Content Contributor Management and Network Effects in a UGC Environment

Kaifu ZHANG
Theodoros EVGENIOU
V. PADMANABHAN
Emile RICHARD
2011/11/DS/MKT
Content Contributor Management and Network Effects in a UGC Environment

Kaifu Zhang*

Theodoros Evgeniou**

V. Padmanabhan***

Emile Richard****

Acknowledgement: The generosity of 1000mercis Research Lab with data and managerial guidance has been invaluable. The assistance of Ilian Mihov on econometric issues is gratefully acknowledged. We thank Hubert Gatignon, Harald Hau, Andrew Stephen and seminar participants at INSEAD Brown Bag Seminar (2009) for their comments and suggestions. Finally, we greatly acknowledge the guidance of the Editor, Area Editor and two anonymous reviewers at Marketing Science.

* PhD Candidate in Marketing at INSEAD, Boulevard de Constance, 77305 Fontainebleau, France. Email: Kaifu.zhang@insead.edu

** Associate Professor of Decision Sciences and Technology Management at INSEAD, Boulevard de Constance, 77305 Fontainebleau, France. Email: Theodoros.evgeniou@insead.edu

*** John H. Loudon Chaired Professor of International Management at INSEAD, 1 Ayer Rajah Avenue, 138676 Singapore. Email: paddy.padmanabhan@insead.edu

**** PhD Candidate at Ecole Normale Supérieure de Cachan, 61 Avenue du Président Wilson, 94230 Cachan, France. Email: emile@millemercis.com

A Working Paper is the author’s intellectual property. It is intended as a means to promote research to interested readers. Its content should not be copied or hosted on any server without written permission from publications.fb@insead.edu Click here to access the INSEAD Working Paper collection
The success of any User-Generated Content (UGC) website depends crucially on its asset of content contributors. How firms should invest in the acquisition and retention of content contributors represents a novel question that is particularly important for these websites. We develop a vector autoregressive (VAR) model to measure the financial values of the retention and acquisition of both content contributors and content consumers. In our empirical application to a C2C marketplace, we find that contributor (seller) acquisition has the largest financial value due to their strong network effects on content consumers (buyers) and other content contributors. However, the wear-in of contributors’ financial values takes longer since the network effects take time to be fully realized. Our simulation-based studies (i) shed light on the value implications of ‘enhancing network effects’, and, (ii) quantify the revenue contributions of marketing newsletter campaigns. Our results indicate enhancing network effects in complementary ways can further increase the marginal benefits of acquisition and retention. We also find that simply tracking click-throughs may vastly underestimate the values of marketing newsletters, in our case by more than a factor of five, which may lead to suboptimal marketing effort allocations.

Keywords: UGC, Content Contributors, VAR, Lifetime Value, Acquisition, Retention, C2C Marketplace, Network Effects
1. Introduction

User-Generated Content (UGC) represents one of the most dynamic media forms on the Internet. Video sharing websites like YouTube, on-line encyclopedias like Wikipedia and Customer-to-Customer (C2C) marketplaces like eBay are rapidly taking over many of their traditional counterparts. The achievements of UGC websites often stand in stark contrast with their sizes. In 1998, the two-year old eBay made revenues of 4.7 million USD with merely 30 employees. Similarly, back in 2006, YouTube was able to serve over 100 million videos daily with 67 employees. The impressive growth of the UGC sector is a testimony to the power and impact of content contributors. In 2008, more than 82 million people in the US created on-line content. It is estimated that content contributors add about 20 hours of video materials every minute on YouTube. Interestingly, a pure C2C website such as the early eBay unlike a traditional B2C website such as Amazon cannot directly decide on its content (e.g., product assortment, inventory, retail display or even pricing). These factors can only be improved through careful acquisition and management of both sellers and buyers. Said differently, the financial success of any UGC website critically depends on its ability to actively manage and grow content contributors in addition to content consumers. Inability to do so was one of the reasons why firms such as Chemdex.com could not reach the point where they could turn a profit.

In this paper, we take a first step at helping firms assess the value of content contributors and

---

1 UGC is a relatively new phenomenon, and no ‘formal’ definition has been given in the literature. People may have different opinions on whether C2C websites like eBay can be classified as UGC applications. In this paper, we use the term UGC website as a rather general term referring to websites that do not provide their key ‘products’ by themselves. In our empirical application, we study a C2C website where all products are listed by users of the website. Although the framework of analysis we propose can be applied to other UGC contexts, we cannot claim the generalizability of our empirical results.

content consumers. In the process, we highlight the importance of understanding the network dynamics that underpin the economic value of contributors to these marketplaces. Our results provide guidance on the marketing actions that can be undertaken by the firm to acquire and retain their contributors and consumers. Empirically, we apply the proposed analysis to a large UGC type website, namely a C2C marketplace. Our framework of analysis can be similarly applied to other UGC settings.

Theoretically, a UGC website can be seen as a two-sided marketplace, with content contributors and content consumers as the two sides of the market, and the website as the platform where the two groups of users interact. The value of contributors is manifested through their network effects on both content consumers and other content contributors. The network effects can be either direct, where the different groups of users derive value from interactions, or indirect, where users benefit from a larger variety of content. Similarly, the presence of more content consumers also benefits the contributors. Thus, the acquisition of content contributors may lead both to more content consumers joining the website and to increased activity of existing consumers and contributors, which in turn translate into financial returns for the website. The retention of existing contributors or consumers will bring value to the website in a similar manner.

Measuring the value of customers has been of long standing interest in the Customer Lifetime Value (CLV) literature. Most traditional models in the CLV literature are based on individual-level cash flow analysis (Fader, Hardie, & Lee, 2005; Gupta et al., 2006; Venkatesan & Kumar, 2004). Cash flow analysis quantifies both the present value of a customer and how this value unfolds over time. Unfortunately, traditional CLV models
eminently suited for one-sided markets are inadequate for UGC settings since they do not accommodate the strong network effects amongst content contributors and content consumers which are idiosyncratic to UGC settings. To address this challenge, a number of recent studies have introduced *persistence modeling* (Dekimpe & Hanssens, 2004) techniques, notably the vector autoregression model (VAR), for the measurement of CLV (see Gupta et al., 2006 for a review).

This paper adds to the above literature and applies the VAR model to measure the average value of both contributors and consumers in a two-sided market setting. We quantify the values of both acquisition and retention of contributors and consumers, the magnitudes of network effects and detail their evolution over time. Based on these empirical estimates we perform VAR-based simulations to answer two key managerial questions: (1) Given the importance of network effects, how can a UGC firm measure the marginal return from specific marketing actions, so as to inform its marketing mix optimization? In our empirical context, we study how the firm can measure the cumulative revenue contribution of e-mail newsletters, beyond just tracking the clicks they get. (2) When the firm wants to enhance network effects, which directions should it focus on? For example, should the firm emphasize promoting the interactions between the consumers, or provide guidance to contributors so that they generate content that is more relevant for the consumers?

We use aggregate-level data from a C2C website where sellers serve as the primary content contributors by designing their virtual shops and showcasing the products, and buyers browse and purchase the products. This setting illustrates the methodology and a set of simulation analyses, which can be applied by any company interested in answering questions about
contributor management in UGC settings. Although the VAR model is not the only model to answer the above questions, it is well suited to our specific empirical setting and gives results that are consistent among themselves.

