THE ROLES OF ONLINE USER GENERATED CONTENT AND THE INTERRELATIONSHIP BETWEEN PRODUCT DEVELOPMENT AND PRODUCT ADOPTION

by

Nguyen-Nga Ho-Dac

A dissertation submitted to the faculty of
The University of Utah
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

in

Business Administration

David Eccles School of Business

The University of Utah

August 2012

The University of Utah Graduate School

STATEMENT OF DISSERTATION APPROVAL

The dissertation of **Nguyen-Nga Ho-Dac**

has been approved by the following supervisory committee members:

William Moore	, Co-Chair	**06/04/2012** Date Approved
Stephen Carson	, Co-Chair	**06/04/2012** Date Approved
Debra Scammon	, Member	**06/13/2012** Date Approved
Abbie Griffin	, Member	**06/04/2012** Date Approved
Rohit Aggarwal	, Member	**06/04/2012** Date Approved

and by **William Hesterly**, Associate Dean of the **David Eccles School of Business**

and by Charles A. Wight, Dean of The Graduate School.

ABSTRACT

Extant studies have been showing that online user generated content (UGC) plays an important role in product adoption. I expand this research by considering the roles of online UGC in both product development and product adoption processes. Moreover, the interrelationship between these two processes is investigated under a theoretical framework of double-sided adverse selection in which both developers and users are uncertain about the future of a product. I model product development and adoption processes over the product lifecycle, capturing the direct and indirect effects of online UGC on both processes. Data for the study are taken from a longitudinal dataset covering 63 months for 11,648 open source projects registered on Sourceforge.net, one of the largest open source software development platforms.

The results of this study reveal that online UGC has a positive effect on both the initiation and the completion of product development activities. Moreover, product adoption helps amplify that positive effect. Those results are consistent with the theory that online UGC can provide ideas and solutions for development activities and that developers capitalize on those ideas and solutions to initiate and to complete development activities. In addition, the results also show that product development and product adoption affect each other positively. Those results are consistent with the theory that product development and adoption signal developers' commitment in developing the

product and users' interest in the product, respectively. Hence, it is suggested that transparency between developers and users creates a positive feedback loop between product development and product adoption. Since this study investigates the behaviors of both developers and users, it implies applications for users, developers, as well as platform owners who want to design a competitive ecosystem to attract and satisfy both users and developers.

TABLE OF CONTENTS

ABSTRACT……. iii

LIST OF TABLES…… vii

ACKNOWLEDGEMENTS……. viii

Chapters

1. INTRODUCTION……. 1

2. THEORECTICAL BACKGROUND ……..……. 5

    2.1 The Roles of Online UGC: From Influencing Users to Affecting Developers …….. 5
    2.2 User Generated Innovation Literature …… 6
    2.3 The Interrelationship between Product Development and Product Adoption: A Double-Sided Adverse Selection Problem …… 8
    2.4 Open Source Software Literature ……..... 9

3. HYPOTHESES ……. 14

    3.1 The Roles of Online UGC in Product Development …….. 14
    3.2 The Interrelationship Between Product Development and Product Adoption Processes …… 15
    3.3 The Moderation Effect of Product Adoption on the Role of Online UGC in Product Development …… 19

4. DATA AND MODELS……. 21

    4.1 Data……. 21
    4.2 Models……. 22
    4.3 Estimation Method…….. 27

5. FINDINGS……. 32

5.1 The Initiation of Product Development Activities ……. 32
5.2 The Completion of Product Development Activities…….. 35
5.3 Product Adoption……..... 38

6. CONCLUSION AND IMPLICATIONS…… 42

6.1 Conclusion and Theoretical Implications ……..... 42
6.2 Managerial Implications……..... 46
6.3 Further Research Directions……..... 50

LIST OF TABLES

4.1. Descriptive statistics ……. 23

4.2 Correlation matrix ……. 23

4.3 Hausman specification test for equation (1)…….. 29

4.4 Hausman specification test for equation (2)…….. 29

4.5 Hausman specification test for equation (3)…….. 30

5.1 The effects of online UGC and downloads on development activities initiated ……. 33

5.2 The effects of online UGC and downloads on development activities completed …….. 33

5.3 The effects of online UGC and development activities on downloads ……. 34

ACKNOWLEDGEMENTS

I would like to gratefully thank Professors William L. Moore and Stephen J. Carson for their guidance and mentorship during my study at the University of Utah. They have not only encouraged me to become an independent researcher but also shown me how to become a good instructor. I would like to thank Professors Abbie Griffin, Debra Scammon, and Rohit Aggarwal who serve in my doctoral committee for their valuable inputs and discussions. I would also like to thank all the Professors at the University of Utah and other institutions who have trained me to become a researcher in the field of marketing.

Finally and importantly, I would like to thank my parents, Khoi Ho-Dac and Hai Nguyen, my wife, Khanh Thai, and all of my beloved for their encouragement, support, and love during my PhD program and always.

CHAPTER 1

INTRODUCTION

Extant research on online user generated content (UGC) has focused mostly on its effect on outcome variables such as product adoption or sales (e.g., Chen, Wu, and Yoon 2004; Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; Li and Hitt 2008; Liu 2006; Zhu and Zhang 2010; Karniouchina 2011; Ho-Dac, Carson, and Moore 2011). These studies suggest that online user generated content in online customer reviews or online discussion forum posts can influence other users' behaviors which, in turn, influence the marketing strategies and tactics of marketers. This stream of research implies that marketers can use various strategies and tactics to make use of online UGC as a communication tool.

This study approaches online UGC from a different angle; how online UGC influences marketers directly. Specifically, this study investigates how online UGC influences the product development process of marketers or product developers. Understanding this phenomenon would provide consumers with a power to shape the product development process and ultimately the product itself for their own good. This study also implies that marketers might use online UGC to accelerate their product development process and serve their customers better. Therefore, this study shifts the

research focus from online UGC as a marketing communication tool to online UGC as a means to increase market efficiency which benefits both consumers and marketers.

Online UGC might influence the product development process if it contains information which is useful for the process and developers utilize that information in their development process. While the effect of online UGC on product development has yet to be examined, the user innovation literature (e.g., von Hippel 1978, 1988, 2005; Bendapudi and Leone 2003; Raymond 1999) suggests that users can contribute to product development in two ways. First, manufacturers can actively solicit users for product improvement ideas or new product opportunities (von Hippel 1988, 2005; Griffin and Hauser 1993). Second, users can actively submit ideas for product development to manufacturers (Moorman and Rust 1999; Lilien et al. 2002). This literature shows that users can be a good source of product development information. However, the efficacy with which the conversation of general users in a natural setting such as online forums or social network sites can generate useful information for product development is unclear. Even if that kind of conversation can actually generate useful information for product development, that information is only influential if developers seek to learn from that kind of conversation.

Whether developers capitalize on the information in online UGC to develop their products may depend on their motivation to do so (e.g., MacInnis and Jaworski 1989). One might expect that developers are motivated by the future product adoption rate. Hence, if users are willing to adopt a product, developers will be willing to learn from online UGC to develop the product and possibly to commit more resources to develop that product. Without an assurance of continued future adoption, developers would not

invest their resources to develop the product, let alone learn from online UGC. On the users' side, they also require an assurance of continued future product development. They would not want to spend time, effort, and money to adopt a product if that product is not going to be supported and continuously developed. A situation in which both sides require mutual assurances to make investment is usually called a double-sided adverse selection problem (Gale 2001).

To reduce that double-sided adverse selection problem, a signaling mechanism is needed to help developers and users signal each other about their intentions. That signaling mechanism would benefit both developers and consumers and increase market efficiency. Recognizing the importance of that mechanism, this study also investigates whether product adoption and product development activities might serve as a signal of future adoption and a signal of commitment to developing the product, respectively. If so, market transparency where developers can observe the adoption activities of users and users can observe the development activities of developers could help reduce the double-sided adverse selection problem.

In short, this study investigates how online UGC influences the product development process and whether product adoption and product development processes might affect each other. Both of these issues aim to increase market efficiency which benefits not only developers and users but also platform owners who need to attract both users and developers to their platforms. For example, platform owners such as Google (Android), Apple (iOS), and Microsoft (Windows Phone) might want to design an ecosystem which is transparent and facilitates the creation of online UGC. Such an ecosystem would suffer less from a double-sided adverse selection problem. The developers in that ecosystem

would be able to develop their product faster and meet the needs of their users better by capitalizing on the information in online UGC. The users in that ecosystem would have the ability to influence the product development process through the content they generated online on social networks or discussion forums, etc.

To study these two issues, I investigate the effects of online UGC on both product development and product adoption processes and examine the interrelationship between the product development and product adoption processes together. To do so, I model product development and product adoption processes over time, capturing the interrelationship between these two processes and the effects of online UGC on both processes. This requires a research context where product development activities, product adoption activities, and online UGC are observable by users and developers, respectively.

Online UGC is usually publicly available. Product adoption is usually observable by developers. However, users usually cannot observe development activities. Therefore, an open source software development platform is chosen to be the research context because in open source software (OSS) development, development activities are usually observable. Data for the study were collected on Sourceforge.net, one of the largest open source software development platforms, where online UGC, adoption activities, and development activities are all publicly observable. In addition, the sample contains only products that have been on the market and that have continuous development activities so we can observe development activities, product adoption, and online UGC together.

