presence of the consumer-to-consumer channel. *Journal of Management Information Systems*, Vol. 25, No. 4, pp. 241–270.
- Forman, C. 2008. Examining the relationship between reviews and sales: The role of reviewer identity disclosure in electronic markets. *Information Systems Research*, Vol. 19, No. 3, pp. 291–313.
- Godes, D., Mayzlin, D. 2004. Using online conversations to study word-of-mouth communication. *Marketing Science*, Vol. 23, No. 4, pp. 545–560.
- Hu, N., Liu, L., Zhang, J. 2008. Do online reviews affect product sales? The role of reviewer characteristics and temporal effects. *Information Technology and Management*, Vol. 9, No. 3, pp. 201–214.
- Hu, N., Pavlou, P. A., Zhang, J. 2006. Can online reviews reveal a product's true quality?: Empirical findings and analytical modeling of online word-of-mouth communication. *Proceedings of the 7th ACM Conference on Electronic Commerce*, pp. 324–330.
- Hu, N., Zhang, J., Pavlou, P. A. 2009. Overcoming the J-shaped distribution of product reviews. *Communications of the ACM*, Vol. 52, No. 10, pp. 144–147.
- Jiang, B., Chen, P. 2007. An economic analysis of online product reviews and ratings. *Working Paper*.
- Kahneman, D., Tversky, A. 1979. Prospect theory: An analysis of decision under risk. *Econometrica*, Vol. 47, No. 2, pp. 263–292.
- Kane, J. S., 1994. A Model of volitional rating behavior. *Human Resource Management Review*, Vol. 4, No. 3, pp. 283–310.
- Kane, J. S., Bernardin, H. J., Villanova, P., Peyrefitte, J. 1995. Stability of rater leniency: Three studies. *The Academy of Management Journal*, Vol. 38, No. 4, pp. 1036–1051.
- Kim, C. J., Piger, J., Startz, R. 2008. Estimation of Markov regime-switching regression models with endogenous switching. *Journal of Econometrics*, Vol. 143, No. 2, pp. 263–273.
- Kuksov, D., Xie, Y. 2010. Pricing, frills, and customer ratings. *Marketing Science*, Vol. 29, No. 5, pp. 925–943.
- Laffont, J., Rey, P., Tirole, J. 1998. Network competition: I. overview and nondiscriminatory pricing. *RAND Journal of Economics*, Vol. 29, No. 1, pp. 1–37.
- Lang, K. R., Vragov, R. 2005. A pricing mechanism for digital content distribution over computer networks. *Journal of Management Information Systems*, Vol. 22, No. 2, pp. 121–139.

- Lee, Y. J., Tan, Y., Hosanager, K. 2009. Do I follow my friends or the crowds? Examining informational cascades in online movie reviews. *Proceedings of the Nineteenth Workshop on Information Technologies and Systems (WITS 2009)*.
- Levy, S. 2007. The future of reading. *Newsweek* (November 17).
- Li, X., Hitt, L. M. 2008. Self-selection and information role of online product reviews. *Information Systems Research*, Vol. 19, No. 4, pp. 456–474.
- Li, X., Hitt, L. M. 2010. Price effects in online product reviews: An analytical model and empirical analysis. *MIS Quarterly*, Vol. 34, No. 4, pp. 809-831.
- Li, Y. 2010. Pricing digital content distribution over heterogeneous channels. *Decision Support Systems*, Vol. 50, No. 1, pp. 243-257.
- Liu, Y. 2006. Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing*, Vol. 70, No. 3, pp. 74–89.
- Mandel, N., Johnson, E. J. 2002. When web pages influence choice: Effects of visual primes on experts and novices. *Journal of Consumer Research*, Vol. 29, No. 2, pp. 235-245.
- Mantena, R., Sankaranarayanan, R., Viswanathan, S. 2010. Platform-based information goods: The economics of exclusivity more options. *Decision Support Systems*, Vol. 50, No. 1, pp. 79-92.
- Miller, C. C., Bosman, J. 2011. E-books outsell print books at Amazon. *The New York Times* (May 19).
- Moe, W., Fader, P. S. 2004. Dynamic conversion behavior at e-commerce sites. *Management Science*, Vol. 50, No. 3, pp. 326–335.
- Moe, W., Schweidel, D. A. 2011. Online product opinions: Incidence, evaluation and evolution. *Marketing Science*, forthcoming.
- Moe, W., Trusov, M. 2011. Measuring the value of social dynamics in online product quality ratings forums. *Journal of Marketing Research*, Vol. 48, No. 3, pp. 444-456.
- Mulhern, Francis J., Leone, Robert P. 1991. Implicit price bundling of retail products: A multiproduct approach to maximizing store profitability. *Journal of Marketing*, Vol. 55, No. 4, pp. 63-76.
- Perdikaki, O., Kesavan, S., Swaminathan, J. 2012. Effect of traffic on sales and conversion rates of retail stores. *Manufacturing & Service Operations Management*, Vol. 14, No. 1, pp. 145–162.
- Rey, P., Tirole, J. 1986. The logic of vertical restraints. *American Economic Review*, Vol. 76, No. 5, pp. 921-939.
- Rochet, J. C., Chone, P. 1998. Ironing, sweeping, and multidimensional screening. *Econometrica*, Vol. 66, No. 4, pp. 783-826.
- Saal, F., Downey, R. G., Lahey, M. A. 1980. Rating the ratings: Assessing the psychometric quality of rating data. *Psychological Bulletin*, Vol. 88, No. 2, pp. 413-428.
- Shapiro, C. 1985. Optimal pricing of experience goods. *The Bell Journal of Economics*, Vol. 14, No. 2, pp. 497-507.

- Sismeiro, C., Bucklin, R. 2004. Modeling purchase behavior at an e-commerce web site: A task-completion approach. *Journal of Marketing Research*, Vol. 41, No. 3, pp. 306-323.
- Spain, W. 2011. Apple App Store hits 10-billion-downloads mark. *MarketWatch.com* (January 22).
- Spence, J. R., Keeping, L. M. 2010. The impact of non-performance information on ratings of job performance: A policy-capturing approach. *Journal of Organizational Behavior*, Vol. 31, pp. 587-608.
- Spengler, J. 1950. Vertical integration and antitrust policy. *Journal of Political Economy*, Vol. 58, No. 4, pp. 347-352.
- Stone, B., Rich, M. 2009. Amazon in big push for new kindle model. *The New York Times* (February 10).
- Sun, M. 2012. How does variance of product quality ratings matter? *Management Science*, Vol. 58, No. 4, pp. 696-707.
- Sundararajan, A. 2004. Nonlinear pricing of information goods. *Management Science*, Vol. 50, No. 12, pp. 1660-1673.
- Tirole, J. 1988. The theory of industrial organization. The MIT Press.
- Trachtenberg, J. A. 2011a. Random house alters e-book pricing strategy. *The Wall Street Journal* (March 1).
- Trachtenberg, J. A. 2011b. E-book prices prop up print siblings. *The Wall Street Journal* (September 12).
- Whinston, M. D. 1990. Tying, foreclosure, and exclusion. *American Economic Review*, Vol. 80, No. 4, pp. 837-859.
- Yu, A., Hu, Y., Fan, M. 2011. Pricing strategies for tied digital contents and devices. *Decision Support Systems*, Vol. 51, No. 3, pp. 405-412.
- Yun, G. J., Donahue, L. M., Dudley, N. M., McFarland, L. A. 2005. Rater personality, rating format, and social context: Implications for performance appraisal ratings. *International Journal of Selection and Assessment*, Vol. 13, No. 2, pp. 97-107.
- Zeman, E. 2012. Apple ahead of Google with 25B App downloads. *InformationWeek* (March 5).
- Zhou, W., Duan, W. 2010. Online user reviews and professional reviews: A Bayesian approach to model mediation and moderation Effects. *ICIS 2010 Proceedings*, Paper 256.
- Zhu, F., Zhang, X. 2010. Impact of online consumer reviews on sales: The moderating role of product and consumer characteristics. *Journal of Marketing*, Vol. 74, No. 2, pp. 133-148.

## APPENDICES

### Appendix A: Additional propositions and proof of Essay 1

**Proposition 1.A1.** *Given the revenue sharing percentage  $s$ , developers' optimal choice of price  $p^*$  and quality level  $q^*$  are:*

i. *Region A1-1: (self-driven) when  $b > b_1^A$  and  $0 < h \leq h_1^A s$ ,*

$$p_1^{A*} = \frac{b(1-r_R)}{k_R(1-b)}, q_1^{A*} = \frac{(1-r_0)(1-b)k_R + bk_U(1-r_R)}{k_R k_U(1-b)};$$

ii. *Region A1-2: (platform owner-driven) when  $b > b_1^A$  and  $h_1^A s < h \leq h_2^A s$ ,*

$$p_2^{A*} = \frac{1}{2} \frac{b^2 k_U^2 s}{h(bk_U + k_R - bk_R)^2} - \frac{(r_R - r_0)b}{bk_U + k_R - bk_R}, q_2^{A*} = \frac{1}{2} \frac{bk_U s}{(bk_U + k_R - bk_R)h};$$

iii. *Region A1-3: (poor marketing) when  $b \leq b_1^A$  and  $0 < h \leq h_3^A s$ ,*

$$p_3^{A*} = \frac{b(1-r_R)}{k_R(1-b)}, q_3^{A*} = \frac{(1-r_0)(1-b)k_R + bk_U(1-r_R)}{k_R k_U(1-b)}.$$

*In all other regions of  $(b, h)$ , developers cannot make non-negative profit. The above thresholds are:*

$$b_1^A = \frac{(2r_R - 1 - r_0)k_R}{(2r_R - 1 - r_0)k_R + k_U(1-r_R)}, h_1^A = \frac{1}{2} \frac{bk_U^2 k_R(1-b)}{(bk_U + k_R - bk_R)((1-r_0)(1-b)k_R + bk_U(1-r_R))},$$

$$h_2^A = \frac{1}{4} \frac{bk_U^2}{(r_R - r_0)(bk_U + k_R - bk_R)}, \text{ and } h_3^A = \frac{k_U^2 bk_R(1-r_R)(1-b)}{((1-b)(1-r_0)k_R + bk_U(1-r_R))^2}.$$