The key results from our analysis for this particular UGC website are summarized below:

- On a per-unit basis, seller acquisition has the largest cumulative financial value as well as network effects, followed by buyer acquisition, seller retention and finally buyer retention. These results, which may seem in contrast to conventional wisdom that often retention is preferred to acquisition (e.g., Reichheld & Sasser Jr, 1990; Sheth & Parvatiyar 1995; Reinartz & Kumar 2000), are a consequence of the significant network effects that characterize two sided markets. In fact, network effects constitute more than 80% of the total revenue contributions of new sellers and new buyers who bring more than 4-5 times revenue than returning sellers and buyers respectively.

- While new sellers and buyers are financially more valuable, it takes longer for their values to unfold, as these are a consequence of network dynamics. As a result, acquiring new sellers and buyers is even more valuable relative to retaining existing sellers and buyers when the firm has a longer planning horizon.

- The long term financial returns from marketing activities such as e-mail newsletters are considerably greater than their short term returns. This is again a consequence of the network effects that they help unleash among the sellers and buyers. For the firm we study, every thousand e-mail newsletters increase revenue by 6-10 Euros over the long term, while their short term marginal contribution is 1-2 Euros. Thus, simply tracing
click-throughs may underestimate the values of marketing newsletters by more than a factor of 5 and lead to suboptimal marketing effort decisions.

- Enhancing the network effects in different directions are complementary decisions in the traditional economics sense (Topkis 1998). The marginal return from each system increases when other systems are implemented or improved. In our empirical context, net of cost, enhancing the network effects of new sellers on other new sellers represents the most profitable direction.

The paper is organized as follows: in Section 2, we review the related literature. In Section 3, we describe our empirical setting, the model setup, specification tests, estimation and identification issues and model fit results. In Section 4, we present our main findings based on Impulse Response Functions (IRFs). In Section 5, we present the simulation results to answer the aforementioned managerial questions. In Section 6, we study the robustness of our results over time as well as across product categories. Finally, in Section 7 we conclude and discuss possible future research directions. In the Appendix, we present some additional analyses, including a Forecast Error Variance Decomposition and a heuristic which decomposes the total values of customers into direct and network components.

2. Literature Review

Our paper is related to several streams of literature. First, it is within the context of the emerging literature on UGC, social networks, and social commerce. Various studies have empirically examined the impact of on-line social media on firms. For example, the literature has examined the relationship between social network structure and the diffusion of products
(Goldenberg, Han, Lehmann, & Hong, 2009), the properties and impact of word of mouth (Trusov, Bucklin, & Pauwels, 2009; Villanueva et al., 2008; Chevalier & Mayzlin, 2006), user behavior (Mathwick, Wiertz, & De Ruyter 2007) and the implications for firm decisions, such as brand building (Muniz Jr and Jensen Schau 2005), and the impact of network formation on social commerce (Stephen & Toubia, 2010). Our paper examines an issue not discussed in this stream: the overarching consumer and contributor asset management problem.

Second, it is related to the marketing and economics literature on two sided markets. Two sided markets have been formally modeled in the theoretical IO literature (Armstrong, 2006; Rochet & Tirole, 2006). One qualitatively robust finding is that when network effects are sufficiently strong, the firm may find it beneficial to subsidize one group of users to foster growth (Fath & Sarvary, 2003). In the empirical literature, cross-group network effects have been found in various contexts (Nair, Chintagunta, & Dubé, 2004; Yao & Mela, 2008; Tucker & Zhang 2009). Our work not only measures the magnitudes of network effects in a UGC setting, but more importantly links network effects to the measurement of the financial values of contributors and consumers. In addition, we explore the managerial implications of enhancing network effects using simulations with empirical estimates from real data.

Third, our study is related to the marketing literature on Customer Life Time Value (CLV) and customer equity management. Marketing researchers have shown that customers are amongst the most important assets of a firm and we now have advanced methodologies for computing the value of this asset (see Gupta et al., (2006) for a recent review of the various modeling approaches). While the traditional approaches are usually based on individual level
models (Fader, Hardie, & Lee, 2005; Gupta et al., 2006; Venkatesan & Kumar, 2004), a number of recent studies have developed aggregate level models, using VAR models (Villanueva, Yoo, & Hanssens, 2008) and alternative methodologies (Gupta, Mela, & Vidal-Sanz, 2009, 2009; Hogan, Lemon, & Libai, 2003). Our study is closely related to Villanueva et al., (2008) both conceptually and methodologically. While Villanueva et al. (2008) measure the value of customers acquired from different channels, we consider the two-sidedness of the market and study both consumers and contributors. In addition, we study both acquisition and retention. Our paper is also conceptually close to Gupta et al., (2009) which derives measures of customer value based on a diffusion model. Unlike that work which considers only acquisition, we focus on both acquisition and retention and study the relative impact of network effects on the values of these two. Moreover, because we use a different methodology, VAR versus a diffusion model, we can study how all effects unfold over time and suggest, among others, different managerial actions depending on the time horizon of a firm. The diffusion model by Gupta et al. (2009) and our VAR model both suggest, among others, the existence of significant network effects.

Finally, methodologically our work relates to the literature that uses persistence modeling techniques such as VAR (Dekimpe & Hanssens, 1995; Pauwels, 2004; Luo, 2009; Trusov, Bucklin, & Pauwels, 2009; Stephen & Toubia, 2010). We follow the standard VAR framework outlined by Dekimpe & Hanssens (2004).
3. Data and Methodology

3.1 Data

We obtain weekly data from a major C2C website in Europe. The website has been operational since 2001, and is currently the leading website of its kind in several European countries with millions of users. The traded product categories range from books and DVDs to electronics, home appliances and furniture. On average, about eighty thousand transactions take place each week and the total transaction value across all products are several million Euros per week. The website is run completely on a C2C basis. In other words, the company itself does not list any product. Sellers come to the website, design their virtual shops, decide the assortment, describe it with detailed product descriptions and set prices. Buyers can browse, purchase and write product and seller reviews. Overall, the website resembles eBay with the exception that the goods are sold via a set price format instead of auctions. Based on discussions with the company’s managers, we designate sellers as the primary group of content contributors on the website.

Our data includes the number of new sellers, returning sellers, new buyers, returning buyers and the total commission earned by the website per week. The observation period lasts from mid-2005 to mid-2009, consisting of a total of 201 weeks. Table 1 summarizes some descriptive statistics. The variables are:

- Number of New Sellers: The number of new sellers who have sold a product for the first time in a given week.
- Number of Returning Sellers: The number of returning sellers who have sold at least one
product in a given week but also in previous weeks.

- **Number of New Buyers:** The number of registered buyers who have purchased a product for the first time in a given week.

- **Number of Returning Buyers:** The number of returning buyers who have purchased products in previous weeks, and have purchased at least one product in the given week.

- **Total Weekly Commission:** The total commission earned by the website in a given week.

  The total commission is a constant fraction of the revenue generated from buyer-seller transactions.

The time evolutions of the key variables are illustrated in Appendix A.6. We use the logged values for all the time series in the model. We denote these variables as \( tLOS_t \) (log value of the number of returning sellers at time \( t \)), \( tLNS_t \) (log value of the number of new sellers at time \( t \)), \( tLOB_t \) (log value of the number of returning buyers at time \( t \)), \( tLNB_t \) (log value of the number of new buyers at time \( t \)), and \( tLC_t \) (log value of commission at time \( t \)).

