JEONGHYE CHOI, SAM K. HUI, and DAVID R. BELL*

For Internet retailers, demand propagation varies not only through time but also over space. The authors develop a Bayesian spatiotemporal model to study two imitation effects in the evolution of demand at an Internet retailer. Building on previous literature, the authors allow imitation behavior to be reflected both in geographic proximity and in demographic similarity. As these imitation effects can be time varying, the authors specify their dynamics using a "polynomial smoother" embedded within the Bayesian framework. They apply the model to new buyers at Netgrocer.com and calibrate it on 45 months of data that span all 1459 zip codes in Pennsylvania. The authors find that the proximity effect is especially strong in the early phases of demand evolution, whereas the similarity effect becomes more important with time. Over time, new buyers are increasingly likely to emerge from new zip codes beyond the "core set" of zip codes that produce the early new buyers, and spatial concentration declines. The authors explore the managerial implications stemming from these findings through a hypothetical "seeding" experiment. They also discuss other implications for Internet retailing practice.

Keywords: imitation, proximity, similarity, long tail, spatiotemporal model

Spatiotemporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer

The Internet has reduced customer access costs for firms and facilitated long-range connections among consumers. Although "location" is a primary determinant of success for traditional retailers (e.g., Huff 1964), Internet retailers are not subject to such constraints. They can attract consumers over a wide geographic area, which means that even physically separated consumers can easily use the same Internet retailing service. This raises important questions about how demand at Internet retailers is likely to evolve not only through time but also over space. In particular, how might consumers "imitate" their peers in their adoption behavior, and what can firms do to expedite the demand process?

A large body of research assumes that, in general, imitation behavior plays an important role in generating demand (see Bass 1969; Hauser, Tellis, and Griffin 2006). Studies that are directly relevant to our research offer two key findings. First, all else being equal, imitation among agents is more likely when they are geographically proximate. Researchers have found consumption externalities for prescribing physicians (Manchanda, Xie, and Youn 2008), competitive effects among retailers in new brand rollout (Bronnenberg and Mela 2004), and possible emulation in trial behavior for an Internet retailer (Bell and Song 2007). Second, the likelihood of imitation is greater among agents who are "similar." These include academics with overlapping research interests (Rosenblat and Mobius 2004), firms with comparable cultural profiles (Albuquerque, Bronnenberg, and Corbett 2007), and people with common overall sociodemographic characteristics (Yang and Allenby 2003).

We contribute to the literature by analyzing the space–time diffusion process as a function of both factors (i.e., proximity and similarity), identifying the relative importance of each over time, and relating our findings to an Internet retailer's new buyer acquisition strategy. Figure 1

*Jeonghye Choi is Assistant Professor of Marketing, Yonsei University (e-mail: jeonghye@yonsei.ac.kr). Sam K. Hui is Assistant Professor of Marketing, Stern School of Business, New York University (e-mail: khui@stern.nyu.edu). David R. Bell is Associate Professor of Marketing, The Wharton School, University of Pennsylvania (e-mail: davidb@wharton.upenn.edu). Jeonghye Choi thanks the Wharton Risk Management and Decision Processes Center for financial support. The authors are grateful to seminar participants at Hong Kong University of Science and Technology, Koç University, University of Arizona, University of Pennsylvania, and the 2007 Marketing Science Conference at Singapore Management University for their comments. Netgrocer.com Chief Executive Officer Lisa Kent generously provided the data. The authors are also grateful to the two anonymous *JMR* reviewers for their constructive advice and suggestions. Michel Wedel served as associate editor for this article. This article was accepted under the former editor, Joel Huber.

ISSN: 0022-2437 (print), 1547-7193 (electronic)

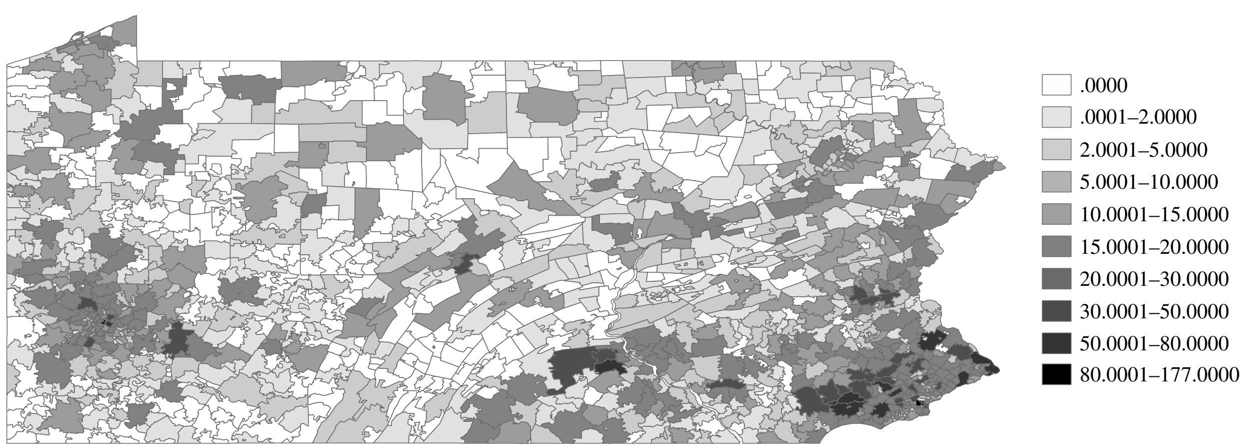
Figure 1

SPATIOTEMPORAL EVOLUTION OF NETGROCER.COM BUYERS IN PENNSYLVANIA

A: Cumulative Number of New Buyers in July 1998

B: Cumulative Number of New Buyers in October 1999

C: Cumulative Number of New Buyers in January 2001



motivates the underlying phenomenon. It shows the cumulative number of new buyers at Netgrocer.com in each zip code in Pennsylvania recorded in 15-month intervals from the inception of the service in May 1997 through January 2001. Three noteworthy patterns appear. First, the evolution of new buyers seems to have begun from two distinct locations and spread to nearby areas (these "hot spots" are Philadelphia and Pittsburgh, the two major cities in

Pennsylvania). Second, the pool of buyers within smaller disaggregate "neighborhoods" intensifies over time. Third, as time progresses, the adopting group expands throughout Pennsylvania such that later areas of sales are physically distant from earlier ones; as a result, the spatial concentration of the new buyers decreases over time.

To analyze the data in Figure 1, we formulate a dynamic Bayesian spatiotemporal Poisson model (e.g., Knorr-Held and Besag 1998) and specify the adoption rate for each region at each period as a function of imitation effects based on proximity and similarity, along with other locally defined covariates. We use a conventional distance-based proximity measure and a demographic similarity metric that mirrors approaches by Rosenblat and Mobious (2004) and Albuquerque, Bronnenberg, and Corbett (2007). To produce efficient estimates of the time-varying coefficients for these variables, we embed a "polynomial smoother" within our Bayesian model using a random-walk prior (Angers and Delampady 1992; Kalyanam and Shively 1998; Wahba 1978; Wedel and Zhang 2004).

Applying our model to the spatiotemporal evolution of new buyers at Netgrocer.com yields three new insights into how demand evolves for an Internet retailer that is geographically unconstrained. First, we find that geographic proximity has the stronger initial impact on the rate at which new buyers are acquired but that its relative importance weakens with time. Therefore, long-term viability is unlikely to be secured through local appeal in hot spots alone. Second, imitation based on demographic similarity, independent of geographic proximity to the preceding buyers, is relatively unimportant early on, but as time progresses, it accounts for a greater number of new buyers who emerge from spatially dispersed places—that is, places that lack sufficient density to be served through conventional means but that, on average, share characteristics with regions containing earlier adopters. This provides a rationale for the decline in spatial concentration of new buyers. The temporal ordering of the importance of the two components—geographic proximity first and demographic similarity second—holds after we control for differences in observed local characteristics (including access to the Internet) and unobserved heterogeneity in the adoption rate. Third, we follow Libai, Muller, and Peres (2005) and use "market seeding" to illustrate possible managerial implications stemming from these results. An initial focus on populous regions should be balanced against acquisition of more remote and dispersed customers.

Our research is subject to the following caveats: First, for reasons of parsimony, data availability, and managerial value, we focus on region-level behavior rather than individual behavior per se. Second, the main purpose of the model is to provide a descriptive analysis of proximity and similarity effects. We do not attempt to build a forecasting model, because this would require a substantially different approach. Finally, the seeding analyses using the imitation coefficients are the best-case scenario given the data and are intended to be illustrative of the potential benefit of the proximity-and-similarity-based strategy.

