

From Density to Destiny: Using Spatial Dimension of Sales Data for Early Prediction of New Product Success

Tal Garber, Jacob Goldenberg

School of Business Administration, The Hebrew University of Jerusalem, Jerusalem, Israel 91905
{tal.garber@intel.com, msgolden@huji.ac.il}

Barak Libai, Eitan Muller

Leon Recanati Graduate School of Business Administration, Tel Aviv University, Tel Aviv, Israel 69978
{libai@post.tau.ac.il, muller@post.tau.ac.il}

One of the main problems associated with early-period assessment of new product success is the lack of sufficient sales data to enable reliable predictions. We show that managers can use spatial dimension of sales data to obtain a predictive assessment of the success of a new product shortly after launch time.

Based on diffusion theory, we expect that for many innovative products, word of mouth and imitation play a significant role in the success of an innovation. Because word-of-mouth spread is often associated with some level of geographical proximity between the parties involved, one can expect “clusters” of adopters to begin to form. Alternatively, if the market reaction is widespread reluctance to adopt the new product, then the word-of-mouth effect is expected to be significantly smaller, leading to a more uniform pattern of sales (assuming that there are no external reasons for clustering). Hence, the less uniform a product’s distribution, the higher its likelihood of generating a “contagion process” and therefore of being a success. This is also true if the underlying baseline distribution is nonuniform, as long as it is an empirical distribution known to the firm.

We use a spatial divergence approach based on cross-entropy divergence measures to determine the “distance” between two distribution functions. Using both simulated and real-life data, we find that this approach has been capable of predicting success in the beginning of the adoption process, correctly predicting 14 of 16 actual product introductions in two product categories. We also discuss the limitations of our approach, among them the possible confusion between natural formation of geodemographic clusters and word-of-mouth-based clusters.

Key words: new products; innovation diffusion; spatial analysis; complexity

History: This paper was received August 1, 2002, and was with the authors 5 months for 3 revisions; processed by Arvind Rangaswamy.

1. Introduction

A marketing manager faces the following problem: S/he has three months’ worth of biweekly sales data on two new products sold in the same geographical areas, targeted at roughly the same segments, using the same distribution channels. Although both time series look similar, the manager feels that one of the two products is not as successful as the other and would like to back up this hunch with quantitative data, as decisions must be made regarding advertising, pricing, and promotion for the two products.

Such dilemmas are quite common, because although much of the research on new product introductions has traditionally focused on improving the initial go/no-go decision, one of the main problems associated with early-period assessments is the lack of data to enable further predictions. The few periods of aggregate sales of monthly data available to marketers render predictions relatively unreliable. For example,

it has been suggested that using diffusion models, such as the well-known Bass model, a stable estimate of the diffusion process parameters requires sales data from introduction through peak of sales (Srinivasan and Mason 1986). Therefore, it has been argued that diffusion models cannot serve as an effective predictive tool for the early years of a product’s life (Kohli et al. 1999, Mahajan et al. 1990a). One promising way to overcome this shortcoming and enable early forecasting is to use orders in the prelaunch stage (Moe and Fader 2002). Another approach is to use market tests such as IRI BehaviorScan tests when these are available (Fader et al. 2003). Unfortunately, in many cases such data are not readily available.

We propose a novel approach to overcome some of the problems associated with early prediction of new product success based on sales data. Instead of using only time series of aggregate sales, we add its spatial dimension, i.e., not only *how much* the product sold

but also *where*. This approach uses more information and requires fewer time-dependent data points to render better prediction. Using our approach, we have been able to predict early on the success of new products in simulations describing the products' growth in space and time, as well as using real data on 16 innovative products.

Spatial data that are often available to marketers have yet to receive the attention they deserve from researchers seeking to predict new product success. While the results in this paper look promising, one should be aware of the limitations of this method. For example, while the method naturally fits spatially homogeneous segments better, it requires more work to apply in cases where natural geodemographic segments exist. We discuss these limitations in the final section of the paper.

2. Pairing Spatial and Temporal Dimensions of Growth

When new products enter a market, they diffuse in time and space. As compared with the wide interest in marketing in the temporal diffusion of new products, scant attention has been paid to the spatial pattern of growth and its relationship to the temporal. Exceptions are Mahajan and Peterson (1979), Allaway et al. (1991), Bronnenberg and Mahajan (2001), and Song and Bell (2002). Recently, studies have been conducted that examine global diffusion of technologies focused on the proximity between countries and geographical regions to explain the temporal patterns of growth (Tellis et al. 2003, Dekimpe et al. 2000, Putsis et al. 1997).

The rationale for the formation of adopter clusters is related to the role of word of mouth and imitation in the diffusion of innovations. Following the diffusion paradigm that views the communication process as the main driver of new product growth, one can see two types of communication effects: external and internal (Mahajan et al. 1990a). While the external effects represent the marketing efforts of the firm, the internal effect represents the influence of previous adopters and provides two concurrent positive reinforcements: positive word of mouth and a source for imitation and legitimation. To a large extent, internal effects constitute the market reaction to the product. If the product is well received, then word of mouth and imitation will carry forth the message, followed by more and more adopters, further feeding the flow of internal influence. Eventually, having reached a sufficient level of internal effects, the product will take off.

Internal effects have been found to be the underlying and driving force of innovation diffusion of many new products, exerting an influence exceeding that of external marketing efforts such as advertising

(Goldenberg et al. 2001, Rogers 1995, Mahajan et al. 1990b). However, for internal effects to take place and personal recommendations to begin circulating, adopters typically must share some form of physical proximity. Indeed, the diffusion literature reports a clear correlation between geographic proximity and the strength and speed of word-of-mouth spread, sometimes labeled the "neighborhood effect" (Baptista 2000, Case 1991, Mahajan and Peterson 1979). It is easy to see, even with a simple simulation, that from a geographical point of view, word-of-mouth effects drive a formation of spatial clusters.