**Table 1:** Descriptive Statistics of the Data (Numbers Represent Weekly Aggregates)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev.</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of New Sellers</td>
<td>1589</td>
<td>374</td>
<td>852</td>
<td>2657</td>
</tr>
<tr>
<td>Number of Returning Sellers</td>
<td>22637</td>
<td>5917</td>
<td>12875</td>
<td>36235</td>
</tr>
<tr>
<td>Number of New Buyers</td>
<td>11401</td>
<td>2747</td>
<td>6548</td>
<td>21198</td>
</tr>
<tr>
<td>Number of Returning Buyers</td>
<td>32068</td>
<td>8800</td>
<td>17444</td>
<td>57250</td>
</tr>
<tr>
<td>Total Weekly Commission (Euros)</td>
<td>269449</td>
<td>70095</td>
<td>150057</td>
<td>508973</td>
</tr>
</tbody>
</table>

A number of caveats apply to our data. First, as in most UGC applications, the boundary between content contributors and content consumers is not clear-cut. For example, in YouTube, even those users who never make a video contribute to the website by leaving
comments, tagging videos and giving ratings. Similarly, in our C2C marketplace, buyers also contribute by writing product reviews. Second, like many website-based studies, we can only count sellers and buyers by registrations on the website, without knowing whether the same person in the real world registered under different IDs. For example, the same person may both buy and sell on the website under different buyer ID and seller ID or even possess multiple buyer IDs and seller IDs. This affects the interpretation of our results in Section 4. Third, by using aggregate level data, all our empirical results are limited to the average values on the entire seller/buyer population. That said, we investigate the heterogeneity in seller/buyer values in Section 6.

3.2 Model Selection and Estimation

We adopt a Vector Autoregression (VAR) model to capture the interdependent evolution of the variables of interest. The evolution of each variable (the numbers of buyers and sellers as well as total commission) are explained by the lag of itself and other variables. By treating each variable as potentially endogenous, the VAR model is particularly suitable to capture the dynamic and complex interdependence between the variables of interest without making stringent identification assumptions. Based on the estimated VAR parameters, simulation techniques can be applied to derive the long term impact of a shock in one variable on all the other variables.

Our analysis follows the standard procedure of VAR modeling, which consists of the following steps: (1) unit-root tests to determine the stationarity of the variables, (2) co-integration test when the variables are non-stationary, (3) deciding model specification
(VAR in levels, VAR in differences, or error-correction (VEC) forms) based on unit-root and co-integration tests results, (4) deciding model specification (number of lags and exogenous variables) based on information criteria. The above procedures are discussed in detail in Dekimpe & Hanssens (2004). Our final step (5) is imposing identification restrictions and deriving the Impulse Response Functions (IRFs).

We first perform the Augmented Dickey-Fuller (ADF) unit-root test to test the null hypothesis of a unit root and the KPSS test (Kwiatkowski, Phillips, Schmidt, & Shin, 1992) to test the null hypothesis of stationarity. The ADF test statistics range from -4.240 to -3.751, all below the 5% critical value of -3.437. Therefore, we reject the null hypothesis of unit-root at the 5% confidence level. The KPSS test statistics range from 0.078 to 0.114, all below the 5% critical value of 0.146. Hence, we cannot reject the null hypothesis of stationarity. We apply the iterative procedure recommended by Enders (2004) to decide whether a time trend should be included in the test. The procedure suggests that a time trend should be included in the ADF tests.

Taken together, the tests agree with each other and suggest that all the variables are trend stationary. Based on the above results, we estimate the following standard-form VAR model in levels (Dekimpe & Hanssens, 2004):

\[
\begin{bmatrix}
    \text{LOS}_t \\
    \text{LNS}_t \\
    \text{LOB}_{t-1} \\
    \text{LNB}_{t-1} \\
    \text{LC}_t
\end{bmatrix} = \bar{\alpha} + \bar{\gamma}t + \sum_{j=1}^{J} B_j \begin{bmatrix}
    \text{LOS}_{t-j} \\
    \text{LNS}_{t-j} \\
    \text{LOB}_{t-j} \\
    \text{LNB}_{t-j} \\
    \text{LC}_{t-j}
\end{bmatrix} + \sum_{k=1}^{4} \bar{v}_k S_{kt} + \sum_{q=2005}^{2009} \bar{\zeta}_q C_{qt} + \bar{\epsilon}_t
\]

(1)

We include the following exogenous variables in the system: a linear time trend \((t)\), seasonal dummies \((S_{kt})\), and dummies for Christmas shopping seasons \((C_{qt})\). Each \(B_j\) is a 5-by-5
matrix. For notational convenience, we index the variables as: 1=Returning Sellers, 2=New Sellers, 3=Returning Buyers, 4=New Buyers, 5=Commission. In the standard-form VAR model, since the right hand side only contains lagged endogenous variables, OLS yields consistent estimates. In Appendix A.1, we explain the underlying structural-form VAR model, which includes theoretically meaningful parameters.

We use the SBIC criterion (as in Dekimpe & Hanssens (1995) and Villanueva et al. (2008)) to select the model parameters. Based on the SBIC criterion, the optimal lag length is set to 1. The linear time trend captures the (unobserved) natural growth of the Internet market over time, a typically needed control for spurious correlations (see Villanueva et al., (2008), Trusov et al., (2009) for similar arguments). We allowed a quadratic term of time trend in the system, which was later excluded because it decreased the SBIC. We finally perform the Portmanteau test (Lütkepohl 2005) to test for residual autocorrelation and cannot reject the null hypothesis of white noise.

We identify the contemporaneous effects from a Cholesky decomposition of the Variance-Covariance Matrix of the residuals. This follows the widely used identification assumptions, sometimes called contemporaneous ordering (Villanueva et al., 2008) or orthogonalizing transformation (Luo, 2009). Based on discussions with the firm, in the basic model, we impose the following ordering of the variables: $$\text{LOS}_t, \text{LNS}_t, \text{LOB}_t, \text{LNB}_t, \text{LC}_t$$.

As a robustness check, we also used an alternative ordering, $$\text{LOS}_t, \text{LOB}_t, \text{LNS}_t, \text{LNB}_t, \text{LC}_t$$, which led to similar results.

We believe these two orderings are the only reasonable choices for the following reasons. First, it is reasonable to assume that the existing users should have contemporaneous effects on the new users but not vice versa. Second, the buyers can observe the number of product postings – which strongly correlates with the number of actual sellers – immediately, but the sellers cannot observe the number of active buyers in the same time period. Therefore, the number of sellers has
4. Empirical Findings

We report our VAR parameter estimates and $R^2$ of each equation in Appendix A.5. As is typically done in the literature, we do not interpret the individual parameters themselves, but focus on the IRFs (see Dekimpe & Hanssens (2004) for this argument and a formal definition of the IRF). The IRFs trace the change in the response variable over time when there is an unexpected shock\(^4\) to the impulse variable. The Cumulative IRF (CIRF) measures the total impact of the unexpected shock in the impulse variable on the response variable. We are interested in the following IRFs:

- When the number of new buyers/sellers is the impulse variable, and the total commission is the response variable, the CIRF measures the revenue contribution of an unexpected addition of a new buyer/seller. We call this the revenue contribution of one instance of acquisition.

- When the number of returning buyers/sellers is the impulse variable, and the total commission is the response variable, the CIRF measures the revenue contribution of an unexpected ‘reactivation’ of an existing buyer/seller. The ‘unexpected reactivation’ of a returning buyer/seller means that due to a shock that is not predicted by the explanatory variable, an otherwise inactive existing buyer/seller becomes active in buying/selling products – which is what we call one instance of retention.