We organize the rest of the article as follows: The next section summarizes extant literature that employs geographic proximity and demographic similarity as proxies for imitation behavior in spatial demand analysis. The following section describes the data and key summary statistics, and the subsequent section specifies our Bayesian spatiotemporal model. We then report our substantive empirical findings. The concluding section outlines a hypothetical seeding experiment and discusses implications for Internet retailers and for future research.

BACKGROUND LITERATURE

We focus on selected empirical evidence from articles in marketing, economics, and sociology that develop proxy measures of geographic proximity and demographic similarity (e.g., among individuals, regions, and firms).

Geographic Proximity

Proximity-based imitation, or "the local neighborhood effect," is largely viewed as arising from either direct social interactions or local emulation among near neighbors. Standard empirical approaches incorporate measures that proxy for imitation or, more broadly, social interactions among physically close people. Goolsbee and Klenow (2002) use the proportion of local households owning computers to show that people in areas with a high proportion of computer ownership are more likely to become first-time buyers, even after controlling for personal traits and local environments. Forman, Ghose, and Wiesenfeld (2008) find that online book sales in a local market are not only associated with the overall disclosure level of user identity-descriptive information but also amplified when disclosure comes from reviewers residing in the same locality.

In addition to being measured through proportions, proximity effects can be investigated using information on pairwise distances or contiguity. Bronnenberg and Mela (2004) employ such measures and find emulation effects among local retailers; namely, new product rollout is influenced by product decisions made by local competitors. Bell and Song (2007) find that new trials of an Internet retailer are related to prior trials in proximate regions.

None of the aforementioned studies measure imitation or social interaction directly. Instead, the observed prior behavior of physically close "neighbors" is used to create measures that, in turn, influence the probability of subsequent action by another individual, firm, or region of interest. Statistically significant effects, in the presence of other controls, are taken as corroborating evidence. Our approach follows this precedent and uses spatially derived proxies to account for the geographic proximity effect.

Demographic Similarity

Fischer (1978) suggests that a resident of Los Angeles has a greater chance of coming into contact with someone from Chicago than with someone from Springfield, even though both Illinois locations are approximately the same physical distance from Los Angeles. This underscores the idea that the propensity for people to interact with and/or imitate one another may not be accounted for solely by physical distances. In line with this idea, many researchers have extended the "neighborhood" construct in ways that depart from a specification based on physical locations. For example, Van Alstyne and Brynjolfsson (2005) point out that "neighborhoods" can be shaped by many dimensions, including interests, preferences, and member characteristics. People agglomerating either in online communities or through their revealed preferences for certain online businesses may exhibit some homogeneity along demographic lines (e.g., by criteria such as occupation, education levels, income, or ethnic grouping).

For example, the way "similarity" is measured is an important empirical and conceptual issue. Agrawal, Kapur, and McHale (2008) define individual-level social proximity using a coethnicity indicator and find a substitution effect between social and spatial proximities. Social proximity provides greater benefit for inventors who are not colocated, whereas spatial proximity does for those who are. Rosenblat and Mobius (2004) define economists' "types" according to academic interests and find that the Internet led to narrower collaborations; for example, labor economists are now less likely to write with economic historians and more likely to coauthor with labor economists who are physically distant. Yang and Allenby (2003) study automobile choice and define people who share similar demographic profiles as "demographic neighbors." A model that accounts for choices by both types of neighbors (i.e., demographic and geographic) is preferred to models that account for either alone.

Other studies have investigated region-level similarity. Conley and Topa (2002) examine spatially clustered unemployment rates in Chicago. Social networks are defined separately for physical distance, race and ethnicity, and occupation using Euclidean distances of the corresponding regional compositions across census tracts. The effects of physical distance and occupation are significant, whereas the effect of race and ethnicity is not. Albuquerque, Bronnenberg, and Corbett (2007) study International Organization for Standardization (ISO) certification diffusion across countries and find that diffusion of ISO 9000 is driven by proximity and trade-based similarity, whereas diffusion of ISO 14000 is driven by proximity and cultural similarity. Building on these studies, and following Rosenblat and Mobius (2004) and Van Alstyne and Brynjolfsson (2005) in particular, we define our similarity measure according to region "types" based on sociodemographic characteristics.

Summary

Prior research has demonstrated that geographic proximity and demographic similarity drive imitation behavior. However, these studies suggest relatively little about how such effects evolve over time. Because different forms of imitation exert different degrees of influence at various stages of the adoption cycle, analyzing their effects in static rather than intertemporal settings may not provide a complete picture of their influence. In this article, we aim to focus on the temporal aspects of geographic proximity and demographic similarity and understand their dynamic influences in driving adoptions of an online retailer.

DATA

New Buyers

We obtained monthly transaction data for new buyers at Netgrocer.com in Pennsylvania from the inception of the service in May 1997 through the end of January 2001. During this period, orders were shipped from a warehouse in New Jersey by FedEx, and customers were charged a fixed shipping fee. The customer file records the order month and shipping zip code for each transaction. To understand how demand evolution varies over space and time, we consider the number of new buyers after aggregating spatially and temporally. Figure 1, Panel C, shows considerable spatial dispersion in the distribution of cumulative new buyers. Figure 2, Panel B, highlights the time dimension of the raw data. It shows that while the overall number of new buyers across zip codes is generally increasing through the 45-month period, there is substantial variability in the overall trend.

Next, we consider the space–time path of the raw data in Figure 1 in greater detail. Table 1 shows summary statistics for the number of buyers per zip code in five-month intervals. The mean number of new buyers per zip code increases over time, but so does the variability across zip codes. That is, the spatial concentration of new customers appears to decrease over time. To examine this more formally, we compute the Getis–Ord G^* statistic (Getis and Ord 1992) each month. The decay in localized concentration of demand supports the observation that, over time, the distribution of new buyers is expanding over space. The considerable spatial and temporal variation in the raw data underscores that when building our model, we must carefully control for regional and temporal baseline effects to accurately measure the demand effects due to imitation.

Regional Characteristics

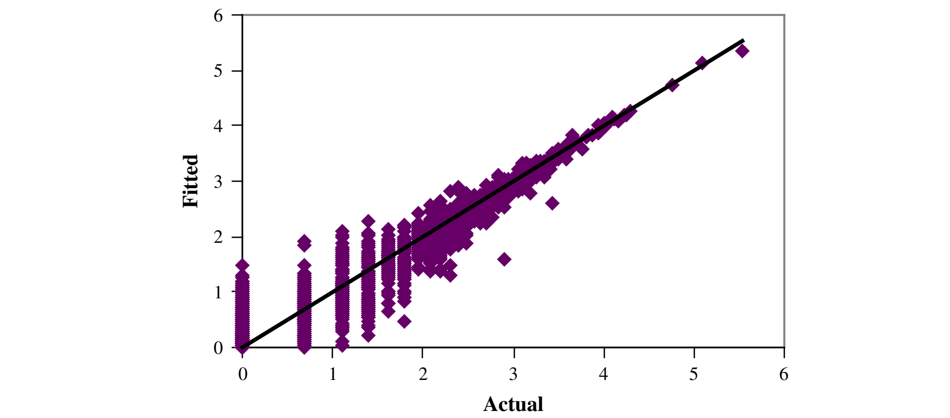
We assemble the data for the imitation proxy variables and the direct measures of regional heterogeneity from three sources: (1) the 2000 U.S. Census, (2) ESRI retailing statistics (esri.com), and (3) the Federal Communications Commission (FCC) broadband access survey. To create empirical measures of local presence for supermarkets and general merchandisers (e.g., Wal-Mart), we count the number located within the focal zip code and the first- and second-order contiguous neighbors. We then compute store density by store type on the basis of land area. Because warehouse clubs are less common, we use a binary indicator for presence within the focal, first-, or second-order contiguous zip codes. Table 2 provides descriptions and summary statistics for all zip code–level variables. For ease of exposition, the variables are classified as pertaining to region-level: (1) local environment, (2) household characteristics, (3) access to retail services, and (4) access to the Internet.

The FCC estimates the number of Internet service providers in each region; however, these data are known to be approximate. Some Internet service providers fail to report their services, and others report a presence in zip codes on the basis of a single customer. Moreover, the data were collected at four discrete times only (December 1999, June 2000, December 2000, and June 2001), three of which are covered by our transaction data. Following Wand's (2003) suggestion, we employ a low-rank thin plate spline smoother to improve the FCC data and provide the complete details in Web Appendix A

Figure 2
AGGREGATE MODEL FITS OVER SPACE AND TIME

A: Fitted Versus Actual Number of New Buyers in Log-Transformation by Zip Code (aggregated over time [months])

B: Fitted Versus Actual Number of New Buyers over Time (aggregated over space [zip codes])



(http://www.marketingpower.com/jmrfeb10).[1] In addition, because the time frame of the broadband access data does not coincide perfectly with the Netgrocer.com data, we impute part of the missing data using a linear interpolation (see also Bell and Song 2007).