However, if the product in question is a "dud," then one can expect the internal effects activity associated with it to be minimal, with no contiguous units buying it. While some consumers will adopt it, mainly as a result of external effects, the effect of their adoption on other consumers will be negligible: Clusters will not form, and spatial distribution will be mainly a result of external effects; i.e., adopters will be randomly distributed in space in what can be expected to be a geographical distribution close to that of a uniform distribution. Thus, a lack of contiguous units adopting the product may be a strong signal for the product's failure, assuming that clusters are not formed due to reasons external to the contagion process.

Figure 1 illustrates our main argument. Figure 1a presents a simulated adoption of a successful product in six discrete time periods in an area with a homogenous population. The adoption of this product is clearly characterized by clusters. In comparison, Figure 1b presents the adoption of a second product, whose distribution of adopters in the same geographical area is relatively uniform.

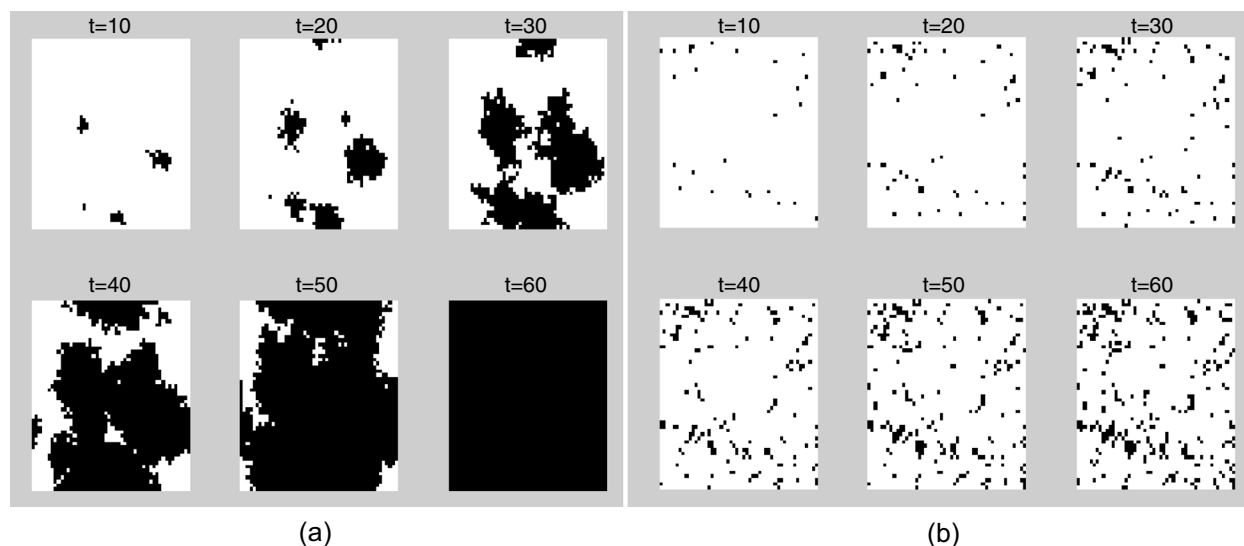
Our claim is that even *before* period 10 it is possible to predict the eventual success/failure of the products by comparing their spatial distribution to the uniform distribution, using an accurate measure of divergence.

3. Small-World and Cross-Entropy

3.1. The Complex Systems Approach

In the first study, we use a simulated environment to evaluate the model. In order to create this environment, we utilize a *complex system approach*. In order to calibrate and validate the proposed method, one must create dyadic sets of data containing successes and failures allowing for reasonable variance within these sets. In a complex systems approach, a system is analyzed through a simulated "would-be world" that allows testing of a wide range of scenarios. One well-known complex system method is *cellular automata*, which has been recently introduced to the marketing literature as described by Goldenberg et al. (2001, 2002). New product growth cellular automata depicts the market as a matrix, the elements of which

Figure 1 Spatial Adoption of Two Products: (a) Successful Product (Clustered) and (b) Failed Product (Uniformly Distributed)



represent adoption of individuals comprised of various discrete cells, where the location of each cell is determined and taken into account in order to render a spatially meaningful process. Each cell interacts with its neighboring cells, with this interaction evolving in time and possibly producing complex behaviors.

Cellular automata’s popularity stems from the fact that despite its parsimony, it generates a wide range of dynamics and growth patterns. However, techniques that model the connection among agents that are *not* necessarily neighbors can also be used to examine the evolution of markets. Among them is the *Small-World* network (see Watts and Strogatz 1998 and, for recent marketing implications, Balakrishnan et al. 2002), a promising technique that has recently drawn considerable attention. An advantage of the Small-World system is that it enables us to describe a social system with a flexible connective structure between networks. According to the Small-World approach, nodes are uniformly distributed on a circle or in a matrix, each connected to its neighbors up to some prespecified range. In addition, a limited number of nonneighboring nodes are allowed for interaction through shortcuts, appearing as strings inside the circle or random shortcuts throughout the matrix, as seen in Figure 2, for a circle network¹.

¹ One of the more colorful examples of a Small-World setting is the Erdős number project. Paul Erdős (1913–1996), a widely traveled and prolific Hungarian mathematician, wrote hundreds of research papers in collaboration with others. His Erdős number is 0. Erdős’ coauthors have Erdős number 1. Coauthors of individuals with Erdős number 1 have Erdős number 2, and so on. For example, one of the authors of this paper has an Erdős number 4, and thus the Erdős number of the other three coauthors is at most 5. The latter might be smaller if there exists a chain of links fewer than 5 connecting them to Erdős. An interesting point to observe—found in the Erdős Number Project homepage—is that almost everyone with a finite Erdős number has a number less than 8.

