JMR
Journal of Marketing Research
Article Postprint
Volume XLV

Bayesian Spatio-Temporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer

Jeonghye Choi, Sam K. Hui, David R. Bell

September 11, 2008

Jeonghye Choi is a Doctoral Candidate at the Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA 19104 (Email: jeonghye@wharton.upenn.edu, Tel: 215 573 0541). Sam K. Hui is an Assistant Professor at the Stern School of Business, New York University, (Email: khui@stern.nyu.edu, Tel: 215 715 7774). David R. Bell is an Associate Professor at the Wharton School, University of Pennsylvania, 3730 Walnut Street, Philadelphia, PA 19104 (Email: davidb@wharton.upenn.edu, Tel: 215 898 8253). Jeonghye Choi thanks the Wharton Risk Management and Decision Processes Center for financial support. We are very grateful to seminar participants at Hong Kong University of Science and Technology, Koc University, The University of Arizona, The Wharton School, and at the 2007 Marketing Science Conference at Singapore Management University for comments. Netgrocer.com CEO Lisa Kent generously provided the data. We are also very grateful to the Editor, Area Editor and two anonymous reviewers for their constructive advice and suggestions.

Bayesian Spatio-Temporal Analysis of Imitation Behavior
Across New Buyers at an Online Grocery Retailer

Abstract

For Internet retailers, demand propagation varies not only through time, but also over space. The authors develop a Bayesian spatio-temporal 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 our Bayesian framework.

The model is applied to new buyers at Netgrocer.com and is calibrated on forty-five months of data that span all 1,459 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. Managerial implications stemming from the findings are explored through a hypothetical "seeding" experiment. Other implications for Internet retailing practice are also discussed.

Keywords: Imitation, Proximity, Similarity, Long Tail, Spatio-Temporal Model

The Internet has reduced customer access costs for firms and facilitated long-range connections among consumers. While “location” is a primary determinant of success for traditional retailers (e.g., Huff 1964), Internet retailers are not subject to the constraint of physical location. They can attract consumers over a wide geographic area which means that even physically-separated consumers can easily utilize 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 consumers may “imitate” their peers in their adoption behavior, and ultimately, what firms might do to expedite the demand process.

A large body of research assumes, in general, that imitation behavior plays an important role in generating demand (see, e.g., Bass 1969; Hauser, Tellis, and Griffin 2006). Studies that are directly relevant to our research offer two key findings. First, all else 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 higher 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 individuals with common overall socio-demographic 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 motivates the underlying phenomenon. It shows the cumulative number of new buyers at Netgrocer.com in each zip code in the state of Pennsylvania recorded in fifteen-month intervals from the inception of the service in May 1997 through January 2001. Three interesting patterns appear. First, the evolution of new buyers seems to have started 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.

[Insert Figure 1 about here]

To analyze the data in Figure 1, we formulate a dynamic Bayesian spatio-temporal Poisson model (e.g., Knorr-Held and Besag 1998), and specify the adoption rate for each region at each time period as a function of imitation effects based on proximity and similarity, along with other locally-defined covariates. We utilize a conventional distance-based proximity measure and a demographic similarity metric that mirrors approaches in 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 spatio-temporal 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. Long term viability is therefore 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 that 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 controlling 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 as this would require a substantially different approach. Lastly, the seeding analyses using the imitation coefficients are the best-case scenario given the data and intended to be illustrate the potential benefit of the *proximity-and-similarity-based* strategy.

The paper is organized 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 spatio-temporal 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 first present 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), and find evidence for imitation behavior.

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 individuals. In Goolsbee and Klenow (2002), the proportion of local households owning computers is used to show that individuals 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 also be investigated using information on pair-wise distances or contiguity. Bronnenberg and Mela (2004) employ measures of this sort and find emulation effects among local retailers—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 above studies measures 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 later 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 individuals to interact with each other and/or to imitate each other might 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. Van Alstyne and Brynjolfsson (2005), for example, point out that "neighborhoods" can be shaped by many dimensions including interests, preferences, and member characteristics. One might expect that individuals agglomerating either in online communities or through their revealed preferences for certain online businesses, should exhibit some homogeneity along demographic lines (e.g., by criteria such as occupation, education levels, income, or ethnic grouping).

