Chapter 2
How Consumer-Generated Advertising Works: An Empirical Agent-Based Simulation

Makoto Mizuno

Abstract Affiliate advertising is a novel form of Internet advertising that enables bloggers to insert advertising for any product in their blog articles and to gain rewards based on consumers' actual responses. To understand how this form of advertising works, we conducted Web-based questionnaire surveys among bloggers, including affiliates, and readers, including buyers. Moreover, we constructed an agent-based model that is empirically validated by the above data. The results of the simulation using this model showed that (1) link structures between affiliates and readers have a remarkable impact on the average revenues of all affiliates, and (2) the presence of random walkers in searching for a better ad mix may increase total revenues for all affiliates compared to the presence of local imitators. Based on these results, we discuss the managerial implications and suggest future avenues for research.

**Keywords** Affiliate advertising • Agent-based modeling • Blogs • Consumer behavior • Empirical validation

2.1 Introduction

Bloggers participating in an affiliate advertising program, often called "affiliates," insert advertising for products they have chosen themselves in their blogs, and gain rewards from advertisers if customers buy the products or click on the advertisements. Advertisers (or their agents) offer a list of products to be advertised and corresponding advertising materials to affiliates, and then pay fees to the affiliates.

M. Mizuno (✉)
School of Commerce, Meiji University, 1-1 Kanda-Surugadai, Chiyoda-ku,
Tokyo 101-8301, Japan
e-mail: makoto@meiji.ac.jp
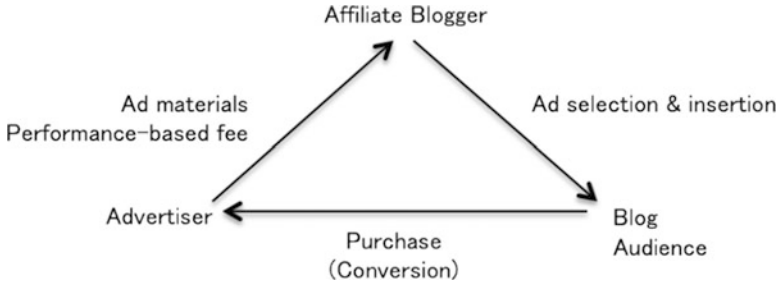
Fig. 2.1 Relationships between agents in affiliate advertising

This relationship is depicted in Fig. 2.1. The affiliate market in Japan has been growing at a moderate rate [8], although the size is limited relative to the entire Internet advertising market in Japan.

A distinctive feature of affiliate advertising is that consumers, who are neither advertisers nor media companies, control advertising insertion; thus, affiliate advertising can be seen as "consumer-generated advertising media." This type of advertising can be interpreted as part of the shift of power in advertising from companies to consumers [6]. Despite the increasing popularity of affiliate advertising, little research has been done on it, except for a few studies [2, 3, 5].

To grasp the state of the art of affiliate advertising, we conducted Web-based questionnaire surveys among bloggers and blog readers. Moreover, to gain insight into the possible outcomes of the complex interactions among bloggers, blog readers, and advertisers, we adopted an agent-based model. We then used our obtained survey data to characterize each agent's preferences and feasible actions under certain constraints.

The problem is the lack of information about the links between affiliates and readers. To address these "missing links," we draw bipartite graphs of the relationships between bloggers and readers, utilizing the survey data. Such links are thought to be a result of choices of blogs by readers, which could be governed by congruence in interests or preferences of both bloggers and readers. Accordingly, we artificially generate possible links based on the various congruences calculated from the survey data.

Given the link structures from the bipartite graphs, we then model and simulate interactions between affiliates and readers. In doing so, we examine how affiliates should act in order to increase revenues. The choice of advertised products (ad mix) is routinized as a strategy. Thus, we consider the following two strategies: local imitation and random walk strategies. By simulating the model and varying the parameters, we examine which conditions and what kind of strategy could yield larger revenues.