- When the number of new buyers/sellers is the impulse variable, and the number of new contemporaneous effects on the number of buyers, but not vice versa. Finally, the numbers of both new and returning buyers/sellers should have contemporaneous effects on the commission, but not vice versa.\(^4\)We use the term ‘unexpected’ in its standard sense in the VAR literature. It means the shock is unpredicted by the exogenous variables, such as time trend and seasonal dummies.
*buyers/sellers* is the response variable, the CIRF measures the *additional users joining the website* over time when there is an unexpected addition of a new user. It is a cumulative measure of new users’ *network effects* on other new users.

- When the *number of returning buyers/sellers* is the impulse variable and the *number of new buyers/sellers* is the response variable, the CIRF measures a type of *network effects* created by the reactivation of an existing user.

- When the *number of new buyers/sellers* is the impulse variable, and the *number of returning buyers/sellers* is the response variable, the conceptual meaning of the CIRF is ambiguous. Following the acquisition of one user at time 0, we may observe more returning users in the subsequent periods. However, we cannot discern whether a returning buyer/seller is the lately acquired buyer/seller herself or an existing buyer/seller who had already joined the website before the acquisition. For this reason, we only focus on new users when we discuss the results about network effects in Sections 4 and 5.2. The network effects of new users on existing users are important but we can not measure them.

Since CIRFs are functions of time, they provide measures about the short term as well as the long term, cumulative effects. For ease of exposition, we transform the CIRF values from percentage terms (derived from the log formulation) back into Euro terms. The transformation is \( \frac{\Delta Y}{\Delta X} = \frac{\bar{Y}}{X\sigma_y} \Delta \ln X \) and is standard in the literature (e.g., Trusov et al., 2009) and is standard in the literature (e.g., Trusov et al., 2009). Finally we multiply the quantity by the mean levels of impulse variable

\[ \frac{\Delta \ln Y}{\Delta \ln X} \]

stands for the IRF derived from the log formulation, where the shock is of one standard deviation size. This number is normalized by the standard deviation \( \sigma_x \). Finally we multiply the quantity by the mean levels of impulse variable.
2009). This allows us to report the financial/network values of one additional new/returning buyer/seller. A similar analysis can be done using percentage terms (e.g. the impact of an unexpected 1% increase of the number of new/returning buyers/sellers). We focus on the actual numbers of customers as they are managerially more relevant.

4.1 Values of Acquisition versus Retention and of Sellers versus Buyers

Our first set of results answers the following question: how much cumulative value (direct cash flow and the indirect network value) does one instance of acquisition or retention, of a seller or a buyer, generate?

Table 2 presents the cumulative long term revenue contributions of buyer/seller acquisition/retention. The firm under study provided range estimates of the marginal acquisition/retention costs. These costs are: 12-60 Euros for seller acquisition, 5-25 Euros for buyer acquisition, 0-10 Euros for seller retention and 0-5 Euros for buyer retention. Together with these cost estimates, the results show that acquisition of sellers or buyers is more valuable than retention of the same type of users by more than a factor of 4-5 in this case. This result at first glance seems in marked contrast with findings in the CLV literature (e.g., Reichheld & Sasser Jr, 1990; Sheth & Parvatiyar 1995; Reinartz & Kumar 2000) wherein often retention is preferred to acquisition even when the costs are comparable. The logic for those results is based on two sided learning (i.e., returning customers tend to become more loyal, less price sensitive and easier to serve over time). Our results reveal that, once

(X) and response variable (Y) to translate the elasticity back into marginal effect.

6 The acquisition cost is partly estimated based on the advertising price on Google AdWord and the estimated conversion rate. The estimation of retention is trickier since retention involves substantial fixed cost investment (designing an Email newsletter campaign and product recommendation system, etc) but the marginal cost is low.
network effects are taken into account, acquisition becomes more valuable because newly acquired users (i.e., content contributors and consumers) generate network effects that have financial returns to the firm significantly beyond the cash flows they generate individually. It should be noted that the cumulative values of acquisition versus retention is firm and data set dependent and our results may not generalize to other data sets. But the essential point is that CLV measures for a business going forward need to consider both the direct financial returns from an individual buyer/seller as well as the additional financial contributions they generate through their network effects on other buyers and sellers.

Table 2\textsuperscript{7}: Cumulative Revenue Contributions of Buyers and Sellers (* indicates \(p<0.05\).)

<table>
<thead>
<tr>
<th></th>
<th>Cumulative Commission Increase</th>
</tr>
</thead>
<tbody>
<tr>
<td>One New Seller</td>
<td>242*</td>
</tr>
<tr>
<td>One New Buyer</td>
<td>94*</td>
</tr>
<tr>
<td>One Returning Seller</td>
<td>45*</td>
</tr>
<tr>
<td>One Returning Buyer</td>
<td>11</td>
</tr>
</tbody>
</table>

We next compare the cumulative values of one buyer versus one seller.\textsuperscript{8} We find that seller acquisition is more valuable than buyer acquisition on a per unit basis, and one instance of seller retention is more valuable than one instance of buyer retention. These results suggest the necessity of developing innovative marketing campaigns targeted at acquiring and retaining sellers – an issue that was previously ignored and is currently being actively pursued.

\textsuperscript{7} These numbers are robust to the change of variable ordering. The alternative ordering of variables \(LOS, LOB, LNS, LNB, LC\) leads to the following estimates (in the same order as in Table 2): 228, 92, 43, 15.

\textsuperscript{8} As discussed in Section 3.2, we cannot observe directly whether it is the same individual behind different buyer and seller IDs. An individual in the real world may both buy and sell online under multiple different IDs. According to the firm’s best estimates, about 50% of sellers are also buyers on the website, and 10% of buyers also sell.
explored by the company.

We perform significance tests based on Monte Carlo simulation (Lütkepohl 2005) with 250 simulations and find that the revenue contribution differences are significant. The test details and results are presented in Appendix A.8. In Appendix A.7, we plot the IRFs with 95% confidence intervals.

Overall, the revenue contribution estimates seem strikingly large compared to the acquisition and retention costs provided by the company. For example, the cost of acquiring one new seller is estimated to be 12-60 Euros while the revenue contribution of one seller is at 242 Euros. This suggests underinvestment by the firm on both acquisition and retention. Interestingly, the cost ranges are on the same order of magnitude as the individual revenue contributions (net of network effects) which we derive in Appendix A.3 as well as the traditional CLV estimates provided by the firm. This suggests that the firm is possibly trying to optimize its acquisition and retention efforts based on the cash-flow measures of lifetime values, but the under-appreciation of network effects leads to severe underinvestment.

Finally, we note that these results should not be interpreted as downplaying the importance of buyers/consumers relative to sellers/contributors. Although the individual revenue contribution of each buyer is small, the overall contribution of buyers is significant as they are a much bigger segment in size relative to the sellers. For example, when the number of new buyers is increased by 1% in a certain week, the total commission cumulatively increases by 4% (of the weekly amount). When the number of new sellers is increased by 1%, the commission only increases by 1.3% cumulatively.
4.2 Network Effects

We now consider the following question: how large are the network effects? For example, what is the impact of new sellers on new buyers, and vice versa?

We measure the network effects as the number of additional users joining the website following one instance of acquisition/retention using CIRFs, as defined above. Network effect is a general term referring to a complex set of more specific effects. For example, observational learning (Tucker & Zhang, 2009) takes place when the fact of some sellers joining the website leads to inference of the profitability of selling in this website, thereby inducing other sellers to join the site as well. New sellers joining the website also enrich the product offering, and lead to greater availability of the popular products. Similarly, buyers have network effects on other buyers and sellers through their product reviews and word-of-mouth (Chevalier & Mayzlin, 2006). Our measure of network effects captures the accumulation of all these effects.

Table 3: Network Effects. * indicates $p<0.05$.