We assess and verify the appropriateness of our approach with reference to additional external sources, including prior literature and alternative data collected in the Current Population Survey (CPS).[2] Application of linear interpolation and spatial smoothing creates a zip code–specific and

Table 1
SUMMARY STATISTICS FOR THE NUMBER OF NEW BUYERS

	M	*SD*	*Minimum*	*Maximum*
May 1997	.001	.037	.000	1.000
October 1997	.008	.090	.000	1.000
March 1998	.055	.282	.000	3.000
August 1998	.198	.631	.000	8.000
January 1999	.065	.322	.000	5.000
June 1999	.083	.350	.000	6.000
November 1999	.212	.602	.000	7.000
April 2000	.220	.607	.000	5.000
September 2000	.219	.823	.000	22.000
January 2001	.235	.802	.000	19.000

[1]We also estimated our model using nonsmoothed broadband data and obtained qualitatively similar results. Full details are available on request.

[2]Household-level Internet usage data were collected as supplementary data in the CPS from 8162 national zip codes in October 1997, December 1998, August 2000, and September 2001. Although the CPS data match nicely with the period for the Netgrocer.com data, they include only 670 (46%) of the zip codes in Pennsylvania. Therefore, we use the spatially smoothed Broadband Access variable derived from the FCC data because this measure can be constructed for all 1459 zip codes in Pennsylvania. The average zip code–level correlations between the CPS data and smoothed Broadband Access are .95 for the total U.S. sample of 8162 zip codes and .97 for the 670 Pennsylvania zip codes. This suggests that the interpolated Broadband Access variable reflects the temporal growth pattern of household-level Internet usage present in the CPS data. Moreover, Bell and Song (2007) demonstrate that a measure constructed from the FCC data is empirically superior to one developed from the CPS data alone.

Table 2
VARIABLE DESCRIPTIONS AND SUMMARY STATISTICS FOR ZIP CODE CHARACTERISTICS

Variable	*Description*	*M*	*SD*
Local Environment			
Population	Total population	8391.600	11,149.400
Population density	Population density	1298.799	3117.592
Population growth	Annual population growth rate from 2000 to 2004	.004	.011
Home value	% of homes valued at $250,000 or more	.060	.108
Urban housing	% of houses with 50 units or more	.018	.056
Land area	Area in square miles	30.607	35.511
Household Characteristics			
Asian	% of Asians	.008	.016
Black	% of blacks	.038	.112
White	% of whites	.938	.130
College	% with bachelors and/or graduate degree	.370	.144
Elderly	% age 65 and above	.156	.041
Wealthy	% of households earning $75,000+	.165	.118
Access to Retail Services			
Density general	Density of general stores within the second-order neighboring zip codes	.107	.251
Density supermarket	Density of supermarkets within the second-order neighboring zip codes	.224	.393
Presence warehouse	Presence of warehouse clubs within the second-order neighboring zip codes	.245	.430
Access to the Internet			
Broadband access	Number of high-speed Internet service providers		
December 1999		1.784	1.320
June 2000		2.060	1.749
December 2000		2.940	2.665
June 2001		2.840	2.773

period-specific measure, which we call Broadband Access. This control for access to the Internet is important to help rule out the alternative hypothesis that space–time evolution of Netgrocer.com new buyers simply mimics the diffusion of Internet access.

Finally, during the period of data collection, Netgrocer.com was not involved in any significant marketing activities in Pennsylvania; thus, this data set offers a unique opportunity to assess imitation effects across space and time, free of explicit marketing interventions. Although we cannot therefore comment on the relationship between local marketing efforts and demand, we can assess the equally important relationship between local characteristics and demand—a relationship of increasing interest (see Forman, Ghose, and Goldfarb 2009; Pauwels and Nelsin 2008; Waldfogel 2007).

MEASURES AND MODEL

Measures of Proximity and Similarity

Competition between Netgrocer.com and offline alternatives is local, so region-level (zip code) sales are of particular managerial relevance and, in general, data that describe regions are widely available and reliable. Thus, our proximity and similarity measures are defined with respect to regions (see also Avery et al. 2008; Brynjolffson, Hu, and Rahman 2008). Moreover, individual-level neighbor covariate information is neither available to work with nor practical. In our model specification, exogenous definition of "neighbors" at the region (zip code) level and influence from the lagged cumulative behavior of neighbors are used to help mitigate the well-known "reflection problem" (Manski 1993, 2000). Manski (1993) emphasizes that to claim imitation effects, two alternatives—contextual (exogenous) effects and correlated effects—should be ruled out. With respect to contextual effects, it is unlikely that some unique exogenous feature of neighboring regions is systematically influencing trial of new buyers in the focal region. Correlated effects—by which the number of new buyers in the focal region is influenced by a similarity in institutional constraints—are also unlikely given our controls for Internet access, retail store availability, and so forth.[3]

We apply standard approaches from the literature that define neighborhood relationships through the use of weighting matrices (Anselin 1988; Bell and Song

[3]Manski (1993, pp. 532–37) provides relevant conditions for identification and estimation of endogenous effects. Possible correlated effects are unlikely for the following reasons: First, Netgrocer.com did not conduct significant marketing activities during the data period. Second, our model controls for access to the Internet and to local retailers. In addition, regional and temporal baselines account for region- and time-specific shocks. Although spatially (and/or demographically) correlated tastes might drive results of imitation behavior, our rich data and specification make this more unlikely than in much of the existing literature. Thus, we have made progress toward addressing the reflection problem, but we cannot entirely rule it out. We thank an anonymous reviewer for these observations.

2007; Bronnenberg and Mela 2004; Yang and Allenby 2003). Specifically, we employ two such matrices: The matrix G captures across-region geographic proximity, and the matrix D captures across-region demographic similarity. For ease of exposition, we assume that there is a finite number of zip codes, n, such that all pairwise relationships can be summarized by an $n \times n$ weighting matrix, G (D), in which each nonnegative element, G_{ij} (D_{ij}), denotes the degree of geographic (demographic) "closeness" of region j to region i. Each weighting matrix is symmetric and row-normalized (row-normalization takes into account relative closeness among neighbors). We also assume that the neighbor relationships do not change over time, as is standard in the previous literature.

Geographic proximity (G). We assume that our measure of across-region proximity is an inverse function of the physical distance in miles, d_{ij}:

$$G_{ij} = \begin{cases} \exp(-\Delta d_{ij}), & i \neq j \\ 0, & i = j. \end{cases} \tag{1}$$

Following Yang and Allenby (2003), we further assume that Δ is equal to 1.[4] The distance-based measure helps control for the large variation of different zip codes in land area and number of contiguous neighbors. We consider alternative proximity matrices based on shared boundaries and contiguity information in Web Appendix B (http://www.marketingpower.com/jmrfeb10).

Demographic similarity (D). Unlike with the measures of physical proximity, there is no single, widely used, straightforward approach for defining similarity. We presume that shared sociodemographic characteristics across regions serve as a proxy for similarity (see Conley and Topa 2002). In other words, if the characteristics of two regions are alike, these regions are more likely to imitate each other, all else being equal. Therefore, we focus on observable characteristics that previous studies have shown to be correlated with levels of imitation—namely, education, income, age, and ethnicity and their corresponding subcategories (e.g., Howard, Raine, and Jones 2001; Katz, Rice, and Aspden 2001; Van Alstyne and Brynjolfsson 2005).

The U.S. Census reports zip code–level information on the percentages of residents in the following educational attainment categories: (1) below high school completion; (2) completed high school, but no university degree; (3) university degree holder; and (4) graduate degree holder. The following categories also are reported for income: (1) below the poverty line, (2) medium income, and (3) income in excess of $75,000 per year. Age categories are as follows: (1) up to age 20, (2) 21–40, (3) 41–65, and (4) age 65 or older. Ethnicity is reported for each region according to the percentage of Asians, blacks, Hispanics, and whites living there. Following Rosenblat and Mobius (2004) and Van Alstyne and Brynjolfsson (2005), we define "profile vectors" that measure the extent of overlap between two regions. The sociodemographic profile vectors have a total of 15 elements (4 each for education, age, and ethnicity and 3 for income). We define pairwise similarity measures as follows:

$$D_{ij} = \begin{cases} \sum_k \min(v_{ik}, v_{jk}), & i \neq j \\ 0, & i = j, \end{cases} \tag{2}$$

where v_{ik} is the kth element of the sociodemographic vector of region i; that is, we sum the minimum values, based on the elementwise comparisons across two sociodemographic vectors, for all $k = 1, 2, \ldots, 15$ elements of their sociodemographic profile. As in the case of physical proximity, we define two alternative measures of demographic similarity in Web Appendix B (http://www.marketingpower.com/jmrfeb10).