**Proposition 1.A2.** *The platform owner's optimal choice of developers' revenue sharing percentage  $s^*$  is:*

i. *Region A2-1: (squeezing) when  $b > b_1^A$  and  $0 < h \leq h_{1\alpha}^A$ ,*

$$s_1^{A*} = \frac{2h(bk_U + k_R - bk_R)((1-r_0)(1-b)k_R + bk_U(1-r_R))}{bk_U^2 k_R(1-b)};$$

ii. *Region A2-2- $\alpha$ : (encouragement) when  $b > b_1^A$  and  $h_{1\alpha}^A < h \leq 2h_2^A / 3$ ,*



$$s_{2\alpha}^{A*} = \frac{1}{2} + \frac{h(r_R - r_0)(bk_U + k_R - bk_R)}{bk_U^2};$$

iii. Region A2-2- $\beta$ : (retention) when  $b > b_1^A$  and  $2h_2^A / 3 < h \leq h_2^A$ ,

$$s_{2\beta}^{A*} = \frac{4h(r_R - r_0)(bk_U + k_R - bk_R)}{bk_U^2};$$

iv. Region A2-3: when  $b \leq b_1^A$  and  $0 < h \leq h_3^A$ ,

$$s_3^{A*} = \frac{h((1-b)(1-r_0)k_R + bk_U(1-r_R))^2}{k_U^2 bk_R(1-r_R)(1-b)}.$$

The threshold:

$$h_{1\alpha}^A = \frac{1}{2} \frac{k_R bk_U^2 (1-b)}{(bk_U + k_R - bk_R)((2-r_R-r_0)(1-b)k_R + 2bk_U(1-r_R))}.$$

### Proof of Proposition 1.1

The developers' objective is

$$\begin{aligned} \max_{q,p} u_D &= ps - hq^2 \\ \text{s.t } E(U) &\geq 0 \end{aligned}$$

Since  $u_D$  is an increasing function of  $p$  and  $E(U)$  is a decreasing function of  $p$ , plus  $u_D$  is an decreasing function of  $q$  and  $E(U)$  is an increasing function of  $q$ , constraint  $E(U) \geq 0$  is always binding. The Lagrange function is

$$L = ps - hq^2 + \lambda \left( -(1-b)p + \frac{b(\min\{1, \max\{0, k(q-p) + r_0\}\} - r_R)}{k} \right).$$

Notice that if  $k(q-p) + r_0 < 0$ , condition  $E(U) = 0$  cannot be satisfied with any non-negative  $p$ . The Lagrange function has two possible cases.

$$\text{Case 1: } L_1 = ps - hq^2 + \lambda \left( -(1-b)p + \frac{b(1-r_R)}{k} \right) + \mu(1 - k(q-p) - r_0), \text{ when } k(q-p) + r_0 = 1.$$

Case 2:  $L_2 = ps - hq^2 + \lambda \left( -(1-b)p + \frac{b(k(q-p) + r_0 - r_R)}{k} \right)$ , when  $k(q-p) + r_0 < 1$ .

We start from Case 2. The first order conditions are:

$$\frac{\partial L_2}{\partial p} = 0, \frac{\partial L_2}{\partial q} = 0, \frac{\partial L_2}{\partial \lambda} = 0 .$$

Solve the first order conditions and then we find

$$p^* = p_2^* = \frac{1}{2} \frac{sb^2}{h} + \frac{b(-r_R + r_0)}{k}, q^* = q_2^* = \frac{1}{2} \frac{sb}{h}, \lambda = s .$$

The Lagrange multiplier  $\lambda = s > 0$ . It is not difficult to verify that the second order condition is satisfied.

Next, we verify the boundary condition  $k(q_2^* - p_2^*) + r_0 < 1$  which is

$$k \left( \frac{1}{2} \frac{sb}{h} - \frac{1}{2} \frac{sb^2}{h} - \frac{b(-r_R + r_0)}{k} \right) + r_0 - 1 < 0 .$$

Solve for  $h$  and then we find

$$h > h_1 s = \frac{1}{2} \frac{bks(1-b)}{br_0 - br_R - r_0 + 1} , \text{ where } h_1 = \frac{1}{2} \frac{bk(1-b)}{br_0 - br_R - r_0 + 1} .$$

It is worth noticing that  $br_0 - br_R - r_0 + 1 > 0$ . Short proof is given below. Define

$$f(b) = br_0 - br_R - r_0 + 1 .$$

Since  $\frac{\partial f}{\partial b} = r_0 - r_R < 0$  (recall that we assume  $r_0 < r_R$ ) and  $f(1) = 1 - r_R > 0$ , we have  $f(b) > 0$

Meanwhile, notice that price  $p$  has to be positive and utility  $u_D^*$  has to be non-negative. Therefore solve  $p^* > 0$  and  $u_D^* \geq 0$  and then we find

$$h \leq h_2 s = \frac{1}{4} \frac{bks}{r_R - r_0} , \text{ where } h_2 = \frac{1}{4} \frac{bk}{r_R - r_0} .$$

Notice that the conditions  $h > h_1 s$  and  $h \leq h_2 s$  need to hold simultaneously to make solution set of Case 2 non-empty. So we have

$$h_2 > h_1.$$

Solve  $h_2 > h_1$  and then we find

$$b > b_1 = 1 - \frac{1 - r_R}{r_R - r_0}.$$

Next, to derive the optimal  $p, q$  for the Case 1, we solve the first order conditions

$$\frac{\partial L_1}{\partial p} = 0, \frac{\partial L_1}{\partial q} = 0, \frac{\partial L_1}{\partial \lambda} = 0, \frac{\partial L_1}{\partial \mu} = 0.$$

We find

$$p^* = p_1^* = \frac{b(1 - r_R)}{k(1 - b)}, \quad q^* = q_1^* = \frac{br_0 - r_R b - r_0 + 1}{k(1 - b)}.$$

The Lagrange multipliers are

$$\lambda = -\frac{-ks + bks + 2h - 2hr_0 + 2bhr_0 - 2hbr_R}{k(1 - b)^2}, \mu = \frac{2h(br_0 - br_R - r_0 + 1)}{k^2(1 - b)}.$$

Since  $br_0 - br_R - r_0 + 1 > 0$ , we have  $\mu > 0$ . We need to secure  $\lambda > 0$ , which is

$$-ks + bks + 2h - 2hr_0 + 2bhr_0 - 2hbr_R < 0.$$

Solve for  $h$  and then we find  $h < \frac{h_1 s}{b}$ . Notice that this condition is satisfied when  $h \leq h_1 s$  is satisfied. It is

not difficult to verify that the second order condition is satisfied. Similarly, by solving  $u_D^* \geq 0$  in Case 1,

we find

$$h \leq h_3 s = \frac{b(1 - r_R)sk(1 - b)}{(br_0 - r_R b - r_0 + 1)^2}, \text{ where } h_3 = \frac{kb(1 - r_R)(1 - b)}{(br_0 - r_R b - r_0 + 1)^2}.$$

It is not difficult to show that  $h_3 > h_1$  when  $b > b_1$ , and  $h_3 \leq h_1$  when  $b \leq b_1$ .

We  $u_D^*$  in Case 1 by  $u_{D1}^*$  and denote the  $u_D^*$  in Case 2 by  $u_{D2}^*$ . We have the following:

$$u_{D1}^* = \frac{b(-1 + r_R)s}{k(-1 + b)} - \frac{h(1 - br_R - r_0 + br_0)^2}{k^2(-1 + b)^2},$$

$$u_{D2}^* = \frac{1}{4} \frac{bs(4hr_0 - 4hr_R + bks)}{hk}.$$

We find

$$u_{D1}^* - u_{D2}^* = -\frac{1}{4} \frac{(b^2ks + 2h - bks - 2hr_0 - 2hbr_R + 2bhr_0)^2}{k^2(1-b)^2h} \leq 0.$$

The above inequality shows that  $u_{D2}^*$  is always preferred when it is available. Therefore, when  $b > b_1$ , if  $h_2s \geq h > h_1s$ ,  $p_2^*$  and  $q_2^*$  are the optimal choices; if  $h \leq h_1s$ ,  $p_1^*$  and  $q_1^*$  are the optimal ones. When  $b \leq b_1$ , Case 2 becomes unavailable (recall that  $h_3 \leq h_1$  when  $b \leq b_1$ ). So we find that when  $b \leq b_1$  and  $h \leq h_3s$ , the optimal choices are  $p_3^*$  and  $q_3^*$  in the expressions shown in Proposition 1.1.

### Proof of Proposition 1.2

The platform owner's objective is

$$\max_s u_p = p(1-s).$$

We solve it by dividing it into the following cases.

When  $h \leq h_1s$  and  $b > b_1$ , the objective is

$$\begin{aligned} \max_s u_p &= \frac{b(1-r_R)}{k(1-b)}(1-s) \\ \text{s.t } h &\leq h_1s = \frac{1}{2} \frac{bks(1-b)}{br_0 - br_R - r_0 + 1} \end{aligned}$$

Solve it and we find that when  $h \leq h_1$ , the optimal  $s^*$  is

$$s^* = s_1^* = \frac{2h(1+br_0 - br_R - r_0)}{bk(1-b)}.$$

When  $h_2s \geq h > h_1s$  and  $b > b_1$ , the objective is

$$\begin{aligned} \max_s u_p &= \left( \frac{1}{2} \frac{sb^2}{h} + \frac{b(-r_R + r_0)}{k} \right) (1-s) \\ \text{s.t } \frac{1}{4} \frac{bks}{r_R - r_0} &\geq h > \frac{1}{2} \frac{bks(1-b)}{br_0 - br_R - r_0 + 1} \end{aligned}$$

Solve it and we find that when  $2h_2/3 \geq h > h_{1\alpha}$ , the optimal  $s^*$  is

$$s^* = s_{2\alpha}^* = \frac{1}{2} + \frac{h(r_R - r_0)}{bk}.$$

When  $h_2 \geq h > 2h_2/3$ , the optimal  $s^*$  is

$$s^* = s_{2\beta}^* = \frac{4h(r_R - r_0)}{bk}.$$

We denote the optimal  $u_p^*$  when  $h \leq h_{1s}$  by  $u_{p1}^*$ . We denote the optimal  $u_p^*$  when  $2h_2/3 \geq h > h_{1\alpha}$  by  $u_{p2\alpha}^*$ . We find

$$u_{p2\alpha}^* - u_{p1}^* = \frac{1}{8} \frac{(-2hr_R + 4h - 2hbr_R - bk + b^2k + 2hbr_0 - 2hr_0)^2}{k^2(1-b)^2 h} > 0.$$

This inequality means that the platform owner always prefers  $u_{p2\alpha}^*$  when it is available. It is not difficult to show  $h_{1\alpha} < h_1$ . Therefore, we find that  $u_{p2\alpha}^* > u_{p1}^*$  when  $h > h_{1\alpha}$ , and  $u_{p1}^* \geq u_{p2\alpha}^*$  when  $h \leq h_{1\alpha}$ .