In a Small-World network, consumers can take on two values: 0 for nonadopters and 1 for adopters. The rules that define transitions of potential adopters from state 0 to state 1 are classified into two types:

- *External Factors*: probability p exists, such that in a certain time period, an individual will be affected by external influence factors such as advertising, to adopt the innovative product.
- *Internal Factors*: probability q exists, such that during a single time period, an individual will be affected by an interaction with a *single* other individual who has already adopted the product.

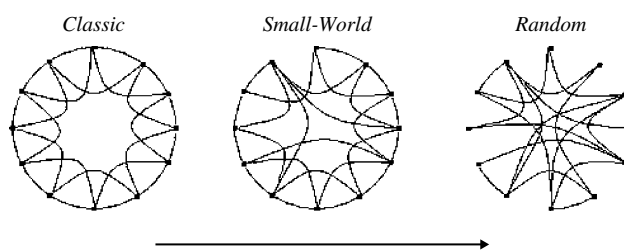
Note that these assumptions directly correspond to those of the aggregate-level Bass model.

The time-dependent (noncumulative) individual probability of adoption, $\text{prob}(t)$, given that the individual has not yet adopted, is based on the following binomial formula:

$$\text{prob}(t) = 1 - (1 - p)(1 - q)^{v(t)+r(t)}, \quad (1)$$

where $v(t)$ is the number of previous adopters with whom the individual maintains contact in the vicinity of the cell under consideration and $r(t)$ are adopters out of his/her *weak-ties contacts* (outside the vicinity). Thus Small-World introduces some random communications between cells that are not in

Figure 2 Small-World Networks



Increasing randomness by adding shortcut links between remote nodes

proximity to each other, thereby in turn introducing more complexity into the network generated by cellular automata modeling. We will report the simulation results using Small-World. However, the same analyses reported below were also performed using cellular automata, with similar results.

3.2. Spatial Divergence Measurements

We are interested in measuring the divergence between the spatial distribution of the product in question and uniform distribution. A rich literature exists concerning divergence measures between two distributions (Johnson and Sinanovic 2001, Lee 1999). A widely used group of such measures is the Ali-Silvey class, which measures the expectations of the likelihood ratio of the two distributions. Consider two probability functions f_1 and f_2 . A (discrete) Ali-Silvey divergence measure is represented by:

$$d(f_1, f_2) = \sum_x f_2(x)g[f_1(x)/f_2(x)], \quad (2)$$

where g is a convex function (see Csiszar 1991). Of this class, the most commonly used divergence measure is the *cross-entropy* divergence (Kullback 1997) in which the function g of Equation (2) is the following: $g(x) = x \log x$. Thus the cross-entropy divergence can be represented by:

$$CE(f_1 | f_2) = \sum_x f_1(x) \log[f_1(x)/f_2(x)]. \quad (3)$$

As the logarithmic function in Equation (3) can be represented as $\log(f_1) - \log(f_2)$, *cross-entropy* sums up the distance between any two points in the two distributions weighted by the probability that these points could occur. To overcome the lack of symmetry of the above measure, another measure in the Ali-Silvey class can be formed (Lee 1999): The *Jensen-Shannon* measure (JS) is obtained from Equation (3) by the following: $JS(f_1, f_2) = CE(f_1 | (f_1 + f_2)/2) + CE(f_2 | (f_1 + f_2)/2)$. In this paper we have performed the calculations on our models using both measures of cross-entropy and the Jensen-Shannon divergence. We found virtually no difference between the predictive abilities of the two measures, and thus we report our results using cross-entropy only.

In order to operationalize the divergence measure, we divide the geographical area under consideration into smaller subsets, called “windows.” The density of adoptions for each window would be the number of adopters in the window divided by the total market potential. Two opposing considerations should be taken into account when determining the window’s size. The first is the adequacy of the representation of the distribution by the complete set of windows. For this purpose, a high resolution is needed and a smaller window is required. Yet decreasing the window’s size causes the number of individuals inside the window to decrease

accordingly, causing a possible decrease in the accuracy of the estimation. However, in order to reduce the noise of the estimated distribution curve, a larger window is required. These two demands are contradictory, and thus it is important to define an efficient window size. One recommended approach is to employ *Parzen Windows* of equal size—proportional to the square root of the potential market—as the area unit of analysis, a widely used technique considered an efficient method for the construction of density functions (Parzen 1962). In the appendix, we show how forecasting accuracy increases when we move from coarse partition to finer partitions.

4. Cross-Entropy as a Predictor of Success

We ran 100 Small-World simulations of 2,500 potential adopters each, with varying p and q values so as to cover various possible ranges of adoption curves (from complete failures to a strong noticeable takeoff). The parameters p and q were chosen to comply with findings on values of aggregate diffusion, transformed to an individual-level grid, where p ranged from 0.0001 to 0.04, and q from 0.0001 to 0.03 (see Sultan et al. 1990 for aggregate diffusion modeling results and standard diffusion parameters; Goldenberg et al. 2001, 2002 for a discussion of the transformation of parameters to individual-level complex methods).

In order to use the Small-World model efficiently, one must decide on the proportion of the distant links from the entire set of links. We have chosen the classic level of 5% as an upper limit, because beyond this level, the social system becomes similar to a random network and less comparable to real-life social systems (Amaral et al. 2000). One common approach is to estimate the percentage of adopters associated with a takeoff. Rogers (1995) suggests that takeoff typically occurs when about 16% of the market potential adopts. Indeed, the 16% representing the cumulative number of innovators and early adopters according to Rogers’ adopter categories is sometimes an accepted number with regard to the anticipated end of the Introduction Stage (Moore and Pessemier 1993).

