Moshe Givon, Vijay Mahajan, & Eitan Muller

Software Piracy: Estimation of Lost Sales and the Impact on Software Diffusion

Software piracy by users has been identified as the worst problem facing the software industry today. Software piracy permits the shadow diffusion of a software parallel to its legal diffusion in the marketplace, increasing its user base over time. Because of this software shadow diffusion, a software firm loses potential profits, access to a significant proportion of the software user base, opportunities for cross-selling, and marketing its other products and new generations of the software. However, shadow diffusion may influence the legal diffusion of the software. Software pirates may influence potential software users to adopt the software, and some of these adopters may become buyers. A diffusion modeling approach is suggested to track shadow diffusion and the legal diffusion of a software over time. The approach enables management to estimate (1) the pirated adoptions over time and (2) the percentage of legal adoptions due to the influence of pirates. The modeling approach is applied to study the diffusion of two types of software (spreadsheets and word processors) in the United Kingdom. The results suggest that although six of every seven software users utilized pirated copies, these pirates were responsible for generating more than 80% of new software buyers, thereby significantly influencing the legal diffusion of the software. The implications of these results are discussed.

In an alarming article, *Business Week* (1992) estimated that in 1991 some of the most competitive U.S. industries lost global sales estimated at as much as $17 billion to pirates. These industries included pharmaceuticals, software, movies, sound recordings, and books. However, among these industries the software industry was identified as having lost the most sales to pirates, $9 of $17 billion.

Software piracy is not limited to the international markets. The Software Publishers Association (SPA), which is the primary trade association for the software industry, has estimated that illegal software cost the industry $1.5 billion in the United States in 1993 (*Fortune* 1994). An independent study commissioned by the Business Software Alliance (BSA), an international organization of software firms, has called software piracy the industry's worst problem (Anthes 1993).

It has been estimated by the SPA that there is one illegal copy of software in use for every one and one-half legal copies (Goldman 1992); that is, 40% (one of 2.5) of the software used in the United States is illegal (*Fortune* 1994).

Despite this alarming figure, this is perhaps the lowest ratio of illegal to legal software use anywhere in the world. It has been reported that such estimates are worse in many European countries (e.g., 80% of the software used in Italy and 86% of the software used in Spain is illegal) and much worse in many countries in Latin America, the Middle East, and Asia (Goldman 1992). The BSA reports that worldwide losses from pirated business software edged up to $12.8 billion in 1993 from $12 billion in 1992 (*Austin American-Statesman* 1994).

The above statistics clearly suggest that software piracy is harmful to the software firms. By losing sales they forego potential profits. By not knowing the users of pirated software they also lose opportunities to cross-sell their other products, market new generations of software, and capitalize on any suggestions from pirates for improving the software or developing new products. A frustrating question for the firms is how to manage the product when they do not know a significant number of their product users.

Is piracy always harmful to firms? Conner and Rumelt (1991) have argued that software piracy may not be harmful for certain types of software, where the value a user derives from the software depends on the user base. In the presence of this positive consumption or network externality (Katz and Shapiro 1986), the utility of the software increases with piracy because it increases the number of other individuals using it.

For such software, piracy protection strategies raise the cost of pirating, causing some would-be pirates to buy and

Moshe Givon is Associate Professor of Marketing, and Eitan Muller is Professor of Marketing, Recanati Graduate School of Business Administration, Tel Aviv University, Israel. Vijay Mahajan is the John P. Harbin Centennial Chair in Business, Graduate School of Business, The University of Texas. The authors thank the editor, three anonymous reviewers, and R. Venkatesh, Don Lehmann, Seenu Srinivasan, Subal C. Kumbhakar, Pradeep Chintagunta, Barry Bayus, and seminar participants at Carnegie Mellon University, Emory University, George Mason University, Stanford University, the University of Texas at Austin, the University of Houston, the University of Illinois, the University of Iowa, and the University of Southern California for their helpful comments.

Journal of Marketing
Vol. 59 (January 1995), 29–37

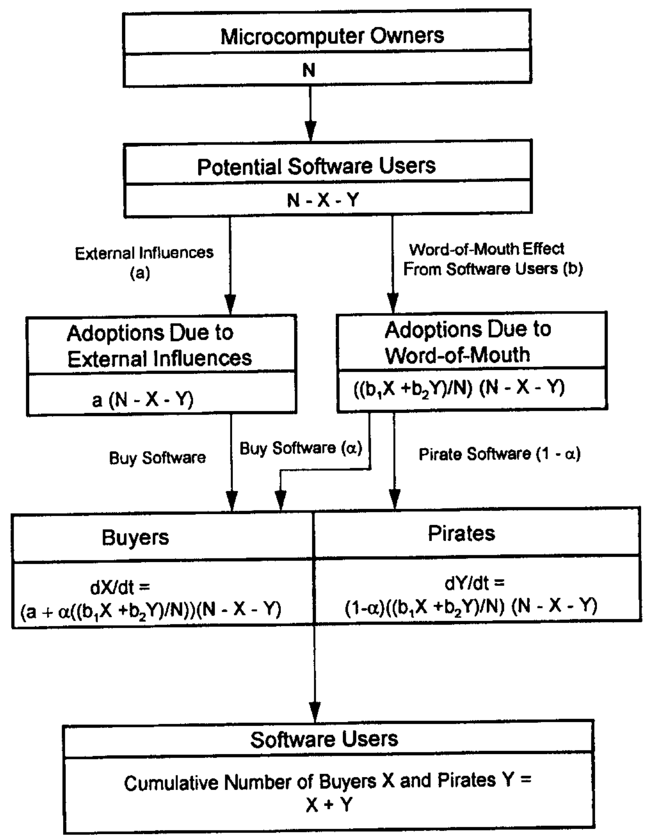
others to do without the product. The resultant smaller user base produces a lower software value and may actually reduce profits and induce the firm to increase the price. In fact, piracy is an efficient form of sampling for firms that want to increase their user base because copies wind up in the hands of the right customers who are willing to pay all costs associated with making and distributing these copies.