The way in which "similarity" is measured is an important empirical and conceptual issue. Agrawal, Kapur, and McHale (2008), for instance, define individual-level social proximity using a co-ethnicity indicator and find a substitution effect between social and spatial proximities. Social proximity provides greater benefit for inventors who are not co-located, 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, e.g., labor economists are now less likely to write with economic historians and more likely to co-author with labor economists who are physically distant. Yang and Allenby (2003) study automobile choice and define individuals 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 ones 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 of ISO certification diffusion across countries and find that diffusion of ISO9000 is driven by proximity and trade-based similarity, whereas diffusion of ISO14000 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 socio-demographic characteristics.

Summary. Prior research demonstrates that geographic proximity and demographic similarity drive imitation behavior. These studies say relatively little, however, about how such effects evolve over time. Since 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 paper, we aim to focus on the temporal aspects of geographic proximity and demographic similarity and understand their dynamic influences in driving adoptions of the online retailer.

DATA

New Buyers. We obtained monthly transaction data for new buyers at Netgrocer.com in the state of 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 via 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 over time, we consider the number of new buyers after aggregating spatially and temporally. The final (bottom right) map in Figure 1 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 forty-five 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 more 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.

[Insert Figure 2 and Table 1 about here]

Regional Characteristics. The data for the imitation proxy variables and the direct measures of regional heterogeneity are assembled from three sources: (1) the 2000 US 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, based on the land area. Since 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.

[Insert Table 2 about here]

The FCC estimates the number of Internet service providers (ISPs) in each region, however, these data are known to be approximate. Some ISPs 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 time periods only (December 1999, June 2000, December 2000, June 2001), three of which are covered by our transaction data. Following the suggestion of Wand (2003), we therefore employ a low-rank thin plate spline smoother to improve the FCC data, and provide the complete details in Appendix I.[1] In addition, since the timeframe 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 and time-period specific measure which we call "Broadband Access." This control for access to the Internet is important in order to help rule out the alternative hypothesis that space-time evolution of Netgrocer.com new buyers simply mimics the diffusion of Internet access.

Finally, it is important to note that during the period of data collection, Netgrocer.com was not involved in any significant marketing activities in Pennsylvania; thus, this dataset offers a unique opportunity for us to assess imitation effects across space and time, free of explicit marketing interventions. While 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 2008; 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 data that describe regions are widely available and generally reliable. Hence, 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 nor practical to work with. 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 in order to help mitigate the well-known “reflection problem” (Manski 1993; 2000). Manski (1993) emphasizes that in order 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—where 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 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, assume there is a finite number of zip codes, n, such that all pair-wise relations 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). Our measure of across-region proximity is assumed to be 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 Δ to be equal to one.[4] The distance-based measure helps control for the fact that different zip codes vary greatly in land area and number of contiguous neighbors. Alternative proximity matrices based on shared boundaries and contiguity information are considered in Appendix II.

Demographic Similarity (*D*). Unlike with the measures of physical proximity, there is no single widely-used and straightforward approach to defining similarity. We presume that shared socio-demographic 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, everything else constant. We therefore 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 US Census reports zip-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. Similarly for income: (1) below the poverty line, (2) medium income, and (3) income in excess of $75,000 per year. Age categories are: (1) up to 20 years old, (2) 21 to 40, (3) 41 to 65, and (4) more than 65 years old. 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 socio-demographic profile vectors have a total of fifteen elements (four for education, age, and ethnicity, and three for income). Pair-wise similarity measures are defined as

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

where v_{ik} is the *k*-th element of the socio-demographic vector of region *i*; i.e., we sum the minimum values, based on the element-wise comparisons across two socio-demographic vectors for all $k = 1, 2, \ldots, 15$ elements of their socio-demographic profile. As in the case of physical proximity, two alternative measures of demographic similarity are defined in Appendix II.