I discuss the theoretical background and the hypotheses in the next two chapters. Then, I describe the data, the models, and the estimation method which is followed by the findings and then the conclusion and implications chapters.

CHAPTER 2

THEORECTICAL BACKGROUND

2.1 The Roles of Online UGC: From Influencing Users to Affecting Developers

An emerging literature in marketing has investigated the effects of online customer reviews (the terms *user* and *customer* are used interchangeably in this study), a form of online UGC, on sales (e.g., Chen, Wu, and Yoon 2004; Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; Li and Hitt 2008; Liu 2006; Zhu and Zhang 2010; Ho-Dac, Carson, and Moore 2011) and brand equity (Ho-Dac, Carson, and Moore 2011). Those studies have found that online customer reviews positively affect sales because online customer reviews reduce information search costs (Chen, Wu, and Yoon 2004) and provide information to other customers (Liu 2006; Luan and Neslin 2009; Zhu and Zhang 2010). Zhu and Zhang (2010) also found that product and consumer characteristics moderate the effect of online customer reviews on sales. Moreover, Ho-Dac, Carson, and Moore (2011) demonstrated that the impact of online customer reviews on sales is diminished as the brand becomes stronger and that a large number of positive online customer reviews (whether for the leading model or for all the models in a product line) can also increase the brand's strength. Whether investigating the impact of online UGC

on sales or on brand equity, these studies shed light on the role of online UGC in influencing other users.

While the role of online UGC in influencing other users has been empirically demonstrated, the role of online UGC in affecting developers (i.e., product development) has not been studied. Whether online UGC can benefit the product development process of developers depends on two conditions. First, the information generated from the discussion must be a rich source of product development ideas. Second, developers or producers must actively search for those product development ideas in virtual customer environments to capitalize on those ideas; if developers do not utilize them, the ideas would not have any impact on product development. To explore the possibility that information generated from user discussion online is a rich source of product development ideas, the next section below discusses the user generated innovation literature.

2.2 User Generated Innovation Literature

The user generated innovation literature (e.g., von Hippel 1978, 1988, 2005; Bendapudi and Leone 2003; Raymond 1999) suggests that users can contribute to product innovation. Obtaining customer input for purposes of product development has been in practice for years (von Hippel 1988, 2005; Griffin and Hauser 1993). Manufacturers have turned to end users for suggestions for new products and for advice on the development of existing models. Salespeople have often been charged with the responsibility of collecting feedback from customers to be used in the product development process. These practices suggest that customers have the potential to provide innovative information to

develop or modify a product. However, structured mechanisms employed by firms such as market surveys and focus groups to tap customer knowledge usually involve only a minority of customers (Nambisan 2002). The advance of technology, especially the Internet, makes acquiring knowledge from a large population possible. For example, crowd-sourcing can generate high quality ideas in terms of novelty and customer benefit (Poetz and Schreier, forthcoming).

In addition to the manufacturer active paradigm which underlies the practices discussed above, von Hippel (1978, 1988) suggested a user active paradigm in which customers actively develop innovative ideas and send requests to manufacturers. Many studies have found evidence to support this paradigm (e.g., Moorman and Rust 1999; Lilien et al. 2002). Several tools such as design palettes, user design, and configurators have been developed to utilize customer innovation (Dahan and Hauser 2002, Thomke and von Hippel 2002, von Hippel 2001). Urban and Hauser (2004) also demonstrated that virtual advisors that capture customers' queries can be used to identify new product opportunities.

This user innovation literature, including both the manufacturer active and user active paradigms, demonstrates that users are informative for product development. However, in the case of online UGC, manufacturers do not actively solicit users for innovative ideas nor do users actively submit their ideas or requests to manufacturers. Rather, it is the discussion among many people in a natural setting which enhances the richness of user information (Nambisan 2002). One of the objectives of this study is to investigate the effect of online user discussion on product development.

When product development and product adoption occur concurrently (e.g., product upgrades or updates, service, etc), the effect of online UGC on product adoption may complicate the effect of online UGC on product development if product adoption and product development interact with each other. The next section discusses the possible interaction between product development and product adoption process.

2.3 The Interrelationship between Product Development and Product Adoption: A Double-Sided Adverse Selection Problem

In situations where product development and product adoption processes occur concurrently, both developers and users are uncertain about the future of a product. On one hand, developers do not know if users are going to continue adopting the product so they hesitate to invest more effort developing it. On the other hand, users do not know if developers are going to continue developing the product. Therefore, they would not want to spend time, effort, and resources to adopt a product that is likely to be discontinued or abandoned. This is known as a double-sided adverse selection problem (see Gale 2001 for a review of double-sided adverse selection) where both users and developers have private information.

If there is a mechanism that lets both developers and users reveal their private information, the problem is solved. However, the private information here is future intentions; neither developers nor users can reveal this, but they can signal their future intentions through their current behaviors. In order to make their signals credible, developers and users must incur a private cost (Wilson 1977; Riley 1979). For example, current adoption may be a signal for future adoption because users must spend time,

effort, and money (in most of the cases) to adopt a product. Similarly, current development activities may be a signal for developers' commitment because developers need to invest their time, effort, and capital to develop a product. Therefore, the double-sided adverse selection argument suggests that developers may take users' activities as signals for users' future intention to adjust their investments. On the other hand, users may take developers' activities as signals for developer future commitment to adjust their product adoption. Moreover, a transparent market where users and developers can see each others' activities would facilitate that signaling mechanism. To study the interrelationship between product adoption and product development, an open source software platform was chosen as a research context because in such a platform, developers' activities and users' activities occur concurrently and transparently. The next section discusses the open source software literature.

2.4 Open Source Software Literature

An open source software platform such as Sourceforge.net provides an excellent context in which to study the impact of online UGC on product development as well as the interrelationship between product development and product adoption because one can observe both user and developer activities concurrently. Computer science and engineering research aside, the extant open source software (OSS) literature can be classified into four (overlapping) streams: the motivation of developers, social networks, the productivity of developers, and the determinants of OSS success.

The motivation stream (e.g., Hars and Ou 2002; Lakhani and Wolf 2005; Lerner and Tirole 2002) tries to explain why developers commit their resources to develop OSS

projects for free (without immediate monetary rewards). These studies reveal that software developers are motivated by other rewards rather than money. Those rewards can be both intrinsic such as personal fulfillment and altruism and extrinsic such as career enhancement and status.

The social network stream studies the structure and evolution of the collaboration network among developers. Singh and Tan (2011) investigated the stability and efficiency of the communication structure among developers and found that the stability of the structure depends on self utility maximization while the efficiency of the structure depends on group utility maximization. Many other studies have extracted the parameters of the organization and evolution of the OSS developer collaboration network (e.g., Xu, Christley, and Madey forthcoming; Xu, Gao, Christley, and Madey 2005; Xu and Madey 2004; Gao, Freeh, and Madey 2003; Madey, Freeh, and Tynan 2002a&b). This social network stream has a strong influence on the two other streams, the productivity of developers and the determinants of OSS success.

The productivity stream focuses on how to improve the productivity of developers. Most of the studies in this stream investigate the impact of various social network parameters on productivity of developers. For example, Singh (2010) showed the effect of developer networks on the productivity of developers. Singh, Tan, and Mookerjee (2011) found that groups with stronger internal cohesion are more productive while external cohesion has an inverse U-shaped impact on productivity. Singh, Tan, and Youn (2011) demonstrated that developers can learn from their own experience and from their peers to improve their productivity.

The OSS success stream investigates the determinants of OSS success. Many studies in this stream utilize social network analysis to explain how relationships among developers influence OSS success (e.g., Grewal, Lilien, and Mallapragada 2006; Singh, Tan, and Mookerjee 2007; Mallapragada, Grewal, and Lilien 2008) and how founders' social capital accelerates OSS's time-to-market (Mallapragada, Grewal, and Lilien 2008 and 2011). In addition to the social network analysis, Mallapragada, Grewal, and Lilien (2011) found that bug reports (users report bugs and problems that they have experienced when using a product to developers) accelerate time-to-market. However, O'Hern et al. (2011) showed that bug reports and feature requests (users send requests to developers to add more features to a product) have no significant effect on product improvement which is measured by the number of code lines written by developers. They also found that patch submits (codes written by users) even have negative effects on product improvement. The latter two studies documented the impact of users' direct contribution to OSS development. In other words, they extended the user active paradigm in the user generated innovation literature to the OSS context.

While OSS has been studied quite extensively, the content generated by users through their conversations on online forums and their effect on product development have not been investigated. This study investigates the effect of online UGC in the form of online forum discussions on the product development process. This study is different from the user active paradigm and its extension in OSS context (e.g., Mallapragada, Grewal, and Lilien 2011; O'Hern et al. 2011) in many aspects. While O'Hern et al. (2011) claimed that users' direct contribution to OSS development is another type of UGC, the concept of online UGC in this study is more traditional. It is the content generated by users in

online forum discussion and does not include users' direct contribution to product development such as codes submitted by users.

The distinction between UGC and users' direct contribution is quite substantial. In the former case, users only generate information in online forums. Only when developers actively seek for and utilize that information, can online UGC affect the product development process. In the latter case, users contribute directly to the product development process.

The implications of this study are also different from that of the user active paradigm and its extension in OSS context (e.g., Mallapragada, Grewal, and Lilien 2011; O'Hern et al. 2011). The implications of this study are for both developers and users: whether users generate useful information for the product development process in online forums and whether developers capitalize on that information in their development process. The implication of the latter is whether users' direct contribution helps or hurts the product development process.