A Bayesian Spatiotemporal Model of New Buyers

Given the sparseness of the adoption data (see Table 1 and Figure 1), our model must take into account significant sampling error to accurately estimate the role of imitation behavior. Toward this end, we specify our model in two levels, as is standard in Bayesian generalized linear models (Gelman et al. 2003). In the first level, we assume that the number of new buyers in zip code i at time t follows a Poisson distribution with (latent) rate parameter λ_{it}; we then model λ_{it} as a function of imitation behavior and other controls. Formally, we specify this as follows:

$$y_{it} \sim \text{Poisson}(\lambda_{it}), \tag{3}$$

where y_{it} denotes the number of new buyers in zip code i during month t.

We justify the Poisson assumption in Equation 3 on both theoretical and empirical grounds. In Web Appendix C (http://www.marketingpower.com/jmrfeb10), we outline a mathematical argument (adapted from Knorr-Held and Besag 1998; Ross 1996) that the Poisson approximation is valid under the assumption that adoption is sparse and within-period imitation is limited. In the next section, using a posterior predictive check (Gelman et al. 2003), we show empirically that the Poisson distribution provides an excellent fit to the raw data. Finally, the Poisson distribution has been used in other instances in which events are rare (e.g., to model the spread of new species [see Wikle and Hooten 2006] or the number of new patients infected by a rare disease [Knorr-Held and Besag 1998]).

Next, in the second level, we model latent adoption rates λ_{it} as a function of region-level characteristics, temporal baseline effects, and geographic and demographic imitation effects:

$$\log(\lambda_{it}) = \log(n_{it}) + \gamma_i + \zeta_t + \beta_t^W z_{it} + \beta_t^G G_{(i)}\vec{z}_t + \beta_t^D D_{(i)}\vec{z}_t + \varepsilon_{it}, \tag{4}$$

$$\gamma_i = \vec{x}_i'\vec{\tau} + \tilde{\gamma}_i, \; \tilde{\gamma}_i \sim N(0, \sigma_\gamma^2), \text{ and} \tag{5}$$

$$\beta_t^W, \beta_t^G, \beta_t^D \geq 0 \; \forall t, \tag{6}$$

where n_{it} denotes the number of people in region i yet to try the service at time t and serves as an offset variable (Agresti

[4]We make this assumption for reasons of computational tractability and consistency with the previous literature (e.g., Claude 2002; LeSage and Pace 2005; Yang and Allenby 2003). To demonstrate that our empirical findings are robust to this assumption, we defined two additional proximity measures with an inverse function of half ($\Delta = .5$) and twice ($\Delta = 2$) the geographic distance and then reestimated the models. Both measures provide consistent model estimates and, thus, the same qualitative insights. We thank an anonymous reviewer for suggesting this check.

2002; Rabe-Hesketh and Skrondal 2005) and γ_i and ζ_t are regional and temporal baselines, respectively. The regional baseline, γ_i, comprises two terms: observed heterogeneity explained by $\vec{x}_i'\vec{\tau}$, a vector of standardized region-level characteristics and the corresponding coefficients vector, and remaining unobserved heterogeneity captured by $\tilde{\gamma}_i$.[5] The terms $G_{(i)}$ and $D_{(i)}$ denote the ith rows of matrices G and D, respectively, and z_{it} denotes the (log-) cumulative number of buyers in region i before time t. The coefficients β_t^W, β_t^G, and β_t^D denote the strength of within-region imitation (W), across-region imitation due to geographic proximity (G), and across-region imitation due to demographic similarity (D), respectively. The error terms, ε_{it}, are assumed to be independent and normally distributed with a mean of 0 and a variance of σ_ε^2, allowing for overdispersion.

We are interested in the final three terms for imitation in Equation 4. The term $\beta_t^W z_{it}$ represents the within–zip code imitation effect due to prior buyers in the same zip code. The row vector $G_{(i)}$ ($D_{(i)}$) measures geographic (demographic) "closeness" of region i to all other regions (see Equations 1 and 2). Postmultiplication by the vector of neighbors' cumulative and lagged numbers of new buyers (i.e., $G_{(i)}\vec{z}_t$ and $D_{(i)}\vec{z}_t$) produces a scalar variable that captures the aggregate time-varying influence of geographic and demographic neighbors on region i at time t. The parameters β_t^G and β_t^D capture imitation effects based on geographic proximity and demographic similarity, respectively.

Given the nature of our data, we are unable to disentangle—except in an *ex post* analysis of marginal effects—whether current users within a region are propagating positive or negative information about Netgrocer.com. Because only nonperishable branded products (e.g., paper products, canned food) were sold during the data collection period, potential new buyers should have been able to assess product quality *ex ante*. Prices were also known. Thus, negative information was most likely related to delivery, which was handled by FedEx. Therefore, we postulate that the more cumulative buyers there are, the greater is the number of new buyers that will emerge. Equation 6 reflects this restriction, which assumes that all three imitation coefficients are nonnegative. These restrictions are of a theoretical nature only; they play no role in the actual empirical application. The estimated imitation coefficients are bounded far away from 0, making this restriction irrelevant (in the "Conclusion" section, we sketch an extension of our model that could accommodate both positive and negative influence).

Prior Specification and Smoothing

The main substantive goal of this research is to understand the relative magnitudes of proximity- and similarity-based imitation effects and how they vary over time. From a model estimation standpoint, our goal is to obtain efficient estimates for β_t^W, β_t^G, and β_t^D. To this end, we embed a "polynomial smoother," commonly used in frequentist nonparametric statistics, into our Bayesian model (Angers and Delampady 1992; Kalyanam and Shively 1998; Wahba 1978; Wedel and Zhang 2004). A smoother enables us to take observations from neighboring periods into account when making inference about a certain period. When making inference about an estimate at time t, we take into account information from periods $t-1, t-2, \ldots$ (and also $t+1, t+2, \ldots$) in polynomially decreasing weights, which enables us to borrow strength from other periods to improve estimation efficiency. The smoother produces estimates that vary smoothly over time, which is consistent with our intuition about how imitation coefficients should evolve. It also provides several key statistical advantages (see Web Appendix D at http://www.marketingpower.com/jmrfeb10).

We specify a Gaussian random-walk prior on our time-varying coefficients. For $t > 1$:[6]

$$(7)\qquad \zeta_t \sim N(\zeta_{t-1}, \sigma_\zeta^2),$$

$$(8)\qquad \beta_t^W \sim N(\beta_{t-1}^W, \sigma_W^2),$$

$$(9)\qquad \beta_t^G \sim N(\beta_{t-1}^G, \sigma_G^2),\ \text{and}$$

$$(10)\qquad \beta_t^D \sim N(\beta_{t-1}^D, \sigma_D^2).$$

We specify standard proper conjugate priors for all the other parameters in the model. We use a Markov chain Monte Carlo procedure to sample from the posterior distributions (see Web Appendix E at http://www.marketingpower.com/jmrfeb10).

EMPIRICAL FINDINGS

We first compare our model with reduced models and demonstrate the adequacy of our model in describing both the spatial and the temporal dimensions of the data. We then present time-varying imitation parameter estimates (for other control variables, see Web Appendix F at http://www.marketingpower.com/jmrfeb10), interpret them, and discuss implications for market seeding and why the spatial concentration of new buyers declines over time.