A Bayesian Spatio-Temporal 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 in order to accurately estimate the role of imitation behavior. Towards 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

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

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 Appendix III we outline a mathematical argument (adapted from Knorr-Held and Besag 1998 and Ross 1996) that the Poisson approximation is valid under the assumption that adoption is sparse and within-period imitation is limited. We show empirically in the next section using a posterior predictive check (Gelman et al. 2003), that the Poisson distribution provides an excellent fit to the raw data. Finally, the Poisson distribution has been used in other instances where events are rare, e.g., to model the spread of new species (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, latent adoption rates λ_{it} are modeled as a function of region-level characteristics, temporal baseline effects, and geographic and demographic imitation effects

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

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

(6) $$\beta_t^W, \beta_t^G, \beta_t^D \geq 0 \quad \forall t$$

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 2002; Rabe-Hesketh and Skrondal 2005). μ_i and ζ_t are regional and temporal baselines, respectively. The regional baseline, γ_i, is comprised of 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] $G_{(i)}$ and $D_{(i)}$ denote the i-th rows of the matrices G and D, respectively, while z_{it} as denotes the (log-) cumulative number of buyers in region i prior to 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 mean 0 and variance σ_ε^2, allowing for over-dispersion.

We are interested in the final three terms for imitation in equation (4). $\beta_t^W z_{it}$ 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)). Post-multiplication 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. Since only non-perishable branded products (e.g., paper products, canned food, etc.) were sold during the data collection period, potential new buyers should have been able to assess product quality ex ante. Prices were also known. Hence, negative information was most likely to relate to delivery, which was handled by Federal Express. Therefore, we postulate the more cumulative buyers there are, the greater the number of new buyers that will emerge. Equation (6) reflects this restriction which assumes that all three imitation coefficients are non-negative. 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 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 allows us to take observations from neighboring time periods into account when making inference about a certain time period. When making inference about an estimate at time *t*, we take into account information from periods *t*-1, *t*-2, … (and also *t*+1, *t*+2, …) in polynomially decreasing weights, thereby allowing 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 in fact evolve. It also provides several key statistical advantages (see Appendix IV).

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

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

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

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

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

Standard proper conjugate priors are specified for all the other parameters in the model. An MCMC procedure is used to sample from the posterior distributions (see Appendix V).

EMPIRICAL FINDINGS

We first compare our model to reduced models and demonstrate the adequacy of our model in describing both the spatial and temporal dimensions of the data. We then present time-varying imitation parameter estimates (see Appendix VI for other control variables), interpret them, and discuss implications for market seeding and why the spatial concentration of new buyers declines over time.

Model Fits and Validation

The full model is compared, using marginal log-likelihood (Chib 1995; Chib and Jeliazkov 2001), to reduced models that "turn off" imitation effects based on proximity and similarity. The marginal log-likelihood for the full model with the proximity and similarity effects is -70,324, which is higher than 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} to the posterior predictive distribution of $\hat{y}_{it}$ (Gelman et al. 2003). Figure 2 panels (a) and (b) indicate an adequate model fit on the spatial and temporal dimensions after aggregating over time and space, respectively. Importantly, accurate spatial fit is obtained 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). The posterior means and 95% posterior intervals for these parameters together with the temporal baseline ζ_t are shown in Figure 3. There is significant non-stationarity 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 Lilien 1997; Van den Bulte and Joshi 2007). 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.

[Insert Figure 3 about here]

Further insights come from examining how the marginal effects of imitation vary across space and time. The marginal effect of imitation at region *i* at time *t* can be assessed by looking at the model-based expected number of new buyers $E(y_{it})$ compared to the expected number of new buyers (under the full model) with the imitation coefficients (β_t^W, β_t^G and β_t^D) set equal to 0. To assess the marginal effect of imitation across *space* we aggregate the 1,459 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 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 less likely to be subject to imitation effects based on proximity.

[Insert Figure 4 about here]

Figure 5 shows the marginal effects of imitation over *time*, by again comparing the expected number of buyers over time under the full model versus the expected number of buyers with imitation coefficients set to 0. The relative contribution of the imitation effects increases over time. This finding is intuitive as the larger the cumulative number of existing customers, the greater the potential for imitation of all types. This again underscores the importance of the installed base of new buyers for the ongoing acquisition of additional new buyers.