When  $b \leq b_1$ , the objective function is

$$\begin{aligned} \max_s u_p &= \frac{b(1-r_R)}{k(1-b)}(1-s) \\ \text{s.t } h \leq h_3 &= \frac{b(1-r_R)sk(1-b)}{(br_0 - r_Rb - r_0 + 1)^2} \end{aligned}$$

Solve it and we find that when  $0 < h \leq h_3$ , the optimal  $s^*$  is

$$s^* = s_3^* = \frac{h(1+br_0 - br_R - r_0)^2}{b(1-r_R)k(1-b)}.$$

### Proof of Corollary 1.3.1 and Corollary 1.3.2

Results with respect to  $p_1^{A*}$  and  $p_3^{A*}$  are straightforward to prove. Therefore we focus on  $p_2^{A*}$ . For simplicity of notation, we denote  $p_2^{A*}$  by  $p^*$  in this subsection. We derive

$$\frac{\partial p^*}{\partial k_U} = \frac{\left( ((r_0 - r_R)h - k_U s)(b-1)k_R - bhk_U(r_0 - r_R) \right) b^2}{\left( (1-b)k_R + bk_U \right)^3 h} = \frac{f_1(h)b^2}{\left( (1-b)k_R + bk_U \right)^3 h}.$$

In order to prove  $\frac{\partial p^*}{\partial k_U} > 0$ , we just need to prove

$$f_1(h) = \left( (r_0 - r_R)h - k_U s \right) (b - 1) k_R - b h k_U (r_0 - r_R) > 0.$$

Since  $\frac{\partial f_1}{\partial h} = (r_R - r_0)(k_R - b k_R + b k_U) > 0$  and  $f_1(0) = k_R k_U s (1 - b) > 0$ , we prove  $\frac{\partial p^*}{\partial k_U} > 0$ .

With respect to  $k_R$  we derive

$$\frac{\partial p^*}{\partial k_R} = \frac{b \left( (k_R - k_U)(r_0 - r_R)h - k_U^2 s \right) b - h k_R (r_0 - r_R) (1 - b)}{\left( (k_R - k_U)b + k_R \right)^3 h}.$$

In order to prove  $\frac{\partial p^*}{\partial k_R} < 0$ , we just need to prove

$$\left( (k_R - k_U)(r_0 - r_R)h - k_U^2 s \right) b - h k_R (r_0 - r_R) < 0.$$

Solve for  $h$  and then we find  $h < \frac{b k_U^2 s}{(r_R - r_0)(k_R - b k_R + b k_U)}$ . Because in Region A1-2 we have

$$h < \frac{1}{4} \frac{b k_U^2 s}{(r_R - r_0)(k_R - b k_R + b k_U)} < \frac{b k_U^2 s}{(r_R - r_0)(k_R - b k_R + b k_U)},$$

we prove  $\frac{\partial p^*}{\partial k_R} < 0$ .

### Proof of Corollary 1.3.4

We denote  $s_1^{A*}$  by  $s^*$  in this subsection. With respect to  $k_U$  we derive

$$\frac{\partial s^*}{\partial k_U} = \frac{h \left( (4(1 - r_0)k_R + 2k_U(r_R - 2 + r_0))b - 4(1 - r_0)k_R \right)}{k_U^3 b} = \frac{h f_1(b)}{k_U^3 b}.$$

In order to prove  $\frac{\partial s^*}{\partial k_U} < 0$ , we just need to prove

$$f_1(b) = (4(1 - r_0)k_R + 2k_U(r_R - 2 + r_0))b - 4(1 - r_0)k_R < 0$$

We find  $f_1(0) = -4(1 - r_0)k_R < 0$  and  $f_1(1) = 2k_U(r_R - 2 + r_0) < 0$ . Therefore we prove  $\frac{\partial s^*}{\partial k_U} < 0$ .

With respect to  $k_R$ , we derive

$$\frac{\partial S^*}{\partial k_R} = \frac{h\left(\left(2(r_0-1)k_R^2 + 2k_U^2(1-r_R)\right)b^2 + 4k_R^2(1-r_0)b - 2k_R^2(1-r_0)\right)}{bk_U^2k_R^2(-1+b)}.$$

We define

$$f_2(b) = \left(2(r_0-1)k_R^2 + 2k_U^2(1-r_R)\right)b^2 + 4k_R^2(1-r_0)b - 2k_R^2(1-r_0).$$

Then we derive  $\frac{\partial^2 f_2}{\partial b^2} = 2(r_0-1)k_R^2 + 2k_U^2(1-r_R)$ . It is not difficult to prove  $\frac{\partial^2 f_2}{\partial b^2} < 0$  given our assumptions  $r_R > r_0, k_R > k_U$ . So  $f_2(b)$  is concave.

We find  $f_2(1) = k_U^2(1-r_R) > 0$  and  $f_2(0) = -k_R^2(1-r_0) < 0$ . Next, we derive

$$f_2(b_1^A) = \frac{k_R^2k_U^2(1-r_R)(r_R-r_0)(-r_0+4r_R-3)}{(-2k_Rr_R+k_R+k_Rr_0-k_U+k_Ur_R)^2}.$$

Notice that when  $-r_0+4r_R-3 \geq 0$ , which is  $r_R \geq \hat{r}_R = \frac{(r_0+3)}{4}$ , we have  $f_2(b_1^A) \geq 0$ . Recall  $f_2(1) > 0$ .

Therefore for any  $b$  in  $(b_1^A, 1)$ , we have  $f_2(b) > 0$ . Therefore we prove  $\frac{\partial S^*}{\partial k_R} < 0$  under this situation.

When  $-r_0+4r_R-3 < 0$ , which is  $r_R < \hat{r}_R = \frac{(r_0+3)}{4}$ , we have  $f_2(b_1^A) < 0$ . Recall that  $f_2(1) > 0$  and

$f_2$  is concave. There exists a unique  $\hat{b}$  in  $(b_1^A, 1)$  which makes  $f_2(b) = 0$ . Solve for  $\hat{b}$  and then we find

$$\hat{b} = \frac{k_R}{k_R + \sqrt{\frac{1-r_R}{1-r_0}}k_U}.$$

Therefore, when  $b < \hat{b}$ , we have  $f_2(b) < 0$  which leads to  $\frac{\partial S^*}{\partial k_R} > 0$ ; when  $b > \hat{b}$ , we have  $f_2(b) > 0$

which leads to  $\frac{\partial S^*}{\partial k_R} < 0$ .

Since  $k_R$  is our primary interest instead of  $b$ , we solve conditions for  $k_R$  from  $b < \hat{b}$ . From  $b < \hat{b}$ , we find

$$k_R > \hat{k}_R = \sqrt{\frac{1-r_R}{1-r_0}} \frac{bk_U}{(1-b)}.$$

From  $b \geq \hat{b}$ , we find

$$k_R \leq \hat{k}_R = \sqrt{\frac{1-r_R}{1-r_0}} \frac{bk_U}{(1-b)}.$$

### Proof of Corollary 1.3.5

For the squeezing region, we have

$$u_P^* = \frac{b(1-r_R)}{k_R(1-b)} \left( 1 - \frac{2h(bk_U + k_R - bk_R)((1-r_0)(1-b)k_R + bk_U(1-r_R))}{bk_U^2 k_R(1-b)} \right).$$

We derive

$$\frac{\partial u_P^*}{\partial k_U} = \frac{2(-bk_U r_R - bk_U r_0 + 2bk_U + 2k_R r_0 b - 2bk_R + 2k_R - 2k_R r_0)h(1-r_R)}{k_R k_U^3 (1-b)}.$$

In order to prove  $\frac{\partial u_P^*}{\partial k_U} > 0$ , we just need to prove

$$f_1(k_R) = -bk_U r_R - bk_U r_0 + 2bk_U + 2k_R r_0 b - 2bk_R + 2k_R - 2k_R r_0 > 0.$$

We find  $\frac{\partial f_1}{\partial k_R} = 2(1-r_0)(1-b) > 0$  and  $f_1(k_R)|_{k_R=k_U} = k_U(r_0 b - b r_R - 2r_0 + 2)$ . Next, we find

$$f_1(k_R)|_{k_R=k_U, b=1} = k_U(2-r_0-r_R) > 0 \text{ and } \frac{\partial f_1(k_R)}{\partial b}|_{k_R=k_U} = k_U(r_0 - r_R) < 0,$$

So we have  $f_1(k_R)|_{k_R=k_U} > 0$ . Recall that we assume  $k_R > k_U$ . Therefore we prove  $\frac{\partial u_P^*}{\partial k_U} > 0$ .

Next, we consider  $\frac{\partial u_P^*}{\partial k_R}$ . We derive



$$\frac{\partial u_P^*}{\partial k_R} = \frac{b(1-r_R)f_2(h)}{k_U k_R^3 (1-b)^2}$$

where

$$f_2(h) = (2bk_R r_0 - 4bk_R + 4bk_U - 4br_R k_U - 2k_R r_0 + 4k_R - 2k_R r_R + 2k_R br_R)h - k_U k_R + k_R bk_U.$$

We find  $f_2(0) = -k_U k_R (1-b) < 0$  and  $f_2(h_{1\alpha}^A) = -\frac{(1-b)^2 k_U k_R^2}{bk_U + k_R - bk_R} < 0$ . Thus we prove  $\frac{\partial u_P^*}{\partial k_R} < 0$ .

With respect to  $r_0$  we find

$$\frac{\partial u_P^*}{\partial r_0} = \frac{2(1-r_R)h(bk_U + k_R - bk_R)}{k_R(1-b)k_U^2} > 0.$$

With respect to  $r_R$  we derive

$$\frac{\partial u_P^*}{\partial r_R} = -\frac{f_3(h)}{k_R^2(1-b)^2 k_U^2}$$

where  $f_3(h)$  is a linear function of  $h$ . We find  $f_3(0) = bk_U^2 k_R (1-b) > 0$  and

$$f_3(h_{1\alpha}^A) = 2(1-r_R)(1-b)k_R(bk_U + k_R - bk_R)h_{41} > 0.$$

Therefore we prove  $\frac{\partial u_P^*}{\partial r_R} < 0$ .

Next, we consider the encouragement region. We derive

$$u_P^* = \frac{1}{2} \frac{\left( \left( (k_R - k_U)(-r_R + r_0)h - \frac{1}{2}k_U^2 \right) b - hk_R(-r_R + r_0) \right)^2}{k_U^2 \left( (-k_U + k_R)b - k_R \right)^2 h}.$$

We first consider  $\frac{\partial u_P^*}{\partial k_R}$ . We derive

$$\frac{\partial u_P^*}{\partial k_R} = \frac{1}{4} \frac{(2hk_R r_R - 2k_U bhr_0 - 2hk_R r_0 - 2hk_R br_R + 2hk_R r_0 b - bk_U^2 + 2hbk_U r_R)b(1-b)}{h(bk_U + k_R - bk_R)^3} = \frac{b(1-b)f_4(h)}{4h(bk_U + k_R - bk_R)^3}$$

In order to prove  $\frac{\partial u_P^*}{\partial k_R} < 0$ , we just need to prove

$$f_4(h) = 2hk_R r_R - 2k_U bhr_0 - 2hk_R r_0 - 2hk_R br_R + 2hk_R r_0 b - bk_U^2 + 2hbk_U r_R < 0.$$

Solve it and then we find

$$h < \frac{1}{2} \frac{bk_U^2}{(r_R - r_0)(bk_U + k_R - bk_R)} = 2h_2^A.$$

This condition is satisfied when  $h \leq \frac{2}{3} h_2^A$ . Therefore, we prove  $\frac{\partial u_P^*}{\partial k_R} < 0$ .