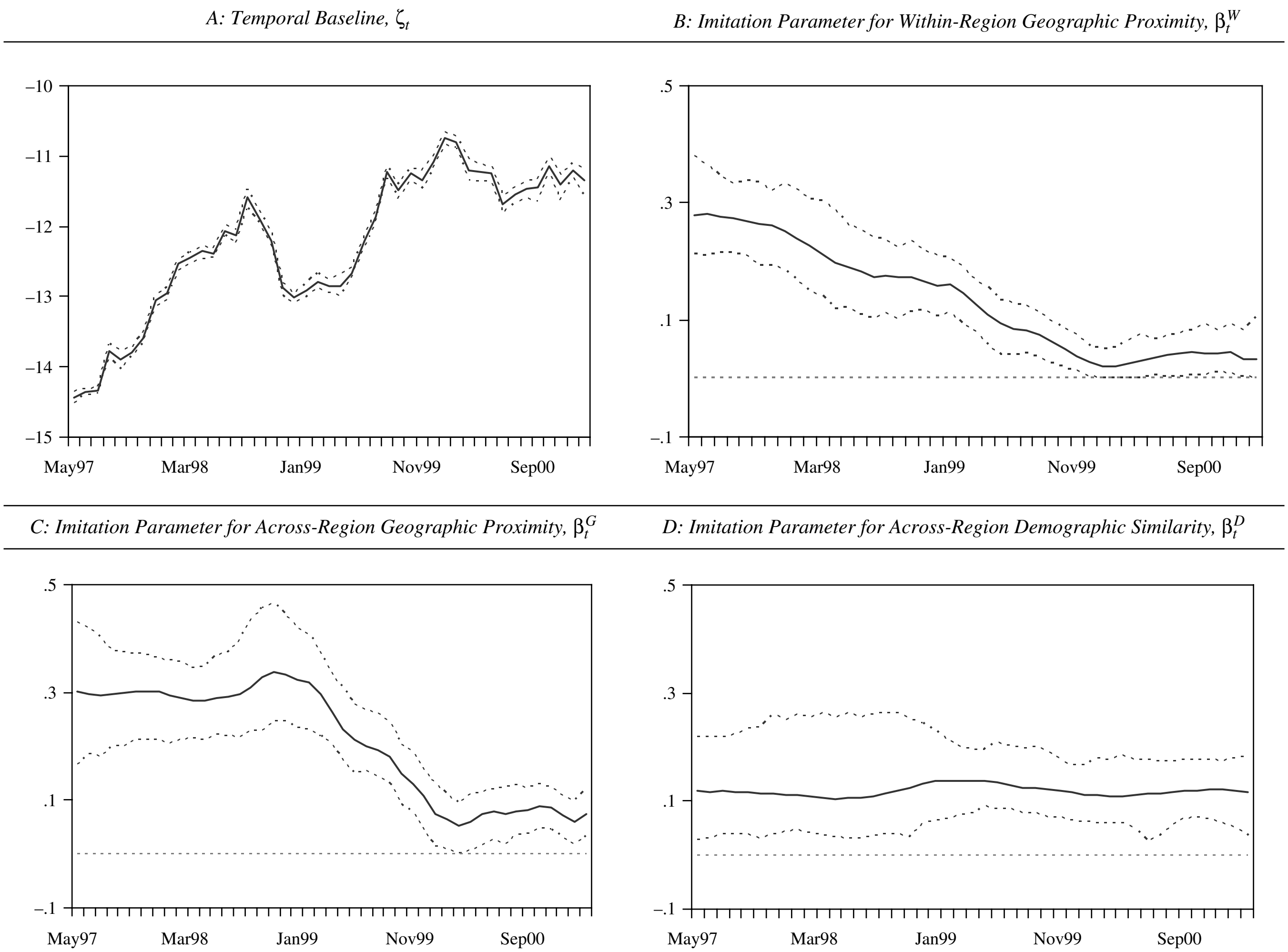
Model Fits and Validation

Using marginal log-likelihood (Chib 1995; Chib and Jeliazkov 2001), we compare the full model with reduced models that "turn off" imitation effects based on proximity and similarity. The marginal log-likelihood for the full

[5]We could specify that these random effects are spatially correlated, for example, using a conditional autoregressive formulation (Cressie 1993). However, Albuquerque, Bronnenberg, and Corbett (2007) find that incorporating spatially correlated errors does not improve their model's performance. Thus, we retain the i.i.d. specification. We could also specify a more general model with demographically correlated random effects—for example, as a joint distribution across zip codes with correlation in demographic space. We thank the associate editor for this observation.

[6]Giving these temporal parameters, independent diffuse normal distributions (i.e., $N[0, 100^2]$) is undesirable for two reasons. First, because these parameters measure the strength of imitation over time, they would be expected to vary smoothly over time, instead of jumping around in a rather haphazard manner. Second, the independence assumption of the prior distributions fails to "borrow strength" across the different periods when estimating these parameters, thus reducing estimation efficiency (Rossi and Allenby 2003). This latter aspect is particularly important for our data, which are fairly sparse, with small numbers of buyers over space and time (see Table 1).

Figure 3
POSTERIOR MEANS AND 95% POSTERIOR INTERVALS FOR THE TEMPORAL BASELINE (ζ_t) AND THE TIME-VARYING IMITATION PARAMETERS (β_t^W, β_t^G, AND β_t^D)



model with the proximity and similarity effects is −70,324, which is higher than those for the model with neither effect (i.e., $\beta_t^G = \beta_t^D = 0$), the model with proximity only (i.e., $\beta_t^G = 0$), and the model with similarity only (i.e., $\beta_t^D = 0$).[7] To assess overall fit to the raw data (y_{it}), we also compare the actual distribution of y_{it} with the posterior predictive distribution of $\hat{y}_{it}$ (Gelman et al. 2003). Figure 2, Panels A and B, indicates an adequate model fit on the spatial and temporal dimensions after aggregating over time and space, respectively. Importantly, we obtain accurate spatial fit not only in the regions with high demand but also in the spatially distant regions with relatively sparse sales.

Parameter Estimates and Interpretation

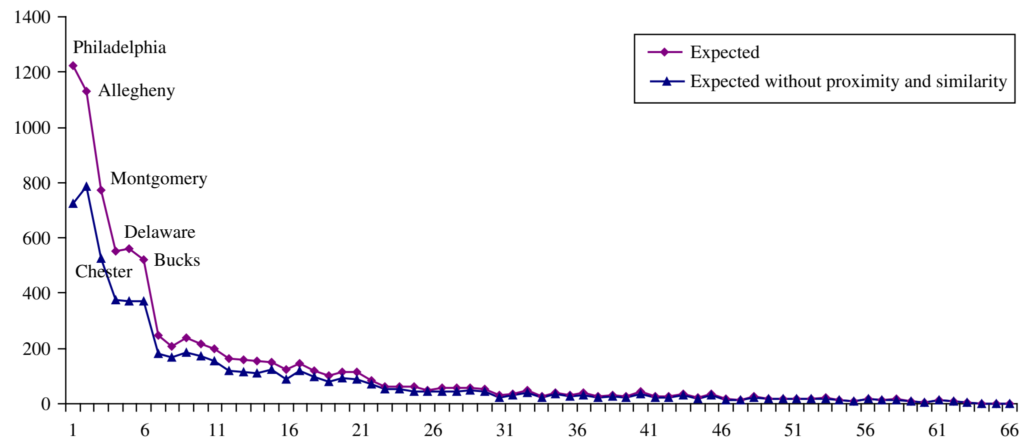
Time-varying coefficients of imitation (β_t^W, β_t^G, and β_t^D). Figure 3 shows the posterior means and 95% posterior intervals for these parameters together with the temporal baseline ζ_t. There is significant nonstationarity in the imitation parameters: β_t^W and β_t^G tend to decay over time, while β_t^D stays somewhat constant. The decay in β^W and β^G is consistent with the decreasing imitation parameter estimate in the Bass model as a data window is extended (Van den Bulte and Joshi 2007; Van den Bulte and Lilien 1997). The decay in the two proximity coefficients offsets the increase in log-cumulative new buyers in the focal region (z_{it}) and contiguous regions ($\vec{z}_t$). The relative constancy of the similarity coefficient indicates that new buyers continue to emerge from disparate and physically distant regions. One interpretation is that new-buyer acquisition through proximity "taps out," while new-buyer acquisition through similarity holds at a "steady" rate of accumulation. An Internet retailer's survival may depend on the ability to acquire similar types of customers from a wide-ranging area.

Further insights come from examining how the marginal effects of imitation vary across space and time. We can assess the marginal effect of imitation at region i at time t by examining the model-based expected number of new

[7]We also compared the full model with three reduced models using Newton and Raftery's (1994) procedure and obtained the same qualitative results. We thank an anonymous reviewer for suggesting Chib (1995).