Our objective is to further complement the arguments advanced by Conner and Rumelt (1991). Based on innovation diffusion models we argue that for certain types of software, where the word-of-mouth interaction among users and potential users is critical to the growth of the user base over time, pirates play an important role in converting potential users into users of the software, many of whom legally purchase the software. That is, pirates play a dominant role in the generation of buyers over the life cycle of the software.

We achieve our goal by proposing a diffusion modeling approach to capture the growth of a software over time, explicitly taking into consideration the influence of pirates on the software diffusion. Because it requires only legal adoption data, which is easily obtained from industry or company sources for its implementation, the suggested approach enables us to (1) estimate the sales lost to piracy over the life of the software and (2) assess the influence of pirates in generating buyers and vice versa. The approach is illustrated by analyzing the unit sales data of two types of software: spreadsheets and word processors for the United Kingdom market.

This article is organized as follows: The first section details the conceptual and analytical underpinnings of the modeling approach, followed by a section on the analyses of the diffusion of the two software products in the United Kingdom. It concludes with the implications of the analyses, limitations, and further extensions of the model.

FIGURE 1
Buyers and Pirates in the Diffusion Process of Microcomputer Software



The Software Piracy Diffusion Model

Diffusion models focus on the development of a life cycle curve and serve to depict and predict first-purchase sales of innovations (Bass 1969; Mahajan, Muller, and Bass 1990). The behavioral rationale underlying the development of diffusion models is that an innovation is first adopted by a select group of buyers, called innovators, who then influence others by word of mouth to adopt the innovation.

In the presence of piracy, we face a situation where in addition to the legal diffusion of a software, there is also a parallel shadow diffusion. Therefore, the actual user base for the software at any time is composed of both buyers and pirates. This user base can influence the potential users to adopt the software. They can either buy the software or copy it from a user who may also have either purchased the software or copied it from someone else.

What the software firms observe is the legal diffusion of the software through their sales data. However, in reality, any attempt to study the diffusion of the software will be incomplete without understanding the shadow diffusion and its influence on the legal diffusion and vice versa. How can we estimate the rate of the shadow diffusion and its influence on the legal diffusion from data that are available only on legal diffusion?

To answer this question and establish the foundation for the proposed diffusion model, Figure 1 depicts a diffusion process diagram. As shown in this figure, all microcomputer owners at any time are possible potential users of software (Parker 1992). Following Bass (1969) (see also Mahajan, Muller, and Bass 1990), we postulate that two mechanisms convert these potential software users to software users: external influences such as advertising and promotions and internal influences due to the word-of-mouth interactions between software users and potential software users.

Because the software users include both buyers and pirates, in principle it can be argued that the word-of-mouth influence of buyers versus pirates on potential software users should differ. There may be several reasons for this, such as (1) it may be easier psychologically to copy from a pirate, (2) buyers may have a more credible influence, and (3) buyers may put more effort into learning to use the product and use it more extensively; as a result they may have more opportunity to influence others.

On the other hand, it could be argued that the intensity of the word-of-mouth influence of buyers versus pirates on potential software users should not be different. The pirates do not go around publicizing that their software is pirated, and there is no a priori reason to believe that their influence on potential software users is greater or less than that of the buyers.

In the absence of any empirical studies to guide us on this issue, we assume in Figure 1 that the word-of-mouth influence of buyers and pirates on potential software users may not be the same.

We further postulate in Figure 1 that potential software users who are predominantly converted by external influences become buyers of the software, that is, they do not pirate the software.[1] However, of those software users who adopt because of word-of-mouth influence, a fraction (α) of them will purchase the software and the remainder ($1 - \alpha$) will pirate it. When tracked over time, these diffusion dynamics provide estimates for both legal and shadow diffusion.

To mathematically express the diffusion dynamics in Figure 1, let:

$N(t)$ = cumulative number of microcomputer owners at time t
$X(t)$ = cumulative number of buyers of the software at time t
$Y(t)$ = cumulative number of pirates at time t
a = coefficient of external influence
b_1 = coefficient of imitation representing the word-of-mouth influence of *buyers* on potential software users
b_2 = coefficient of imitation representing the word-of-mouth influence of *pirates* on potential software users
α = coefficient representing proportion of individuals influenced by word of mouth who purchase the software

Using the above designations, Figure 1 also includes various analytical expressions for different diffusion flows. (Argument t is dropped in Figure 1 for convenience.) As noted in the figure, at any time t, $a(N–X–Y)$ individuals buy the software due to the external influences. Similarly, of all the individuals who adopt the software because of word-of-mouth influence, that is, $\left(\frac{b_1 X + b_2 Y}{N}\right)(N–X–Y)$, proportion α purchases the software and $(1 - \alpha)$ pirates it.

Summarizing these flows, the following differential equations represent the diffusion dynamics over time for the legal diffusion as well as the shadow diffusion:

Legal Diffusion:

$$(1)\quad \frac{dX(t)}{dt} = [a + \alpha \frac{(b_1X(t) + b_2Y(t))}{N(t)}]\,[N(t) - X(t) - Y(t)]$$

Shadow Diffusion:

$$(2)\quad \frac{dY(t)}{dt} = (1 - \alpha) \frac{(b_1X(t) + b_2Y(t))}{N(t)}]\,[N(t) - X(t) - Y(t)]$$

and $X(0) = 0$ and $Y(0) = 0$.