Then we consider  $\frac{\partial u_P^*}{\partial k_U}$ . We find

$$\frac{\partial u_P^*}{\partial k_U} = \frac{1}{4} \frac{f_4(h) f_5(h)}{k_U^3 h (k_U b + k_R - bk_R)^3}.$$

We already proved  $f_4(h) < 0$ .  $f_5(h)$  is a linear function of  $h$ . We find  $f_5(0) = -bk_U^2 k_R (1-b) < 0$  and

$$\frac{\partial f_5}{\partial h} = -2(bk_U + k_R - bk_R)^2 (r_R - r_0) < 0. \text{ So } f_5(h) < 0. \text{ Therefore we prove } \frac{\partial u_P^*}{\partial k_U} > 0.$$

Regarding  $\frac{\partial u_P^*}{\partial r_0}$ , we derive

$$\frac{\partial u_P^*}{\partial r_0} = -\frac{1}{2} \frac{f_4(h)}{(bk_U + k_R - bk_R) k_U^2}.$$

Since we know  $f_4(h) < 0$ , we prove  $\frac{\partial u_P^*}{\partial r_0} > 0$ .

Regarding  $\frac{\partial u_P^*}{\partial r_R}$ , we derive

$$\frac{\partial u_P^*}{\partial r_R} = \frac{1}{2} \frac{f_4(h)}{(bk_U + k_R - bk_R) k_U^2}.$$

Since we know  $f_4(h) < 0$ , we prove  $\frac{\partial u_P^*}{\partial r_R} < 0$ .

Next, we consider the retention region. We derive the platform owner's profit

$$u_P^* = -\frac{(-bk_U^2 + 4hbk_U r_R + 4hk_R r_R - 4hbk_R r_R - 4k_U bhr_0 - 4hk_R r_0 + 4hbk_R r_0)(-r_R + r_0)}{k_U^2(-bk_U - k_R + bk_R)}.$$

Regarding  $\frac{\partial u_P^*}{\partial k_R}$ , we find

$$\frac{\partial u_P^*}{\partial k_R} = -\frac{(r_R - r_0)b(1-b)}{(bk_U + k_R - bk_R)^2} < 0.$$

Regarding  $\frac{\partial u_P^*}{\partial k_U}$ , we derive

$$\frac{\partial u_P^*}{\partial k_U} = -\frac{(r_R - r_0)f_5(h)}{k_U^3(k_U b + k_R - bk_R)^2}.$$

Regarding  $f_5(h)$ , we find

$$f_5\left(\frac{2}{3}h_2^A\right) = -\frac{1}{3}k_U^2 b(4k_R - 4bk_R + bk_U) < 0 \text{ and } \frac{\partial f_5}{\partial h} = -8(bk_U + k_R - bk_R)^2(r_R - r_0) < 0.$$

Therefore  $f_5(h) < 0$ . Therefore, we prove  $\frac{\partial u_P^*}{\partial k_U} > 0$ .

Regarding  $\frac{\partial u_P^*}{\partial r_0}$ , we derive

$$\frac{\partial u_P^*}{\partial r_0} = \frac{f_6(h)}{k_U^2(bk_U + k_R - bk_R)}$$

where

$$f_6(h) = 8hbk_U r_R - 8k_U bhr_0 + 8hk_R r_R - 8hk_R r_0 - 8hbk_R r_R + 8hbk_R r_0 - bk_U^2.$$

In order to prove  $\frac{\partial u_P^*}{\partial r_0} > 0$ , we just need to prove  $f_6(h) > 0$ , which yields

$$h > \frac{1}{8} \frac{bk_U^2}{(r_R - r_0)(bk_U + k_R - bk_R)} = \frac{1}{2} h_2^A.$$

Since we have  $h > \frac{2}{3} h_2^A$  in the retention region, we find  $f_6(h) > 0$  and hence  $\frac{\partial u_P^*}{\partial r_0} > 0$ .

Regarding  $\frac{\partial u_P^*}{\partial r_R}$ , we derive

$$\frac{\partial u_P^*}{\partial r_R} = -\frac{f_6(h)}{k_U^2(bk_U + k_R - bk_R)}.$$

Since  $f_6(h) > 0$ , we prove  $\frac{\partial u_P^*}{\partial r_R} < 0$

Summarizing the above, we find that for all the squeezing, encouragement and retention region:

$$\frac{\partial u_P^*}{\partial k_U} > 0, \frac{\partial u_P^*}{\partial k_R} < 0, \frac{\partial u_P^*}{\partial r_0} > 0 \text{ and } \frac{\partial u_P^*}{\partial r_R} < 0.$$

### Proof of Corollary 1.4.1 and Corollary 1.4.2

The social welfare in the encouragement region is

$$W = \frac{1}{4} \frac{b^2 k_U s (2k_R - 2bk_R + 2bk_U - k_U s)}{h(k_R - bk_R + bk_U)^2}$$

where

$$s = \frac{1}{2} + \frac{h(r_R - r_0)(bk_U + k_R - bk_R)}{bk_U^2}.$$

First we derive

$$\frac{\partial W}{\partial r_R} = -\frac{1}{4} \frac{2k_U(k_R - k_U)b^2 + \left( \left( (2r_0 - 2r_R)h - 2k_U \right) k_R - 2k_U \left( -\frac{1}{2}k_U + h(r_0 - r_R) \right) \right) b - 2hk_R(r_0 - r_R)}{k_U^2(k_R - bk_R + bk_U)}.$$

In order to prove  $\frac{\partial W}{\partial r_R} > 0$ , we just need to prove

$$f_1(k_R) = 2k_U(k_R - k_U)b^2 + \left( \left( (2r_0 - 2r_R)h - 2k_U \right) k_R - 2k_U \left( -\frac{1}{2}k_U + h(r_0 - r_R) \right) \right) b - 2hk_R(r_0 - r_R) < 0.$$

Notice that  $f_1$  is a linear function of  $k_R$ . Recall  $k_R > k_U$ . So we just need to prove  $\frac{\partial f_1}{\partial k_R} < 0$  and

$f_1(k_R)|_{k_R=k_U} < 0$ . Solve the conditions  $\frac{\partial f_1}{\partial k_R} < 0$  and  $f_1(k_R)|_{k_R=k_U} < 0$ . We find that when  $h < \frac{bk_U}{r_R - r_0}$ , both

conditions are satisfied.

Since we have  $h \leq \frac{2}{3}h_2^A$  in the encouragement region, we just need to prove  $\frac{bk_U}{r_R - r_0} > \frac{2}{3}h_2^A$ . It is

equivalent to prove  $-3k_R + 3bk_R - 3bk_U + k_U < 0$ . Define  $f_2(b) = -3k_R + 3bk_R - 3bk_U + k_U$ . We find

$\frac{\partial f_2}{\partial b} = 3k_R - 3k_U > 0$  and  $f_2(1) = -2k_U < 0$ . Thus we prove that  $h < \frac{bk_U}{r_R - r_0}$  is always satisfied in the

encouragement region. Therefore we prove  $\frac{\partial W}{\partial r_R} > 0$ .

Second, we derive

$$\frac{\partial W}{\partial r_0} = \frac{1}{4} \frac{f_1(k_R)}{(k_R - bk_R + bk_U)k_U^2}.$$

Since we already proved that  $f_1(k_R) < 0$ , we prove  $\frac{\partial W}{\partial r_0} < 0$ .

Third, we derive

$$\frac{\partial W}{\partial k_R} = \frac{1}{8} \frac{b(1-b)f_1(k_R)}{h(bk_U + k_R - bk_R)^3}.$$

Since we proved  $f_1(k_R) < 0$ , we prove  $\frac{\partial W}{\partial k_R} < 0$ .

Fourth we derive

$$\frac{\partial W}{\partial k_U} = \frac{1}{8} \frac{-f_1(k_R)f_3(h)}{hk_U^3(bk_U + k_R - bk_R)^3}$$

where  $f_3(h) = -2(bk_U - k_Rb + k_R)^2(r_R - r_0)h + bk_Rk_U^2(1-b)$ .

The sign of  $\frac{\partial W}{\partial k_U}$  is the same as  $f_3(h)$ . When  $h < h_{2\alpha}^A = \frac{1}{2} \frac{k_R b k_U^2 (1-b)}{(-b k_U + k_R b - k_R)^2 (r_R - r_0)}$ , we find

$f_3(h) > 0$  and hence  $\frac{\partial W}{\partial k_U} > 0$ . When  $h > h_{2\alpha}^A$ , we find  $f_3(h) < 0$  and hence  $\frac{\partial W}{\partial k_U} < 0$ .

When  $h > \frac{2}{3} h_2^A$ , the social welfare in the retention region is

$$W = \frac{1}{4} \frac{b^2 k_U s (2k_R - 2b k_R + 2b k_U - k_U s)}{h (k_R - b k_R + b k_U)^2}$$

where  $s = \frac{4h(r_R - r_0)(b k_U + k_R - b k_R)}{b k_U^2}$ . For the above  $W = \frac{1}{4} \frac{b^2 k_U s (2k_R - 2b k_R + 2b k_U - k_U s)}{h (k_R - b k_R + b k_U)^2}$ , it is not

difficult to prove  $\frac{\partial W}{\partial r_R} > 0$ ,  $\frac{\partial W}{\partial r_0} < 0$ ,  $\frac{\partial W}{\partial k_U} < 0$  and  $\frac{\partial W}{\partial k_R} = 0$ .