The rest of this chapter is organized as follows. In Sect. 2.2, we show the key findings from the surveys related to our modeling. In Sect. 2.3, we describe the formulation of our model. In Sects. 2.4 and 2.5, we describe the simulation procedure and present the results. Lastly, we discuss the managerial implications and suggest future research avenues.

2.2 Empirical Study

First, we show the key results from the Web-based questionnaire surveys. The first survey was conducted on February 2011 among bloggers who posted blog entries at least once a month. The respondents are categorized as either affiliates or non-affiliates based on whether they participate in any affiliate program or not. The subsequent survey was conducted on December 2011 among blog readers who read blogs at least once within the most recent month. The readers are categorized as either buyers or a non-buyers based on whether they bought advertised products or not within the recent half year. These four classes of respondents are sampled independently and at random from a large panel provided by a Web-centric research company in Japan ($N = 361$ for affiliates; $N = 361$ for non-affiliates; $N = 412$ for buyers; and $N = 412$ for non-buyers).

Although the surveys yielded many interesting findings [4], we only show those relevant to our modeling. Table 2.1 presents the findings regarding the reasons why affiliate bloggers participated in any affiliate program (note that the respondents were allowed to choose multiple answers).

This result suggests that affiliates are strongly motivated by monetary rewards, at least when they begin the program. Table 2.2 shows the distribution of monthly affiliate rewards. It shows that more than half of the affiliates earned small or very small monetary rewards despite their initial motivation.

About 70 % of the affiliates adopted some means to increase rewards. As shown in Table 2.3, the three main means are changing the advertising mix, changing the

Table 2.1 Reasons for participating in affiliate advertising programs (multiple answers)

Reason	%
To gain monetary rewards	64.0
To start a blog free of charge	50.0
To recommend favorite products to readers	29.0
To provide useful information to readers	13.0

Table 2.2 Distribution of average monthly revenues for affiliates

Revenue	%
Higher than 50,000 yen (about $500)	2.3
10,000–50,000 yen (about $100–$500)	5.0
5,000–10,000 yen (about $50–$100)	3.0
1,000–5,000 yen (about $10–$50)	12.5
100–1,000 yen (about $1–$10)	24.1
Lower than 100–1,000 yen (lower than $1–$10)	51.5
Unknown	1.7

Table 2.3 Means to increase affiliate rewards (multiple answers)

Means	%
Advertise products that fit the blog	39.0
Align the content of the blog with the ads	37.0
Increase traffic to the blog	36.0
No means in particular	28.0

content or theme of the blog, and increasing the number of blog visitors. Among these options, changing the content of the blog seems the riskiest because doing so might disappoint current readers of the blogs or alter the identity of the blogs. On the other hand, changing the ad mix seems to be the least risky option. Hence, this study focuses on examining the effect of the ad mix in blogs, keeping the content of the blog constant.

2.3 Model

In modeling the affiliate advertising process, we model the behaviors of two classes of agents: bloggers who participate in an affiliate program, and thus were classified as affiliates in the blogger survey, and blog readers who bought products by clicking on affiliate ads, and thus were classified as buyers in the reader survey. These agents are linked across classes by a bipartite graph, but they are not directly linked within the same class.

2.3.1 Affiliate Bloggers' Behavior

In our model, affiliates are allowed to choose the ad mix, which is represented by the following K-dimensional vector:

$$\mathbf{a}_{it} = (a_{i1t}, a_{i2t}, \ldots, a_{ikt}, \ldots, a_{iKt}) \tag{2.1}$$

where a_{ijt} is a variable that equals 1 if product $k(= 1, 2, \ldots, K)$ is advertised in blog $i(= 1, 2, \ldots, I)$ at time $t(= 1, 2, \ldots, T)$, and 0 otherwise. On the other hand, affiliates have central themes or topics in their blogs, represented by the J-dimensional vector

$$\mathbf{b}_i = (b_{i1}, b_{i2}, \ldots, b_{ij}, \ldots, b_{iJ}) \tag{2.2}$$

where b_{ij} is a variable that equals 1 if topic $j(= 1, 2, \ldots, J)$ is covered in blog i, and 0 otherwise. Here, it is assumed that the sets of topics never vary over time in each blog. The components of vectors $\mathbf{a}$ and $\mathbf{b}$ are determined based on our questionnaire surveys.