Unlike other studies on OSS which usually measure product development by the number of code lines written (e.g., O'Hern et al. 2011), this study investigates the number of development activities initiated and the number of development activities completed. First, a development activity represents a product improvement. When one development activity is completed, a bug is fixed or a feature is added. Different development activity requires different number of code lines written to be completed. Developers' skill also affects the number of code lines needed to complete an activity. Better developers usually write fewer code lines to accomplish one specific activity. Therefore, a development activity reflects a visible change in a product but a code line does not. Second, I can

separate the effect of online UGC on the initiation (ideas) and completion (solutions) of product development activities to provide more detailed suggestions to users and developers. Specifically, this study reveals whether developers should look for ideas or solutions or both in online UGC and whether users can influence what will be done or how it will be done or both.

Moreover, this study investigates the effects of online UGC on product development and adoption concurrently. The concurrent processes of product development and adoption also raise a question as to whether developers and users adjust their activities based on the activities on the other side.

CHAPTER 3

HYPOTHESES

3.1 The Roles of Online UGC in Product Development

The user generated innovation literature (e.g., von Hippel 1978, 1988, 2005; Bendapudi and Leone 2003; Raymond 1999) has confirmed that manufacturers can actively ask for innovation ideas from customers (von Hippel 1988, 2005; Griffin and Hauser 1993; Poetz and Schreier, forthcoming) and that customers (usually lead customers or highly skilled customers) can actively submit their ideas to manufacturers (e.g., Moorman and Rust 1999; Lilien et al. 2002) to improve product development. This literature suggests that customers or users can be a good source of product development ideas. Therefore, one may expect that the conversation among users on online forums can generate ideas which are useful for product development. If developers capitalize on those ideas to initiate development activities, the number of online forum posts should positively affect the initiation of product development activities in the next period.

H1: The number of forum posts increases the number of product development activities initiated in the next period.

In addition to providing developers with ideas to initiate new development activities, the content generated on online forums may also trigger developers to find solutions to

the development activities they are working on or even provide them with the solutions. Hence, the number of online forum posts should positively affect the completion of product development activities in the next period.

H2: The number of forum posts increases the number of product development activities completed in the next period.

3.2 The Interrelationship Between Product Development and Product Adoption Processes

While online UGC may provide developers with ideas and solutions to initiate and complete development activities, the number of downloads (a measure of product adoption) contains no direct information for product development. However, the number of downloads may have a signaling effect on developers which remedies half (on the developer side) of the double-sided adverse selection problem between developers and users. Usually, developers do not know if and how many users will continue to adopt their products but they know the number of downloads in the past. If they take the number of downloads in the past as a signal of future adoption, they would adjust their investment and efforts in both initiating new development activities and completing current development activities accordingly to the number of downloads in the past. Therefore:

H3: The number of downloads increases the number of development activities initiated in the next period.

H4: The number of downloads increases the number of development activities completed in the next period.

Unlike the number of downloads, the number of development activities initiated and the number of development activities completed may convey additional information beyond signals of developers' commitment. Therefore, the effects of product development on product adoption are more complex. Figure 3.1 shows an example of the breakdown of these effects. On the far left, four development activities have been initiated but only two of them have been completed. The improvements resulted from these two completed development activities are realized in the new version when it is released. This means that the entire product improvements resulted from product development activities are only realized when they are integrated into a new released version. Therefore, the effect of new versions released on product adoption controls for all the impact of product improvement on product adoption.

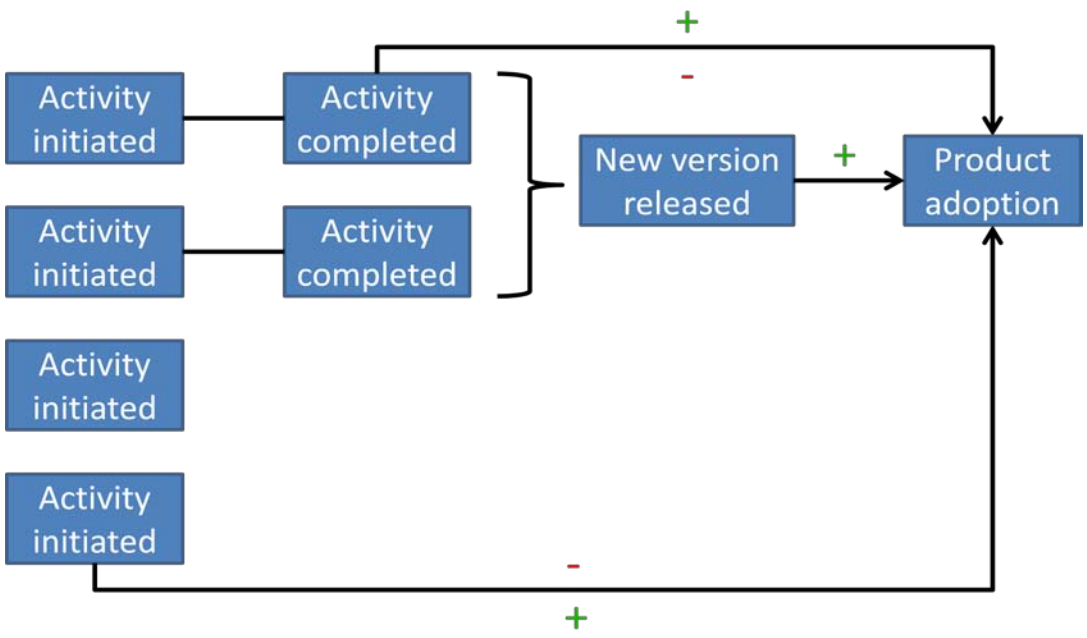

Figure 3.1. The effects of product development on product adoption.

After controlling for the impact of product improvement on product adoption, the completion of development activities could hold off product adoption because users could wait for a new version which has product improvement resulted from the completed development activities before adopting the product. This argument would suggest that the number of development activities completed decreases the number of downloads in the next period.

On the other hand, the completion of development activities might signal the commitment of developers in developing the product that remedies the other half (on the user side) of the double-sided adverse selection problem between developers and users. Usually, users do not know if and how much effort developers will commit to develop a specific product, but they know how much effort developers have committed to develop the product in the past based on the number of development activities completed in the past. If they take the figure in the past as a signal of future commitment, they would adjust their adoption according to that figure. This signaling argument would suggest that the number of development activities completed increases the number of downloads in the next period.

Given the two contradictory predictions of the two arguments above, I state the formal hypothesis in favor of the signaling argument. If the empirical analysis supports the formal hypothesis, the signaling effect is stronger than the hold off effect and vice versa.

H5: The number of development activities completed has a positive effect on the number of downloads in the next period.

After controlling for the impacts of product improvement and the number of development activities completed on product adoption, the initiation of development activities could indicate that there are problems which need to be resolved and/or features that need to be added for the product to function properly. Hence, they may increase the uncertainty of the product performance which raises customer perceived risk. Because customer perceived risk has an adverse effect on adoption of a product (Dowling and Staelin 1994), development activities initiated might cause users to hold off on downloading and wait for the improvement. This argument would suggest that the number of development activities initiated decreases the number of downloads in the next period.

On the other hand, the number of development activities initiated may have a signaling effect on users which remedy the other half (on the user side) of the double-sided adverse selection problem between developers and users in a similar manner to the number of development activities completed. This signaling argument would suggest that the number of development activities initiated increases the number of downloads in the next period.

Given the two contradictory predictions of the two arguments above, I state the formal hypothesis in favor of the signaling argument. If the empirical analysis supports the formal hypothesis, the signaling effect is stronger than the hold off effect and vice versa.

H6: The number of development activities initiated has a positive effect on the number of downloads in the next period.

3.3 The Moderation Effect of Product Adoption on the Role of Online UGC in Product Development

The two hypotheses H1 and H2 assume that developers process the information in the content generated by users on online forums to initiate and to complete their development activities. Many studies on information processing have demonstrated that more motivated individuals process information more effectively (e.g., MacInnis and Jaworski 1989; MacInnis, Moorman, and Jaworski 1991). Hence, the extent to which developers process the information in online UGC to initiate and to complete their development activities should be amplified by their motivation. Motivation is "the drives, urges, wishes, or desires which initiate the sequence of events known as behavior" (Bayton 1958, p.282). In this context, motivation is the drives, urges, wishes, or desires of developers to process online UGC to capitalize on ideas and solutions for product development. As discussed earlier, developers may take the number of downloads as a signal of future product adoption so their motivation to process online UGC to develop a product should increase as the number of downloads increases because they want to maximize the utility of their efforts. Therefore, if the signaling argument holds true, the number of downloads should amplify the effect of online UGC on product development in the next period.

H7: The interaction between the number of forum posts and the number of downloads has a positive effect on development activities initiated in the next period.

H8: The interaction between the number of forum posts and the number of downloads has a positive effect on development activities completed in the next period.

The next chapter describes the data, the model specification, and the estimation process to test the hypotheses presented in this section.

CHAPTER 4

DATA AND MODELS

4.1 Data

I use data on Sourceforge.net, one of the largest open source software development platforms, to investigate the hypotheses. As a platform, Sourceforge.net provides software developers an environment to coordinate their development activities. New development activities initiated and development activities completed are recorded by a tracker system. The number of current development activities that the development team is working on is the difference between the cumulative number of activities initiated and the cumulative number of activities completed. When accumulating enough improvement through bug fixes, features added, and performance enhancement, a development team may decide to release a new version of the product which incorporates those changes. Any changes of the number of developers in each project are also documented.