Thus, a growth process is deemed a success if 16% of the market is obtained before a specified time T . Note that the determination of the time T may differ for various growth processes. However, for the durables studied in the meta-analysis reported by Sultan et al. (1990), takeoff was typically reported prior to ten periods. Thus we consider failure all cases in which one of the following two events takes place²:

² We also used experienced judges to visually determine the existence of takeoff within a certain time period, in the spirit of Golder and Tellis (1997). The results in terms of correct prediction using cross-entropy were similar and are available from the authors.

Table 1 Logistic Regression Confusion Table

	Predicted success (%)	Predicted failure (%)
Observed success	93.1	6.9
Observed failure	14.3	85.7

Note. Average correct predictions = 90.0%.

(i) 16% of the population has not adopted within ten periods ($T = 10$)

(ii) The peak in sales occurs prior to the time at which 16% of the potential adopters have adopted the product.

For each process, the cross-entropy was calculated, and a logistic regression performed on the success measure and the cross-entropy measure of divergence. The logistic regression correctly classified 90% of the cases, as shown in Table 1.

In addition, we performed a boot-strapped hold-out prediction by dropping one product at a time and predicting its success/failure based on the parameters of the logistic regression. The results remained precisely the same with average correct predictions of 90.0%. In order to further test the stability of our regression model, we performed a split-half analysis (the holdout sample was 50% and selected randomly). We used the logistic regression model with its parameters to make predictions on the holdout sample. Again, there was no change in the confusion table, and prediction level remained exactly the same.

To better understand how the two processes of a successful product and a failure differ in terms of cross-entropy, consider Figure 3, which illustrates the time-dependent cross-entropy calculation for the two products represented in Figures 1a and 1b. The difference in cross-entropy is noticeable early in the growth process. As the process continues, the difference between the cross-entropies decreases, and in this example, diminishes as sales approach the peak level for the two cases of success and failure.

The phenomenon depicted in Figure 3 may look counterintuitive: Typically, one can expect that as

more sales data are accumulated, better predictions regarding a product’s success will be generated. Here, however, better separation between successful and nonsuccessful processes is observed at earlier stages. The reason for this behavior lies in the intrinsic development of clusters through word of mouth. In the case of a successful product, initially small kernels of adopters form, as can be observed in Figure 1. The kernels grow to form clusters, and later on clusters start to merge. At a certain point in time, it becomes harder to distinguish between cluster-based and noncluster-based growth processes, as the distribution of buyers reverts to a spatially near-uniform distribution. In fact, toward the end of the process, clusters of nonadopters have formed (as opposed to a uniform distribution of nonadopters in the failure case), and the difference between the cross-entropies of the two processes is once again distinguishable.

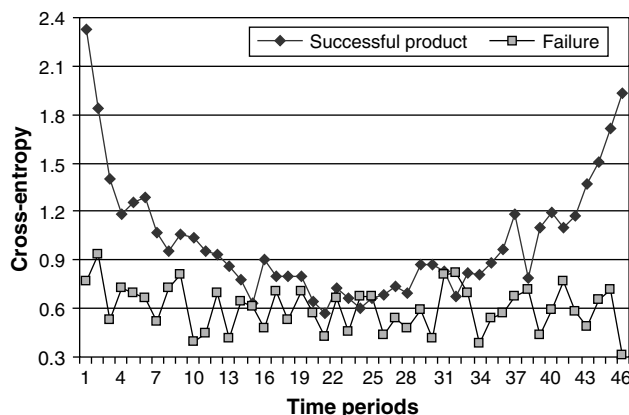
5. Field Tests

5.1. Supermarket Products

Obtaining the data needed to test the divergence method in a real-life application is not simple. First, one needs to obtain growth data of a product, as opposed to only aggregate data, in a number of geographic locations. Second, for discriminating validity purposes, one needs to obtain spatial sales data on failures, which are generally hard to obtain. We were able to obtain data on the sales of eight new health and personal hygiene products sold in a dominant supermarket chain in a Mediterranean country. The data include monthly sales during the first year. While the general product categories are relatively mature, as is the case with most supermarket products, the products themselves were considered innovative relative to the category. Each product was launched simultaneously in each of the chain’s locations. The successful products had a pattern of a rapid, monotonic increase in sales. In contrast, the failed products did not sell well based on the chain’s standards, and very moderate growth was observed. At the time, these products were under consideration for removal from the shelves and cessation of their distribution.

Because we were limited to data provided by the supermarket chain, data could not be broken down to optimal window (Parzen windows) size. Rather, it was coded based on the retailer’s data into 12 regions. Consequently, window size had to be constrained by the weight and area of the retailer distribution regions. Each window was weighted according to the corresponding area’s relative population size. For cross-entropy calculation purposes, each region was designated as a window. Before comparing the probabilities of success in a manner similar to the previous study, we can examine the cross-entropy differences between the successful and failed products to see if a

Figure 3 Cross-Entropy Measures of the Two Products



difference is sufficiently great to discriminate between the cases. What we find is a pattern—similar to that in Figure 3—in which the successful products have a *declining* cross-entropy measure, while the failures have a *consistently low* cross-entropy measure. Figure 4 presents the cross-entropy measure each month for one successful product and one failed product. Both graphs use the same scale. The difference between failure and success is indeed pronounced in the first periods (about 50 times higher in the first period for the successful product). As can be seen, the early-period cross-entropy pattern of these cases is similar to the cross-entropy pattern for the Small-World case in Figure 3, obtaining its maximum level at the beginning of the process.