The affiliates' choice of an ad mix is affected by many factors, including non-monetary motivations; however, in this chapter, we focus only on the affiliates' monetary motivation, which is not unrealistic given the results from our blogger survey (Table 2.1). In addition, we assume that affiliates search for a better ad mix if they are not satisfied with their monetary rewards.

The most naïve strategy for seeking a better ad mix may be the random walk, which requires little rationality or thought by agents. Another strategy is the local

imitation, in which slightly more rational affiliates adopt a strategy in which they imitate the most successful action of their peers, for which the scope for search is bounded. More specifically, these two strategies are formulated as follows:

Random Walk. If an agent's revenue at time t is lower than the revenue at time $t − 1$, the agent replaces a product in the ad mix with another one, both chosen at random. The number of products advertised simultaneously in each blog is kept at the initial level indicated in the survey data. Of affiliate agents, $g * 100\%$ adopt this strategy.

Local Imitation. If the revenue at time t is lower than the revenue at time $t − 1$, the agent searches for some very similar blogs in terms of topics covered, and then copy the ad mix of the blog that earns the largest revenue. The similarity between blog i and i' is calculated by Eq. (2.3):

$$SIML_{ii'} = 1 - \sum_{j=1}^{J} |b_{ij} - b_{i'j}| / J \qquad (2.3)$$

It may seem unrealistic to assume that affiliates can know each other's revenues; however, they can estimate the revenues by looking at the number of visits for a blog, which is easily available through indicators embedded in the front page of a blog or through ranking sites. Thus, from such traffic data, bloggers can roughly estimate which blogs have a similar number of clicks and conversion rates as theirs. $(1 − g) * 100\%$ of affiliate agents adopt this strategy.

2.3.2 Readers' Behavior

Blog readers in our model are potential buyers of products advertised in affiliate ads. They have preferences for blog topics and advertised products, which are represented, respectively, as follows:

$$\mathbf{u}_h = (u_{h1}, u_{h2}, \ldots, u_{hj}, \ldots, u_{hJ}) \qquad (2.4)$$

$$\mathbf{v}_h = (v_{h1}, v_{h2}, \ldots, v_{hk}, \ldots, v_{hK}) \qquad (2.5)$$

where u_{hj} equals 1 if buyer h has a preference for topic j, and 0 otherwise, and v_{hk} equals 1 if buyer h has a preference for product j, and 0 otherwise. These factors were also determined based on our questionnaire surveys.

Before reading blog entries or purchasing a product, readers have to choose which blogs to visit. Unfortunately, our surveys did not collect information on which affiliate blog each buyer has visited; the reason is that the blogger and reader respondents were independently sampled, and thus, it would be difficult to connect the blogs to the readers and vice versa. Alternatively, we artificially spanned

"links" between these two groups, utilizing the empirical data as much as possible. Fortunately, our surveys collected information that we could use to match agents based on preferences in topics and products. We explain this matching process below.

Firstly, we reasonably expected that readers will be inclined to visit blogs where the topics are the most congruent to their preference. The extent of this congruence (topic congruence) is calculated by Eq. (2.6):

$$TOPIC_{ih} = 1 - \sum_{j=1}^{J} |b_{ij} - u_{hj}| / J \qquad (2.6)$$

Likewise, we can consider links based on the congruence between advertised and preferred products (product congruence). The extent of this congruence is calculated by Eq. (2.7):

$$PROD_{ih} = 1 - \sum_{k=1}^{K} |a_{ik} - v_{hk}| / K \qquad (2.7)$$

Although it seems unrealistic to assume that readers choose blogs based on the advertising of their favorite products, product congruence could be a good proxy for similarity in taste for consumption. In addition, if affiliates are able to find an ad mix that is most congruent to the target readers' product preferences, the resulting links will maximize this congruence. In that sense, this congruence could measure the "optimality" of the affiliates' actions.