Sourceforge.net has a system of online forums where users can discuss various issues related to each product. Each product has its own forums. Online UGC is measured by the number of monthly posts for each product. The number of downloads of each product is also recorded. All of these data are publicly observable.

Data for all projects which were on Sourceforge.net from November 1999 to January 2005 were retrieved from the Sourcefore.net database hosted at the University of Notre Dame (Madey 2012). The raw data file contains the data of 122,205 projects. Among those projects, only 11,648 projects (9.53%) were active in both development activities and adoption activities. Other projects which were not active in either development activities or adoption activities were excluded from the dataset. This resulted in a longitudinal dataset of 11,648 projects over 63 months. The descriptive statistics of the dataset are presented in Table 4.1. The correlation matrix is presented in Table 4.2.

4.2 Models

To test hypotheses H1, H3, and H7, the number of development activities *initiated* is regressed on the number of forum posts, the number of downloads, and their interaction in the last period. I also control for the effect of the number of developers in each project on the initiation of new development activities of that project. When more developers join in the development team, more initiations are expected because there are more people to develop the product, but this effect should exhibit a declining return to scale because a large team might suffer from coordination problems. Project age might also have an effect on the initiation of development activities. As a project is developing, more and more activities may be initiated over time. However, when it reaches a mature point where little improvement can be made or when developers lose interest in the project, fewer and fewer activities should be initiated over time.

Table 4.1 Descriptive statistics.

Variables	Mean	Std. Dev.	Min	Max
Number of downloads	1,886.012	44,778.420	0	7,605,172
Number of forum posts	3.841	40.286	0	4,375
Number of developers	3.992	6.128	1	276
Number of new versions	0.320	1.283	0	180
Number of development activities initiated	2.030	23.247	0	2,611
Number of development activities completed	1.696	24.005	0	2,557
Number of incomplete activities	10.474	39.905	0	2,963

(Monthly per project data)

Table 4.2 Correlation matrix.

		1	2	3	4	5	6	7	8
1	Ln(number of forum posts)	1.00							
2	Ln(number of downloads)	0.38	1.00						
3	Ln(number of development activities initiated)	0.44	0.45	1.00					
4	Ln(number of development activities completed)	0.37	0.35	0.79	1.00				
5	Ln(number of developers)	0.15	0.27	0.29	0.25	1.00			
6	Ln(age)	-0.06	-0.15	-0.13	-0.12	-0.02	1.00		
7	Ln(number of incomplete activities)	0.25	0.41	0.49	0.30	0.39	0.03	1.00	
8	Ln(number of new versions)	0.22	0.29	0.34	0.39	0.09	-0.09	0.05	1.00
9	Ln(cumulative number of downloads)	0.27	0.77	0.29	0.20	0.30	0.06	0.44	0.08

$$(1)\quad \ln(AI_{it}) = \alpha_0 + \alpha_P \ln(P_{i,t-1}) + \alpha_D \ln(D_{i,t-1}) + \alpha_{PD} \ln(P_{i,t-1}) * \ln(D_{i,t-1}) + \alpha_{Dev} \ln(Dev_{it}) + \alpha_{Dev2} [\ln(Dev_{it})]^2 + \alpha_d \ln(d_{it}) + \alpha_{d2} [\ln(d_{it})]^2 + \mu_i + \varepsilon_{it}$$

- AI_{it} is the number of development activities for product *i* {1,…, 11,648}initiated in month *t* {1,…,63}.
- $P_{i,t-1}$ is the number of forum posts for product *i* in month *t-1*.
- $D_{i,t-1}$ is the number of downloads of product *i* in month *t-1*.
- Dev_{it} is the number of developers of product *i* in month *t*.
- d_{it} is the age (in months) of product *i* at the end of month *t*.
- μ_i is the time-invariant product-specific effect of product *i* which captures differences across the products, such as working operation system, programming language, license type, and type of software, etc.
- ε_{it} is an idiosyncratic error.

To test hypotheses H2, H4, and H8, the number of development activities *completed* is regressed on the number of forum posts, the number of downloads, and their interaction in the last period. I also control for the effect of the number of incomplete development activities of a project at the end of the last period on the completion of development activities of that project. Incomplete activities are development activities that have been initiated in previous periods but have not been completed in the current period. The developers might complete more development activities when there are more incomplete activities because they have more development activities to work on. However, too many incomplete activities may discourage developers from finishing those

activities. In addition, when more developers join in the development team, there are more people to complete the current incomplete activities so the completion rate should increase. However, this effect should exhibit a declining return to scale because a large team might suffer from co-ordination problems. Project age might also have an effect on the number of activities completed. As a project is developing, more and more activities should be completed. However, when it reaches a mature point where there is little improvement that can be made or when developers lose interest in the project, fewer and fewer activities should be completed over time.

$$(2)\quad Ln(AC_{it}) = \beta_0 + \beta_P ln(P_{i,t-1}) + \beta_D ln(D_{i,t-1}) + \beta_{PD} ln(P_{i,t-1})*ln(D_{i,t-1}) + \beta_I ln(I_{i,t-1}) + \beta_{I2}[ln(I_{i,t-1})]^2 + \beta_{Dev} ln(Dev_{it}) + \beta_{Dev2}[ln(Dev_{it})]^2 + \beta_d ln(d_{it}) + \beta_{d2}[ln(d_{it})]^2 + \eta_i + \nu_{it}$$

- AC_{it} is the number of development activities for product i {1,…, 11,648}completed in month t {1,…,63}.
- $I_{i,t-1}$ is the number of incomplete development activities (activities that have been initiated but not yet finished) of product i at the end of month $t-1$.
- η_i is the time-invariant product-specific effect of product i which captures differences across the products, such as working operation system, programming language, license type, and type of software, etc.
- ν_{it} is an idiosyncratic error.

To test hypotheses H5 and H6, the number of downloads is regressed on the number of development activities initiated and the number of development activities completed in the last period. To control for the effect of product improvement on the number of

downloads, I include the number of new versions of a product released in the previous period. It is expected that users download more after new versions are released to benefit from the improvement. New users may also want to wait for a new version that addresses problems thus delaying their downloading the product. Because all of the improvements are not integrated into a product until a new version is released, after controlling for the effect of new versions released on the number of downloads, the positive (or negative) effects of the number of development activities initiated and the number of development activities completed on the number of downloads (if any) are not the effects of product improvement on the number of downloads but are signaling (or hold off) effects.

The number of downloads of a product in the current period should also be impacted by the cumulative number of downloads of that product at the end of the previous period which represents the user base of the product. A larger user base should lead to better continuous adoption and new adoption. Project age might also have an effect on the number of downloads. As a project is developing, more and more people may adopt the product. However, when the user base reaches a mature point where there are no more new adopters or when users lose interest in the project, fewer and fewer new downloads should occur over time.

$$(3) \qquad \ln(D_{it}) = \gamma_0 + \gamma_P \ln(P_{i,t-1}) + \gamma_{AI} \ln(AI_{i,t-1}) + \gamma_{AC} \ln(AC_{i,t-1}) + \gamma_R \ln(R_{i,t-1}) + \gamma_{\Sigma D} \ln\left(\sum_{j=1}^{t-1} D_{i,j}\right) + \gamma_d \ln(d_{it}) + \gamma_{d2} [\ln(d_{it})]^2 + \pi_i + \xi_{it}$$

- D_{it} is the number of downloads of product i {1,…, 11,648} in month t {1,…,63}.
- $R_{i,t-1}$ is the number of new versions of product i released in month $t-1$.

- π_i is the time-invariant product-specific effect of product *i* which captures differences across the products, such as working operation system, programming language, license type, and type of software, etc.
- ξ_{it} is an idiosyncratic error.

4.3 Estimation Method

The interaction term in equations (1) and (2) is constructed from the mean centered values of the number of forum posts and the number of downloads. This transformation helps interpret the estimation results easier. After the transformation, the main effects of the number of forum posts on the initiation and the completion of development activities can be interpreted as the effects of the number of forum posts on the initiation and the completion of development activities when the number of downloads is at its mean value (zero after the transformation) and vice versa.

It is possible that unobserved (by the researcher) product characteristics (such as working operation system, programming language, license type, and type of software, etc.) influence the number of development activities initiated, the number of development activities completed, and the number of downloads. Therefore, μ_i may be correlated with $\ln(D_{i,t-1})$ and $\ln(P_{i,t-1})*\ln(D_{i,t-1})$ in equation (1); η_i may be correlated with $\beta_D\ln(D_{i,t-1})$ and $\ln(P_{i,t-1})*\ln(D_{i,t-1})$ in equation (2); and π_i may be correlated with $\ln(AI_{i,t-1})$ and $\ln(AC_{i,t-1})$ in equation (3). These possible endogeneity problems may prevent the use of random effect estimation of equations (1), (2), and (3) which requires the assumption that all explanatory variables are strictly exogenous with respect to the individual effects for the estimation to be consistent (Mundlak 1978). Fixed effect estimation does not suffer from

these endogeneity problems but is less efficient than random effects estimation.