### Proof of Corollary 1.4.3

The social welfare in the squeezing region is

$$W = \frac{b((1-r_0)(1-b)k_R + b k_U(1-r_R))}{k_R k_U (1-b)} - \frac{h((1-r_0)(1-b)k_R + b k_U(1-r_R))^2}{k_R^2 k_U^2 (1-b)^2}.$$

We define

$$f_1(h) = 2h b k_U (1-r_R) - (b k_U - 2h(1-r_0))(1-b)k_R$$

We derive

$$\frac{\partial W}{\partial r_0} = \frac{f_1(h)}{k_R (1-b) k_U^2}, \quad \frac{\partial W}{\partial r_R} = \frac{b f_1(h)}{k_U k_R^2 (1-b)^2}, \quad \frac{\partial W}{\partial k_R} = \frac{b(1-r_R) f_1(h)}{k_U k_R^3 (1-b)^2}, \quad \frac{\partial W}{\partial k_U} = \frac{(1-r_0) f_1(h)}{k_R (1-b) k_U^3}$$

Therefore, we just need to prove  $f_1(h) < 0$ . Notice that  $f_1(h)$  is a linear function of  $h$  and

$f_1(0) = -b k_U (1-b) k_R < 0$ . Then we just need to prove  $f_1(h_{1\alpha}^A) < 0$ . Doing a little algebra we find that it

is equivalent to prove

$$f_2(k_R) = -(1-b)^2 (2-r_R-r_0) k_R^2 + k_U (1-b) (r_0 b - 4b + 3r_R b + 1-r_0) k_R + b k_U^2 (1-r_R) (1-2b) < 0.$$

We find  $f_2(k_R)|_{k_R=k_U} = -k_U^2(1-r_R) < 0$ . Recall  $k_R > k_U$ . Therefore we just need to prove  $\frac{\partial f_2}{\partial k_R}|_{k_R=k_U} < 0$ .

We derive  $\frac{\partial f_2}{\partial k_R}|_{k_R=k_U} = -k_U(1-b)(r_0b - r_Rb + 3 - r_0 - 2r_R)$ . It is not difficult to prove that when  $b \in (0,1)$

we have  $r_0b - r_Rb + 3 - r_0 - 2r_R > 0$ . Therefore  $\frac{\partial f_2}{\partial k_R}|_{k_R=k_U} < 0$ . Therefore we have  $f_2(k_R) < 0$ . Therefore we

have  $f_1(h) < 0$ . Therefore Corollary 1.4.3 is proved.

### Proof of Proposition 1.3

When  $b > b_1$ , the platform owner's total revenue from developers in  $(0, h_1s]$  is

$$u_{P1} = \int_0^{h_1s} \frac{b(1-r_R)}{k(1-b)}(1-s)dh;$$

the platform owner's total revenue from developers in  $(h_1s, h_2s]$  is

$$u_{P2} = \int_{h_1s}^{h_2s} \left( \frac{1}{2} \frac{sb^2}{h} + \frac{b(-r_R + r_0)}{k} \right) (1-s)dh.$$

Thus, the total revenue is

$$u_P = u_{P1} + u_{P2} = b^2s(1-s) \left( \frac{1}{4} + \frac{1}{2} \ln \left( \frac{1}{2} \frac{br_0 - br_R - r_0 + 1}{(r_R - r_0)(1-b)} \right) \right).$$

We find the optimal  $s^* = \frac{1}{2}$ .

When  $b \leq b_1$ , the platform owner's total revenue from developers in  $(0, h_3s]$  is

$$u_P = \int_0^{h_3s} \frac{b(1-r_R)}{k(1-b)}(1-s)dh = \frac{b^2(1-r_R)^2}{(br_0 - r_Rb - r_0 + 1)^2} s(1-s).$$

We find the optimal  $s^* = \frac{1}{2}$ . Therefore the overall optimal  $s^* = \frac{1}{2}$ .

### Proof of Proposition 1.4

In this screening game, the uninformed party, the platform owner, moves first. The unobserved information is  $h$ . The platform owner's task is to choose a set of parameters  $\{s_L, s_H, \hat{r}\}$  to achieve a separating equilibrium. The criterion is to maximize the platform owner's profit while at the same time satisfying the developers' individual rationality (IR) constraints and incentive compatibility (IC) constraints.

Given  $0 < h_L \leq h_{1\alpha}$  and  $h_{1\alpha} < h_H \leq \frac{2}{3}h_2$ , from Proposition 1.2 we derive the following platform owner's optimal revenue sharing percentages for developer  $h_L$  and  $h_H$  respectively.

$$s_L = \frac{2h_L(1 + br_0 - br_R - r_0)}{bk(1-b)}$$

$$s_H = \frac{1}{2} + \frac{h_H(r_R - r_0)}{bk}$$

It is not difficult to prove that  $s_H > s_L$  and  $s_H$  and  $s_L$  maximize the platform's profit. Now the question is whether we can find a corresponding  $\hat{r}$  to guarantee both the IR and IC constraints. It is not difficult to prove that given  $h_L$  and  $h_H$  being in the scopes specified in the proposition, IR constraints are automatically satisfied if  $s_L, s_H$  follow the above expressions. We focus on finding a  $\hat{r}$  which satisfies IC constraints.

First we consider type  $h_L$  developer's choice. Define

$$f_L(r, s) = \frac{b(r - r_R)s}{k(1-b)} - \frac{h_L(r + br_0 - br_R - r_0)^2}{k^2(1-b)^2}$$

where  $r$  is app's rating from high valuation consumers group. Proposition 1.1 and 1.2 show that if type  $h_L$  developer chooses  $s_L$ , the developer's optimal profit is  $f_L(1, s_L)$ .

If type  $h_L$  developer chooses  $s_H$ , we can prove that her optimal choice of rating is  $r^* = \hat{r}$ . Hence the optimal profit is  $f_L(\hat{r}, s_H)$ . Short proof is given below.



We show  $f_L(r, s)$  is increasing in  $r$  in  $[r_R, 1]$ . We derive

$$\frac{\partial f_L}{\partial r} = -\frac{-bks + b^2ks + 2h_L r + 2h_L r_0 b - 2h_L b r_R - 2h_L r_0}{k^2(-1+b)^2}.$$

We just need to prove  $-bks + b^2ks + 2h_L r + 2h_L r_0 b - 2h_L b r_R - 2h_L r_0 < 0$  which yields

$$h_L < \frac{1}{2} \frac{bks(1-b)}{br_0 - br_R - r_0 + r}.$$

Notice that we have  $h_L \leq h_1 s = \frac{1}{2} \frac{bks(1-b)}{br_0 - br_R - r_0 + 1}$ . Thus  $\frac{\partial f_L}{\partial r} > 0$ .

Since  $\hat{r}$  is the maximum allowed rating for choosing  $s_H$ , we prove  $r^* = \hat{r}$ . Therefore the IC constraint for  $h_L$  developer is  $f_L(1, s_L) \geq f_L(\hat{r}, s_H)$ .

Next, we verify that the  $\hat{r}$  we proposed,

$$\hat{r} = r_0 - \frac{(r_R - r_0)(1+b)}{2} - \frac{1}{4} \frac{bk(1-b)}{h_H},$$

satisfies  $f_L(1, s_L) > f_L(\hat{r}, s_H)$ . Define

$$F(h_H, h_L) = 16h_H^2 k^2 (1-b)^2 (f_L(1, s_L) - f_L(\hat{r}, s_H))$$

We want to prove that for any feasible  $(h_H, h_L)$ , condition  $F(h_H, h_L) > 0$  holds. First, we prove  $\frac{\partial F}{\partial h_L} > 0$ .

We derive

$$\begin{aligned} \frac{\partial F}{\partial h_L} = & (16 + 24r_0 r_R + 56r_R^2 b + 20r_R^2 b^2 + 56r_0^2 b + 20r_0^2 b^2 - 12r_0^2 - 112r_0 b r_R - 40r_0 b^2 r_R + 4r_R^2 - 32r_R) h_H^2 + \\ & (8b^2 k r_R - 8b^2 k r_0 - 12b^3 k r_R + 12b^3 k r_0 + 4b k r_R - 4b k r_0) h_H - 2b^3 k^2 + b^2 k^2 + b^4 k^2 \end{aligned}$$

Notice that  $\frac{\partial F}{\partial h_L}$  is a quadratic function of  $h_H$ . Define  $f_1(h_H) = \frac{\partial F}{\partial h_L}$ . We derive

$$f_1'(b) = \frac{\partial^2 f_1}{\partial h_H^2} = 20(r_0 - r_R)^2 b^2 + 56(r_0 - r_R)^2 b + 16 + 24r_0 r_R - 12r_0^2 + 4r_R^2 - 32r_R.$$

We find  $f'(b)|_{b=b_1} = 4(4r_0 - 7r_R + 3)^2 \geq 0$ . Since  $b > b_1$  and  $\frac{\partial f'}{\partial b} = (40b + 56)(r_0 - r_R)^2 > 0$ , we prove

$\frac{\partial^2 f_1}{\partial h_H^2} > 0$ . We also find  $\frac{\partial f_1}{\partial h_H}|_{h_H=0} = 4bk(3b+1)(1-b)(r_R - r_0) > 0$  and  $f_1(0) = b^2 k^2 (1-b)^2 > 0$ . Thus we

prove  $f_1(h_H) > 0$  when  $h_H > 0$ . Therefore we prove  $\frac{\partial F}{\partial h_L} > 0$ .

Second, we prove  $\frac{\partial F}{\partial h_H} > 0$ . Similarly we treat  $\frac{\partial F}{\partial h_H}$  as a quadratic function of  $h_H$  and define

$f_2(h_H) = \frac{\partial F}{\partial h_H}$ . We find  $\frac{\partial^2 f_2}{\partial h_H^2} = 24(r_0 - r_R)^2(b+3)(1-b) > 0$ . Next we derive

$$\begin{aligned} \frac{\partial f_2}{\partial h_H} = & (32 + 112r_0^2b - 80r_Rb^2r_0 + 112r_R^2b + 48r_Rr_0 + 40r_0^2b^2 + 40r_R^2b^2 - 24r_0^2 - 64r_R - 224r_Rbr_0 + 8r_R^2)h_L + \\ & 32bkr_R - 32b^2kr_R + 32b^2kr_0 - 32bkr_0 \end{aligned}$$

Notice that  $\frac{\partial f_2}{\partial h_H}$  is a function of  $h_L$ . Define  $f_3(h_L) = \frac{\partial f_2}{\partial h_H}$ . We find

$$f_3(0) = 32bk(r_R - r_0)(1-b) > 0,$$

$$\frac{\partial f_3}{\partial h_L} = 40(r_0 - r_R)^2b^2 + 112(r_0 - r_R)^2b + 32 + 8r_R^2 - 24r_0^2 + 48r_Rr_0 - 64r_R.$$

Follow the similar procedure as  $\frac{\partial^2 f_1}{\partial h_H^2}$ , we find  $\frac{\partial f_3}{\partial h_L} > 0$  when  $b > b_1$ . Thus we prove  $\frac{\partial f_2}{\partial h_H} > 0$ .

Combined with  $\frac{\partial^2 f_2}{\partial h_H^2} > 0$ , the last condition we need to prove is  $f_2(0) > 0$ . We derive

$$f_2(0) = 2kb(b-1)(kb^2 - 6h_Lbr_R + 6h_Lbr_0 - kb + 2h_Lr_0 - 2h_Lr_R).$$

We just need to prove  $f_4(h_L) = kb^2 - 6h_Lbr_R + 6h_Lbr_0 - kb + 2h_Lr_0 - 2h_Lr_R < 0$ . Notice that we find

$f_4(0) = kb(b-1) < 0$  and  $\frac{\partial f_4}{\partial h_L} = 2(r_0 - r_R)(3b+1) < 0$ . So we have  $f_4(h_L) < 0$ . Therefore  $f_2(0) > 0$ .