The links maximizing topic congruence and product congruence are, respectively, extremes of a continuum of the possible links between the agents. Since there is an infinite number of factors affecting these links, more realistic links would not converge into such extremes, rather existing somewhere in between them.

We consider that the realistic links between affiliates and readers can be determined by combining all of the above criteria. Hence, we use Eq. (2.8) as a meta-level criterion:

$$C_{ih} = w_1 RAND_{ih} + w_2 TOPIC_{ih} + w_3 PROD_{ih} \qquad (2.8)$$

where $RAND_{ih}$ is a random disturbance term uniformly distributed within $(0, 1)$, $w_1 + w_2 + w_3 = 1$, and w_1, w_2 and w_3 are all nonnegative. With the in-degree or out-degree constraints of each node in the bipartite graphs, the affiliate to which each reader is linked is selected so as to maximize Eq. (2.8). The constraints are given by the surveys: the number of favorite blogs is used to determine the out-degree of a reader, and the page view and the frequency of blog visits are used to determine the in-degree of an affiliate.

After visiting and reading a blog, readers decide whether to buy a product advertised on that blog. For simplicity, we assume that readers are exposed to

all advertising inserted in the blog and consider buying a product that best fits their preference. We also assume that if there are many great alternatives, they only choose one at random. The readers' purchase volume is determined by the constraints assigned to them, based on the frequencies of blog visits and shopping reported in the survey.

2.3.3 Assignment of Agents

The agents in our model are generated based on the questionnaire surveys: bloggers who participate in any affiliate program are classified as affiliates, and blog readers who have previously purchased a product by clicking on affiliate advertising are classified as buyers.

In assigning agents in the model to respondents in the survey, we find a gap between desirable agent size and actual sample size. To ensure the scalability of the model, we apply a method of bootstrapping [1] in the context of an agent-based modeling. That is, we generate the size of the bootstrap sample by repeated random sampling from a survey data with replacement; correspondingly, we can draw distributions of the sample statistics or the aggregate outcomes emerging from simulation. This approach also provides more reliable information to avoid sample bias.

2.4 Simulation

Firstly, we repeatedly generated agents by bootstrapping, resulting in 50 samples involving 100 affiliates and 1,000 readers. Secondly, for each bootstrap sample, we constructed different link structures corresponding to the following sets of parameters in Eq. (2.8): $\mathbf{w} = (w_1, w_2, w_3) = (1, 0, 0)$, $(1/2, 1/2, 0)$, $(0, 1, 0)$, $(1/2, 0, 1/2)$, $(1/3, 1/3, 1/3)$, $(0, 1/2, 1/2)$, or $(0, 0, 1)$, where the first component represents whether to consider the strength of random noise in link formation; the second, topic congruence; and the third, product congruence.

Each of the affiliate characteristics $\mathbf{a}$ and $\mathbf{b}$ (see Eqs. (2.1) and (2.2)) is given by the products (11 categories, e.g., books, foods/beverages) that the respondents of affiliates in our questionnaire survey have advertised and by the topics (29 genres, e.g., cooking, social events, travel) of blog entries that they have posted, respectively. On the other hand, each of the reader characteristics $\mathbf{u}$ and $\mathbf{v}$ (see Eqs. (2.4) and (2.5)) is given by the topics that the respondents of blog readers in the survey prefer to read and by the products that they have bought via any affiliate, respectively.

Thirdly, given the link structure and the bootstrap sample, we performed a simulation, varying the proportion of the random walkers and local imitators: $g = 0.1, 0.3, 0.5, 0.7$, or 0.9. For a set of these parameters, a simulation is repeated

20 times to cancel out accidental variations. The time horizon is set to $t = [1, 30]$. In imitating others, affiliates following the local imitation strategy choose the five agents whose blogs are most similar to theirs, based on the calculations from Eq. (2.3).