Figure 4

EXPECTED NUMBER OF BUYERS IN EACH COUNTY (AGGREGATED OVER TIME) UNDER OUR FULL MODEL VERSUS SETTING IMITATION COEFFICIENTS TO 0



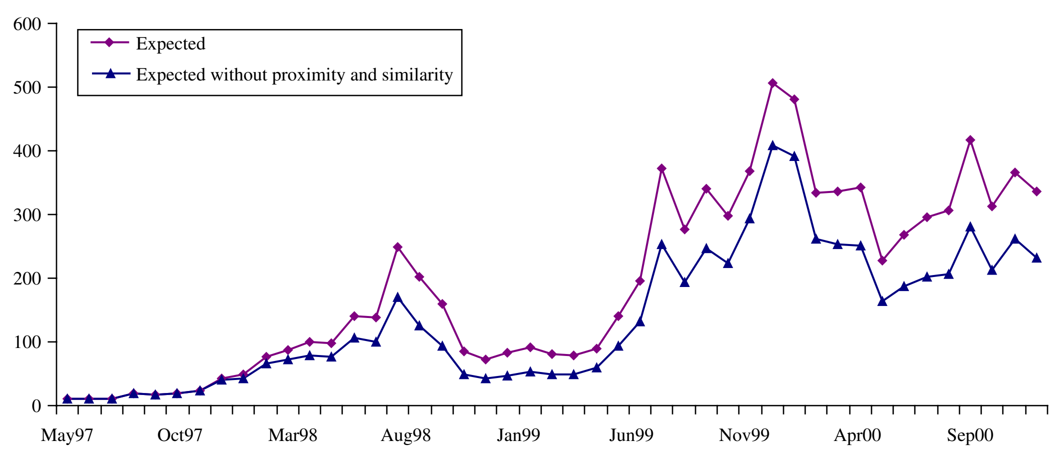
Notes: The gap is the marginal effect of imitation under our model framework.

buyers $E(y_{it})$ compared with the expected number of new buyers (under the full model) when the imitation coefficients (β_t^W, β_t^G, and β_t^D) are set equal to 0. To assess the marginal effect of imitation across space, we aggregate the 1459 zip codes to their corresponding county, which results in 67 different counties. Figure 4 shows the expected number of buyers in each county under the full model versus the expected number when the imitation coefficients are set to 0. The gap between the two expected values indicates the marginal effect of imitation in that county. The location of each county on the x-axis is given by its rank in terms of number of new buyers. To avoid clutter, we identify by name only the top six counties (Philadelphia is the number-one county, and Allegheny, which includes Pittsburgh, is the number-two county). The marginal effect of imitation is not uniform but rather varies significantly even among the well-performing counties. For example, while Philadelphia shows more than a 40% contribution of imitation behavior to the total number of buyers, Allegheny shows only 30%. This could be because Allegheny is more spatially isolated from other well-performing areas (i.e., Philadelphia, Montgomery, Chester, Delaware, and Bucks) and therefore is less likely to be subject to imitation effects based on proximity.

Figure 5 shows the marginal effects of imitation over time by again comparing the expected number of buyers over time under the full model with the expected number of buyers when imitation coefficients set to 0. The relative

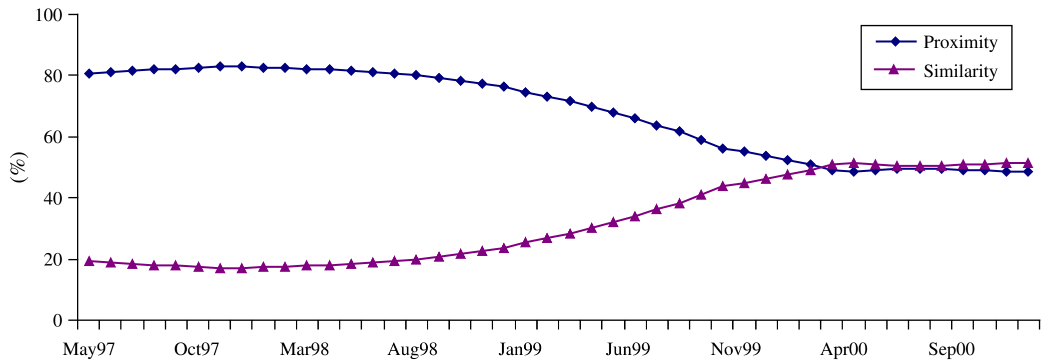
Figure 5

EXPECTED NUMBER OF NEW BUYERS IN EACH MONTH (AGGREGATED OVER ZIP CODES) UNDER OUR FULL MODEL VERSUS SETTING IMITATION COEFFICIENTS TO 0



Notes: The gap is the marginal effect of imitation under our model framework.

Figure 6
RELATIVE MAGNITUDES OF THE PROXIMITY AND SIMILARITY EFFECTS OVER TIME



contribution of the imitation effects increases over time. This finding is intuitive: The larger the cumulative number of existing customers, the greater is the potential for imitation of all types. Again, this underscores the importance of the installed base of new buyers for the ongoing acquisition of additional new buyers.

Proximity and similarity. Imitation effects for a focal zip code have three components: (1) the within–zip code effect of prior new buyers on the current period rate, (2) the across–zip code geographic proximity effect of prior new buyers in contiguous neighbors, and (3) the across–zip code demographic similarity effect. Because the first two components are based on short-range physical proximity and their relative magnitudes are relatively stable over space and time (the ratio of within- and across-proximity effects is approximately .5), we now combine them as one overall effect called "proximity" and compare it with the similarity effect.

Figure 6 plots the relative magnitudes of the "proximity" and "similarity" effects over time. The proximity effect is relatively more important initially; however, after approximately 30 months, the similarity effect becomes just as important. This model-based insight complements the observed decreasing spatial concentration of new buyers implied by Figure 1. Initially, new buyers begin to emerge in hot-spot areas (e.g., Philadelphia, Pittsburgh) and areas that are geographically proximate areas to hot spots. Subsequently, new buyers increasingly emerge from new zip codes beyond the "core set" of zip codes that produce the early new buyers. The similarity effect plays a more significant role in explaining new buyers in laggard areas that are "similar" to previously successful areas. Despite the larger similarity effect in subsequent periods being aggregated over space, its ultimate multiplicative effect in laggard areas does not generate as many new buyers in total as the proximity effect does early on in high-popularity areas. The effect is nevertheless important. This is because it helps drive orders from spatially dispersed customers who are small in number individually but who collectively account for a significant percentage of total sales.

Market Seeding

Our findings suggest that the firm can influence the space–time demand trajectory through judicious market seeding (see also Godes and Mayzlin 2009). To explore this possibility, we perform hypothetical simulations based on our model parameters and compare and contrast alternative seeding approaches. To perform this analysis, we assume that (1) the firm knows all the imitation coefficients beforehand (perhaps from using an "analogous product" in an approach common for Bass imitation coefficients; see Lilien and Rangaswamy 2004), (2) the imitation coefficients are invariant to the firm's seeding actions, and (3) costs are equivalent across scenarios. Because validating these assumptions requires data that are beyond our sample, we must stress that the analyses presented here are purely conceptual and are intended to be treated only as a springboard for further research.[8]

With this caveat in mind, we explore the following "seeding" scenario: Suppose that the firm considers seeding new buyers in month t. It then faces the decision of where these new buyers should be "planted" or allocated. Candidate zip codes are selected in accordance with the seeding policy, and one buyer is added to each zip code in that month. We compare how many new buyers the alternative time t seeding strategies bring to Netgrocer.com from month $t+1$ onward. Following terminology in Libai, Muller, and Peres (2005) and in accordance with their study, we compare and contrast the following four strategies (the first three draw on their work directly):

1. *Support-the-weak strategy*: The firm seeds new buyers in regions with the greatest remaining "market potential"—that is, current performance is relatively "weak" compared with what might be expected. A common heuristic is that the market potential is roughly proportional to population size,

[8]The seeding experiment using the imitation parameter estimates is parallel to an oracle test in statistics and data mining, which attempts to derive the best result given perfect knowledge of the parameters. If imitation estimates need to be predicted, the proximity-and-similarity-based strategy would not perform as well as it does here. Therefore, the proximity-and-similarity-based strategy in this article should be interpreted as the best-case scenario.

so we pick candidate regions according to the population yet to adopt at time t.

2. *Support-the-strong strategy*: The firm seeds in the historically (up to time t) best regions (i.e., those that have demonstrated "strong" performance to date).
3. *Uniform strategy*: The firm seeds new buyers randomly across regions regardless of market potential (based on population) or historical performance.
4. *Proximity-and-similarity-based strategy*: The firm seeds by choosing new zip codes that are the most responsive in month t when the combined impact of both effects is taken into account.

By December 1997, approximately eight months after the Web site was launched, 105 zip codes in Pennsylvania had at least one buyer. We implement our seeding experiment immediately thereafter. January 1998 is the first month available for seeding. For month t, we seed one new buyer into 50 regions selected by each strategy outlined previously and simulate expected trajectories of incremental buyers that should result from this one-time seeding. As an illustration, Figure 7, Panel A, shows the trajectory of incremental new buyers from the April 1998 seeding. In July 1998, for example, the 50 buyers seeded in April 1998 by the support-the-weak strategy have generated 3 new buyers.