Summarizing the above we prove  $\frac{\partial F}{\partial h_H} > 0$ .

So far we have proved  $\frac{\partial F}{\partial h_H} > 0$  and  $\frac{\partial F}{\partial h_L} > 0$ . Thus, we derive

$$\inf \{F(h_H, h_L)\} = F(h_{1\alpha}, 0).$$

It is not difficult to verify the second order condition. We find

$$F(h_{1\alpha}, 0) = \frac{4b^3k^3(1-b)^3(1+r_R-2r_0)(1+br_0-br_R-r_0)}{(2+br_0-br_R-r_0-r_R)^3} > 0.$$

Therefore we prove  $f_L(1, s_L) > f_L(\hat{r}, s_H)$ .

The last condition we need to secure the equilibrium is to show  $\hat{r}$  does not bind the type  $h_H$  developer's optimal rating choice so that the IC constraint for the type  $h_H$  developer is satisfied. When

$s_H = \frac{1}{2} + \frac{h_H(r_R - r_0)}{bk}$ , the type  $h_H$  developer would choose the platform owner-driven region. It is not

difficult to prove that the corresponding optimal  $r^*$  is no greater than  $\hat{r}$ . Therefore the IC constraint is satisfied.

## Appendix B: Proof of propositions in Essay 2

### Proof of Proposition 2.1

Denote the retailer's total profit by  $\pi_P$ . We solve the first order conditions. From  $\frac{\partial \pi_P}{\partial p_D} = 0$ , we derive

$$\frac{1}{4} \frac{-2bp_E + p_E^2 - 2p_Dm + 2bp_F - p_F^2 + 2tm}{tm} - \frac{1}{2} \frac{\frac{(p_E - w_E)(b - p_E)}{m} + p_D - c_D}{t} = 0.$$

Solve for  $p_D$  and we have

$$p_D = \frac{1}{4} \frac{-4bp_E + 3p_E^2 + 2bp_F - p_F^2 + 2tm + 2w_Eb - 2w_Ep_E + 2c_Dm}{m}.$$

Substitute it into  $\frac{\partial \pi_P}{\partial p_E} = 0$  and then we derive

$$-\frac{1}{8} \frac{(p_E - w_E)(-p_E^2 + 2bp_F - p_F^2 + 2tm - 2w_Eb + 2w_Ep_E - 2c_Dm)}{m^2t} = 0$$

So we obtain

$$p_E = w_E \text{ or } p_E = w_E \pm \sqrt{w_E^2 + 2bp_F - p_F^2 + 2tm - 2w_E b - 2c_D m} \text{ (if feasible).}$$

Substitute  $p_E = w_E \pm \sqrt{w_E^2 + 2bp_F - p_F^2 + 2tm - 2w_E b - 2c_D m}$  into  $\pi_p$  and then we find

$$(i) \text{ when } p_E = w_E - \sqrt{w_E^2 + 2bp_F - p_F^2 + 2tm - 2w_E b - 2c_D m}$$

$$\pi_p = -\frac{1}{2} \frac{\left( (w_E - b) \sqrt{A + 2m(t - c_R)} - A + (c_D + p_D - 2t)m \right)^2}{m^2 t} \leq 0$$

where  $A = w_E^2 - p_F^2 + 2b(p_F - w_E)$  ;

$$(ii) \text{ when } p_E = w_E + \sqrt{w_E^2 + 2bp_F - p_F^2 + 2tm - 2w_E b - 2c_D m}$$

$$\pi_p = -\frac{1}{2} \frac{\left( (w_E - b) \sqrt{A + 2m(t - c_D)} + A + (2t - c_D - p_D)m \right)^2}{m^2 t} \leq 0$$

Both of them are non-positive. So  $p_E = w_E \pm \sqrt{w_E^2 + 2bp_F - p_F^2 + 2tm - 2w_E b - 2c_D m}$  cannot be the global maximum solutions if  $p_E = w_E$  and the corresponding  $p_D$  lead to positive  $\pi_p$ . Substitute  $p_E^* = w_E$  and the corresponding  $p_D^*$  into  $\pi_p$  and then we find

$$\pi_p = \frac{1}{32} \frac{(A + 2m(t - c_R))^2}{m^2 t} \geq 0.$$

Finally we compare  $p_E^* = w_E$  with the boundary situation  $p_E^* = 0$ . We find

$$\pi_p \Big|_{p_E^* = w_E} - \pi_p \Big|_{p_E^* = 0} = w_E^2 \geq 0.$$

So we have the optimal choice of e-book retail price  $p_E^*$  and e-reader price  $p_D^*$  shown in Proposition 2.1.

The condition  $(t - c_D)/(2t) + (p_F - w_E)(2b - w_E - p_F)/(4tm) > 0$  comes from the second order condition.

We denote the 2x2 Hessian matrix by

$$H = \begin{bmatrix} \frac{\partial^2 \pi_p}{\partial p_E^2} & \frac{\partial^2 \pi_p}{\partial p_E \partial p_D} \\ \frac{\partial^2 \pi_p}{\partial p_D \partial p_E} & \frac{\partial^2 \pi_p}{\partial p_D^2} \end{bmatrix}.$$

Evaluate  $H$  at optimal point  $(p_E^*, p_D^*)$ . We find

$$H_{11} = \frac{\partial^2 \pi_p}{\partial p_E^2} = -\frac{1 - 18w_E b + 9w_E^2 + 2bp_F - p_F^2 + 2tm - 2c_D m + 8b^2}{m^2 t},$$

$$|H| = \frac{1}{8} \frac{-2w_E b + w_E^2 + 2bp_F - p_F^2 + 2tm - 2c_D m}{m^2 t^2}.$$

Doing a little algebra for  $|H| > 0$ , we derive the precondition

$$(t - c_D) / (2t) + (p_F - w_E)(2b - w_E - p_F) / (4tm) > 0.$$

Meanwhile, it is not difficult to prove that when the above precondition is satisfied, condition  $H_{11} < 0$  holds.

Such precondition has a direct economic interpretation. Consider the e-book market size  $s_E(p_E, p_D)$  as a function of  $p_E$  and  $p_D$ . The precondition is essentially  $s_E(w_E, c_D) > 0$ . It indicates that the e-book market size is greater than zero when the e-reader's price  $p_D$  is set to its cost  $c_D$  and the e-book retail price  $p_E$  is set to the publisher's wholesale price  $w_E$ . Intuitively this condition holds in real business practice.

### Proof of Proposition 2.2

We denote the publisher's total profit under the retailer's optimal choice  $(p_E^*, p_D^*)$  by  $\hat{\pi}_p$ . The publisher's optimal choice on e-book's wholesale price  $w_E^*$  is derived by solving the first order condition  $\partial \hat{\pi}_p / \partial w_E = 0$ . Since  $\hat{\pi}_p$  is a quartic function of  $w_E^*$  we don't have closed form solution. We use the implicit function method to help us find out the directional results regarding  $w_E^*$ .

First we derive

$$\frac{\partial^2 \hat{\pi}_p}{\partial w_E^2} = -\frac{1}{2m} - \frac{1}{t} \left( \frac{1}{4} \frac{-2p_F^2 - 9w_E b + 6w_E^2 + 3bp_F - c_F b - 2c_D m + 2c_A b + c_A p_F + c_F p_F + 2b^2 - 3c_A w_E}{m^2} \right).$$

To have  $\partial^2 \hat{\pi}_p / \partial w_E^2 < 0$ , we need to have  $t$  to be sufficiently large:

$$t > c_D - \frac{1}{2} \frac{-2p_F^2 - 9w_E b + 6w_E^2 + 3bp_F - c_F b + 2c_A b + c_A p_F + c_F p_F + 2b^2 - 3c_A w_E}{m}.$$

Assume that  $t$  is sufficiently large such that  $\partial^2 \hat{\pi}_p / \partial w_E^2 < 0$ . According to implicit function method, we derive

$$\frac{dw_E^*}{dc_F} = - \left( \frac{\partial^2 \hat{\pi}_p}{\partial w_E \partial c_F} \right) / \left( \frac{\partial^2 \hat{\pi}_p}{\partial w_E^2} \right).$$

So  $dw_E^* / dc_F$  has the same sign as  $\partial^2 \hat{\pi}_p / (\partial w_E \partial c_F)$ . We find

$$\frac{\partial^2 \hat{\pi}_p}{\partial w_E \partial c_F} = -\frac{1}{4} \frac{(b - p_F)(b - w_E)}{m^2 t} < 0.$$

Therefore, we prove  $dw_E^* / dc_F < 0$ .

For the optimal e-reader's price, we know from Proposition 2.1

$$p_D^* = (t + c_D) / 2 + (p_F - w_E^*)(2b - p_F - w_E^*) / (4m).$$

Since  $dp_D^* / dc_F = (\partial p_D^* / \partial w_E^*) / (dw_E^* / dc_F) + \partial p_D^* / \partial c_F$ , we find

$$\frac{\partial p_D^*}{\partial w_E^*} = -\frac{b - w_E}{2m} < 0, \quad \frac{\partial p_D^*}{\partial c_F} = 0.$$

Therefore we prove  $dp_D^* / dc_F > 0$ .

### Proof of Proposition 2.3

Define  $f_1(w_E) = \partial \hat{\pi}_p / \partial w_E$  which is a cubic function of  $w_E$ . Denote  $\bar{w}_E = (b + c_A) / 2$ . We derive

$$f_1(c_A) = -\frac{1}{8} \frac{(b - c_A)(4c_A b - 4bp_F + 2c_F b - 2tm - 2c_A p_F + 2c_D m + 3p_F^2 - c_A^2 - 2c_F p_F)}{m^2 t}$$

$$f_1(\bar{w}_E) = -\frac{1}{32} \frac{(b - c_A)f_2(c_A)}{m^2 t}$$

where  $f_2(c_A) = c_A^2 + 2(b - 2p_F)c_A + (b - 2p_F)^2 + 4c_F(b - p_F)$  which is a function of  $c_A$ . Notice that  $f_1$  is a cubic function of  $w_E$ . Plus, it is not difficult to prove that  $f_1(c_A) > 0$  when the precondition  $(t - c_D)/(2t) + (p_F - w_E)(2b - w_E - p_F)/(4tm) > 0$  is satisfied. Therefore we only need to prove  $f_1(\bar{w}_E) < 0$  and verify the second order condition. Condition  $f_1(\bar{w}_E) < 0$  requires  $f_2(c_A) > 0$ . We find that

$$\min_{c_A} f_2(c_A) \text{ is equal to } 4c_F(b - p_F) > 0 \text{ at } c_A = 2p_F - b.$$

Second order condition can also be verified. Therefore, we prove  $c_A < w_E^* < \bar{w}_E = (b + c_A)/2$ .