It should be noted that the summation of equations (1) and (2) represents the total diffusion of the software. In fact, if $Z(t) = X(t) + Y(t)$ or $\frac{dZ(t)}{dt} = \frac{dX(t)}{dt} + \frac{dY(t)}{dt}$ and $b_1 = b_2 = b$, then the summation of equations (1) and (2) yields the Bass model, with a time-varying market potential for the diffusion of the software, that is,

$$(3)\quad \frac{dZ(t)}{dt} = [a + b\frac{Z(t)}{N(t)}]\,[N(t) - Z(t)]$$

In the absence of piracy, $\alpha = 1$ and $Y(t) = 0$ for all t.[2] Also, in this case, Equation (1) reduces to the Bass model with a time-varying market potential for the diffusion of the software, that is,

$$(4)\quad \frac{dX(t)}{dt} = [a + \frac{b_1X(t)}{N(t)}]\,[N(t) - X(t)]$$

Therefore, the value of the parameter α moderates the split between legal and shadow diffusion.

The implementation of equations (1) and (2) requires estimation of a, b_1, b_2, and α. Once these coefficients have been estimated, equations (1) and (2) can be used to estimate the evolution of legal diffusion and shadow diffusion by tracking the ratios $\left(\frac{dX(t)/dt}{dZ(t)/dt}\right)$ and $\left(\frac{dY(t)/dt}{dZ(t)/dt}\right)$ over time.

Furthermore, the ratio $\dfrac{[\alpha\, b_2 \left(\frac{Y(t)}{N(t)}\right)(N(t) - X(t) - Y(t))]}{\frac{dX(t)}{dt}}$ from Equation (1) yields the proportion of buyers who purchase a software because of the influence of pirates. Similarly, the ratio $\dfrac{[(1 - \alpha)\, b_1 \left(\frac{X(t)}{N(t)}\right)(N(t) - X(t) - Y(t))]}{\frac{dY(t)}{dt}}$ from Equation (2) yields the proportion of pirates who adopt the software because of the influence of buyers.[3]

Software Piracy in The United Kingdom

To illustrate the application of the proposed diffusion modeling approach—equations (1) and (2)—to estimate the lev-

[1]Some of the software users who adopt it due to external influences may also pirate the software. The empirical studies (e.g., Mahajan, Muller, and Srivastava 1990) that use such diffusion models report a very low value of the coefficient of external influence. Consequently, the proportion of the potential market that adopts the software due to external influences is also quite small, .2% to 2.8% (Mahajan, Muller, and Srivastava 1990). In our empirical investigation of the U.K. data, we relaxed this assumption and obtained a value of zero for the proportion of these adopters that pirate the software.

[2]When $\alpha = 1$, $\frac{dY(t)}{dt} = 0$ in Equation (2). This implies that $Y(t)$ is a constant. Because $Y(0) = 0$, the value of this constant is zero. That is, $Y(t) = 0$.

[3]It is important to note here that Nascimento and Vanhonacker (1988) have suggested a diffusion model to examine the pricing issue for consumer products that can be reproduced. In our notation, if we assume that N_1 = the eventual number of buyers and N_2 = the eventual number of pirates, their model states the following equations for buyers and pirates:

$$\frac{dX(t)}{dt} = (a_1 + b_1 (X(t) + Y(t)))\,(N_1 - X(t)), \text{ and}$$

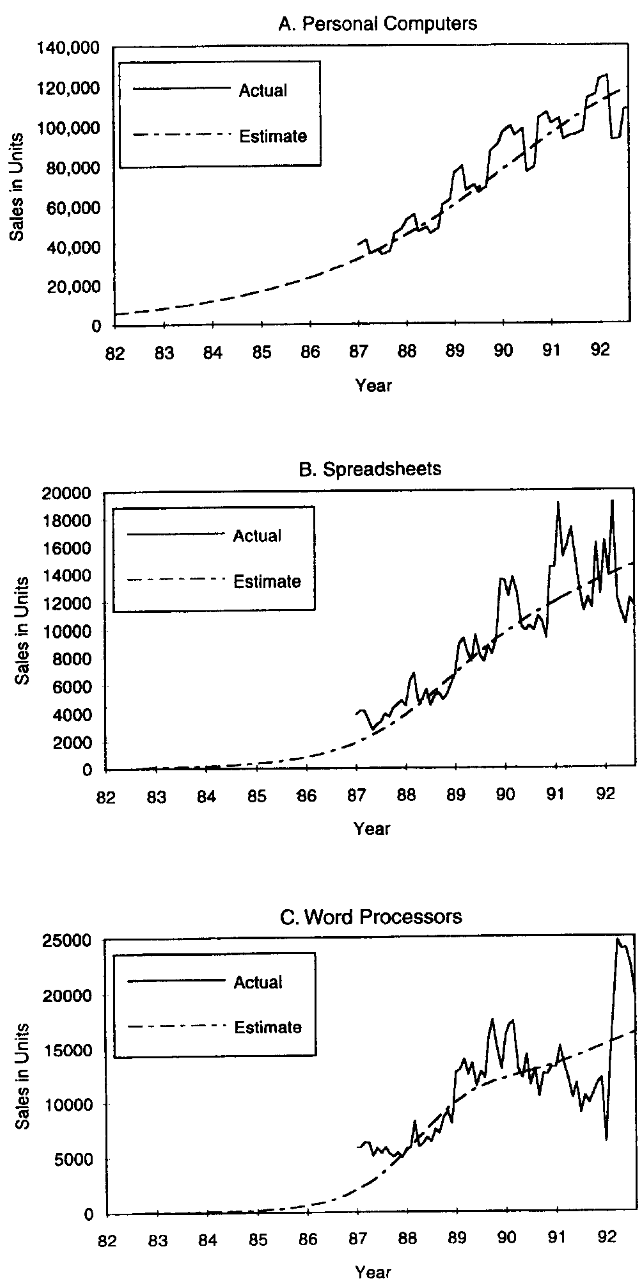
$$\frac{dY(t)}{dt} = (a_2 + b_2 (X(t) + Y(t)))\,(N_2 - Y(t))$$