<table>
<thead>
<tr>
<th></th>
<th>Additional New Sellers</th>
<th>Additional New Buyers</th>
</tr>
</thead>
<tbody>
<tr>
<td>One New Seller</td>
<td>2.8*⁹</td>
<td>10.8*</td>
</tr>
<tr>
<td>One New Buyer</td>
<td>0.7*</td>
<td>5.6*</td>
</tr>
<tr>
<td>One Returning Seller ('One instance of Retention')</td>
<td>0.2*</td>
<td>2.3*</td>
</tr>
<tr>
<td>One Returning Buyer ('One instance of Retention')</td>
<td>0.04</td>
<td>-0.8</td>
</tr>
</tbody>
</table>

Table 3 summarizes our results. The network effects reported here correspond to the

---

⁹ The cumulative CIRF is 3.8 when the number of new sellers is both the impulse and the response variable. Thus the additional number of new sellers (excluding the unexpected initial acquisition) equals 3.8-1=2.8.
converged cumulative effects, taken as the CIRF at the sixteenth week. For example, the network effects of new sellers on additional new buyers correspond to 10.8. This means that following the unexpected acquisition of one new seller, the number of new buyers joining the website will be 10.8 more compared to the scenario where the unexpected seller acquisition does not take place.

Table 3 indicates that the acquisition of new sellers has the largest network effects, which underpins the large financial value of seller acquisition highlighted earlier. Acquisition of new buyers has the second largest network effects. Retentions of sellers and buyers have smaller network effects. Discussion with the company reveals that the large differences in network effects between new and returning sellers/buyers may be explained by a seller/buyer’s propensity to engage in so called ‘non-transaction activities’. For example, during the first weeks after joining the website, sellers spend considerable effort in designing their virtual shops and similarly, buyers are more likely to write product reviews. While these activities do not generate transactions directly, they significantly increase the network benefits a new seller/buyer generates. On the other hand, returning sellers engage relatively less on these non-transaction activities.

The above findings reveal that, in agreement with Table 2, the most valuable events (seller and buyer acquisition) also have the largest network effects. This suggests that network effects are the key driver behind the large values of seller and buyer acquisition. The magnitudes of networks effects indicate that in our C2C marketplace, the valuation of sellers and buyers cannot be done through a simple cash flow analysis, where the impact of one
seller/buyer on other sellers and buyers is ignored. In fact, given that one new seller will bring in a total of 3.8 sellers, the cash flow generated by one seller is about 25% of her total revenue contribution. Similarly, the cash flow generated by one new buyer is less than about one sixth of her total revenue contribution. We further develop this reasoning in Appendix A.3 to decompose the values of buyers and sellers into a direct component and a network effects component.

The network effects of a new seller/buyer on existing users cannot be precisely quantified, for the reasons explained in the beginning of Section 4. However, our results do indicate the existence of such effects. Our IRFs show that when there is an unexpected unit shock to the number of new sellers at time 0, the number of returning sellers at time 1 increases by 1.7, at least 0.7 of whom must have joined the market before the acquisition takes place. By time 1, the number of returning buyers has increased by 3, at least 2.6 of whom must have already been in the marketplace before the seller acquisition since the acquisition only leads to about 0.4 additional new buyers at time 0. Therefore, existing sellers and buyers are indeed more likely to return to the website because of the acquisition of a new seller.

Traditionally, acquisition and retention are often posed as distinct managerial decisions between which the firm has to allocate its resources. With network effects the acquisition and retention decisions become, as expected, more interdependent. In some circumstances, acquisition of new users can be a way to retain existing users. We further develop this argument in Appendix A.4.

---

10 This observation is in accordance with the conclusions of (Gupta, Mela, & Vidal-Sanz, 2009).
4.3 Accumulation of Values over Time

We now move to the following question: how do the financial values and network effects realize over time?

Table 2 showed the long-run cumulative revenue contributions of seller/buyer acquisition/retention. We also obtain their short term contributions by examining the IRF in the 1st week. The short run revenue contributions are 29 Euros for one new seller, 14 Euros for one new buyer, 28 Euros for one returning seller and 9 Euros for one returning buyer. These indicate that the difference in revenue contribution between retention and acquisition is much smaller in the short run. For example, in the first week, retaining an existing seller generates about the same amount of revenues as one new seller (28 Euros vs. 29 Euros). In the long run (cumulatively), however, acquiring one seller brings in five times more revenue than retaining one seller.

To shed more light on how the values accumulate over time, we calculate the number of weeks it takes for the cumulative revenue contributions and network effects to wear-in. It takes 11 weeks for a newly acquired seller to realize 95% of her cumulative revenue contribution, and 6 weeks for a retained seller to realize 95% of her cumulative revenue contribution. The wear-in periods for both revenue contributions and network effects are shown in Table 4.

The results indicate that although buyer and seller acquisitions are more valuable in the long run, it takes longer for these impacts to be realized. What underlies the long wear-in periods of the value of seller and buyer acquisition? Network effects serve as one explanation. The
values of new sellers and buyers are composed of direct effects and even larger indirect network effects. For example, acquiring a new seller may bring in more sellers by word-of-mouth, and these sellers in turn attract more buyers by enriching the product variety of the website as a whole. These indirect effects may be large in magnitude, but they also take time to realize, as buyers and sellers need time to observe and react to the growth of the website.

**Table 4: Wear-in Periods of Revenue Contributions and Network Effects.**

<table>
<thead>
<tr>
<th></th>
<th>Revenue contributions reach 95% of their cumulative levels</th>
<th>Network effects reach 95% of their cumulative levels (new sellers, new buyers)</th>
</tr>
</thead>
<tbody>
<tr>
<td>One New Seller</td>
<td>11</td>
<td>(8, 11)</td>
</tr>
<tr>
<td>One New Buyer</td>
<td>12</td>
<td>(11, 12)</td>
</tr>
<tr>
<td>One Returning Seller</td>
<td></td>
<td></td>
</tr>
<tr>
<td>('One instance of Retention')</td>
<td></td>
<td>(6, 5)</td>
</tr>
<tr>
<td>One Returning Buyer</td>
<td></td>
<td></td>
</tr>
<tr>
<td>('One instance of Retention')</td>
<td></td>
<td>(1,1)</td>
</tr>
</tbody>
</table>

Logically, the complete wear-in of the network effects of a user must precede the complete wear-in of her value. For example, it takes 8-12 weeks for the network effects of new users to reach 95% of their cumulative level, and it takes approximately the same length of time for the revenue contributions of new buyers and new sellers to wear-in. The long wear-in periods for network effects indicate that network effects indeed serve as explanations for the long wear-in periods for the revenue contributions of buyer/seller acquisition. Taken together, the results in Tables 2, 3 and 4 establish a consistent positive relationship between the values,

---

11 The wear-in periods of network effects are shown in pairs: the first number stands for the wear-in period of the network effect on new sellers, the second number stands for the effects on new buyers.

12 Recall that the cumulative revenue contribution of one returning buyer at week 16 (Table 2) is not significantly different from zero. We report the time period it takes for the value to reach 95% of the maximal cumulative value which is significant.
network effects and lengths of wear-in periods for buyers and sellers. In short:

- There is a relation between a user’s revenue contributions, network effects and wear-in periods. Seller/buyer acquisitions have large values, large network effects, but it takes longer for their values to fully realize.

- The wear-in of network effects precedes the wear-in of revenue contributions.