Using the same approach of Proposition 2.2, we find that  $\partial w_E^* / \partial p_F$  has the same sign as  $\partial^2 \hat{\pi}_P / (\partial w_E \partial p_F)$ . We derive

$$\frac{\partial^2 \hat{\pi}_P}{\partial w_E \partial p_F} = -\frac{1}{4} \frac{3b - 4p_F + c_A + c_F}{m^2 t} \cdot w_E + \frac{1}{4} \frac{2b^2 - 3bp_F + 2c_A b + bc_F - c_A p_F}{m^2 t}$$

which is a linear function of  $w_E$ . Define  $f_3(w_E) = \partial^2 \hat{\pi}_P / (\partial w_E \partial p_F)$ . Recall that we already proved  $w_E^* < \bar{w}_E = (b + c_A)/2$ . Therefore, solve  $f_3(0) > 0$  and  $f_3(\bar{w}_E) > 0$ . We find that the condition for  $f_3(w_E) > 0$  in  $(0, \bar{w}_E)$  is  $c_F \leq p_F \leq p_{th1} = (b + c_A + c_F)/2$ . Next we solve  $f_3(0) < 0$  and  $f_3(\bar{w}_E) < 0$  and we find that the condition for  $f_3(w_E) < 0$  is  $b > p_F \geq p_{th2} = b(2b + 2c_A + c_F)/(3b + c_A)$ .

#### Proof of Proposition 2.4

Recall  $\pi_R = (rp_E q_E + (p_D - c_D))s_E$  where

$$s_E = (t - p_D)/(2t) + (p_F - p_E)(2b - p_E - p_D)/(4tm).$$

Solve  $\partial \pi_R / \partial p_D = 0$  and then we find

$$p_D^* = (t + c_D)/2 + (p_F - p_E)(2b - p_E - p_F)/(4m) - rp_E(b - p_E)/(2m).$$

### Proof of Proposition 2.5

The proof of Proposition 2.5 is similar to Proposition 2.3. For simplicity of notation, in this subsection all notations are the ones in the agency model. For example,  $\hat{\pi}_p$  represents the publisher's profit in the agency model. Define  $f_4(p_E) = \partial \hat{\pi}_p / \partial p_E$  which is a cubic function of  $p_E$ . We derive

$$f_4(c_A) = \frac{1}{8} \frac{(b(1-r) + 2c_A r - c_A)A}{m^2 t}$$

where

$$A = 2m(t - c_D) - (4c_A b r - 2c_F b + 4b p_F - 4c_A b - 4c_A^2 r + 2c_A p_F - 3p_F^2 + 2c_F p_F + c_A^2).$$

Similar to Proposition 2.3, when precondition  $(t - c_D)/(2t) + (p_F - w_E)(2b - w_E - p_F)/(4tm) > 0$  holds, we have  $A > 0$ . Also it is not difficult to prove  $b(1-r) + 2c_A r - c_A > 0$ . Therefore we have  $f_4(c_A) > 0$ .

Since  $f_4(p_E)$  is a cubic function of  $p_E$ , we only need to prove  $f_4(\bar{p}_E) < 0$  and verify the second order condition. Define  $\bar{p}_E = (b + c_A/(1-r))/2$ . We derive

$$f_4(\bar{p}_E) = \frac{1}{32} \frac{f_5(r)f_6(r)}{m^2(1-r)^2 t}$$

where

$$f_5(r) = (c_A + r b - b - 2c_A r),$$

$$f_6(c_A) = (1-r)^2 b^2 + 2(c_A - 2p_F + 2c_F)(1-r)b + 4(1-r)p_F^2 - 4(c_A + c_F)(1-r)p_F + c_A^2.$$

We find that  $f_5(0) < 0$  and  $f_5(1) < 0$ . So we have  $f_5(r) < 0$ . We can also prove that  $\min_{c_A} f_6$  is equal to  $4(1-r)(c_F - p_F r)(b - p_F)$ . Therefore, when  $c_F - p_F r > 0$  which is  $r < c_F / p_F$ , we have  $f_6(c_A) > 0$ . Given both  $f_5(r) < 0$  and  $f_6(c_A) > 0$ , we prove  $f_4(\bar{p}_E) < 0$ . Recall  $f_4(p_E) = \partial \hat{\pi}_p / \partial p_E$ . The second order condition is not difficult to verify. Combine  $f_4(c_A) > 0$ ,  $f_4(\bar{p}_E) < 0$  and  $\partial^2 \hat{\pi}_p / \partial p_E^2 < 0$ , we prove that  $c_A < p_E^* < \bar{p}_E$  when  $r < c_F / p_F$ .



### Proof of Proposition 2.6 and Proposition 2.7

The procedure of proving Proposition 2.6 is similar to Proposition 2.2. For simplicity of notation, in this section of proof all notations are the ones in the agency model, for example,  $\hat{\pi}_p$  represents the publisher's profit in the agency model. We derive

$$\frac{\partial^2 \hat{\pi}_p}{\partial p_E \partial c_F} = \frac{1}{4} \frac{(b - p_F)(p_E - 2p_E r + rb - b)}{m^2 t}.$$

It is not hard to prove that  $(p_E - 2p_E r + rb - b) < 0$  when  $0 < r < 1$ . Similar to Proposition 2.2, when  $t$  is sufficiently large, we have  $\partial^2 \hat{\pi}_p / \partial w_E^2 < 0$ . Therefore  $\partial p_E^* / \partial c_F < 0$ .

The proof of Proposition 2.7 is very similar to Proposition 2.3. Define  $f_7(p_E) = \partial^2 \hat{\pi}_p / (\partial p_E \partial p_F)$ . Recall that in Proposition 2.5 we proved that  $p_E^* < \bar{p}_E = (b + c_A / (1 - r)) / 2$ . Then we solve  $f_7(0) > 0$  and  $f_7(\bar{p}_E) > 0$ . We find the condition for  $f_7(p_E) > 0$  in  $(0, \bar{p}_E)$  is  $c_F \leq p_F \leq p_{th1} = (b + c_A + c_F) / 2$ .

Then we solve  $f_7(0) < 0$  and  $f_7(\bar{p}_E) < 0$ . We find the condition for  $f_7(p_E) < 0$  is  $b > p_F \geq \tilde{p}_{th2} = b(c_A + (1 - r)(2b + c_A + c_F)) / (c_A + 3(1 - r)b)$ .

### Proof of Proposition 2.8

In this subsection of proof, we denote the publisher's profit in the agency model's SPNE by  $\hat{\pi}_p^A$ . We denote the publisher's profit in the wholesale model's SPNE by  $\hat{\pi}_p^W$ . Align the two first order conditions

$$\begin{cases} \partial \hat{\pi}_p^A / \partial p_E = 0 \\ \partial \hat{\pi}_p^W / \partial w_E = 0 \end{cases}$$

We notice that there is a common term in both equations,  $m(t - c_D) + p_F(b - p_F / 2)$ . We define

$$A = m(t - c_D) + p_F(b - p_F / 2).$$

We eliminate the common  $A$  from both equations and then obtain a new equation  $f_8(p_E^{A*}, w_E^{W*}) = 0$ .

Since  $w_E^{W*} = p_E^{W*}$  in the wholesale model's SPNE, in order to prove  $p_E^{A*} > p_E^{W*}$  we just need to prove

$p_E^{A^*} > w_E^{W^*}$ . Define  $w_E^{W^*} = kp_E^{A^*}$  and substitute it into  $f_8(p_E^{A^*}, w_E^{W^*}) = 0$ . This transforms  $f_8(p_E^{A^*}, w_E^{W^*}) = 0$  into  $f_8(k) = 0$ . Define  $f_9(k) = m^2 t \cdot f_8(k)$ . Then we just need to prove that there exists one and only one root of  $f_9(k)$  in  $[0, 1]$ . We prove it through proving  $f_9(0) < 0$ ,  $f_9(1) > 0$  and  $\partial f_9 / \partial k > 0$  in  $0 < k \leq 1$ . In the following, for simplicity of notation, we denote  $p_E^{A^*}$  by  $p_E$  and denote  $w_E^{W^*}$  by  $w_E$  in the following proof.

First we want to prove  $\partial f_9 / \partial k > 0$  in  $0 < k \leq 1$ . Derive

$$\frac{\partial f_9}{\partial k} = -\frac{1}{4} \frac{p_E \cdot f_{11}(p_E) \cdot f_{10}(p_E)}{(b - 2kp_E + c_A)^2}$$

where

$$f_{11}(p_E) = -8k^3 p_E^3 + (9k^2 c_A + 15k^2 b) p_E^2 + (-12kc_A b - 9kb^2 - 3kc_A^2) p_E + c_F b^2 - p_F^2 c_A + 2b^3 - bc_F p_F + c_A^2 p_F + 2c_A^2 b - c_F bc_A + c_F p_F c_A - b^2 p_F + 4c_A b^2 + bp_F^2,$$

$$f_{10}(p_E) = rb - b - c_A + 2p_E - 2p_E r.$$

We find that  $\partial f_{10}(p_E) / \partial p_E = 2(1 - r) > 0$  and  $f_{10}(\bar{p}_E) = 0$ . In Proposition 2.5 we have proved  $p_E < \bar{p}_E = (b + c_A / (1 - r)) / 2$ . Therefore we have  $f_{10}(\bar{p}_E) < 0$ . Next we prove  $f_{11}(p_E) > 0$ . In Proposition 2.3 we have proved  $w_E < \bar{w}_E = (b + c_A) / 2$ . Since we have  $w_E = kp_E$ , we find  $p_E < \hat{p}_E = (b + c_A) / (2k)$ . Then we derive

$$f_{12}(p_E) = \frac{\partial^2 f_{11}}{\partial p_E^2} = -48k^3 p_E + 18k^2 c_A + 30k^2 b,$$

$$f_{13}(p_E) = \frac{\partial f_{11}}{\partial p_E} = -24k^3 p_E^2 + 2(9k^2 c_A + 15k^2 b) p_E - 12kc_A b - 9kb^2 - 3kc_A^2.$$

Recall  $\hat{p}_E > p_E > c_A$  from the proof of Proposition 2.5. We first find  $f_{12}(\hat{p}_E) = 6k^2(b - c_A) > 0$ . Second, when  $0 < k \leq 1$ , we find  $f_{12}(c_A) = 6k^2(3c_A(1 - k) + 5(b - c_A k)) > 0$ . Therefore  $f_{12}(p_E) > 0$ . Next, we find  $f_{13}(\hat{p}_E) = 0$  and  $f_{13}(c_A) = -3k(b + c_A - 2c_A k)(3b + c_A - 4c_A k)$ . It is obvious that when

$0 < k \leq 1$ , we have  $f_{13}(c_A) < 0$ . Together with  $f_{12}(p_E) = \partial f_{13} / \partial p_E > 0$ , we prove  $f_{13}(p_E) < 0$ . Therefore, in order to prove  $f_{11}(p_E) > 0$ , we just need to prove  $f_{11}(\hat{p}_E) > 0$ . We derive

$$f_{11}(\hat{p}_E) = \frac{1}{4}(b - c_A)f_{14}(c_A)$$

where  $f_{14}(c_A) = c_A^2 + (2b - 4p_F)c_A + b^2 + 4p_F^2 + 4c_F b - 4bp_F - 4c_F p_F$ . We just need to prove  $f_{14}(c_A) > 0$ .