As compared to our model, where we assume that buyers and pirates come from the same population of potential software users, their formulations assume that buyers and pirates come from two distinct market populations. Nascimento and Vanhonacker (1988) have provided no empirical support for their model. However, we believe that their model is difficult to implement. Because $Y(t)$, the number of pirates at time t is not known, the model requires information on N_1 and N_2 to estimate it. In reality, unless one makes some assumptions about how the eventual market is going to be split between N_1 and N_2, it may not be possible to implement their model.

els of legal and shadow diffusion for software products and to assess the influence of pirates on legal diffusion, we examine the diffusion of spreadsheets and word processors in the United Kingdom.

The monthly shipment data on DOS-based microcomputers and these two software products were bought from MAID (Market Analysis Information Database) in London. These data, for the 68 periods from January 1987 through August 1992, are included in Table 1 and shown plotted in Figure 2.

FIGURE 2
Actual and Estimated Legal Sales of PCs, Spreadsheets, and Word Processors in the U.K.



We assume that the DOS-based microcomputers were introduced in the U.K. in October 1981, and the two software products were introduced in October 1982. Industry sources estimate that these two software products have been widely pirated in the U.K. market. This is also evident from the trends in the levels of units sold for the microcomputers and the two software products in Table 1 and Figure 2.

For example, despite the popularity of these two types of software, the monthly unit sales for the microcomputers ranged from 76,000 to 124,000 from January 1990 through August 1992, whereas the respective monthly unit sales for the two software products for the same period ranged from 10,000 to 24,000. Is there any piracy? How much? Did piracy help the legal sales? These are the questions that we intend to investigate by estimating the coefficients a, b_1, b_2, and α in equations (1) and (2).

If one assumes that all the microcomputer owners will either buy the two software products through legal channels or use pirated copies, one can obtain the number of pirated units by subtracting the cumulative number of software units sold from the cumulative number of microcomputer owners, that is, $Y(t) = N(t) - X(t)$.

In reality, however, not all the microcomputer owners will buy the software products. Hence, $X(t) + Y(t)$, the total number of users of software in equations (1) and (2) will be less than N(t), the total number of microcomputer owners. Over time, however, $(N(t) - X(t) - Y(t))$ will decrease and eventually may approach zero.

The level of piracy can also be measured by conducting a survey of microcomputer owners. However, because the ratio $\frac{Y(t)}{X(t)}$ changes over time, one must repeatedly conduct surveys to assess this value. Furthermore, these surveys may not be able to assess the impact of the word-of-mouth influence on sales. In this respect, the diffusion model proposed in this article is easy to implement, because it requires only the sales data.

Parameter Estimation

To estimate the parameters a, b_1, b_2, and α over time T using equations (1) and (2), we need data on N(t), the cumulative number of microcomputer owners at time t, and X(t), the cumulative number of software buyers at time t. However, because data are only available from January 1987, before we can use equations (1) and (2) we must have an estimate of the cumulative number of microcomputer owners in January 1987.

To generate this estimate, we first fit the Bass model to the available data on microcomputers and then used it to forecast unit adoptions before January 1987. In the discrete form, the Bass model states that:

$$\hat{n}(t) = (\hat{p} + \hat{q}\,\frac{N(t-1)}{\hat{m}})\,(\hat{m} - N(t-1)) \tag{5}$$

where $\hat{n}(t)$ is the predicted noncumulative number of microcomputer buyers at time t, $\hat{p}$ and $\hat{q}$ are estimated coefficients of innovation and imitation, respectively, and $\hat{m}$ is the estimated constant market potential.

Note that $N(0) = 0$ and $t = 1$ is the month of introduction of microcomputers. The parameter estimates $\hat{p}$, $\hat{q}$, and $\hat{m}$

TABLE 1
Actual Monthly Legal Sales of PCs, Spreadsheets, and Word Processors in the U.K.

Month	Year	PCs	Spreadsheets	Word Processors	Month	Year	PCs	Spreadsheets	Word Processors
1	87	40,164	3,811	5,815	1	90	96,098	13,481	16,075
2	87	41,303	4,119	5,814	2	90	97,866	12,336	17,063
3	87	42,466	4,134	6,294	3	90	99,627	13,711	17,409
4	87	35,525	3,456	6,224	4	90	94,725	12,547	12,965
5	87	36,513	2,739	5,004	5	90	96,351	10,199	12,182
6	87	37,521	3,130	5,755	6	90	97,964	9,857	14,309
7	87	34,961	3,282	5,299	7	90	76,556	10,228	11,549
8	87	35,913	3,913	5,780	8	90	77,770	9,881	12,983
9	87	36,884	3,659	5,209	9	90	78,969	10,870	10,487
10	87	45,856	4,290	4,969	10	90	103,640	10,453	12,580
11	87	47,076	4,552	5,336	11	90	105,142	9,309	12,466
12	87	48,319	4,849	4,874	12	90	106,618	14,425	13,133
1	88	52,560	4,458	5,667	1	91	100,641	14,428	13,135
2	88	53,924	6,219	5,764	2	91	101,958	19,068	15,094
3	88	55,310	6,840	8,227	3	91	103,241	15,162	13,375
4	88	46,498	4,806	5,847	4	91	92,712	16,010	12,024
5	88	47,670	4,908	6,178	5	91	93,787	17,292	10,306
6	88	48,860	5,629	6,699	6	91	94,827	15,126	11,709
7	88	45,769	4,483	6,285	7	91	95,080	12,832	8,943
8	88	46,887	5,166	7,459	8	91	96,037	11,274	10,510
9	88	48,020	5,430	7,096	9	91	96,953	12,250	9,853
10	88	60,047	4,912	8,603	10	91	113,410	11,494	10,782
11	88	61,462	5,280	9,096	11	91	114,371	16,130	11,635
12	88	62,893	6,033	8,002	12	91	115,279	12,505	12,158
1	89	76,022	6,745	12,645	1	92	123,043	16,313	6,249
2	89	77,745	8,874	12,891	2	92	123,886	13,836	13,719
3	89	79,483	9,335	13,916	3	92	124,667	19,094	19,116
4	89	67,270	8,366	12,422	4	92	92,329	12,205	24,610
5	89	68,730	7,654	13,544	5	92	92,808	11,144	23,890
6	89	70,198	9,543	11,448	6	92	93,238	10,273	23,971
7	89	66,232	8,051	12,750	7	92	107,933	12,130	22,614
8	89	67,601	7,621	12,138	8	92	108,312	11,655	19,814
9	89	68,974	8,833	15,916					
10	89	86,915	8,176	17,508					
11	89	88,615	9,276	14,972					
12	89	90,315	13,510	12,930					