This figure demonstrates how a clear difference in divergence helps to differentiate between successes and failures by qualitative visual assessment. Yet a more rigorous approach should call for a quantitative tool that would help to convert cross-entropy results into probabilities of success in an unambiguous manner. Thus we return to Small-World simulation, this time as a tool to translate the empirical results into success probabilities. We ran Small-World and connected success or failure status (as the depen-

dent variable) to cross-entropy divergence measure (as an independent variable) through logistic regression. As will be explained next, using the resultant logistic regression function, we can determine the probability of success for each cross-entropy value, including real-life cases.

We must, however, take into account that unlike the simulations in which the area unit was based on the Parzen windows, real-life data may present itself as geographical areas with varying sizes, market potentials, and concentrations of adopters. Hence, the Small-World analysis should use spatial units calibrated to take into account actual size and potential. Consequently, since the actual market was divided into twelve geographical areas, the simulated space was divided into twelve windows of relative sizes to match the actual areas of distribution. We calculated cross-entropy for each simulated process again, taking into account the new window sizes. We performed a logistic regression that uses the new results with a dependent variable of success/failure to produce a function that describes the success probability as a function of the cross-entropy in the field case at hand. Thus, when a marketing manager wishes to estimate the probability for success for a specific case, s/he should perform the following procedure:

(1) First, calculate the cross entropy of the spatial distribution of the sales, based on the windows made possible by the data available.

(2) Next one needs to form the function that will translate the cross-entropy values into probabilities of success. In order to do this one has to follow these three steps:

(2a) Use the set of Small-World simulations with the following modification. The windows used to calculate cross-entropy are now changed from optimal (Parzen) size to windows that match the window numbers and sizes in the real-life case of Stage 1. The result is the same set of processes but with new cross-entropy values.

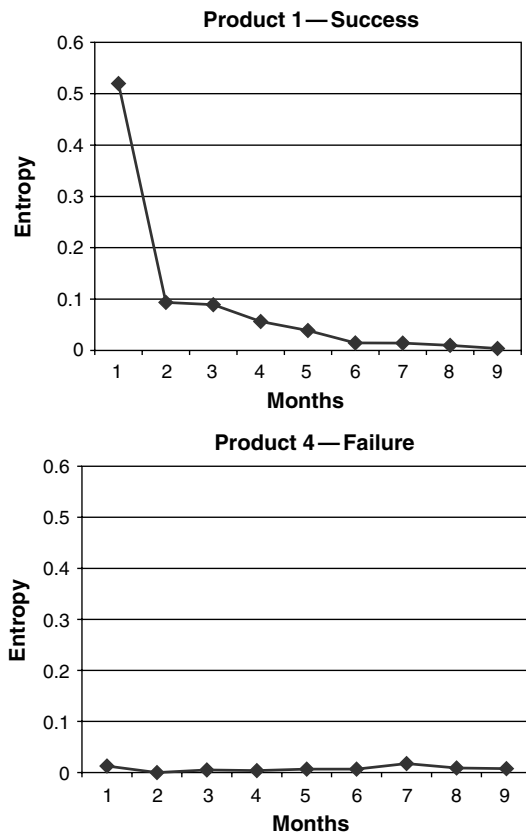
(2b) Create a data set for the logistic regression: Each point is actually one Small-World process, the independent variable is the cross entropy measures early in the process and the dependent variable indicates whether the product is a success or failure.

(2c) Run the logistic regression to form the relationship between cross entropy values and probability of success.

(3) Now, going back to the cross entropy of the real-life case, and using the relationship determined in Stage (2c), find the probability of success for this specific case.

The calculation of the cross-entropy of the Small-World simulation for the 12 regions of the super-market chain again revealed a phase transition-like phenomenon in which there was clear distinction

Figure 4 Cross-Entropy, Calculated Monthly for the Two Products



between success and failure. Again, high values of cross-entropy indicate a high probability of success, and low values indicate a low probability of success. The transition between these two regimes is abrupt, with the threshold value that discriminates between success and failure found to be around 0.3. The fact that this measure is sufficiently robust, even when the window sizes are far from optimal, suggests that cluster formation may be an eminent factor of success. The correct predictions of the logistic regression (based on the Small-World cases) are 80%. This value is naturally not as high compared to the values achieved when using optimal window size in the simulation study (see *The Effect of Number of Windows Reduction* in the appendix). Table 2 presents the predictions of the logistic regression.

The first column of Table 2 represents the product, while the second column indicates the field observations based on the growth of the product’s sales during the nine-month period. The third column shows cross-entropy value for the first month. It can be seen that the cross-entropy of a successful product is in order of magnitude larger than that of a failure. The last column shows the prediction of success probability, calculated through the logistic regression. Thus, for example, for product 1, a success, the cross-entropy measure of 0.52 yields a 99.9% probability of success in the logistic regression function.

Note that the fifth case (denoted with a double star) is a successful product that was predicted to be a failure. In a discussion with the executives of the distribution chain, the possibility was raised that the success of the product was so phenomenal, that after one month the diffusion covered too large an area, causing the cross-entropy to drop sooner than the other products. If correct, this means that regarding a rapid diffusion of a highly successful product, the cross-entropy value must be measured earlier than the first month.

Table 2 Supermarket Product Category: Cross-Entropy Calculations for Successful and Failed Products

Product	Outcome	Cross-entropy (first month)	Probability of success* (%)
1	Success	0.52	99.9
2	Success	0.51	99.9
3	Success	0.31	99.8
4	Success	0.30	99.8
5	Success**	0.03	0.6
6	Failure	0.01	0.4
7	Failure	0.02	0.4
8	Failure	0.08	6.5

* Computed from the logistic regression whose independent variable is the cross-entropy of the Small-World and whose dependent variable is success/failure.