[Insert Figure 5 about here]

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. Since the first two components are based on short-range physical proximity and their relative magnitudes are relatively stable over space and over time (the ratio of *within*- and *across*- proximity effects is about 0.5), 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, from about thirty months out 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 start to emerge in hot spot areas (such as Philadelphia and Pittsburgh) and areas that are geographically proximate areas to hot spots. Later on, 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 later time 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 very important. This is because it helps drive orders from spatially-dispersed customers who are small in number individually, but collectively account for a significant percentage of total sales.

[Insert Figure 6 about here]

Market Seeding

Our findings suggest that the firm can influence the space-time demand trajectory through judicious market seeding (see also Godes and Mayzlin 2008). 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. Since validating these assumptions requires data that are beyond our sample, we must stress that the analyses presented here are purely conceptual and intended only to be treated as a springboard for future research.[8]

With this caveat in mind, we explore the following "seeding" scenario. Suppose 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 $t + 1$ onwards. 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", i.e., current performance is relatively "weak" compared to what might be expected. A common heuristic is that the market potential is roughly proportional to population size so we pick candidate regions according to 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 website 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 above and simulate expected trajectories of incremental buyers that should result from this one-time seeding. As an illustration, the trajectory of incremental new buyers from the April 1998 seeding is shown in Figure 7 panel (a). In July 1998, for example, the 50 buyers seeded in April 1998 by the *support-the-weak* strategy have generated 3 new buyers.

Among the strategies of Libai, Muller, and Peres (2005), the s*upport-the-weak* strategy shows the best performance early on (prior to January 1999), but later on it does not perform as well as it fails to target potential markets that are spatially-dispersed. With time the *proximity-and-similarity-based* strategy performs best as the similarity effect starts to impact 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 about here]

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). “Jan 2000 Seeding”, for example, 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; spatially-dispersed efforts are generally superior to spatially-clustered efforts. 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 *support-the-weak* very early on can outperform a *uniform* strategy continuously applied. This is because the model shows that very early on the proximity effect plays a significant role and the *support-the-weak* strategy (based on relatively under-performing areas with relatively large populations) can pick up zip codes with good potential for proximity effects. The *support-the-weak* strategy however fails to pick up spatially-dispersed markets and therefore its performance deteriorates fast with time. A switch from *support-the-weak* to *uniform* strategies might engender better performance. Unfortunately, it is “hard-to-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 as 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 (b)). Our finding highlights the insight that serving many small pools of somewhat *similar* buyers, who are spatially distant from each other, can be important to an Internet retailer as 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 on 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 paper, 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 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 Future Research.

Our study focuses on a description of the behavior of new buyers only, and does not explicitly measure the interactions among individuals. These limitations open several opportunities for future research, including the four areas described below.

- *Forecasting*: In this paper, we focus on building a descriptive model instead of a forecasting model. Moving from description to forecasting requires a different model formulation. In particular, one may want to utilize a Bayesian dynamic model (e.g., Bass et al. 2007; West and Harrison 1997) and assess its market seeding performance.

- *Incorporating social networks by demographic types:* We measure the demographic similarity by the extent of shared socio-demographic characteristics. One could allow for separate social networks by demographic types and examine which demographic network drives imitation (e.g., Conley and Topa 2002). One could also expand a model with demographically-correlated random effects in demographic space.
- *Incorporating WOM valence*: We have assumed, similar to Albuquerque, Bronnenberg, and Corbett (2007), that there is non-negative imitation, which could be driven in part by positive word-of-mouth from the earlier buyers. One interesting extension would be to allow for negative influence (e.g., Godes and Mayzlin 2004).
- *Incorporating marketing activities*: A unique aspect of our data is the absence of significant marketing efforts. We can thus 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 suggested in Bass et al. (2007). Moreover, one could build on the approach in Jank and Kannan (2005) who find significant spatial correlation in individual-level preference for PDF and print forms of books, and that this impacts 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*, forthcoming.

Agresti, Alan (2002), *Categorical Data Analysis*, Wiley: New York, NY.

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), 451-468.

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

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*, Kluwer: Boston, MA.

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.