2.5 Results

In order to confirm that the outcomes predicted by our simulation are distributed in a unimodal and almost symmetric manner (e.g., unlike a power law), we average out the outcomes using a set of parameters to summarize the results. As Fig. 2.2 indicates, the average cumulative revenue per affiliate at $t = 30$ is remarkably high when $w_3 = 1$, which indicates link structures based on product congruence rather than topic congruence or random noises. Also, the proportion of random walkers, g, shows a positive effect on the average revenue of all agents in most cases. We can find the threshold between $g = 0.1$ and 0.3, except for $\mathbf{w} = (1, 0, 0)$, where only random disturbance is dominant. This suggests the effectiveness of the random walk strategy in boosting average revenue for all affiliates. In particular, if random walkers were eliminated from the market beyond a certain line (that is, if g is less than between 0.1 and 0.3), the market would contract drastically.

Figure 2.3 shows the temporal changes in the average revenue of all affiliates by a set of parameters. Below are some of the interesting findings:

1. Average revenue is sufficiently high at the initial period when $w_3 = 1$ (links based on product congruence), and is very low otherwise. This result is understandable since maximizing product congruence is equivalent to maximizing
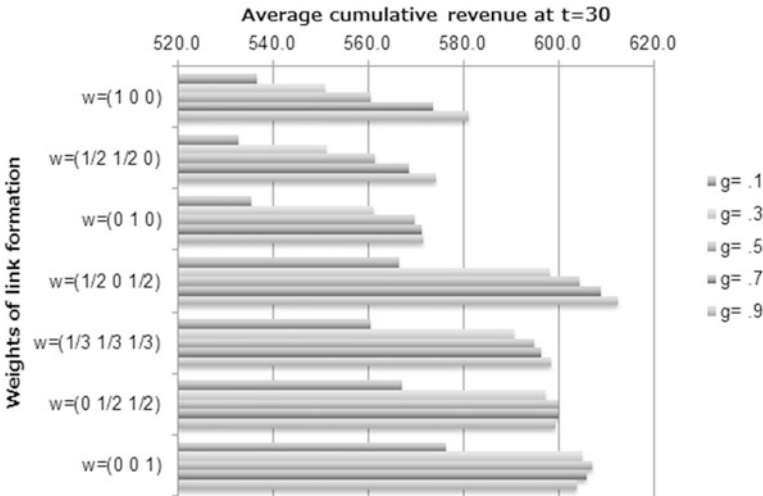

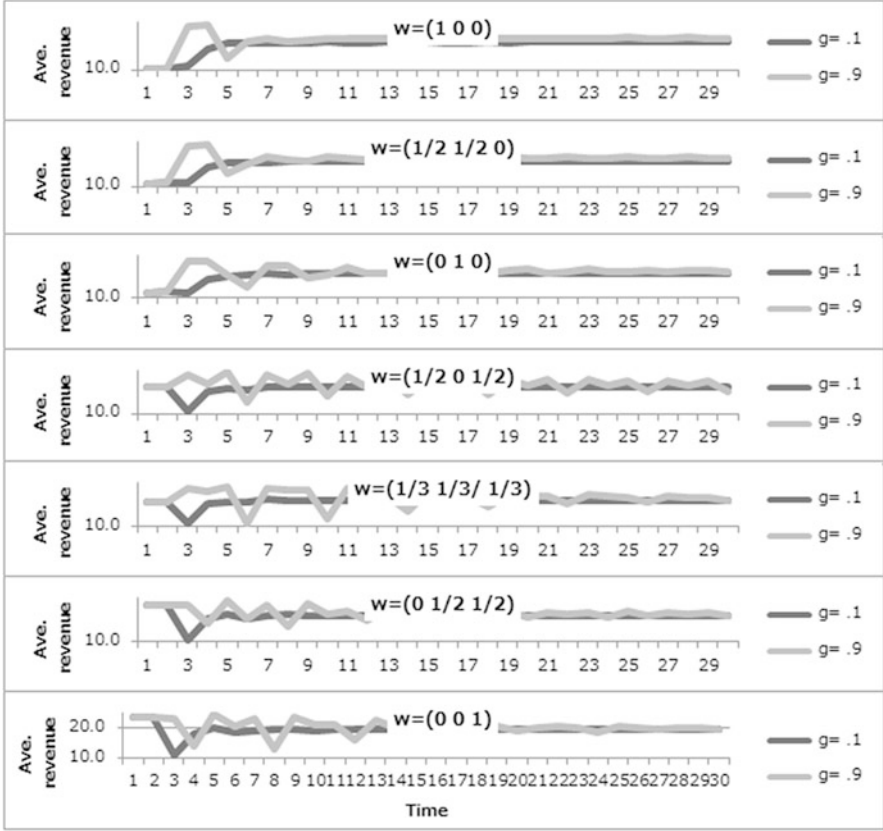**Fig. 2.2** Average cumulative revenues at $t = 30$