To determine whether fixed effect estimation or random effect estimation should be applied to estimate equations (1), (2), and (3), three Hausman (1978) specification tests were conducted. The null hypothesis is that there are no systematic differences between the efficient estimator (random effect estimator) and the consistent estimator (fixed effect estimator). If there is no difference, random effect estimation should be used because of its efficiency. Otherwise, random estimation is not consistent so fixed effect estimation should be used.

The results of the Hausman specification tests in Tables 4.3, 4.4, and 4.5 show that equations (1), (2), and (3) should be estimated by a fixed effect estimation method because their random effect estimators are not consistent due to the correlation between the time-invariant individual effects (μ_i, η_i, and π_i) and the independent variables. An eyeball comparison also confirms the differences between the random effect estimators and the fixed effect estimators in Tables 4.3, 4.4, and 4.5. For example, the fixed effect estimator of the impact of the number of developers on the number of activities initiated is significantly negative while the random effect estimator is not significant (Table 4.3). The fixed effect estimator of the impact of project age on the number of downloads is significantly positive while the random effect estimator is not significant (Table 4.5).

Because fixed effect estimation eliminates all the time-invariant individual specific effects, it eliminates any problems which may arise because of the possible correlation among them. In other words, any possible correlations among μ_i, η_i, and π_i should not cause any problems when one estimates equations (1), (2), and (3) by a fixed effect estimation method.

Table 4.3 Hausman specification test for equation (1).

Dependent variable: Ln(number of activities initiated)$_{it}$	Fixed effect estimation		Random effect estimation		Hausman's test
	Coef.	p	Coeff.	p	
Ln(number of forum posts)$_{i,t-1}$	0.093	0.000	0.100	0.000	
Ln(number of downloads)$_{i,t-1}$	0.026	0.000	0.039	0.000	
Ln(number of forum posts)$_{i,t-1}$ *Ln(number of downloads)$_{i,t-1}$	0.022	0.000	0.024	0.000	
Ln(number of developers)$_{it}$	-0.016	0.018	0.004	0.543	
[Ln(number of developers)$_{it}$]^2	0.026	0.000	0.037	0.000	$\chi^2(7) = 2952.57$
Ln(age)$_{it}$	0.577	0.000	0.558	0.000	p-value= 0.000
[Ln(age)$_{it}$]^2	-0.115	0.000	-0.111	0.000	
Constant	-0.229	0.000	-0.227	0.000	
R^2 within		0.068		0.066	
R^2 between		0.382		0.415	
R^2 overall		0.283		0.312	

Table 4.4 Hausman specification test for equation (2).

Dependent variable: Ln(number of activities completed)$_{it}$	Fixed effect estimation		Random effect estimation		Hausman's test
	Coeff.	p	Coeff.	p	
Ln(number of forum posts)$_{i,t-1}$	0.070	0.000	0.077	0.000	
Ln(number of downloads)$_{i,t-1}$	0.009	0.000	0.017	0.000	
Ln(number of forum posts)$_{i,t-1}$ *Ln(number of downloads)$_{i,t-1}$	0.015	0.000	0.017	0.000	
Ln(number of incomplete activities)$_{i,t-1}$	0.220	0.000	0.171	0.000	
[Ln(number of incomplete activities)$_{i,t-1}$]^2	-0.036	0.000	-0.019	0.000	$\chi^2(9) = 4440.50$
Ln(number of developers)$_{it}$	-0.065	0.000	-0.059	0.000	
[Ln(number of developers)$_{it}$]^2	0.040	0.000	0.050	0.000	p-value= 0.000
Ln(age)$_{it}$	0.448	0.000	0.406	0.000	
[Ln(age)$_{it}$]^2	-0.096	0.000	-0.090	0.000	
Constant	-0.258	0.000	-0.170	0.000	
R^2 within		0.050		0.047	
R^2 between		0.251		0.336	
R^2 overall		0.166		0.220	

Table 4.5 Hausman specification test for equation (3).

Dependent variable: $\text{Ln(number of downloads)}_{it}$	Fixed effect estimation		Random effect estimation		Hausman's test
	Coeff.	p	Coeff.	p	
$\text{Ln(number of forum posts)}_{i,t-1}$	0.120	0.000	0.155	0.000	
$\text{Ln(number of activities initiated)}_{i,t-1}$	0.189	0.000	0.238	0.000	
$\text{Ln(number of activities completed)}_{i,t-1}$	0.041	0.000	0.043	0.000	
$\text{Ln(number of new versions)}_{i,t-1}$	0.311	0.000	0.374	0.000	$\chi^2(7) = 14307.14$
$\text{Ln(cumulative number of downloads)}_{i,t-1}$	0.522	0.000	0.616	0.000	p-value= 0.000
Ln(age)_{it}	0.350	0.000	0.016	0.767	
$[\text{Ln(age)}_{it}]^2$	-0.316	0.000	-0.278	0.000	
Constant	3.293	0.000	3.588	0.000	
R^2 within		0.197		0.194	
R^2 between		0.689		0.711	
R^2 overall		0.626		0.644	

After eliminating the individual specific effects, there is still a possible problem of serial correlation within idiosyncratic errors ε_{it}, ν_{it}, and ξ_{it}. Even though this serial correlation does not make the estimated coefficients biased, it may cause bias in standard errors (Wooldridge 2001). Therefore, standard errors are estimated by Huber-White sandwich variance component estimation (Wooldridge 2009; Arellano 2003) which is consistent when there is serial correlation in ε_{it}, ν_{it}, and ξ_{it} and when the residuals are not identically distributed over the panel (heteroskedasticity). In short, equations (1), (2), and (3) are estimated by a fixed effect estimation method and the standard errors are corrected by Huber-White sandwich variance component estimation. This method yields consistent estimated coefficients and robust estimated standard errors. Variance inflation factors are

also calculated to make sure that multicollinearity is not a problem. The next chapter presents and discusses the results of the estimation of the three equations.

CHAPTER 5

FINDINGS

The results of equations (1), (2), and (3) are presented in Tables 5.1, 5.2, and 5.3. Specifically, Table 5.1 presents the effects of the number of forum posts, the number of downloads, their interaction, and control variables on the number of development activities initiated. Table 5.2 presents the effects of the number of forum posts, the number of downloads, their interaction, and control variables on the number of development activities completed. Table 5.3 presents the effects of the number of forum posts, the number of development activities initiated, the number of development activities completed, and control variables on the number of downloads. The results in these three tables underlie my discussion about the role of online UGC in the initiation and completion of product development activities and the interrelationship between product development and product adoption processes.

5.1 The Initiation of Product Development Activities

The results of equation (1) without and with the interaction between the number of downloads and the number of forum posts are presented in Table 5.1. The F test shows that including the interaction improves the goodness-of-fit of the equation significantly.

Table 5.1 The effects of online UGC and downloads on development activities initiated.

Dependent variable: Ln(number of activities initiated)$_{it}$	Without interaction		Full estimation	
	Coefficients	p-values	Coefficients	p-values
Ln(number of forum posts)$_{i,t-1}$	**0.145**	0.000	**0.093**	0.000
Ln(number of downloads)$_{i,t-1}$	**0.025**	0.000	**0.026**	0.000
Ln(number of forum posts)$_{i,t-1}$* Ln(number of downloads)$_{i,t-1}$			**0.022**	0.000
Ln(number of developers)$_{it}$	-0.016	0.526	-0.016	0.541
[Ln(number of developers)$_{it}$]^2	**0.026**	0.036	**0.026**	0.041
Ln(age)$_{it}$	**0.580**	0.000	**0.577**	0.000
[Ln(age)$_{it}$]^2	**-0.114**	0.000	**-0.115**	0.000
Constant	-0.241	0.058	-0.229	0.070
R^2 within		0.064		0.068
R^2 between		0.378		0.382
R^2 overall		0.276		0.283
F	2923 > 6.63 (0.01 critical value)			

Table 5.2 The effects of online UGC and downloads on development activities completed.

Dependent variable: Ln(number of activities completed)$_{it}$	Without interaction		Full estimation	
	Coefficients	p-values	Coefficients	p-values
Ln(number of forum posts)$_{i,t-1}$	**0.105**	0.000	**0.070**	0.000
Ln(number of downloads)$_{i,t-1}$	**0.008**	0.000	**0.009**	0.000
Ln(number of forum posts)$_{i,t-1}$* Ln(number of downloads)$_{i,t-1}$			**0.015**	0.000
Ln(number of incomplete activities)$_{i,t-1}$	**0.218**	0.000	**0.220**	0.000
[Ln(number of incomplete activities)$_{i,t-1}$]^2	**-0.035**	0.000	**-0.036**	0.000
Ln(number of developers)$_{it}$	**-0.065**	0.010	**-0.065**	0.010
[Ln(number of developers)$_{it}$]^2	**0.040**	0.001	**0.040**	0.001
Ln(age)$_{it}$	**0.451**	0.000	**0.448**	0.000
[Ln(age)$_{it}$]^2	**-0.096**	0.000	**-0.096**	0.000
Constant	**-0.264**	0.019	**-0.258**	0.021
R^2 within		0.048		0.050
R^2 between		0.242		0.251
R^2 overall		0.158		0.166
F	2871 > 6.63 (0.01 critical value)			

Table 5.3 The effects of online UGC and development activities on downloads.