We find  $\min_{c_A} f_{14} = 4c_F(b - p_F) > 0$ . Therefore, we prove  $f_{11}(p_E) > 0$ . Together with  $f_{10}(\bar{p}_E) < 0$ , we

prove  $\partial f_9 / \partial k > 0$  in  $0 < k \leq 1$ .

Second, we want to prove  $f_9(0) < 0$ . Treat  $f_9(0)$  as a function of  $p_E$ . Define  $f_{15}(p_E) = f_9(0)$ . We derive

$$\begin{aligned} f_{15}(p_E) &= \frac{(1-2r)(-1+r)}{2} p_E^3 + \frac{3}{8}(-7rb + 3b + 4r^2b - 2c_A r + c_A) p_E^2 - \frac{A}{2c_A + 2b} p_E \\ &+ \frac{1}{4} \frac{c_A r b (-bp_F - c_F p_F + p_F^2 - c_A p_F + c_F b)}{c_A + b} \end{aligned}$$

where

$$\begin{aligned} A &= (1-r)^2 b^3 + \left( (r^2 - 3r + 2)c_A - \frac{1}{2}p_F + \frac{1}{2}c_F \right) b^2 + \left( (1-r)c_A^2 + \left( c_F \left( r - \frac{1}{2} \right) - p_F r \right) c_A + \frac{1}{2}p_F (p_F - c_F) \right) b \\ &+ (p_F - c_A - c_F) p_F c_A \left( r - \frac{1}{2} \right) \end{aligned}$$

Notice that it is a cubic function of  $p_E$ . Derive  $f_{15}(c_A) = -\frac{1}{8} \frac{c_A f_{17}(r) f_{16}(b)}{c_A + b}$  where

$$f_{16}(b) = -4(1-r)b^2 + (2p_F - 3c_A - 2c_F)b - 2p_F^2 - 4c_A^2 r + 2c_F p_F + c_A^2 + 2c_A p_F.$$

$$f_{17}(r) = rb - 2c_A r + c_A - b$$

Notice that  $f_{16}(b)$  is a quadratic function of  $b$ . Recall  $b > p_F$ . We find that  $\partial f_{16}(b) / \partial b < 0$  when

$r < 3/4$  and  $b > p_F$ . We also find  $f_{16}(p_F) = -4p_F^2(1-r) - c_A(p_F - c_A) - 4c_A^2 r < 0$ . Therefore, when

$r < 3/4$  we have  $f_{16}(b) < 0$ . We find that  $f_{17}(0) = c_A - b < 0$  and  $f_{17}(1) = -c_A < 0$ . Therefore we have  $f_{17}(r) < 0$ . Therefore we have  $f_{15}(c_A) < 0$ .

Next, we derive  $f_{15}(\bar{p}_E) = \frac{1}{32} \frac{f_{17}(r)f_{18}(c_A)}{(1-r)^2}$  where

$$f_{18}(c_A) = c_A^2 + 2(1-r)(b - 2p_F)c_A - (1-r)(rb^2 - 4c_Fb - b^2 - 4p_F^2 + 4c_Fp_F + 4bp_F),$$

We find that  $\min_{c_A} f_{18}(c_A)$  is equal to  $-4(1-r)(p_Fr - c_F)(b - p_F)$ . When  $r < c_F / p_F$ , we have  $-4(1-r)(p_Fr - c_F)(b - p_F) > 0$ . Therefore when  $r < c_F / p_F$ , we have  $f_{18}(c_A) > 0$ . Therefore we find that when  $r < c_F / p_F$ , we have  $f_{15}(\bar{p}_E) < 0$ .

Notice that  $f_{15}(p_E) = f_9(0)$  and  $f_{15}(p_E)$  is a cubic function of  $p_E$ . Now we have  $f_{15}(c_A) < 0$  and  $f_{15}(\bar{p}_E) < 0$ . In order to prove  $f_9(0) < 0$ , the last condition we need to prove is  $\partial f_{15}^2 / \partial p_E^2 > 0$  in  $p_E \in (c_A, \bar{p}_E)$ . Define  $f_{19}(p_E) = \partial f_{15}^2 / \partial p_E^2$ . We find that  $f_{19}(p_E)$  is a linear function of  $p_E$ . We also find that  $f_{19}(c_A) = \frac{3}{4}(-3 + 4r)f_{17}(r) > 0$  when  $r < 3/4$ , and  $f_{19}(\bar{p}_E) = -\frac{3}{4}c_A + \frac{3}{4}b + \frac{3}{2}c_Ar - \frac{3}{4}rb > 0$  when  $0 < r < 1$ . Therefore we prove  $f_{19}(p_E) > 0$ . Therefore we prove  $\partial f_{15}^2 / \partial p_E^2 > 0$ . Therefore we prove  $f_9(0) < 0$ .

Next we want to prove  $f_9(1) > 0$ .  $k=1$  implies  $p_E = w_E < \bar{w}_E = (b + c_A) / 2$ . We derive

$$f_9(1) = -\frac{1}{8} \frac{1}{b + c_A - 2p_E} \cdot f_{20}(r)$$

where  $f_{20}(r)$  is a quadratic function of  $r$ . To prove  $f_9(1) > 0$ , we just need to prove  $f_{20}(r) < 0$ . We derive  $f_{20}(0) = 0$  ( $f_{20}(0) = 0$  essentially confirms that the agency model is coincided with the wholesale model when  $r = 0$ ). We also find that  $\partial f_{20}^2 / \partial r^2 = 4p_E(b - p_E)(b - 2p_E)(b - 2p_E + c_A) < 0$  when

$c_A < p_F/6$  and  $c_F > 4p_F/9$ . Therefore, in order to prove  $f_9(1) > 0$  we just need to prove that

$f_{20}(1) < 0$ . We define

$$\begin{aligned} f_{21}(p_E) = f_{20}(1) = & -10p_E^3c_A + (12c_Ab + 4c_Fp_F + 4bp_F - 4c_Fb - 4p_F^2 + 4c_Ap_F + 6c_A^2)p_E^2 \\ & + (-4b^2c_A - 2bc_Fp_F - 2b^2p_F + 2c_Fb^2 - 4c_A^2p_F + 4p_F^2c_A - 4c_Fp_Fc_A + 2bp_F^2 - 4c_A^2b - 6bc_Ap_F + 4c_Fbc_A)p_E \\ & - 2c_Fb^2c_A + 2bc_Fp_Fc_A - 2bp_F^2c_A + 2b^2c_Ap_F + 2bc_A^2p_F \end{aligned}$$

$f_{21}(p_E)$  is a cubic function of  $p_E$ . We find that  $f_{21}(c_A) = -4c_A^2(b - c_A)^2 < 0$  and

$f_{21}(\bar{w}_E) = -\frac{1}{4}c_A(b - c_A)f_{14}(c_A) < 0$ . We also find that  $\partial^3 f_{21}(p_E)/\partial p_E^3 < 0$ . Therefore, we just need to

prove that  $\partial f_{21}(p_E)/\partial p_E|_{p_E=\bar{w}_E} > 0$ . Define  $f_{22}(b) = \partial f_{21}(p_E)/\partial p_E|_{p_E=\bar{w}_E}$  and notice that  $f_{22}(b)$  is a

quadratic function of  $b$ . We find  $f_{22}(b) > 0$  when  $b > p_F$ . Therefore, we prove  $f_9(1) > 0$ . Therefore,

we prove  $p_E^{A*} > p_E^{W*}$ .

Regarding the e-reader price, we have

$$\begin{aligned} p_D^{W*} = & -\frac{1}{4} \frac{2w_E^{W*}b - (w_E^{W*})^2 - 2bp_F + p_F^2 - 2tm - 2c_Dm}{m}, \\ p_D^{A*} = & -\frac{1}{4} \frac{2bp_E^{A*} - (p_E^{A*})^2 - 2bp_F + p_F^2 - 2tm + 2rp_E^{A*}b - 2r(p_E^{A*})^2 - 2c_Dm}{m}. \end{aligned}$$

We derive  $p_D^{A*} - p_D^{W*} = -\frac{1}{4}(p_E^{A*} - w_E^{W*})(2b - p_E^{A*} - w_E^{W*}) - \frac{1}{2}rp_E^{A*}(b - p_E^{A*})$ . We find that when  $p_E^{A*} > w_E^{W*}$ ,

we have  $p_D^{A*} < p_D^{W*}$ .

### Appendix C: Additional tables in Essay 3

**Table 3.A1** Correlation table

	<i>reputation</i>	<i>productnmb</i>	<i>avgprice</i>	<i>mkttool</i>	<i>goodsqualr</i>	<i>imreplyratio</i>	<i>buyeripv</i>	<i>ipvuv</i>	<i>puv</i>
<i>reputation</i>	1								
<i>productnmb</i>	0.4163	1							
<i>avgprice</i>	-0.013	-0.0541	1						
<i>mkttool</i>	0.267	0.1335	0.043	1					
<i>goodsqualr</i>	0.5695	0.3263	-0.0663	0.1618	1				
<i>imreplyratio</i>	0.2619	0.1131	-0.0152	0.097	0.2087	1			
<i>buyeripv</i>	0.0415	0.1306	-0.0307	0.0204	0.0889	0.0069	1		
<i>ipvuv</i>	0.3314	0.2461	0.0161	0.1769	0.1806	0.1277	-0.011	1	
<i>puv</i>	0.3771	0.1994	-0.0132	0.1765	0.2097	0.122	0.0097	0.3418	1

**VITA**

Lin Hao received Ph.D. in Business Administration in June 2012 from University of Washington at Seattle. He received his Bachelor and Master degree in Electrical Engineering from Tsinghua University, Beijing, China. His research interests include mobile commerce, pricing in the emerging electronic markets, online retail strategies and business analytics. He joined University of Notre Dame as Assistant Professor of Management in July 2012.