In Appendix A.2, we also explore the above insights using the Forecast Error Variance Decomposition (Enders, 2004) technique. In Appendix A.4, we show that the differential wear-in periods of retention and acquisition have implications for the firm’s resource allocation decisions under different planning horizons. For example, a firm with a longer planning horizon should shift more resources to seller acquisition from seller retention compared to the case of a company with a shorter planning horizon.

5. Managerial Implications

We use VAR based simulations to highlight the implications for two sets of managerial actions – (1) Sending Marketing Newsletters and (2) Network Effect Enhancement. Optimizing the level of marketing mixes (in this case sending marketing newsletters) is an essential issue for any type of company, and C2C marketplaces or UGC firms are no exceptions. Enhancing network effects is a novel but central decision for C2C marketplaces and UGC based companies. Specifically, we consider the following:

1. Sending Marketing Newsletters:

- When the firm wants to optimize the number of marketing newsletters sent in each week,
how should it quantify the marginal benefit of sending more newsletters? We compare
the value of newsletters estimated with two different approaches.

2. Enhancing network effects:

- Given the importance of network effects, what is the marginal benefit of enhancing
  network effects? When the firm wants to enhance network effects, which directions
  should it focus on? For example, should the firm focus on promoting the interactions
  between the buyers, or on implementing WOM referral programs among the sellers?

5.1 Sending Marketing Newsletters

The results thus far have quantified the financial values of new and returning users. We now
examine the impact of the firm’s marketing actions on the size of the buyer/seller networks
and the firm’s commission. Optimizing the promotional activities is a problem of great
managerial interest to the firm and we focus in this section on marketing newsletters which
the firm considers as their primary marketing tool. We develop a simulation methodology to
calculate the marginal increase in revenue when 1000 marketing newsletters are sent. The
marginal revenue increase can be used together with the marginal cost information to inform
marketing mix optimization.\textsuperscript{13}

The marketing newsletters come in various formats and have rich content focusing on either
buyers or sellers. Each marketing newsletter campaign targets a designated subset of
registered buyers/sellers chosen by the firm. For example, buyer-targeted newsletters

\textsuperscript{13} Clearly, there are many other possible marketing actions an online marketplace can consider, such as sales contests, price
promotion (delivered through coupons), or online advertising targeted at potential buyers and sellers. Based on data
availability and significance for this firm, we use the case of marketing newsletters as an illustrative example, but a similar
analysis can be applied to other marketing mix contexts.
encourage purchase as well as word of mouth by emailing existing users coupon codes that they can use or forward to their friends. Similarly, seller-targeted newsletters provide sellers with tips on pricing, product assortment and virtual shop design. We obtain the weekly numbers of buyer-targeted and seller-targeted newsletters (as defined by the company) from the beginning of 2007 to the end of our observational period (a total of 137 observations). The average number of buyer-targeted newsletters during this period is 5895532 per week, and 655937 for the seller-targeted ones. We include the logged values as two additional exogenous variables $LSL_t, LBL_t$ in the VAR system, which leads to the following model specification:

$$\begin{bmatrix} LSL_t^j \\ LNS_t^j \\ LOB_t^j \\ LNB_t^j \\ LC_t^j \end{bmatrix} = \alpha + \beta_t + \sum_{j=1}^{J} B_j \begin{bmatrix} LSL_{t-j}^j \\ LNS_{t-j}^j \\ LOB_{t-j}^j \\ LNB_{t-j}^j \\ LC_{t-j}^j \end{bmatrix} + H \begin{bmatrix} LSL_t \\ LBL_t \end{bmatrix} + \sum_{k=1}^{4} \bar{\nu}_k S_{kt} + \sum_{q=2005}^{2009} \bar{\lambda}_q C_{qt} + \bar{\epsilon}_t$$

where $H$ is a 5x2 matrix capturing the ‘newsletter elasticities’ of the performance variables (numbers of new/returning buyers/sellers and commission) at their current levels. Our discussion with the firm reveals that $LSL_t$ and $LBL_t$ are not correlated with $\epsilon_t$, since the planning of a newsletter campaign (choice of the targeted audience, design of E-mail content etc) takes time and is finalized before week $t$. Thus $LSL_t, LBL_t$ can be justified as exogenous to the left hand side variables. We do not include the lagged numbers of newsletters based on the fact that almost none of the users click on a newsletter that was sent in the previous week. In addition, we test whether the omission of $LSL_t, LBL_t$ in model (1) leads to a

---

14 The firm started tracking the number of marketing letters at the beginning of 2007.
misspecification problem. 15

We start with the contemporaneous effects of marketing newsletters on the performance variables. 16 Table 5 shows the transpose of the estimated matrix $H$, namely the contemporaneous newsletter elasticities of the corresponding variables. As expected, all elasticities are positive.

The finding that newsletters have positive effects on commission is particularly relevant. Assuming the marginal cost of sending newsletters online is zero, 17 an optimizing firm should choose the number of newsletters such that sending more letters has a zero marginal effect on commission. In the case of this firm, the newsletter elasticities of commission are small but positive, indicating that the firm might further benefit by sending more newsletters each week. As expected from our results on network effects, newsletters ‘targeted’ at one group also have positive effects on the number of users in the other group. For example, the number of new buyers may increase when seller-targeted newsletters lead to better virtual shop visibility.

15 If our basic model (1) is mis-specified because of the omission of marketing mix variables, the estimates in Section 4 will suffer from an omitted variable bias. We estimate the VAR system with and without the numbers of marketing newsletters on the same sample starting from 2007. The values of a new seller, a new buyer, a returning seller, and a returning buyer are 302, 72, 29, and 15 respectively without including the number of marketing newsletters; the values are 302, 81, 32, 12 when the number of newsletters are included. These estimates are not significantly different. This suggests that the estimates in the basic model are not biased because of the omission of the marketing mix variables. This also justifies that the number of marketing newsletters is not correlated with the lagged value of the endogenous variables. All results based on model (2) are available from the authors.

16 A caveat applies to all our findings below. In general, the marginal effects of newsletters should be decreasing as more newsletters are sent since consumer attention saturates. All our findings are thus limited to the current levels of newsletters and performance variables.

17 Discussion with the firm reveals that designing newsletters involve substantial costs. But once the design is finished, sending more newsletters involve little additional cost. At least locally, we can assume that the marginal cost of newsletters is close to zero.
Table 5: Effects of Marketing Newsletters: In Elasticity Terms. *** p<0.05, ** p<0.1, * p<0.15

<table>
<thead>
<tr>
<th>Newsletter Type</th>
<th>Commission</th>
<th>Number of New Sellers</th>
<th>Number of Returning Sellers</th>
<th>Number of New Buyers</th>
<th>Number of Returning Buyers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buyer Targeted</td>
<td>0.0368*</td>
<td>0.0388</td>
<td>0.0249*</td>
<td>0.0434*</td>
<td>0.0370**</td>
</tr>
<tr>
<td>Seller Targeted</td>
<td>0.0026***</td>
<td>0.0029***</td>
<td>0.0017***</td>
<td>0.0031***</td>
<td>0.0023***</td>
</tr>
</tbody>
</table>

We can derive the contemporaneous marginal impact of newsletters on revenues from the estimated elasticities.\(^{18}\) The results are presented in Table 6. To inform the firm’s marketing mix optimization, however, the contemporaneous elasticities may not sufficiently estimate the newsletters’ cumulative long term impact. When marketing newsletters help acquire sellers and buyers, these users would continue to generate value due to their network effects on other users. Put differently, the impact of a newsletter may persist even if the original newsletter no longer gets any clicks. Quantifying the total revenue contribution of sending one buyer/seller newsletter would require a procedure similar to that of the IRF simulation. The cumulative financial impact of the additional newsletter message can be simulated by tracking the long term impact of the initial impact over time:

\[
v_j = \sum_{t=1}^{5} \left( \gamma_{ji} \sum_{i=0}^{\infty} \phi_{i5}(t) \right)
\]

where \( \phi_{i5}(t) \) is the IRF function with commission as the response variable, and \( \gamma_{ji} \) is the contemporaneous effect of newsletter type \( j \) on endogenous variable \( i \). The \( \gamma_{ji} \) parameters are identified from matrix \( H \) following the procedure in Appendix A.1. Thus, the value of

---

\(^{18}\) According to the standard formula \( \frac{\Delta Y}{\Delta X} = \frac{\bar{Y}}{\overline{X} \sigma_x} \cdot \frac{\Delta \ln Y}{\Delta \ln X} \) which translates elasticities into marginal effects.
newsletters can be calculated by simulating a shock in the number of newsletters at time 0, which is equivalent to a shock vector that affects all endogenous variables contemporaneously. The cumulative impact $v_j$ is thus a weighted average of the five CIRFs where total commission is the response variable.