Dependent variable: Ln(number of downloads)$_{it}$	Coefficients	p-values
Ln(number of forum posts)$_{i,t-1}$	**0.120**	0.000
Ln(number of activities initiated)$_{i,t-1}$	**0.189**	0.000
Ln(number of activities completed)$_{i,t-1}$	**0.041**	0.000
Ln(number of new versions)$_{i,t-1}$	**0.311**	0.000
Ln(cumulative number of downloads)$_{i,t-1}$	**0.522**	0.000
Ln(age)$_{it}$	**0.350**	0.011
[Ln(age)$_{it}$]^2	**-0.316**	0.000
Constant	**3.293**	0.000
R^2 within		0.197
R^2 between		0.689
R^2 overall		0.626

The full estimation results show that the number of forum posts has a positive impact on the number of new product development activities initiated in the next period, supporting the main-effect relationship in hypothesis 1. This is consistent with the theory that (1) the content of forum posts provides developers with useful information for the initiation of product development activities *and* (2) developers capitalize on the content of forum posts to initiate development activities for their products.

In addition, the number of downloads increases the number of new product development activities initiated in the next period, supporting hypothesis 3. This positive effect of the number of downloads on the number of development activities initiated is consistent with the theory that developers take the number of downloads in the past as a signal of future adoption and then adjust their efforts in initiating new development activities in order to meet their users' expectation in the future. This signaling mechanism helps developers reduce the double-sided adverse selection problem.

Moreover, the interaction between the number of forum posts and the number of downloads has a positive effect on the initiation of development activities, supporting hypothesis 7. This is consistent with the theory that forum posts contain useful information for product development activity initiation and the number of downloads motivates developers to capitalize on this source of information to start new development activities. The more the downloads, the more motivated the developers are to capitalize on the information contained in the forum posts to start new development activities.

While the number of developers has no significant impact on the number of development activities initiated, its quadratic term has a positive impact on the initiation of development activities. Overall, the number of developers has an increasing impact on the number of development activities initiated. There is no declining return to scale here, perhaps because starting a new development activity does not require much co-ordination.

The effect of project age on the initiation of development activities reflects a typical project life cycle. The number of development activities initiated increases over time and then declines when the project reaches a mature point when little improvement can be made or when developers lose interest in the project.

5.2 The Completion of Product Development Activities

The estimation results of equation (2) without and with the interaction between the number of downloads and the number of forum posts are presented in Table 5.2. The F test shows that including the interaction improves the goodness-of-fit of the equation significantly.

The effects of the number of forum posts, the number of downloads, and their interaction on the number of development activities completed (Table 5.2, full estimation) are similar to those on the number of development activities initiated. The number of forum posts has a positive impact on the number of product development activities completed in the next period, supporting the main-effect hypothesis 2. This is consistent with the theory that (1) the content of forum posts provides developers with useful solutions to complete current product development activities *and* (2) developers capitalize on the content of forum posts to complete development activities for their products.

Besides, the number of downloads increases the number of product development activities completed in the next period, supporting hypothesis 4. This positive effect of the number of downloads on the number of development activities completed provides further evidence to support the theory that developers take the number of downloads in the past as a signal of future adoption and then adjust their efforts in completing product development activities. This evidence supports the signaling argument even more strongly than the positive effect of the number of downloads on the number of development activities initiated because, unlike initiating a development activity, completing a development activity requires much more effort and commitment. Hence, the number of downloads in the previous period is a signal of future adoption which is strong enough to commit developers to completing development activities. The positive effect of the number of downloads on both the initiation and completion of development activities demonstrates that the signaling mechanism can help developers reduce the double-sided adverse selection problem.

Moreover, the interaction between the number of forum posts and the number of downloads has a positive effect on the completion of development activities, supporting hypothesis 8. Hence, downloads make development activity completion more likely ‘over and above’ the main effect of forum posts. This supports further the theory that forum posts contain useful solutions for product development activity completion and the number of downloads motivates developers to capitalize on this source of solutions to complete development activities. The more the downloads, the more motivated the developers are to capitalize on the solutions contained in the forum posts to complete development activities.

The number of incomplete activities - activities that have been started but not yet completed - at the end of the previous period also has an effect on the number of development activities completed in the current period. While the number of incomplete activities at the end of the previous period increases the number of development activities completed in the current period, its quadratic term has a negative impact on the completion of development activities. Basically, developers are trying to complete these incomplete activities. The more incomplete activities, the more they can complete in the period, but too many incomplete activities might discourage them from finishing those activities.

Unlike the impact of the number of developers on the initiation of development activities, the number of developers has a negative impact on the number of development activities completed, but its quadratic term has a positive impact on the completion of development activities. This suggests that the co-ordination problem arises very early. When more developers join a project, they slow down the completion of the current

development activities (controlling for initiations via the incomplete activities variable). However, the positive impact of a large team becomes stronger than the negative impact of co-ordination problems when many developers join the team. This phenomenon can be explained by two factors. First, in contrast to starting a new development activity which does not require much coordination, completing a current development activity requires a lot of coordination. Second, loosely organized development teams which are typical in open source software development projects are not effective in co-ordination but gain advantages by the number of developers.

Again, the effect of project age on the completion of development activities reflects a typical project life cycle. The number of development activities completed increases over time and then declines when the project reaches a mature point when there is little improvement that can be made or when developers lose interest in the project.

5.3 Product Adoption

Table 5.3 presents the effects of forum posts, the number of development activities initiated, the number of development activities completed, and other control variables on the number of product downloads. In addition to the effects on development activities, forum posts also affect the number of downloads positively, confirming the results of extant studies on the effect of online UGC on product adoption (e.g., Chen, Wu, and Yoon 2004; Chevalier and Mayzlin 2006; Dellarocas, Zhang, and Awad 2007; Li and Hitt 2008; Liu 2006; Zhu and Zhang 2010; Ho-Dac, Carson, and Moore 2011).

An increase in either the number of development activities initiated or the number of development activities completed drives more downloads in the next period, supporting

hypotheses 5 and 6. Specifically, a 1% increase in the number of development activities initiated (or 0.0203 activity initiated on average) leads to a 0.189% increase in the number of downloads (or 3.565 downloads on average) while a 1% increase in the number of development activities completed (or 0.01696 activity completed on average) leads to a 0.041% increase in the number of downloads (or 0.773 download on average). In other words, on average, one more new development activity initiated leads to an increase of 175.616 downloads while one more completed development activity leads to an increase of 45.578 downloads.

Note that the effect of new versions released on the number of downloads has been controlled for. Including the number of new versions of a product released in equation (3) practically controls the effect of product improvement on the number of downloads because all of the improvements are not realized until a new version is released. Users do not benefit from product improvement when a development activity is initiated or even when that activity is completed. Only when the improvement from the completion of that activity has been integrated into the product in a new released version, can users benefit from the improvement by downloading and using the new version. In fact, the effect of product improvement on the number of downloads is significantly positive as demonstrated by the positive coefficient of the number of new versions released (Table 5.3).

Therefore, the effects of both the initiation and completion of product development activities on the number of downloads are not due to product improvement but rather due to users' expectations for the commitment of developers to developing the product. In other words, the results suggest that users take the initiation and completion of product

development activities as signals of future commitment and then adjust their downloading accordingly. It is straightforward that users take the number of development activities completed as a signal of developers' commitment to developing the product in the future and then increase their downloading accordingly. The completion of a development activity requires a sizable amount of investment from the developers so it is a credible signal. Moreover, users may expect that the improvement from the completion of that activity will be included in the next version, even if they choose to begin their involvement with the product in its current version.

However, the impact of the initiation of development activities on the number of downloads is more intriguing. First, the initiation of a development activity is not as credible as the completion of a development activity as a signal of developers' commitment because the former does not require much investment from the developers. Second, the initiation of development activities indicates the existence of problems which need to be resolved and/or features that need to be added for the product to function properly. Those problems increase users' perceived risk associated with the product. Because users' perceived risk has an adverse effect on adopting a product (Dowling and Staelin 1994), users should hold off on downloading and wait for the improvement when they see a development activity initiated. The fact that the number of development activities initiated has a positive impact on the number of downloads in the next period suggests that (1) the initiation of a development activity is a credible signal of developers' commitment to developing the product from the users' viewpoint and (2) this signaling effect is stronger than the hold off effect caused by the risk associated with the initiation of that activity.

In short, the positive effects of the number of development activities initiated and the number of development activities completed on the number of downloads in the next period suggests that the signaling mechanism can help users overcome the double-sided adverse selection problem. This leads to another interesting finding. Since forum posts have a positive effect on both the initiation and the completion of development activities (which is amplified by the number of downloads), forum posts also influence the number of downloads indirectly through their role in development activities, in addition to their direct effect on the number of downloads.

The user base which is measured by the cumulative number of downloads at the end of the previous period also plays an important role in product downloading. A larger user base essentially leads to a better continuous adoption and new adoption as demonstrated by a positive coefficient of the cumulative number of downloads.

Similar to development activities, product downloads are also affected by project age through a typical project life cycle. The number of downloads increases over time as the product is evolving and then declines when the user base reaches a mature point when there are no more new adopters or when users lose interest in the product. The next chapter summarizes the main findings and discusses the implications of this study.

CHAPTER 6

CONCLUSION AND IMPLICATIONS

This study investigates how online UGC influences the product development process and whether product adoption and product development activities might serve as a signal of future adoption and a signal of commitment to developing the product, respectively. These influence and signaling effects increase market efficiency, which benefits not only developers and users but also platform owners. After summarizing the main findings, I discuss the implications for users, developers, and platform owners. Even though this study was conducted on data from an open source software development platform, the implications are not limited to the context of OSS but can be generalized into contexts where UGC are available and product development and product adoption processes are observable.