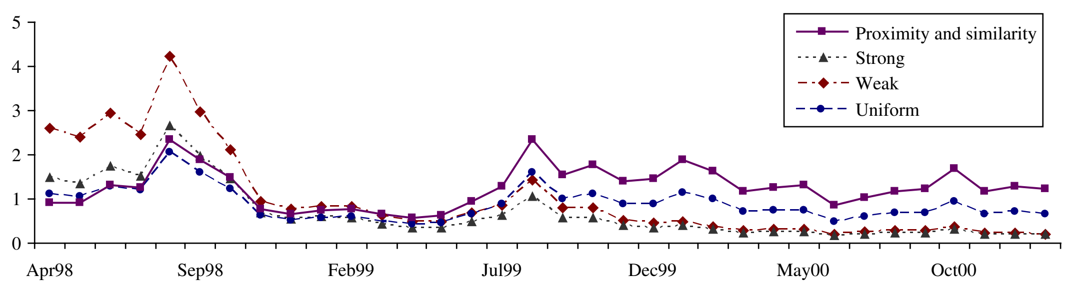
Among Libai, Muller, and Peres's (2005) strategies, the support-the-weak strategy shows the best performance early on (before January 1999), but subsequently it does not perform as well, because it fails to target potential markets that are spatially dispersed. With time, the proximity-and-similarity-based strategy performs best because the similarity effect begins to affect new and distant areas. By adjusting the impact of proximity and similarity effects over time, the proximity-and-similarity-based strategy pinpoints the most promising areas for growth. This natural coordination makes this strategy consistently superior over time.

Figure 7, Panel B, shows the aggregate number of incremental buyers through January 2001 that result from three different one-time seeding months (January 1998, January 1999, and January 2000). For example, "Jan 2000 Seeding" shows that seeding 50 buyers in January 2000 using the proximity-and-similarity-based strategy yields 18 new buyers in total by January 2001. Our findings with respect to the three strategies studied by Libai, Muller, and Peres (2005) are consistent with theirs; in general, spatially dispersed efforts are superior to spatially clustered efforts.

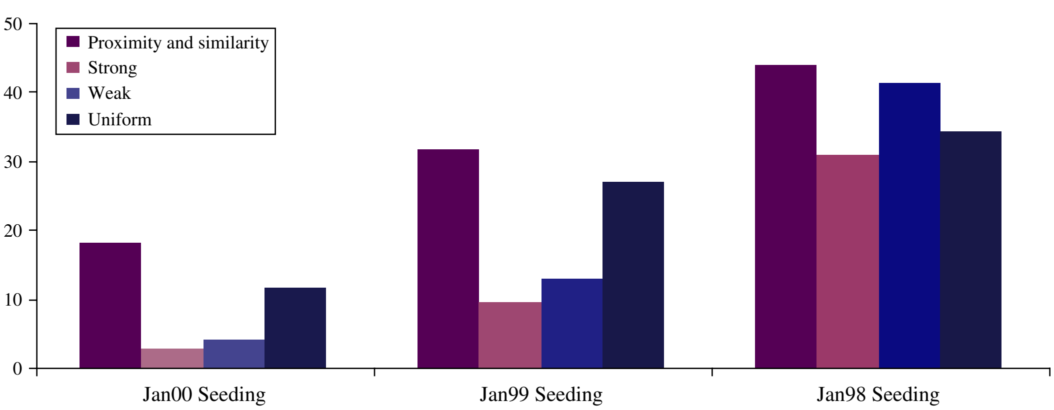
Figure 7

HYPOTHETICAL SEEDING EXPERIMENTS

A: Temporal Trajectory of the Number of Incremental New Buyers from the One-Time Seeding in April 1998 Through January 2001[a]



B: Aggregate Number of Incremental New Buyers Resulting from Three One-Time Seeding Months (in January 1998, January 1999, and January 2000) Through January 2001[b]



[a]50 new buyers were seeded in April 1998.
[b]50 new buyers were seeded in these seeding events.

When seeding is delayed, the support-the-weak strategy has less time to reap the benefit from proximity and its average performance deteriorates. The best overall outcome is induced by the proximity-and-similarity-based approach, and its superiority becomes more evident as the similarity effect gains momentum.

Panels A and B of Figure 7 together provide insight into how to optimize seeding strategies over time. The uniform strategy is the best among the strategies of Libai, Muller, and Peres (2005), but seeding by the support-the-weak strategy very early on can outperform a uniform strategy continuously applied. This is because the model shows that very early on, proximity effect plays a significant role, and the support-the-weak strategy (based on relatively underperforming areas with relatively large populations) can pick up zip codes with good potential for proximity effects. However, the support-the-weak strategy fails to pick up spatially dispersed markets, and therefore its performance quickly deteriorates with time. A switch from support-the-weak to uniform strategies might engender better performance. Unfortunately, it is difficult, if not impossible (from a practical perspective), to predict when to switch strategies.

This implies that Internet retailers in their infancy should perhaps focus initially on populous metropolitan areas. However, this strategy needs to be altered over time to incorporate the similarity effect as local concentration of demand declines. A spatially expanded customer base is likely to be important to an Internet retailer's growth. Our proximity-and-similarity-based strategy is a good candidate to this end because it automatically balances the similarity effect against the proximity effect while avoiding the need to manually switch strategies. Moreover, the relative advantage of this strategy increases the later seeding is started (see Figure 7, Panel B). Our finding highlights the insight that serving many small pools of somewhat similar buyers, who are spatially distant from one another, can be important to an Internet retailer because the relative contribution of these buyers to sales increases over time.

It is widely believed that a firm can offer an almost unlimited product assortment when the product stocking constraint is relaxed and that small sales levels over a large number of products account for substantial aggregate sales, a phenomenon termed "the long tail" (Anderson 2006; Brynjolfsson, Hu, and Simester 2006). Our insight into the importance of the sales distribution over obscure regions (see Balasubramanian 1998) mirrors the importance of the sales distribution over obscure products in the long tail. Here, the benefit comes primarily through the ability to sell in essentially unlimited local markets rather than sell an unlimited product assortment. The Internet retailer with sufficient distribution capabilities (e.g., through use of a third-party expert such as FedEx or UPS) is freed from the constraint of geography and can enjoy the benefit from serving sparse pockets of geographically diverse demand.

CONCLUSION

The vastly expanded trading area of the Internet retailer is perhaps the starkest difference between it and a traditional retailer. As such, it is critical for the Internet retailer to understand how and why demand varies spatially. In this article, we focus on the dynamic role of imitation based on geographic proximity and demographic similarity in generating new buyers over space and time. We find that in the initial phases of demand growth, proximity effects are more prominent. New demand in a local area is influenced by the extent of prior demand not only in the same local area and but also in contiguous and "geographically close" regions. As time progresses, the proximity effect diminishes in relative importance, but it does not dissipate entirely. The similarity effect tends to increase in relative importance over time and is particularly salient to demand generation in spatially dispersed regions with relatively small absolute sales.

LIMITATIONS AND DIRECTIONS FOR FURTHER RESEARCH

Our study focuses on a description of the behavior of new buyers only and does not explicitly measure the interactions among people. These limitations open several opportunities for further research, including the following four areas:

1. *Forecasting*: In this article, we focus on building a descriptive model instead of a forecasting model. Moving from description to forecasting requires a different model formulation. In particular, a researcher may want to use a Bayesian dynamic model (e.g., Bass et al. 2007; West and Harrison 1997) and assess its market seeding performance.
2. *Incorporating social networks by demographic types*: We measure the demographic similarity by the extent of shared sociodemographic characteristics. The researcher could allow for separate social networks by demographic types and examine which demographic network drives imitation (e.g., Conley and Topa 2002). A model could also be expanded with demographically correlated random effects in demographic space.
3. *Incorporating word-of-mouth valence*: Similar to Albuquerque, Bronnenberg, and Corbett (2007), we assumed that there is nonnegative imitation, which could be driven in part by positive word of mouth from the earlier buyers. An interesting extension would be to allow for negative influence (e.g., Godes and Mayzlin 2004).
4. *Incorporating marketing activities*: A unique aspect of our data is the absence of significant marketing efforts. Thus, we can assess the impact of imitation without controlling directly for potential marketing activities (e.g., advertising, promotions). If marketing activities are present, our model can be extended to control for them, perhaps using the method that Bass and colleagues (2007) suggest. Moreover, it may be possible to build on the approach of Jank and Kannan (2005), who find that there is significant spatial correlation in individual-level preference for PDF and print forms of books and that this affects price sensitivity at different geographical locations.