With the above simulation approach we can trace the long term values created by the newsletters, both directly and indirectly. We also report the estimated cumulative revenue contributions of 1000 buyer newsletter messages and 1000 seller newsletter messages in Table 6.

**Table 6: Revenue Contributions of 1000 Newsletter Messages.**

<table>
<thead>
<tr>
<th></th>
<th>Contemporaneous Value ('Value from Clicks')</th>
<th>Cumulative Value ('Value beyond Clicks')</th>
</tr>
</thead>
<tbody>
<tr>
<td>Buyer Targeted Newsletter</td>
<td>1.7 Euros</td>
<td>10.1 Euros</td>
</tr>
<tr>
<td>Seller Targeted Newsletter</td>
<td>1.1 Euros</td>
<td>6.4 Euros</td>
</tr>
</tbody>
</table>

The results indicate that at the current level of newsletter intensity, 1000 additional buyer newsletter are worth on average (cumulatively) 10.1 Euros to this firm, while 1000 additional seller newsletter are worth on average about 6.4 Euros.

The above simulation estimates the cumulative values of marketing newsletters to be about five to six times those calculated from the contemporaneous elasticities. The current approach of the firm to calculate the value of marketing newsletters is to trace the click-through rates of
a marketing newsletter campaign and multiply that with the direct benefits created by the
buyers and sellers who have responded to the newsletters. Thus, it is assumed that an E-mail
which is no longer getting any clicks ceases to generate value. If one holds this assumption,
the values of buyer and seller newsletters would be fully realized in the week they are sent
since most newsletters cease to get any clicks after the first week. Our results indicate that by
ignoring network effects, these simple estimates significantly understate the value of
marketing actions – by more than a factor of 5 in this case. Such underestimation may lead to
suboptimal decisions on newsletter strategies.

5.2 Enhancing Network Effects

Our empirical analysis indicated the importance of network effects. Contributors are valuable
not only because of the cash flow they generate, but also because of their network value: their
presence benefits the other customers/contributors, which translates into indirect
contributions to the website’s revenue.

There is a wide variety of actions that the website can use to enhance the network effects in
the marketplace. For example, the company in our empirical study is considering improving
the following measures:

- Product and Seller Review System: This allows the buyers to share their product
  experiences and score a seller, therefore enhancing the network effect of buyers on other
  buyers.

- Seller Referral Programs: This encourages existing sellers (especially the newly acquired
  sellers) to recommend their selling experience with their friends, in the hope of acquiring
more sellers through word of mouth.

We expect such measures will make seller and buyer acquisitions even more valuable by increasing the network component in their revenue contribution. In the following simulations, we compare the value of one new seller or buyer before and after enhancing the network effects.

Formally, we define the network effect of group $i$ on group $j$, $\alpha_{ji}$, as the elasticity of the number of group $j$ users in time $t$ with respect to the number of group $i$ users in time $t-1$: $\alpha_{ji} = \frac{\partial N_j}{\partial N_{i-1}}N_{i-1}$ with respect to the number of group $i$ users in time $t-1$.

For example, $\alpha_{42}$ denotes the network effect of new sellers on new buyers. When for example $\alpha_{42} = 0.5$, acquiring 1% more new sellers at time $t-1$ leads to 0.5% more new buyers joining the website at $t$, *all other things being equal*.

When $\alpha_{42}$ is increased, for example, from 0.5 to 0.7, acquiring the same percentage of sellers (1%) now leads to even more buyers (0.7%) joining the website at time $t$.

Our simulation approach, in short, corresponds to modifying the $\alpha_{ji}$ structural parameters in the seller/buyer acquisition equations while keeping all other equations unchanged. This simulation methodology is along the lines of VAR-based simulations common in the economics and marketing literatures (Bernanke, Gertler, Watson, Sims, & Friedman, 1997; Primiceri, 2005; Sims & Zha, 2006; Horváth, 2003; Pauwels, 2004), where a number of issues, such as the Lucas critique (Lucas, 1976; Van Heerde, Dekimpe, & P. Putsis Jr, 2005) have been discussed. We refer interested readers to these papers for more detailed discussions. In Appendix A.1, we explain the procedure in greater detail and illustrate why modifying the parameters in the standard form, which are directly estimated, is inappropriate.
We simulate the impact of enhancing network effects by comparing the monetary values of buyer and seller acquisition (measured by CIRFs) before and after a structural parameter $\alpha_j$ has been increased by a certain percentage. We focus on seller and buyer acquisition for the reasons stated in the beginning of Section 4: when the number of returning sellers/buyers is the response variable, the conceptual meaning of CIRFs is not clear.

Thus, we study the impact of modifying $\alpha_{24}, \alpha_{42}, \alpha_{22}$ and $\alpha_{44}$ on values of new buyers and new sellers. Table 7 illustrates the results. We compare the value increases when $\alpha_{42}, \alpha_{24},\alpha_{22}$ or $\alpha_{44}$ are enhanced by 5% respectively. To assess the outcome when several effects are simultaneously improved, we show the case when $\alpha_{24}$ and $\alpha_{42}$ are enhanced by 5% simultaneously (enhancing other pairs of effects leads to qualitatively similar results) and the case when all four effects are enhanced by 5% simultaneously. The results are qualitatively similar when the parameters are changed by different percentages.

<table>
<thead>
<tr>
<th></th>
<th>Before enhancing network effects</th>
<th>Enhancing buyers’ network effects on sellers ($\alpha_{24}$) by 5%</th>
<th>Enhancing sellers’ network effects on buyers ($\alpha_{42}$) by 5%</th>
<th>Enhancing sellers’ network effects on sellers ($\alpha_{22}$) by 5%</th>
<th>Enhancing buyers’ network effects on buyers ($\alpha_{44}$) by 5%</th>
<th>Enhancing both $\alpha_{42}$ and $\alpha_{24}$ by 5%</th>
<th>Enhancing all four effects by 5%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Value of One New Seller</td>
<td>242</td>
<td>251</td>
<td>246</td>
<td>308</td>
<td>248</td>
<td>267</td>
<td>410</td>
</tr>
<tr>
<td>Value of One New Buyer</td>
<td>94</td>
<td>102</td>
<td>95</td>
<td>105</td>
<td>100</td>
<td>105</td>
<td>145</td>
</tr>
</tbody>
</table>

The results reveal several findings. First, in order to increase the value of one group of users, it is sometimes optimal to focus on the other groups. In our example, when the firm wants to improve the value of the buyers, it should focus on network externalities between the new
sellers ($\alpha_{22}$) instead of those of the buyers ($\alpha_{24}$ / $\alpha_{44}$), if the costs are comparable. On the other hand, when we increase the network effects of one type of users, the value of this type of users will also increase the most. For example, when we increase $\alpha_{24}$ or $\alpha_{44}$ (new buyers’ network effects on others), the value of buyers will increase by 8.5% and 6.4% respectively, more than the increase in the value of sellers (3.7% or 2.5%).