** Product 5 was predicted to be a failure, while its actual status is a success.

5.2. Home Furniture Sets

In order to further demonstrate our approach, we obtained further data on eight products in the home furniture category. A large home furniture chain was asked to select four successful new furniture sets and four failures. The marketing manager defined a failure as a set whose sales have not reached a predefined level and whose first stock was not sold out. A successful product was defined as a set all of whose entire first run units have been sold and more have been ordered from the manufacturers. The chain supplied data of monthly regional sales by individual store. The regional data were rich, including dozens of locales, in contrast to the 12 windows of the previous study. The comparison between failures and successes was obtained by computing the success probability for each case in the same way as was performed in the previous study, *mutatis mutandis*. We also corrected the window size for local market size, as described in the appendix. As in the previous studies, the cross-entropy measures produced values that were found to be largest in the first period and then declined with time. Table 3 presents the computed prediction of success probability. It can be seen that the cross-entropy of a successful product is in order of magnitude larger than that of a failure. Note that unlike in Table 2, a monotonic relationship between cross-entropy and probability of success is not found in Table 3. The reason is that the analyzed areas for the furniture sets were different for different sets, and thus the windows used for the analysis were not identical. Thus, for each product we used a different logistic regression as explained earlier. Had the windows been equal, a single logistic regression could be used and a monotonic relationship between cross-entropy and probability of success would have been expected.

Table 3 Home Furniture Sets Product Category: Cross-Entropy Calculations for Successful and Failed Products

Product	Outcome	Cross-entropy (first month)	Probability of success* (%)
1	Success	0.21	91
2	Success	0.27	96
3	Success	0.21	70
4	Success	0.26	98
5	Failure	0.12	0.1
6	Failure**	0.37	92
7	Failure	0.07	23
8	Failure	0.19	1

* Computed from the logistic regression whose independent variable is the cross-entropy and whose dependent variable is the outcome (success/failure).

** Product 6 was predicted to be a success, while its actual status is that of a failure.

6. Discussion and Limitations

This paper demonstrates the ability of the spatial divergence approach to serve as an early-period prediction tool, in both simulated and actual field studies. While the results look promising, as it successfully predicted 14 of 16 cases, caution should be exercised when applying this approach to a variety of market conditions. A key concern is the homogeneity assumption of the spatial distribution of consumers. While this assumption might be reasonable in some cases, there are cases where the baseline distribution might be nonuniform. Hence, when applying the cross-entropy approach to new product prediction, we recommend the following steps:

First, one should determine to what extent the homogeneity assumption deviates from reality in the case at hand. For example, in the supermarket case, the executives who supplied us with the data saw no reasons for clustering based on demographic or other covariates for these specific products and regions.

One way to deal with the question of nonhomogeneity might be found in retrospective analysis of past product introductions. If cross-entropy successfully separates the winners and the losers, we might have an indication that the approach is adequate on an *ad hoc* basis. This means that with the marketing and market conditions at hand, the noise introduced because of nonhomogeneity does not overpower word of mouth-based clustering (this method might be sensitive to dynamic changes in the population, and thus the tests should be performed first on products that were launched in the past few years). Furthermore, the cross-entropy value of the current introduction could be compared to the success and failure groups, i.e., if the current value is significantly closer to the success's past values of cross-entropy, the product is more likely to be a success, and vice versa.

If nonhomogeneity distorts the results, one can take one of two courses of action. One possibility is to conduct the analysis in relatively homogeneous areas for which geodemographic clustering is less of a problem. For example, a certain city and its neighboring rural areas may be analyzed separately. Since our main aim is to predict success and not the spatial pattern that follows it, focusing on a limited area may offer signals of success that are measurable, and match well with the destiny of the new product. The second course of action is not to compare the distribution of actual penetration to the uniform distribution, but rather to another baseline distribution, which represents marketing executives' information on possible clustering based on drivers that are external to the communication process. There are mainly three such variables: size, marketing plans, and innovativeness distribution.

Size and Marketing Plans

In the furniture study presented above, we concentrated on size, correcting the window sizes by calibrating each point of sales by its window size. In the appendix, we present an empirical test for which the underlying distribution is known to be nonuniform due to variations in the number of users in each region—the hybrid corn case. Similar weighting can be done if the firm is aware of different levels of marketing expenditures in various areas.

Innovativeness Distribution

The situation is more complicated if geodemographic clustering may be expected, because propensity to adopt differs in various areas. One possibility is to develop a measure of innovativeness for each area, based on surveys that examine the timing of adoption of innovations in the past in various areas. If differing levels of innovativeness are available for the various areas, managers can use them to build a baseline distribution that is not uniform regarding the propensity to adopt.

We wish to point out that the latter activity of building baseline distributions might be a challenge in many cases. The current literature on patterns of spatial distribution of potential adopters should be augmented with more findings before better generalizations can be made in this regard. We hope that this note will serve as a compelling enough motivation for further research on this key issue.

Acknowledgments

The authors would like to thank Noa Golan, Oded Lowengart, Sarit Moldovan, Steven Shugan, an area editor, and two anonymous reviewers for their helpful comments and suggestions. This study was supported by the Israeli Science Foundation, the George Leven Hi-Tech Management School, Tel Aviv University, and the Kmart and Davidson Centers, the Hebrew University of Jerusalem.