were obtained by minimizing the total sum of squared errors for the microcomputer owners from January 1987 to August 1992, that is,

$$SS(\hat{p}, \hat{q}, \hat{m}) = \sum_{j=t_s}^{T} [n(j) - \hat{n}(j)]^2 \tag{6}$$

where $\hat{n}(j)$ is the predicted number of microcomputer buyers at time j given by Equation (5) and t_s is the first time period for which data are available. We used a quasi-Newton nonlinear least squares algorithm that is included in the NAG Library No. E04JAF (Phillips 1991) to obtain the following estimates from Equation (6): $\hat{p} = .00037$, $\hat{q} = .0316$, $\hat{m} = 15{,}386{,}100$, and $R^2 = .897$. Figure 2A shows the good model fit to the data.[4] These parameter estimates were used in Equation (5) to generate the noncumulative and cumulative number of microcomputer owners before and including January 1987.

Once the estimated value for the cumulative number of microcomputer owners in January 1987 was known, the available data on microcomputer owners and software buyers (see Table 1) were used to obtain estimates for a, b_1, b_2, and α from equations (1) and (2) by minimizing the total sum of squared errors for the software buyers:

$$SS(\hat{a}, \hat{b}_1, \hat{b}_2, \hat{\alpha}) = \sum_{j=t_s}^{T} [x(j) - \hat{x}(j)]^2 \tag{7}$$

where x(j) is the actual and $\hat{x}(j)$ is the predicted (noncumulative) number of software buyers (legal diffusion) at time j. Here again, X(0) = Y(0) = 0, and t = 1 is the time of the software introduction. The predicted number of software buyers was obtained from equations (1) and (2).

It should be noted that the use of equations (1) and (2) requires information on pirates, Y(t), which is not available. However, the two equations (1) and (2) can be technically written as one equation, $\frac{dX(t)}{dt}$, in terms of lagged variables of X(t), eliminating Y(t) altogether. The resultant equation can be then used in Equation (7) to provide $\hat{x}(j)$. (See the Appendix for details.)

We again used the nonlinear least squares algorithm included in the NAG library No. E04JAF to obtain $\hat{a}$, $\hat{b}_1$, $\hat{b}_2$, and $\hat{\alpha}$, from Equation (7). The algorithm searches for the

[4]Because we use monthly data, we get q = .0316. This is approximately equal to .0316 x 12 = .3792 for the annual data. This value of q is consistent with the average value of q = .38 reported by Sultan, Farley, and Lehmann (1990).

TABLE 2
Parameter Estimates for the Diffusion of Word Processors and Spreadsheets in the United Kingdom

A. $b_1 \neq b_2$

Software	Coefficient of external influence a	Coefficient of imitation for buyers b_1	Coefficient of imitation for pirates b_2	Piracy coefficient α	Explained variance R^2
Word Processors	.0002	.13531	.13511	.14380	.563
Spreadsheets	.00069	.09755	.10409	.12065	.788

B. $b_1 = b_2$

Software	Coefficient of external influence a	Coefficient of imitation b	Piracy coefficient α	Explained variance R^2
Word Processors	.0002	.13518	.14378	.563
Spreadsheets	.00063	.10399	.12122	.789

*The analyses in Table A assume that the intensity of the word-of-mouth influence of buyers and pirates on potential software users is not the same, i.e., $b_1 \neq b_2$. The analyses in Table B assume it is the same, i.e., $b_1 = b_2$

values of these parameters that minimize the total sum of squared errors in Equation (7). Because the optimal values can depend on the initial values provided for these parameters in the algorithm, several different initial values of these parameters were used to check the convergence of the final estimates.

Results

Table 2A provides the parameter estimates for the two software products. In this table the coefficients b_1 and b_2 are very similar for each software (e.g., b_1 = .13531 and b_2 = .13511 for word processors). This suggests that the intensity of the word-of-mouth influence of buyers and pirates on potential software users may be the same.

Given these results, equations (1) and (2) were reestimated assuming that $b_1 = b_2 = b$. These estimates were used to assess the impact of piracy on software diffusion in the rest of the analyses in this article.

Table 2B provides the parameter estimates for the two software products, which are identical to the estimates in Table 2A when $b_1 \neq b_2$.[5] Figures 2B and 2C show that the model fits the data. Once the coefficients a, b, and α were known, the unit sales were also projected for the months before the data available period. These are shown plotted in Figure 2.