Fig. 2.3 Time-series trajectory of average revenues

purchase in our model. We expect such a case to be more realistic if the market is at the mature stage, since the affiliates could have enough time to learn their readers' preferences.

2. In earlier periods, just after the initial period, average revenue drastically fluctuates. For instance, at $t = 3$ or 4, it reaches a peak if g is relatively high, that is, when the proportion of random walkers is higher. This result suggests that the presence of agents with "no intelligence," who have put little thought into their ad mix, amplifies short-term fluctuations in the market in earlier periods.

3. In the long term, the average revenues converge regardless of the link structure or proportion of random walkers. Accordingly, the difference in cumulative revenues may depend on the difference in earlier revenues. This dependence becomes more important when calculating the present discounted value with a reasonable discount rate.
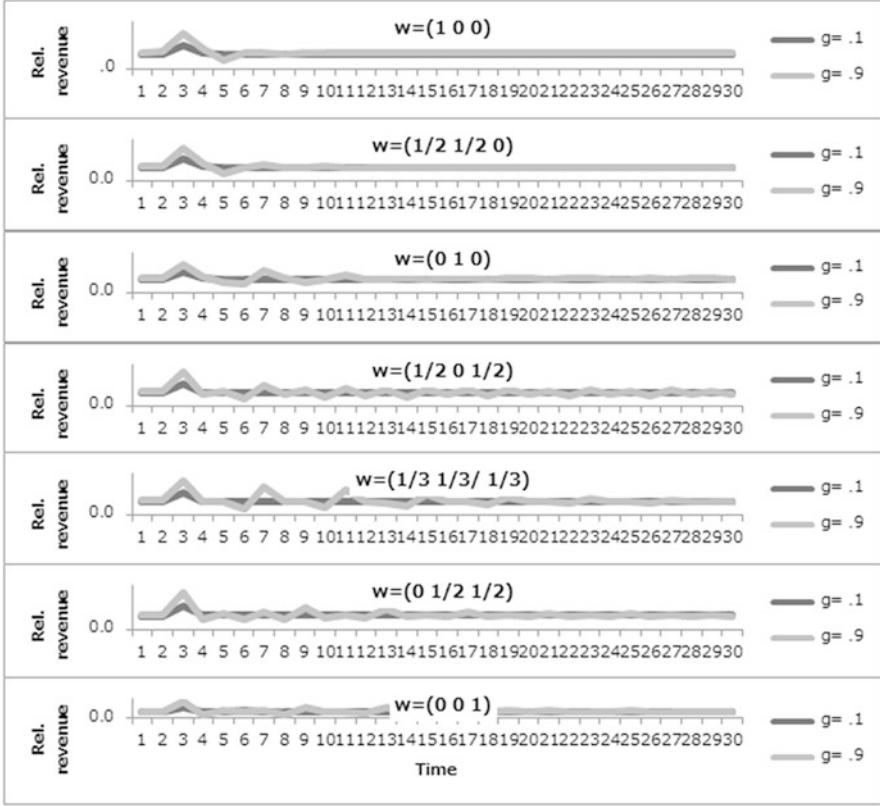
Fig. 2.4 Relative revenues of random walkers vs. local imitators

We compare the revenues of random walkers with those of local imitators. As shown in Fig. 2.4, the ratio between these two types of revenues quickly converges, that is, both types are almost always at the same level except in the earlier periods. This result suggests that the choice of search strategies may have little impact over the long term. Only when the proportion of random walkers exceeds a threshold ($g > 0.1$) do the average revenues of random walkers become higher than those of local imitators in earlier periods.