Appendix. The Spatial Approach with Nonuniform Base Distribution: The Case of Hybrid Corn

In order to examine the usefulness of the spatial divergence approach even when the distribution of consumers is not uniform, we present an empirical analysis of the well-known case of hybrid corn, because the actual distribution of the examined variable, namely the corn acreage planted, is known. Due to the lack of reliable data prior to 1933, cross-entropy was calculated for that year. As opposed to the uniform case, in the hybrid corn case, each U.S. state was considered a window with a unique size. In the first stage, we calculated the acreage for each state, as well as the proportional acreage percentage in each state compared to the total U.S. acreage. We then calculated a probability function representing the adoption of hybrid corn in the U.S. from this data. Thus, when calculating the sum of $f_1(x) \cdot \log(f_1(x)/f_2(x))$ for Illinois, $f_2(x)$ was equated to 0.0754, or the corn acreage in Illinois divided by the total corn acreage in the U.S. (data on the penetration of hybrid

corn and total corn acreage were obtained from USDA agricultural statistics).

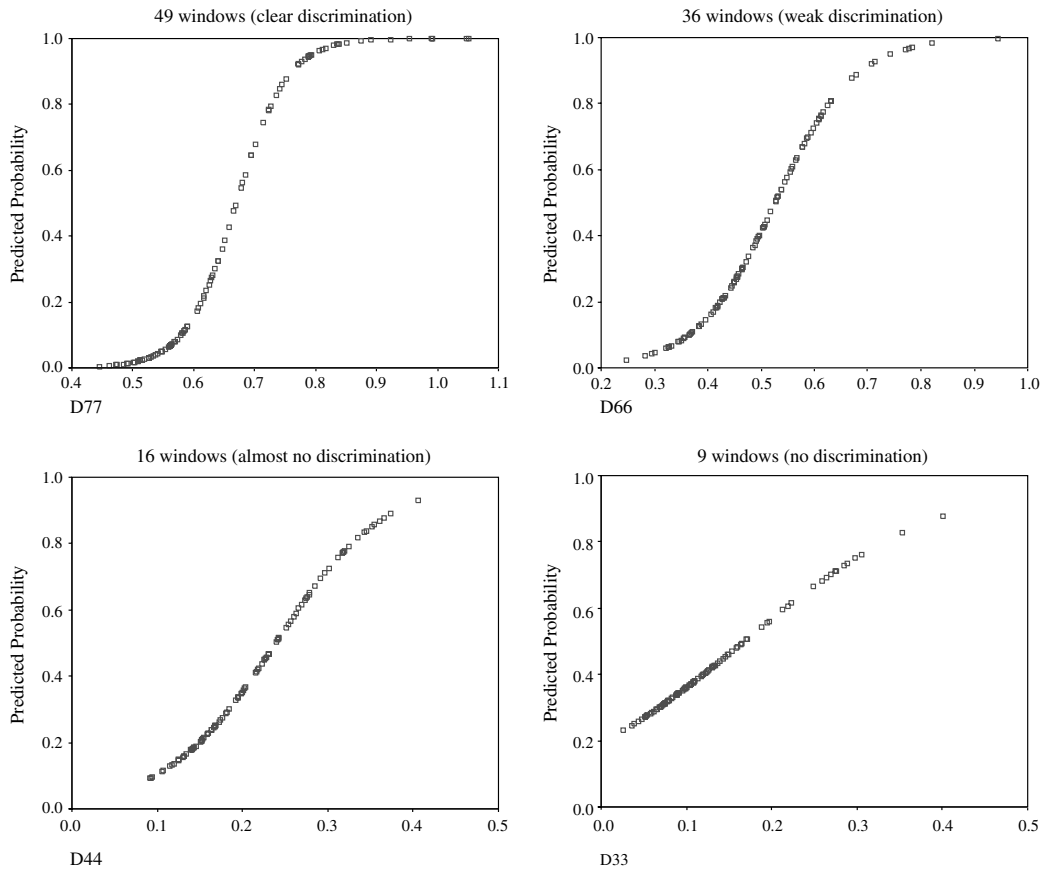
In order to use this measure to predict success, a comparison had to be performed to the cross-entropy values obtained in the paper. Consequently, the simulations should take into account the fact that the area and the windows distribution differ. Consequently, the simulated space world was divided into windows similar to the respective states' sizes. Thus, we performed the Small-World simulations again, and cross-entropy divergence measures for each process were calculated taking into account the new windows' sizes.

Because the Small-World data were now calibrated to represent the world of the hybrid corn, the cross-entropies of both the Small-World and the real case could be matched into the same graph. Hence we could "plug" the cross-entropy value into the logistic curve and use it as the independent variable to read the value of the dependent variable, or the probability of hybrid corn becoming a success. For the cross-entropy that we found (1.63), we computed the probability of success from the logistic regression whose independent variable is the cross-entropy and whose dependent variable is the outcome, and found that the probability of success was 99.9%.

Next we demonstrate the comparison of forecast accuracy as we move from coarse partitions to finer ones. We first performed this test on field data (the hybrid corn case). Instead of using the countries as windows, we used two other classifications (census regions and divisions of the U.S.) with four and nine windows respectively. The four

windows are the standard division of Northeast, Midwest, South, and West, and the nine windows are a finer division, i.e., the subdivision of the South into South Atlantic, East South Central, and West South Central. We simulated the U.S. as a Small-World representation, and calculated the predicted probability for success. In nine windows, the result again was 99% for success. This is probably related to the fact that hybrid corn is indeed a successful innovation. However, in four windows classifications, the predicted probability dropped to 45%, an indication of failure.