In Table 2B α is .14 for word processor and .12 for spreadsheets. These values suggest that of all the potential users who were converted to users at any time t due to word of mouth from users, 86% (i.e., $1 - \alpha$) of word processor users and 88% for spreadsheet users were likely to be pirates. Figure 3A shows the change in the ratio of pirates/buyers over time for both software products. As expected, this ratio grew over time and settled at around 1:6 in the late 1980s; that is, for every buyer who purchased the software there were about six pirates who were also using the software.

Figure 3B shows the percentage of unit sales of buyers due to the influence of pirates. As expected, this percentage grew over time, indicating the overwhelming influence of pirates in converting potential software users to buyers. In fact, from 1988 on, more than 80% of the software purchased by buyers was probably the result of the influence of pirates.

Similarly, Figure 3C shows the percentage of pirated unit adoptions due to the influence of buyers on the potential software users. This percentage decreased over time, stabilizing at around 15% after 1987. As expected, all the pirated adoptions in the beginning of the diffusion process were due to the influence of buyers, because there were not too many pirates.

As the number of pirates increased over time, it exerted a disproportionate influence in converting potential users to pirates.

The growth patterns for the legal and shadow diffusions are depicted in Figure 4, which clearly confirms the strong presence of shadow diffusion or piracy.

Conclusions

This article suggests a diffusion modeling approach to estimating the pirated sales of a software. The model development postulated that although piracy cultivates shadow diffusion of the software parallel to its legal diffusion in the marketplace, the shadow diffusion can have a significant impact on the legal diffusion of the software. Through word-of-mouth interactions, pirates may influence the potential

[5]Table 2 reports results when $\alpha \neq 1$; that is, these analyses suggest the presence of piracy. Because the model that assumes no piracy (i.e., $\alpha = 1$) is nested in equations (1) and (2) and is given by Equation (4), how well does it describe the diffusion of the two software products? While searching for an optimal set of parameter values, the algorithm used for deriving parameter estimates from Equation (7) considers all possible values of α as well. Because it did not suggest optimal parameter estimates of $\alpha = 1$, we fit Equation (4) to the data by setting $\alpha = 1$. Although this model yielded R^2 values comparable to those reported in Table 2, it produced b = 0 for both the software products, suggesting that the diffusion of these two software products was not driven by the word-of-mouth effect. Given this inexplicable result, we concluded that this model did not describe the diffusion of the two software products.

FIGURE 3
Legal and Shadow Diffusion Dynamics for Spreadsheets and Word Processors in the U.K.

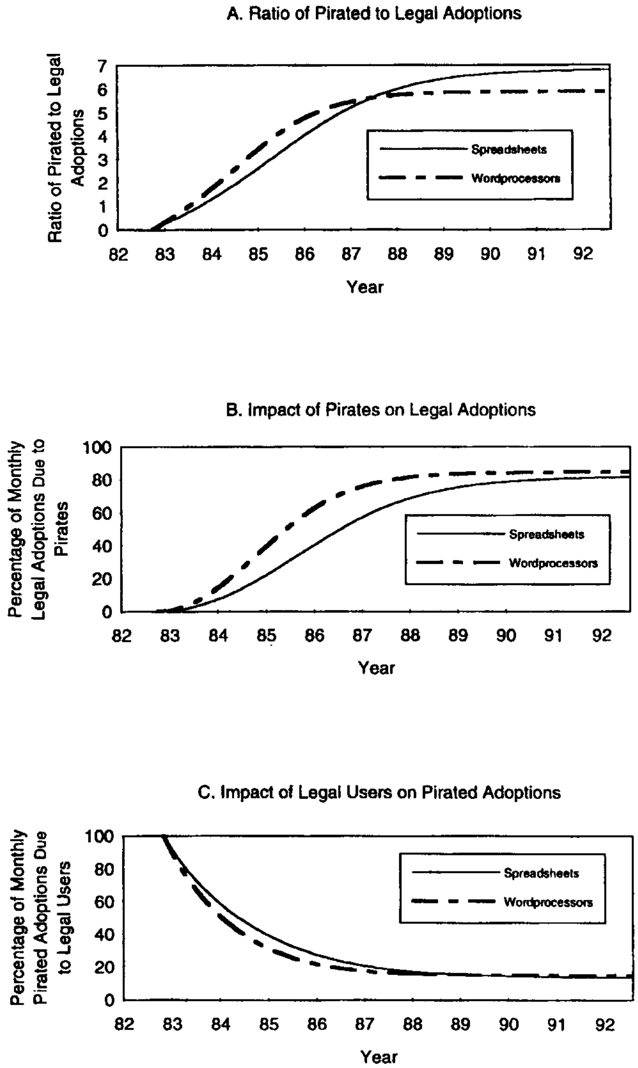
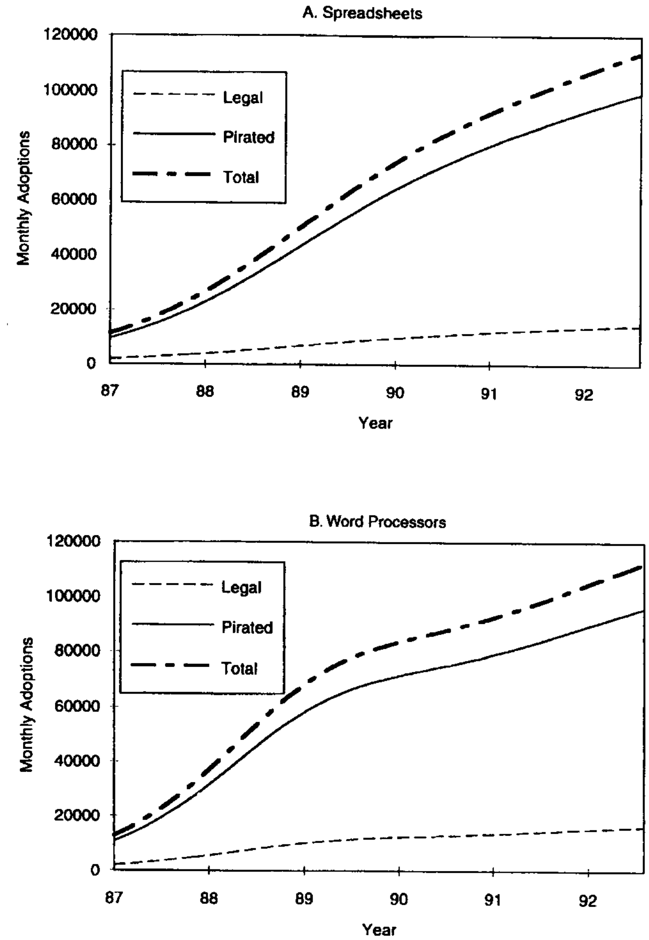