6.1 Conclusion and Theoretical Implications

The results from the estimation of the two equations which model the number of development activities initiated and the number of development activities completed show that the number of forum posts has a positive effect on the initiation and the completion of development activities. Moreover, its interaction with the number of last

period downloads has a positive effect on both the initiation and the completion of development activities. Therefore, it appears that online UGC in online forums can provide developers with product development ideas and solutions. It also suggests that how much developers capitalize on those ideas and solutions to start and to complete product development activities depends on their motivation to do so. That motivation is associated with past product adoption. The more downloads in the last period, the more online UGC is utilized by developers to initiate and to complete their development activities.

In addition to motivating developers to capitalize on online UGC to initiate and complete development activities, the number of downloads in the last period also increases the initiation and completion of product development activities directly. This is because it indicates that users are interested in the product and signals users' likely future downloads, motivating developers to initiate and to complete development activities in general. Therefore, developers tend to invest more to develop those products that receive more downloads in the previous period. This signaling effect helps provide assurance reducing the developers' adverse selection problem.

On the users' side, the results from the estimation of the equation which models the number of downloads show that both the number of development activities initiated and the number completed in the last period increase the number of downloads after controlling for the product improvement effect by the number of new versions released. This suggests that the number of development activities initiated and the number of activities completed indicate developers' commitment to the product development and signal developers' future development and product support so they motivate users to

download the product. Therefore, users are more likely to download if the product has increased the number of development activities initiated or completed in the previous period. This signaling effect helps assure users' reducing the adverse selection problem.

In short, the double-sided adverse selection problem can be eliminated if developers can observe the product adoption rate in the last period and users can observe the initiation and the completion of development activities in the last period. Hence, transparency between developers and users creates a positive feedback loop between product development and product adoption.

Besides these main results, the estimation of the three equations reveals several other findings. As mentioned above, the product improvement effect on the number of downloads is controlled by the number of new versions released. In fact, all product improvements can only be realized after they are integrated into a newly released version. Neither the initiation nor the completion of development activities actually adds anything to the product. The estimation of the last equation shows that the number of new versions released in the last period has a positive effect on the number of downloads.

In addition to the effects on development activities, the number of forum posts also has a positive impact on the number of downloads. This finding is consistent with the results of the extant literature on online UGC that online UGC leads to more product purchases (e.g., Chevalier and Mayzlin 2006; Li and Hitt 2008; Liu 2006; Zhu and Zhang 2010; Ho-Dac, Carson, and Moore 2011).

The number of incomplete development activities at the end of the last period has a positive effect on the completion of development activities but this effect is declining to scale. That means the more incomplete development activities to work with, the more

activities the developers complete but too many incomplete activities discourage them from finishing those activities.

The finding on the impact of the number of developers on development activities is somewhat unexpected. While the more developers lead to more development activities initiated, the number of developers has a negative effect on the number of development activities completed in a period but this negative effect is declining to scale. The difference between the initiation and the completion of development activities is that initiating a development activity does not requires much effort or coordination but completing it needs much more effort and coordination. This finding reveals that development teams in OSS projects may suffer from coordination problems so that their productivity decreases when more developers join the team. However, the team productivity increases when the advantage of having more developers out weights the coordination problem. This pattern might be characterized by an OSS setting in which development teams are loosely organized. Hence, it may not be generalizable to other settings.

The numbers of development activities initiated and development activities completed first increase and then decline over time when the project reaches a mature point where developers lose interest in the project or there is little improvement that can be made. Similarly, the number of downloads first increases and then declines over time when the project reaches a mature point where users lose interest in the project or there are no more new adopters. This pattern reflects a typical project life cycle.

6.2 Managerial Implications

Since this study investigates the behaviors of both developers and users, specifically the effects of online UGC on product development and adoption as well as the interrelationship between product development and adoption processes, it implies applications for users, developers, as well as platform owners who want to design a competitive ecosystem to attract and satisfy both users and developers. Several implications are discussed in the following subsections.

6.2.1 How Can Users Influence the Product Development Process?

It is expected that users want to have the product they need which operates in the way they like. One way to have that product is to influence the product development process. This study demonstrates that users can influence the product development process. They can provide product development ideas and solutions through their conversations in online forums. However, developers need to capitalize on those ideas and solutions to initiate and complete development activities. Otherwise, those ideas and solutions would not be realized. Fortunately, users can motivate developers to do so by purchasing and using the product. When users purchase and use a product, they signal their interest in the product so developers will initiate and complete more development activities in general. In short, users can shape the product development process through online UGC and purchasing behavior so that developers provide their desired products.

6.2.2 Developers Should Capitalize on Online UGC to Develop Products

Online UGC is shown to be a good source of product development ideas and even product development solutions. Developers can speed up their product development process by capitalizing on those ideas to initiate development activities which could improve the product in the way that users want. They can also utilize the solutions in online UGC to complete their current development activities. Doing so speeds up the whole product development process which in turn promotes product adoption through product improvement and signaling effect.

6.2.3 Should I Let Users Know What I Am Doing?

The results of this study imply that product development and product adoption processes influence each other positively due to the signaling effects which remedy both sides of the double-sided adverse selection problem. Usually, developers know how many users buy the product and when so the developers' adverse selection problem is not an actual issue. However, users usually do not know whether and what developers are working on so the users' adverse selection problem is often a problem in reality. To overcome that problem, developers should reveal what they are working on to secure current and future adoption.

In contrast, revealing what developers will not work on also works but in the opposite direction. Therefore, developers must be very careful in this signaling game because telling people what you are working on may imply what you are not committed to. For example, after Nokia announced that it would make the Windows Phone its future platform, many people held off on cell phone upgrades to wait for Nokia Windows

phones. Nokia's Symbian phones lost sales and none of its new Symbian phones sales took off after that announcement.

6.2.4 Why Do Upgrade Promises Work?

Recently, many companies sell their products with upgrade promises. For example, Asus launched its Eee Pad Transformer Prime which originally runs Android 3.2 Honeycomb with a promise that the product will receive a free upgrade to Android 4.0 Ice Cream Sandwich. While this practice could be criticized because consumers would not want to buy a half-baked product or an empty promise, this study suggests that upgrade promises work. Upgrade promises could signal that the companies are working on those upgrades so they could work similarly to the initiation of development activities in this study which leads to more adoption. Upgrade promises could secure current and future adoption as the initiation of development activities does. For example, consumers who want an Ice Cream Sandwich tablet may buy an Eee Pad Transformer Prime which runs Honeycomb with an Ice Cream Sandwich free upgrade promise rather than hold off their purchase until the promise is realized.

6.2.5 Application and Content Markets

Another current trend in the mobile industry is the formation of application and content markets. This trend soon may come to the computer industry. The idea is to bring all the content providers and users of a platform who were previously scattered into a single place. For example, Google launched Android Market to gather all Android application developers and users into one place. In March 2012, it launched Google Play

to unify its Android and Google TV platforms which gather all content providers and users into one place. Google Play is a much broader market than Android Market. It has applications, music, movies, books, and games, etc.

When bringing all content providers and users into one place, the transparency of the market increases. This study suggests that transparency facilitates the signals between users and developers of a product so it accelerates both the development and adoption processes of that product. In a more general sense, developers can observe the adoption behaviors of not only their products but also other products. Similarly, users can observe the development activities for not only their products but also other products. This helps both developers and users direct interest and effort to the right products, making the market more efficient.

6.2.5 Market Efficiency

As discussed in the previous paragraph, transparency increases market efficiency. In addition, the positive impact of online UGC on both product development and adoption processes implies that if platform owners can set up and manage an effective online UGC subsystem, they can accelerate the development and adoption processes of the products in their markets (or ecosystems), making their markets more efficient. Besides encouraging users to generate online content, platform owners also need to motivate developers to access the product development ideas and solutions in online UGC in order to translate them into product improvement. In summary, an effective online UGC subsystem and market transparency help a market become more efficient in both product development and adoption processes which, in turn, increases the competitiveness of the platform.

6.3 Further Research Directions

This study demonstrates that online UGC increases the initiation and the completion of product development activities. It suggests that (1) online UGC can provide developers with development ideas and solutions for the product development process and (2) developers capitalize on those ideas and solutions to initiate and to complete their development activities. However, this study does not distinguish the two issues. Further studies on the contents of online UGC and the behavior of developers may address these two issues separately.

In addition, this study lumps all types of product adoption together. Separating new adoption (first time buying) from continuous adoption (upgrading and updating) might shed more light on the product adoption process and the effects of online UGC and the development process on the adoption process. Similarly, separating the effects of the first post and the following posts of a thread; questions and answers; helps, suggestions, and complains, etc. on the product development process might provide us with more insights of how online UGC influences the product development process.

REFERENCES

Arellano, Manuel (2003), *Panel Data Econometrics*, Oxford: Oxford University Press.

Bayton, James A. (1958), "Motivation, Cognition, Learning-Basic Factors in Consumer Behavior," *Journal of Marketing*, 22, 282-289.

Bendapundi, Neeli and Robert P. Leone (2003), "Psychological Implications of Customer Participation in Co-Production," *Journal of Marketing*, 67 (April), 14-28.