We also performed the test in synthetic data using cellular automata. As suggested by Parzen, the optimal windows number is the square root of the number of individuals (2,500); thus in our case it was 50. In order to stay within a symmetrical environment, we performed all analyses of the simulated data on squares, and thus the near-optimal number is 49 windows. We varied this number above and below 49. We noted three effects: First, a number of windows larger than Parzen's recommendations leads to a decrease in the correct predictions: 83% as opposed to 90%. Second, decreasing the number of windows leads to a noticeable decrease in correct predictions: from 90% for 49 windows, to 80% for 36 windows, 80% for 25 windows, 72% for 16 windows, 65% for nine windows, and 57% for three windows (this last figure is not significantly different than random draw). Third, the clear step-function of the logistic curve, whose purpose is to separate success from failure, slowly metamorphosed from an S-shaped function to a straight line. To illustrate this change, consider the graphs below:



References

- Allaway, Arthur W., Barry L. Mason, William C. Black. 1991. *The Dynamics of Spatial and Temporal Diffusion in a Retail Setting*, Avijit Ghosh, Charles A. Ingene, eds. *Spatial Analysis in Marketing*. JAI Press, Inc., New York.
- Amaral, Luis A., Marisa A. Scala, Marc Barthelemy, Eugene H. Stanley. 2000. Classes of Small-World networks. *Proc. National Acad. Sci. USA* **97**(21) 11149–11152.
- Balakrishnan, Anant, Narzul I. Shaikh, Arvind Rangaswami. 2002. Small-World markets and their implications for the diffusion of innovations. Marketing Science Conference, University of Alberta, Edmonton, Alberta, Canada.
- Baptists, Rui. 2000. Do innovations diffuse foster within geographical clusters? *Internat. J. Indust. Organ.* **18** 515–535.
- Bronnenberg, Bart, Vijay Mahajan. 2001. Unobserved retailer behavior in multimarket data: Joint spatial dependence in market shares and promotion variables. *Marketing Sci.* **20** 284–299.
- Case, Anne C. 1991. Spatial patterns in household demand. *Econometrica* **59**(4) 953–965.
- Csiszar, Imre. 1991. Why least square and maximum entropy? An axiomatic approach to inference for linear inverse problems. *Ann. Statist.* **19** 2032–2066.
- Dekimpe, Marnik G., Philip M. Parker, Miklos Sarvary. 2000. Multimarket and global diffusion. Vijay Mahajan, Eitan Muller, Yoram Wind, eds. *New Product Diffusion Models*. Kluwer Academic Publications, Boston, MA.
- Fader, Peter S., Bruce G. S. Hardie, Chun-Yao Huang. 2003. A dynamic changepoint model for new product sales forecasting. *Marketing Sci.* **23**(1) 50–65.
- Goldenberg, Jacob, Barak Libai, Eitan Muller. 2001. Talk of the network: A complex system look at the underlying process of word of mouth. *Marketing Lett.* **12** 209–221.
- Goldenberg, Jacob, Barak Libai, Eitan Muller. 2002. Riding the saddle: How cross-market communications can create a major slump in sales. *J. Marketing* **66**(2) 1–16.
- Golder, Peter N., Tellis, Gerard T. 1997. Will it every fly? Modeling the takeoff of really new consumer durables. *Marketing Sci.* **16** 256–270.
- Johnson, Don H., Sinan Sinanovic. 2001. Symmetrizing the Kullback-Leibler distance. Working paper, Rice University, Houston, TX.
- Kohli, Rajeev, Donald R. Lehmann, Jae Pae. 1999. Extent and impact of time in new product diffusion. *J. Product Innovation Management* **16** 134–144.
- Kullback, Solomon. 1997. *Information Theory and Statistics*. Books on Mathematics, Dover, New York.
- Lee, Lillian. 1999. Measures of distributional similarity. *Proc. 37th Association for Comput. Linguistics*. MIT Press, Cambridge, MA, 25–32.
- Mahajan, Vijay, Eitan Muller, Frank M. Bass. 1990. New product diffusion models in marketing: A review and directions for research. *J. Marketing* **54** 1–26.
- Mahajan, Vijay, Eitan Muller, Rajendra K. Srivastava. 1990. Determination of adopter categories using innovation diffusion models. *J. Marketing Res.* **27** 37–50.
- Mahajan, Vijay, Robert A. Peterson. 1979. Integrating time and space in technological substitution models. *Tech. Forecasting Social Change* **14** 231–241.
- Moe, Wendy W., Peter S. Fader. 2002. Using advanced purchase orders to forecast new product sales. *Marketing Sci.* **21** 347–364.
- Moore, William L., Edgar A. Pessemier. 1993. *Product Planning and Management*. McGraw-Hill, New York.
- Parzen, Emanuel. 1962. On estimation of a probability density function and model. *Ann. Math. Statist.* **33** 1065–1076.
- Putsis, William, Sridhar Balasubramanian, Edward H. Kaplan, Subrata K. Sen. 1997. Mixing behavior in cross-country diffusion. *Marketing Sci.* **16**(4) 354–369.
- Rogers, Everette M. 1995. *The Diffusion of Innovations*, 4th ed. Free Press, New York.
- Song, Sangyoung, David Bell. 2002. Space-time adoption: The case of internet grocery shopping. Marketing Science Conference, University of Alberta, Edmonton, Alberta, Canada.
- Srinivasan, V., Charlotte Mason. 1986. Nonlinear least square estimation procedure for innovation diffusion models of new product acceptance. *Marketing Sci.* **5** 169–178.
- Sultan, Fareena, John U. Farley, Donald R. Lehman. 1990. A meta-analysis of application of diffusion models. *J. Marketing Res.* **25** 331–341.
- Tellis, Gerard J., Stefan Stremersch, Eden Yin. 2003. The international takeoff of new products: The role of economics, culture, and country innovativeness. *Marketing Sci.* **22** 188–208.
- Watts, Duncan J., Steven. H. Strogatz. 1998. Collective dynamics of ‘Small-World’ networks. *Nature* **393** 440–442.