In addition, we compare the revenues of affiliate leaders, that is those whose cumulative revenues are in the top 10 % by $t = 30$, with those of the rest of the affiliates (followers). Figure 2.5 shows that the difference in trajectories depends mainly on whether the link structure is product-congruence based ($w_3 = 1$) or not ($w_3 = 0$). For product-congruent link structures, the leaders' revenues surge temporally in earlier periods, particularly if the proportion of random walkers is low. For other link structures, the leaders' revenues are high from the beginning, decreasing over time, particularly if the proportion of random walkers is high. In any case, the differences in revenues between leaders and followers disappear over time as the revenues converge.
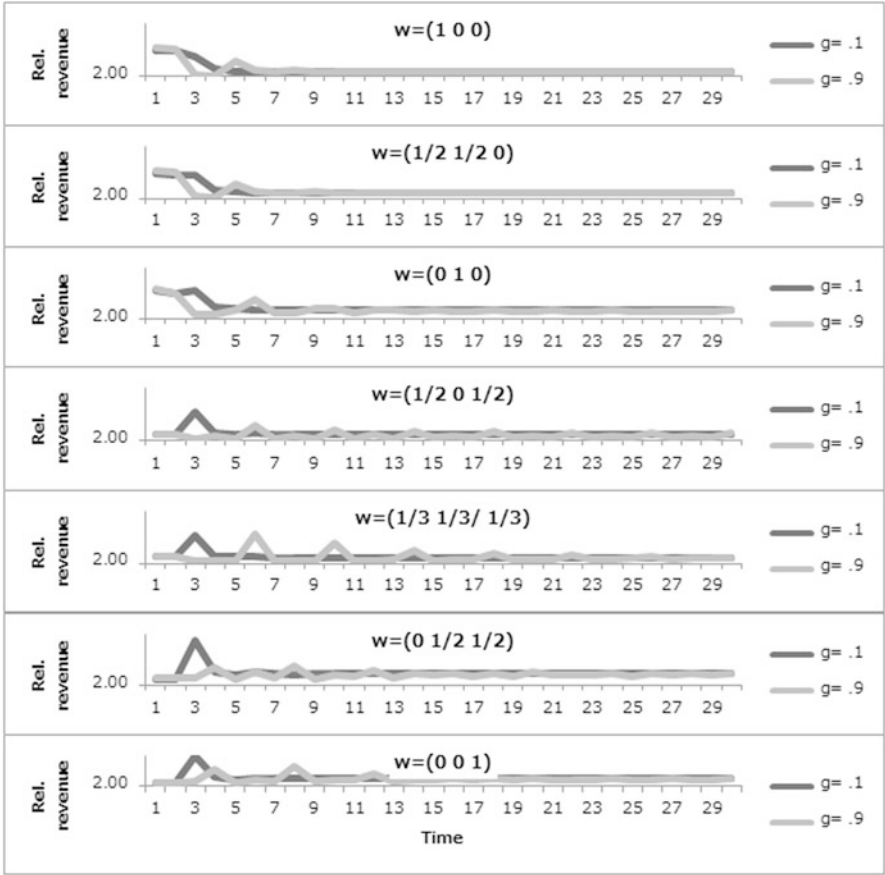
Fig. 2.5 Relative revenues of leaders vs. followers

2.6   Discussion

Although most affiliates participated in affiliate programs with a monetary motivation, their revenues from such programs were very small. One of the possible means for increasing revenues is changing the ad mix in the blogs. In our model, affiliates seek a better ad mix by adopting a more or less bounded-rational strategy, that is, they become either a random walker or a local imitator. On the other hand, readers choose blogs to visit based on their preferences on topics or products with a certain level of noise, and buy a product advertised on blogs based on their preferences for products.