Chen, Pei-Yu, Shin-yi Wu, and Jungsun Yoon (2004), "The Impact of Online Recommendations and Consumer Feedback on Sales," in proceedings of the *International Conference on Information Systems*, 25, 711–24.

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43 (3), 345-354.

Dahan, Ely, John R. Hauser (2002), "The Virtual Customer," *J. Product Innovation Management*, 19(September), 332–354.

Dellarocas, Chrysanthos, Xiaoquan Zhang, and Neveen F. Awad (2007), "Exploring The Value of Online Product Reviews in Forecasting Sales: The Case of Motion Pictures," *Journal of Interactive Marketing*, 21 (4), 23–45.

Dowling, G. and Staelin, R. (1994),"A Model of Perceived-Risk and Risk Handling Activity," *The Journal of Consumer Research*, 21 (June), 119-134.

Gale, Douglas (2001), "Signaling in Markets with Two-Sided Adverse Selection," *Economic Theory*, 18 (2), 391-414.

Gao, Yongquin, Vince Freeh, and Greg Madey (2003), "Analysis and Modeling of the Open Source Software Community," *North American Association for Computational Social and Organizational Science*, Pittsburgh, PA.

Grewal, Rajdeep, Gary Lilien, and Girish Mallapragada (2006), "Location, Location, Location: How Network Embeddedness Affects Project Success in Open Source Systems," *Management Science*, 52 (July), 1043-1056.

Griffin, Abbie and John R. Hauser (1993), "The Voice of the Customer," *Marketing Science*, 12, 1 (Winter), 1-27.

Hars, Alexander and Shaosong Ou (2002), "Working for Free? Motivating for Participation in Open-Source Projects," *International Journal of Electronic Commerce*, 6 (Spring), 25-39.

Hausman, J. A. (1978), "Specification Tests in Econometrics," *Econometrica*, 46, 1251–1271.

Ho-Dac, Nga N., Stephen J. Carson, and William L. Moore (2011), "Online Customer Reviews and Product Sales: The Role of Brand Equity," *American Marketing Association Winter Educators' Conference*, Austin, TX.

Karniouchina, Ekaterina V. (2011), "Impact of Star and Movie buzz on Motion Picture Distribution and Box Office Revenue," *International Journal of Research in Marketing*, 28(1), 62-74.

Lakhani, K. R., and Eric von Hippel (2003), "How Open Source Software Works: 'Free' User-to-User Assistance," *Research Policy*, 32(6), 923-943.

Lakhani, Karim R. and Robert G. Wolf (2005), "Why Hackers Do What They Do: Understanding Motivation and Effort in Free/Open Source Projects," in *Perspectives on Free and Open Source Software*, Joseph Fuller, Brian Fitzgerald, Scott A. Hissam and Karim R. Lakhani, ed. Cambridge, MA: MIT Press, 3-22.

Lerner, Josh. and Jean Tirole (2002), "Some Simple Economics of Open Source," *Journal of Industrial Economics*, 50 (June), 197–234.

Li, Xinxin and Lorin M. Hitt (2008), "Self Selection and Information Role of Online Product Reviews," *Information Systems Research*, 19 (4), 456-474.

Lilien, Gary L., Pamela D. Morrison, Kathleen Searls, Mary Sonnack, Eric Von Hippel (2002), "Performance Assessment of the Lead User Idea-Generation Process for New Product Development," *Management Science*, 48(8), 1042–1059.

Liu, Yong (2006), "Word of Mouth for Movies: Its Dynamics and Impact on Box Office Revenue," *Journal of Marketing*, 70 (3), 74-89.

Luan, Jackie Y. and Scott A. Neslin (2009), "The Development and Impact of Consumer Word of Mouth in New Product Diffusion," Working Paper, Tuck School of Business at Dartmouth.

MacInnis, Deborah J. and Bernard J. Jaworski (1989), "Information Processing from Advertisements: Toward an Integrative Framework," *Journal of Marketing*, 53, 1–23.

MacInnis, Deborah J., Christine Moorman, and Bernard J. Jaworski (1991), "Enhancing and Measuring Consumers' Motivation, Opportunity, and Ability to Process Brand Information from Ads," *Journal of Marketing*, 55, 32 –53.

Madey, Gregory (2012), *The SourceForge Research Data Archive*, University of Notre Dame, (accessed February 05), [available at http://srda.cse.nd.edu/].

Madey, Gregory, Vince Freeh, and Renee Tynan (2002a), "The Open Source Software Development Phenomenon: an Analysis Based on Social Network Theory," *Americas Conference on Information Systems*, Dallas, TX.

Madey, Gregory, Vince Freeh, and Renee Tynan (2002b), "Understanding OSS as a Self-Organizing Process," *24th International Conference on Software Engineering*, Orlando, FL.

Mallapragada, Girish, Rajdeep Grewal, and Gary Lilien (2008), "Born to Win? How Network Structure at Founding Affects the Success of Open Source Development Projects," *ISBM Report*, 3-2008, Pennsylvania State University, University Park, Pennsylvania.

Mallapragada, Girish, Rajdeep Grewal, and Gary Lilien (2011), "User-Generated Open Source Products: Founder's Social Capital and Time-to Product-Release," working paper, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina.

Moorman, Christine, Roland T. Rust (1999), "The Role of Marketing," *Journal of Marketing*, 63(October Special Issue), 180–197.

Mundlak, Yair (1978), "On The Pooling of Time Series and Cross Section Data," *Econometrica*, 46 (1), 69-85.

Nambisan, S. (2002) "Designing Virtual Customer Environments for New Product Development: Toward a Theory," *Academy of Management Review*, 27(3), 392-413.

Nambisan, Satish and Robert A. Baron (2009), "Virtual Customer Environments: Testing a Model of Voluntary Participation in Value Co-creation Activities," *Journal of Product Innovation Management*, 26, 388–406.

O'Hern, Matthew S., Aric Rindfleisch, Kersi D. Antia, and David A. Schweidel (2011), "The Impact of User-Generated Content on Product Innovation," *SSRN Working Paper Series*.

Poetz, Marion K. and Martin Schreier (forthcoming), "The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?" *Journal of Product Innovation Management*.

Raymond, Eric S. (1999), *The Cathedral and The Bazaar*. North Sebastopol, CA: O'Reilly.

Riley, J. (1979), "Informational Equilibrium," *Econometrica*, 47, 331-359.

Singh, Param V. (2010), "The Small World Effect: The Influence of Macro Level Properties of Developer Collaboration Networks on Open Source Project Success," *ACM Transactions of Software Engineering and Methodology*, 20(2), 6:1-6:27.

Singh, Param V. and Yong Tan (2011), "Developer Heterogeneity and Formation of Communication Networks in Open Source Software Projects," *Journal of Management Information Systems*, 27(3), 179-210.

Singh, Param V., Yong Tan, and Nara Youn (2011), "A Hidden Markov Model of Developer Learning Dynamics in Open Source Software Projects," *Information Systems Research*, 22(4).

Singh, Param V., Yong Tan, and Vijay S. Mookerjee (2007), "Social Capital, Structural Holes and Team Composition: Collaborative Networks of the Open Source Software Community," in *Proceedings of the 28th International Conference on Information Systems*.

Singh, Param V., Yong Tan, and Vijay s. Mookerjee (2011), "Network Effects: the Influence of Structural Social Capital on Open Source Project Success," *Management Information Systems Quarterly*, 35(4), 813-829.

Thomke, Stefan and Eric von Hippel (2002), "Customers as Innovators: A New Way to Create Value," *Harvard Business Review*, 80 (April), 74-81.

Urban, Glen L., John R. Hauser (2004), ""Listening In" To Find and Explore New Combinations of Customer Needs," *Journal of Marketing*, 68(April), 72–85.

von Hippel, Eric (1978), "Successful Industrial Products from Customer Ideas," *Journal of Marketing*, 42(1), 39-49.

von Hippel, Eric (1988), *The Sources of Innovation*, New York: Oxford University Press.

von Hippel, Eric (2001), "User Toolkits for Innovation," *J. Product Innovation Management*, 18(4), 247–257.

von Hippel, Eric (2005), *Democratizing Innovation*. Cambridge, MA: MIT Press.

Wilson, C (1977), "A Model of Insurance Markets with Incomplete Information," *Journal of Economic Theory*, 16, 167-207.

Wooldridge, Jeffrey M. (2001), *Econometric Analysis of Cross Section and Panel Data*. Cambridge, MA: MIT Press.

Wooldridge, Jeffrey M. (2009), *Introductory Econometrics: A Modern Approach*, 4th ed. Cincinnati, OH: South-Western.

Xu, Jin and Gregory Madey (2004), "Exploration of the Open Source Software Community," *North American Association for Computational Social and Organizational Science*, Pittsburgh, PA.

Xu, Jin, Scott Christley, and Gregory Madey (forthcoming), "Application of Social Network Analysis to the Study of Open Source Software", *The Economics of Open Source Software Development*, eds. Jürgen Bitzer and Philipp J.H. Schröder, Elsevier Press.

Xu, Jin, Yongqin Gao, Scott Christley, and Gregory Madey (2005), "A Topological Analysis of the Open Source Software Development Community," *38th Hawaii International Conference on System Sciences*.

Zhu, Feng and Xiaoquan (Michael) Zhang (2010), "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing*, 74 (March), 133–148.