FIGURE 4
Composition of Legal, Pirated, and Total Adoptions of Spreadsheets and Word Processors in the U.K.



users to adopt software, and some of these adopters may purchase the software. In fact, when tracked over time, these diffusion dynamics permit estimation of the sales lost to piracy and the influence of pirates in generating buyers for the software.

The modeling approach was demonstrated to analyze the diffusion of spreadsheets and word processors in the United Kingdom. The results indicated that since the late 1980s, out of every seven software users, six had pirated copies. On the other hand, the pirates significantly influenced the potential users to adopt this software. In fact, they contributed to generating more than 80% of the unit sales for these two types of software.

Limitations and Extensions

The modeling approach, the data used, and the results derived are not without limitations. However, these limitations suggest opportunities for further research. Like the other diffusion studies (Mahajan, Muller, and Bass 1990), data used in this article are at an aggregate level. We assume one adoption (legal or illegal) per microcomputer owner and consider first purchases only.

The simple model does not include site licenses (multiple users per software adoption), software upgrades, or possession of multiple microcomputers (new or second unit) per user. Because we use monthly data, there may be some seasonality in the data. Furthermore, because our data begin in January 1987, it was necessary to use the estimated model to compute adoptions before that date. This retrospective assumed that the earlier years followed the life cycle curve specified by the Bass model (see Figure 2A).

In Figure 3A, the curve depicting the ratio of pirated to legal adoptions increases over time, reaching an asymptotic value. Is it possible for this curve to rise and then fall? That is, can the ratio of pirates to buyers increase or decrease over time? Note from equations (1) and (2) that the ratio of pirates to buyers is given by:

$$(8) \qquad \frac{dY/dt}{dX/dt} = \frac{(1-\alpha)}{a\left(\dfrac{N}{b_1X + b_2Y}\right) + \alpha}$$

If $b_1 = b_2$ and because a, and α are constants, the ratio $\frac{N}{X+Y}$ controls the ratio of pirates to buyers. Because this ratio, $\frac{N}{X+Y}$, is greater than one (eventually approaching one), the ratio of pirates to buyers takes the shape shown in Figure 3A, with the asymptote given by $\frac{1-\alpha}{\frac{a}{b}+\alpha}$. The ratio of pirates to buyers can increase or decrease over time only if α and the other parameters change over time. However, we assume in our analysis that all the parameters are invariant with time.

Figure 1 assumes that the intensity of the word-of-mouth effect of buyers (or pirates) on potential software users (i.e., coefficients b_1 and b_2) is the same whether potential software users convert to buyers or pirates. By creating two additional parameters, b_3 and b_4, it is possible to relax this assumption. This requires replacing coefficients b_1 and b_2 in Equation (2) by b_3 and b_4. However, this will create an identification problem because α and $1 - \alpha$ will be indistinguishable from the b coefficients in their respective equations.

Our results suggest that six out of seven software users of spreadsheets and word processors in the U.K. utilized pirated copies. An important question is how close the model's estimates of pirated units are to the actual pirated units. Because we do not have data on pirated units, we cannot validate this result. However, industry reports mentioned in the introduction of this article and the respectable model fit results reported in Table 2 provide some credibility to the model and the reported results.

Managerial Implications

Given our results, two questions can be raised: (1) What should firms do if their products are subject to piracy? (2) How should the firms change their marketing mix strategies under such conditions?

Answers to these questions clearly depend upon the answer to the following question: Is piracy always bad? The results for the United Kingdom suggest that although the piracy of spreadsheets and word processors was abundant, the pirates helped significantly in creating legal penetration of these two types of software. Hence, any attempt to stop or restrict shadow diffusion of the software could also have significantly slowed their legal penetration. In this respect, our empirical results corroborate the analytical results reported by Conner and Rumelt (1991); for software whose value to the users increases with the expansion in the user base, software protection strategies can be more harmful to the firms than piracy.

Our objective in this article is not to advocate piracy. However, our results for the United Kingdom suggest that software piracy has an impact on the shape and form of the legal product life cycle of a software.

Therefore, instead of destroying shadow diffusion of a software, firms whose products are subject to piracy may be well advised to examine marketing mix mechanisms that can facilitate the conversion of shadow diffusion into legal diffusion. These mechanisms may include differential pricing strategies, limited and self-destructing software codes, bundling of software, sharing of software (shareware), installation of software in the hardware itself, software clubs, and self-help software books. These examples are primarily provided to highlight the point that instead of using punitive mechanisms (e.g., lawsuits, raids, or protection strategies) that might destroy shadow diffusion, firms can use creative marketing mechanisms to convert shadow diffusion to their advantage.

Consider, for example, a firm that is contemplating installation of a protective device to stop piracy. In the presence of shadow diffusion, it faces the critical issue of the optimal timing for the installation of the protective device. At a first glance, it seems that bcause of the relative low cost of such devices and the large number of pirates, it makes sense for the firm to launch the product with the protective device already installed.