The key results from our simulation are summarized as follows:

1. Average long-term revenues for affiliates are higher if the link structure is more product congruent and the proportion of random walkers exceeds a certain threshold.

2. The effect of the proportion of random walkers on total average revenues fluctuates only in earlier periods; in other words, the presence of random walkers contributes to the increase in total revenues through this fluctuation.
3. There is no difference in the average revenues between random walkers and local imitators; the difference is stable over time except in very early periods. The presence of random walkers has little impact on the competitiveness of each agent.
4. Except for the fluctuation in earlier periods, the difference in average revenues between leaders (top 10 % of affiliates based on earnings) and followers (all other affiliates) is stable over time and does not vary by link structure or the presence of random walkers.

Interestingly, the presence of random walkers, despite the little thought they put into their ad mix, positively affects not only their revenues but also those of all other affiliates in the entire market, resulting in little difference in competitiveness. A possible reason for this outcome is that a random-walk search guarantees a diverse ad mix that sufficiently satisfies readers' heterogeneous preferences. This suggests that a key factor in earning more from an affiliate program is maintaining diversity among the ad mixes of blogs.

Hence, for affiliates who wish to increase revenue, the results suggest not imitating peers who have seemingly effective ad mixes, but rather adopting a distinctive ad mix, as if searching at random. Consequently, if the proportion of random walkers in the market increases, the expected revenues for all affiliates will also increase. This scenario illustrates a positive-sum game with no trade-off among affiliates.

This study also has implications for researchers interested in empirical validation for agent-based modeling [7]. In this study, we calibrated each agent's parameters based on the questionnaire surveys for bloggers and readers. To map the empirical data according to the agents in the model, we extend the method of bootstrapping, which has been developed in statistics [1].

Lastly, we suggest further avenues for research. In this paper, we simplified many factors in order to keep the model parsimonious. However, we could further extend this simplification to deal with more complex or dynamic phenomena. For instance, we can consider other types of strategies for each affiliate for improving the ad mix. Moreover, we could allow affiliates to change the topics or themes of their blogs in order to enter or exit the market. At the same time, we could allow readers to change their preferences for blogs and advertised products. If such changes occur in converging into a steady state reported as above, some longer standing fluctuations might emerge.

As a further extension, we can consider agents who put greater thought into or use emotions to determine their strategy or behavior. When buying something from affiliate advertising, readers' decisions might be influenced somehow by emotions such as sympathy, reciprocity, trust, or anxiety. From a managerial viewpoint, our findings are expected to help integrate affiliate advertising with other advertising media and marketing practices.

References

1. Efron B, Tibshirani RJ (1993) An introduction to the bootstrap. Chapman & Hall, New York
2. Hoffman DL, Novak TP (2000) How to acquire customers on the web. Harvard Bus Rev 78: 179–183
3. Libai B, Biyalogorsky E, Gerstner E (2003) Setting referral fees in affiliate marketing. J Serv Res US 5:303–315
4. Mizuno M (2012) A study on media characteristics and effectiveness measurement of affiliate advertising: in search of ad planning in the age of consumer-driven advertising (in Japanese). Report for a Grant from the Yoshida Hideo Memorial Foundation
5. Papatla P, Bhatnagar A (2002) Choosing the right mix of on-line affiliates: how do you select the best? J Advert 16:69–81
6. Thompson DV, Malaviya P (2011) Consumer response to user-generated advertising: the effects of disclosing consumers as ad creators. Working Paper Series Report No. 11–106, Marketing Science Institute
7. Windrum P, Fagiolo G, Moneta A (2007) Empirical validation of agent-based models: alternatives and prospects. J Artif Soc Soc Simul 10. http://jasss.soc.surrey.ac.uk/10/2/8.html
8. Yano Research Institute (2011) Research Report on the Affiliate Market (in Japanese). http://www.yanoict.com/yzreport/161