Moreover, enhancing the network effects in different directions are complementary decisions. For example, enhancing $\alpha_{24}$ by 5% increases the value of a seller by 251-242=9 Euros. Enhancing $\alpha_{42}$ by 5% increases the value of a seller by 246-242=4 Euros. Moreover, enhancing both $\alpha_{42}$ and $\alpha_{24}$ by 5% simultaneously will increase the value of a seller by 267-242=25 Euros, which is greater than 9+4=13 Euros. We obtain similar results when other pairs of network effects are improved simultaneously. These findings have clear managerial implications. The C2C firm in our context has the options of implementing seller referral system and product review system. Since these systems tend to focus on different types of network effects, the marginal return from implementing/improving each system increases when other systems are already implemented or improved. Taking cost into account, when the cost functions are convex, the firm should focus on the most profitable network effect (sellers on sellers in the C2C website we study) when cost is low and ‘spread’ out its investment when the cost becomes higher.

Finally, as expected, increasing network effects overall leads to large increase in the values of seller and buyer acquisition. In the C2C website we study, enhancing sellers’ network effects on other sellers represents the most profitable direction.

19 See (Topkis 1998) for a formal definition and implications of complementarities in actions.
6. Further Analyses and Robustness Checks

The above analysis is based on a VAR model estimated from aggregated data and over the entire observation period. In this section, we present further analyses that examine the evolution of VAR estimates over time and the heterogeneity across product categories. While the results about time evolution and cross category heterogeneity are of their own interest, these analyses also serve as robustness checks of the qualitative findings from our basic model.

6.1 Time Evolution of Values and Network Effects

We first examine whether the revenue contributions of buyers and sellers change over time during our observation period. To answer this question, we estimate the VAR model on 100 consecutive time windows, with the first time window corresponding to the 1st to the 101st observations (weeks) and the last time window corresponding to the 101st to the 201st observations. This process is similar to the moving window analysis in (Pauwels & Hanssens, 2007). The procedure yields 100 different VAR estimates, with which we can perform IRF simulations to track the evolution of seller/buyer values over time. We plot the revenue contributions of new/returning sellers/buyers against the number of time window in Figure 1.20

The figure suggests that our qualitative findings (that sellers make higher revenue contributions than buyers and that acquisition makes higher revenue contribution than retention) remain valid over time. Pair-wise significance tests by Monte-Carlo simulation

---

20 The relative small sample size and the large number of parameters give less efficient estimates. We plot the 95% confidence interval derived from Monte Carlo simulation for each value of CIRF.
confirm that the estimated revenue contributions are not significantly different in any two time windows. The qualitative finding that new sellers bring in the most number of other sellers and buyers is also robust over time (details not shown for brevity). The company considers these findings reasonable, since 2005-2009 has been a period of stable growth for the website.

Figure 1. Time Evolution of Revenue Contributions.21

6.2 Segment-level Analysis

Unobserved heterogeneity and aggregation bias is a particularly important concern for dynamic models such as VAR based on aggregate data (see Pauwels et al (2004) for a review).

---

21 The solid lines correspond to the estimates and the dashed lines correspond to 95% confidence intervals.
To explore this issue, we perform a segment-level analysis to examine the relative values of buyers/sellers in different product categories. We estimate VAR models for six different product categories. We report the estimated revenue contributions in Table 8.

Overall, the results indicate that there is considerable heterogeneity in the value of buyers and sellers across product categories. Yet, our main results qualitatively hold in all segments. In particular, the revenue contribution of sellers is higher than that of buyers, and the revenue contribution of new sellers/buyers is higher than that of returning sellers/buyers. We test these via Monte Carlo tests with 250 replications. The results are shown in Appendix A.8. Note that unlike the results from the basic model, new buyers are not necessarily more valuable than returning sellers across all product categories.

Table 8: Segment-Level Analysis: * indicates $p<.05$.

<table>
<thead>
<tr>
<th>Product Category</th>
<th>New Seller</th>
<th>Returning Seller</th>
<th>New Buyer</th>
<th>Returning Buyer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Main Model (as in Table 2)</td>
<td>242*</td>
<td>45*</td>
<td>94*</td>
<td>11</td>
</tr>
<tr>
<td>Books, Magazines</td>
<td>248*</td>
<td>45*</td>
<td>65*</td>
<td>-1</td>
</tr>
<tr>
<td>Videos and CDs</td>
<td>366*</td>
<td>52*</td>
<td>41*</td>
<td>-0.8</td>
</tr>
<tr>
<td>Computer and Peripheral</td>
<td>238*</td>
<td>82*</td>
<td>27*</td>
<td>13*</td>
</tr>
<tr>
<td>Household Goods</td>
<td>173*</td>
<td>74*</td>
<td>13*</td>
<td>6*</td>
</tr>
<tr>
<td>Furniture</td>
<td>125*</td>
<td>64*</td>
<td>25*</td>
<td>6*</td>
</tr>
</tbody>
</table>

A cautionary note applies when we try to interpret the estimated values. The segment-level analysis neglects any between segments effect, and the estimated value for each segment
should not be directly compared to the results obtained from the main model.

7. Concluding Remarks

In this paper, we take a first step in studying the ‘content contributor management’ problem for a UGC website. We developed an empirical framework based on persistence modeling techniques and applied it to the context of a C2C marketplace. The empirical results indicate how the network effects, which we quantify, affect the valuations of buyers and sellers. The most valuable users also tend to have the largest network effects on other users. For example, for the particular company we study, acquiring one new seller is worth on average about 20 times more than retaining one existing buyer, and about 6 times more than retaining one existing seller. In general network effects constitute more than 80% of the total revenue contributions of new sellers and new buyers. The simulations shed light on the firm’s marketing mix choices as well as ‘enhancing network effects’ decisions. For example we find that for the particular company we study, measuring the value of email marketing campaigns using only the immediate cash flow of such campaigns (for example when receivers click on links in the email they receive) can underestimate by more than a factor of 5 the value of such campaigns. We also find that if this firm wants to improve the value of its buyers, it should focus on network effects between its sellers instead of those between its buyers. For example, when costs are comparable, a “seller referral program” focusing on sellers can increase the buyers’ value more than a “product review system” that focuses on buyers. Our methodology allows UGC companies to reach more sound, and possibly unexpected, managerial conclusions through a better measurement of all relevant quantities.
A number of directions for future research are possible. First, obtaining data over the entire life cycle of a website can lead to additional insights on how the values of seller/buyer acquisition/retention evolve over the stages of a website’s life cycle. The values of sellers and buyers may differ considerably for a young website and for a more mature website. Second, while our approach measures the magnitudes of overall network effects further disentangling the different sources of network effects (such as word of mouth, observational learning, product assortment enrichment, direct social interaction, etc) can shed light on the effectiveness of more specific managerial actions. For example, when is a seller scoring system preferred to a product review system as an instrument to enhance buyer-buyer and buyer-seller network effects? Finally, using data from other UGC firms could allow us to study the generalizability of our results or to extend and improve them. In addition, alternative methodologies (such as diffusion models or controlled experiments) can be used to assess the impact of specific managerial actions.

References