REFERENCES

Agrawal, Ajay, Devesh Kapur, and John McHale (2008), "How Do Spatial and Social Proximity Influence Knowledge Flows? Evidence from Patent Data," *Journal of Urban Economics*, 64 (2), 258–69.

Agresti, Alan (2002), *Categorical Data Analysis*. New York: John Wiley & Sons.

Albuquerque, Paulo, Bart J. Bronnenberg, and Charles J. Corbett (2007), "A Spatiotemporal Analysis of the Global Diffusion of ISO 9000 and ISO 14000 Certification," *Management Science*, 53 (3), 451–68.

Anderson, Chris (2006), *The Long Tail: Why the Future of Business Is Selling Less of More*. New York: Hyperion.

Angers, Jean-Francois and Mohan Delampady (1992), "Hierarchical Bayesian Curve Fitting and Smoothing," *The Canadian Journal of Statistics*, 20 (1), 35–49.

Anselin, Luc (1988), *Spatial Econometrics: Methods and Models*. Boston: Kluwer.

Avery, Jill, Thomas J. Steenburgh, John Deighton, and Mary Caravella (2008), "Adding Bricks to Clicks: The Effects of Store Openings on Sales through Direct Channels," working paper, Harvard Business School, Harvard University.

Balasubramanian, Sridhar (1998), "Mail Versus Mall: A Strategic Analysis of Competition Between Direct Marketers and Conventional Retailers," *Marketing Science*, 17 (3), 181–95.

Bass, Frank (1969), "A New Product Growth Model for Consumer Durables," *Management Science*, 15 (5), 215–27.

———, Norris Bruce, Sumit Majundar, and B.P.S. Murthi (2007), "Wearout Effects of Different Advertising Themes: A Dynamic Bayesian Model of the Advertising–Sales Relationship," *Marketing Science*, 26 (2), 179–95.

Bell, David R. and Sangyoung Song (2007), "Neighborhood Effects and Trial on the Internet: Evidence from Online Grocery Retailing," *Quantitative Marketing and Economics*, 5 (4), 361–400.

Bronnenberg, Bart J. and Carl F. Mela (2004), "Market Roll-Out and Retailer Adoption of New Brands," *Marketing Science*, 23 (4), 500–518.

Brynjolfsson, Eric, Yu (Jeffrey) Hu, and Mohammad S. Rahman (2008), "Battle of the Retail Channels: How Product Selection and Geography Drive Cross-Channel Competition," working paper, Sloan School of Management, Massachusetts Institute of Technology.

———, ———, and Duncan Simester (2006), "Goodbye Pareto Principle, Hello Long Tail," working paper, Sloan School of Management, Massachusetts Institute of Technology.

Chib, Siddhartha (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90 (432), 1313–21.

——— and Ivan Jeliazkov (2001), "Marginal Likelihood From the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96 (453), 270–81.

Claude, Besner (2002), "A Spatial Autoregressive Specification with a Comparable Sales Weighting Scheme," *Journal of Real Estate Research*, 24 (2), 193–211.

Conley, Timothy G. and Giorgio Topa (2002), "Socio-Economic Distance and Spatial Patterns in Unemployment," *Journal of Econometrics*, 17 (4), 303–327.

Cressie, Noel (1993), *Statistics for Spatial Data*. New York: John Wiley & Sons.

Fischer, Claude S. (1978), "Urban-to-Rural Diffusion of Opinions in Contemporary America," *American Journal of Sociology*, 84 (1), 151–59.

Forman, Chris, Anindya Ghose, and Avi Goldfarb (2009), "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, 55 (1), 47–57.

———, ———, and Batia Wiesenfeld (2008), "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, 19 (3), 291–313.

Gelman, Andrew, John B. Carlin, Hal S. Stern, and Donald B. Rubin (2003), *Bayesian Data Analysis*, 2d ed. New York: Chapman & Hall.

Getis, Arthur and J. Keith Ord (1992), "The Analysis of Spatial Association by Use of Distance Statistics," *Geographical Analysis*, 24 (3), 189–206.

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.

——— and ——— (2009), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, 28 (4), 721–39.

Goolsbee, Austan and Peter J. Klenow (2002), "Evidence of Learning and Network Externalities in the Diffusion of Home Computers," *Journal of Law and Economics*, 45 (2), 317–42.

Hastie, Trevor, Robert Tibshirani, and Jerome Friedman (2001), *The Elements of Statistical Learning*. New York: Springer.

Hauser, John, Gerard J. Tellis, and Abbie Griffin (2006), "Research on Innovation: A Review and Agenda for Marketing Science," *Marketing Science*, 25 (6), 687–717.

Howard, Philip E., Lee Raine, and Steve Jones (2001), "Access, Civic Involvement, and Social Interaction," *American Behavioral Scientist*, 45 (3), 382–404.

Huff, David L. (1964), "Defining and Estimating a Trade Area," *Journal of Marketing*, 28 (July), 34–38.

Jank, Wolfgang and P.K. Kannan (2005), "Understanding Geographic Markets of Online Firms Using Spatial Models of Customer Choice," *Marketing Science*, 24 (4), 623–34.

Kalyanam, Kirthi and Thomas S. Shively (1998), "Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach," *Journal of Marketing Research*, 35 (February), 16–29.

Katz, James E., Ronald E. Rice, and Philip Aspden (2001), "The Internet, 1995–2000," *American Behavioral Scientist*, 45 (3), 405–419.

Knorr-Held, Leonhard and Julian Besag (1998), "Modeling Risk from a Disease in Time and Space," *Statistics in Medicine*, 17 (18), 2045–2060.

LeSage, J.P. and R.K. Pace (2005), "Matrix Exponential Spatial Specification," *Journal of Econometrics*, 140 (1), 190–214.

Libai, Barak, Eitan Muller, and Renana Peres (2005), "The Role of Seeding in Multi-Market Entry," *International Journal of Research in Marketing*, 22 (4), 375–93.

Lilien, Gary L. and Arvind Rangaswamy (2004), *Marketing Engineering*, 2d ed. Victoria, BC, Canada: Trafford.

Manchanda, Puneet, Ying Xie, and Nara Youn (2008), "The Role of Targeted Communication and Contagion in Product Adoption," *Marketing Science*, 27 (6), 961–76.

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60 (3), 531–42.

——— (2000), "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14 (3), 115–36.

Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of Royal Statistical Society: Series B*, 56 (1), 41–42.

Pauwels, Koen and Scott A. Neslin (2008), "Building Bricks and Mortar: The Impact of Opening Physical Stores in a Multichannel Environment," MSI Working Paper No. 08-102.

Rabe-Hesketh, Sophia and Anders Skrondal (2005), *Multilevel and Longitudinal Modeling Using Stata*. College Station, TX: Statacorp LP.

Rosenblat, Tanya S. and Markus M. Mobius (2004), "Getting Closer or Drifting Apart?" *Quarterly Journal of Economics*, 119 (3), 971–1009.

Ross, Sheldon M. (1996), *Stochastic Processes*, 2d ed. New York: John Wiley & Sons.

Rossi, Peter E. and Greg M. Allenby (2003), "Bayesian Statistics and Marketing," *Marketing Science*, 22 (3), 304–328.

Van Alstyne, Marshall and Erik Brynjolfsson (2005), "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities," *Management Science*, 51 (6), 851–68.

Van den Bulte, Christophe and Yogesh V. Joshi (2007), "New Product Diffusion with Influentials and Imitators," *Marketing Science*, 26 (3), 400–421.

——— and Gary L. Lilien (1997), "Bias and Systematic Change in the Parameter Estimates of Macro-Level Diffusion Models," *Marketing Science*, 16 (4), 338–53.

Wahba, Grace (1978), "Improper Priors, Spline Smoothing and the Problem of Guarding Against Model Errors in Regression," *Journal of the Royal Statistical Society: Series B*, 40 (3), 364–72.

Waldfogel, Joel (2007), *The Tyranny of the Market*. Cambridge, MA: Harvard University Press.

Wand, Matt P. (2003), "Smoothing and Mixed Models," *Computational Statistics*, 18 (2), 223–49.

Wedel, Michel and Jie Zhang (2004), "Analyzing Brand Competition Across Subcategories," *Journal of Marketing Research*, 41 (November), 448–56.

Wikle, Christopher K. and Mevin B. Hooten (2006), "Hierarchical Bayesian Spatio-Temporal Models for Population Spread," in *Applications of Computational Statistics in the Environmental Sciences: Hierarchical Bayes and MCMC Methods*, J.S. Clark and A. Gelfand, eds. Oxford: Oxford University Press, 144–69.

Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (August), 282–94.