However, given the very sizable impact of pirates on legal buyers (Figure 3B), the firm will find it best in such cases to wait and take advantage of this positive influence of pirates on buyers and introduce the protective device at a later stage, possibly with a new generation of the software. In addition, the sheer size of the pirate group might act as a barrier to entry for any potential competitor, who might decide either to delay or even cancel the introduction of its new product.

It should also be noted that the existing innovation diffusion literature totally ignores the existence of shadow diffusion. Innovation diffusion studies that relate the propensity to pirate to demographic and religious characteristics, social communication structures, and the effectiveness of various marketing instruments, including price, will be helpful in providing further guidelines to firms that face piracy (e.g., Dhebar and Oren 1985; Nascimento and Vanhonaker 1988; Solomon and O'Brien 1991).

Appendix

Technically, the two equations (1) and (2) could be written as one equation, dX(t)/dt, in terms of lagged variables of X(t), altogether eliminating Y(t). To clarify this point, when x(t) and y(t) represent noncumulative monthly adoptions, the equations (1) and (2) in discrete form can be rewritten as:

$$x(t) = \left[a + \alpha\,\frac{[b_1X(t-1) + b_2Y(t-1)]}{N(t)}\right](N(t) - X(t-1) - Y(t-1))$$

$$y(t) = (1-\alpha)\,\frac{[b_1X(t-1) + b_2Y(t-1)]}{N(t)}\,(N(t) - X(t-1) - Y(t-1))$$

We start with X(0) = Y(0) = Y(1) = 0, conditions that indicate piracy lags legal adoption.

$$X(1) = x(1) = aN(1)$$

$$Y(2) = Y(1) + y(2) = (1 - \alpha)\,\frac{[b_1X(1) + b_2Y(1)]}{N(2)}\,(N(2) - X(1) - Y(1))$$

$$= (1 - \alpha\,\frac{[b_1aN(1)]}{N(2)}\,(N(2) - X(1))$$

$$x(2) = \left[a + \alpha\,\frac{[b_1X(1) + b_2Y(1)]}{N(2)}\right](N(2) - X(1) - Y(1))$$

$$= \left[a + \alpha\,\frac{[b_1X(1)]}{N(2)}\right](N(2) - X(1))$$

Note that x(2) is expressed in terms of X only; we fit the model on x(t). Similar expressions for Y's, x's, and X's can be written for other time periods. This simplified deterministic formulation of Y(t) is a direct offshoot of the nonavailability of data on shadow diffusion, which is typical given the illegality of piracy.

We recognize that a superior estimation procedure would be one that uses data on shadow diffusion and includes error terms both in equations (1) and (2). However, because Y(t) appears as an independent variable in Equation (1) and X(t) appears in Equation (2), the parameters must be estimated, treating the two equations as a system with correlated errors and cross-equation restrictions. Because we do not have data on shadow diffusion and could not implement this procedure, we did not try to ascertain standard errors for the estimates reported in Table 2.

In view of the deterministic formulation of Y(t), the only exogenous variable in our model is X(t), the cumulative adoptions of spreadsheets or word processors, and the endogenous variable is x(t), the noncumulative adoptions at time t. Therefore, the model is identified. In addition, for any value of $\alpha\ b_i$, there is a unique value of $(1 - \alpha)\ b_i$ in equations (1) and (2).

REFERENCES

Anthes, Gary H. (1993), "Study Cites Software Industry Growth, Piracy Problems," *Computerworld*, (March 29), 119.

Austin American-Statesman (1994), "Software Piracy Costs $12.8 Billion," (April 29), E4.

Bass, Frank M. (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15 (January), 215–27.

Business Week (1992), "The Patent Pirates Are Finally Walking the Plank," (February 17), 125–27.

Conner, Kathleen Reavis and Richard P. Rumelt, (1991), "Software Piracy: An Analysis of Protection Strategies," *Management Science*, 37 (February), 125–39.

Dhebar, Anirudh, and Shmuel S. Oren, (1985), "Optimal Dynamic Pricing for Expanded Networks," *Marketing Science*, 4 (Fall), 336–51.

Fortune (1994), "Freeze, It's the Cyber Fuzz," (May 2), 12

Goldman, Neal D. (1992), "Software Theft: A $4.5 Billion Headache," *Business Forum*, (Spring), 10–12.

Katz, Michael and Carl Shapiro (1986), "Technology Adoption in the Presence of Network Externalities," *Journal of Political Economy*, 94, 822–41.

Mahajan, Vijay, Eitan Muller, and Frank M. Bass (1990), "New Product Diffusion Models in Marketing: A Review and Directions for Research," *Journal of Marketing*, 54 (January), 1–26.

———, ———, and Rajendra K. Srivastava (1990), "Determination of Adopter Categories by Using Innovation Diffusion Models," *Journal of MarketingResearch*, 27 (February), 37–50.

Nascimento, Fernando and Wilfried R. Vanhonacker (1988), "Optimal Strategic Pricing of Reproducible Consumer Products," *Management Science*, 34 (August), 921–37.

Parker, Philip M. (1992), "Price Elasticity Dynamics Over the Adoption Life Cycle," *Journal of Marketing Research*, 29 (August), 358–67.

Phillips, Jen (1991), *The Nag Library: A Beginner's Guide.* Oxford, England: Clarendon Press.

Solomon, Susan L. and James A. O'Brien (1991), "The Effect of Demographic Factors on Attitudes Toward Software Piracy," in *Ethical Issues in Information Systems*, R. Dejoie, G. Fowler, and D. Paradice, eds. Boston, MA: Boyd & Fraser.

Sultan, Fareena, John U. Farley, and Donald Lehmann (1990), "A Meta-Analysis of Applications of Diffusion Models," *Journal of Marketing Research*, 27 (February), 70–77.