Balasubramanian, Sridhar (1998), “Mail versus Mall: A Strategic Analysis of Competition between Direct Marketers and Conventional Retailers,” *Marketing Science*, 17 (3), 181-195.

Barbour, A. D., Lars Holst, and Svante Janson (1992), *Poisson Approximation*, Clarendon: Oxford, England.

Bass, Frank (1969), “A New Product Growth Model for Consumer Durables,” *Management Science*, 15 (5), 215-227.

________, 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), 179-195

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, MIT.

Casella, George and Edward I. George (1992), "Explaining the Gibbs sampler," *American Statistician*, 46 (3), 167-174.

Chib, Siddhartha (1995), "Marginal Likelihood from the Gibbs Output," *Journal of the American Statistical Association*, 90 (432), 1313-1321.

________ and Ivan Jeliazkov (2001), "Marginal Likelihood From the Metropolis-Hastings Output," *Journal of the American Statistical Association*, 96 (453), 270-281.

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*, Wiley: New York, NY.

Fischer, Claude S. (1978), "Urban-to-Rural Diffusion of Opinions in Contemporary America," *American Journal of Sociology*, 84 (1), 151-159.

Forman, Chris, Anindya Ghose, and Avi Goldfarb (2008), "Competition Between Local and Electronic Markets: How the Benefit of Buying Online Depends on Where You Live," *Management Science*, forthcoming.

________, ________, and Batia Wiesenfeld (2008), "Examining the Relationship Between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research*, forthcoming.

French, Jonathan L., Erin E. Kammann, and Matt P. Wand (2001), "Semiparametric Nonlinear Mixed-Effects Models and Their Applications: Comment," *Journal of the American Statistical Association*, 96 (456), 1285-1288.

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

Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545-560.

________ and ________ (2008), "Firm-Created Word-of-Mouth Communication: Evidence from a Field Test," *Marketing Science*, forthcoming.

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), 317-342.

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

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

Hastings, W. Keith (1970), "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57 (1), 97-109.

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 (3), 34-38.

Jank, Wolfgang, P. K. Kannan (2005), "Understanding Geographic Markets of Online Firms Using Spatial Models of Customer Choice," *Marketing Science*, 24 (4), 623-634.

Kalyanam, Kirthi and Thomas S. Shively (1998), "Estimating Irregular Pricing Effects: A Stochastic Spline Regression Approach," *Journal of Marketing Research*, 35 (1), 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), 375-393.

Lilien, Gary L. and Arvind Rangaswamy (2004), *Marketing Engineering, 2nd Edition*, Trafford: Victoria, B.C.

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

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60 (3), 531-542.

________ (2000), "Economic Analysis of Social Interactions," *Journal of Economic Perspectives*, 14 (3), 115-136.

Molenberghs, Geert and Geert Verbeke (2006), *Models for Discrete Longitudinal Data*, Springer-Verlag: New York, NY.

Newton, Michael A. and Adrian E. Raftery (1994), "Approximate Bayesian Inference by the Weighted Likelihood Bootstrap," *Journal of Royal Statistical Society Series B*, 56, 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 08-102, Cambridge, MA.

Rabe-Hesketh, Sophia and Anders Skrondal (2006), *Multilevel and Longitudinal Modeling Using Stata*, Statacorp LP: College Station, TX.

Ravishanker, Nalini, and Dipak K. Dey (2002), *A First Course in Linear Model Theory*, Chapman and Hall, Boca Raton, FL.

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, 2nd Edition*, Wiley, New York.

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

Simonoff, J. S. (1986), *Smoothing Methods in Statistics*, Springer, New York.

Van Alstyne, Marshall and Erik Brynjolfsson (2005), "Global Village or Cyber-Balkans? Modeling and Measuring the Integration of Electronic Communities," *Management Science*, 51 (6), 851-868.

Van den Bulte, Christophe and Gary L. Lilien (1997), "Bias and Systematic Change in the Parameter Estimates of Macro-Level Diffusion Models," *Marketing Science*, 16 (4), 338–353.

________ and Yogesh V. Joshi (2007), "New Product Diffusion with Influentials and Imitators," *Marketing Science*, 26 (3), 400-21.

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), 364-372.

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

Wand, Matt P. (2003), "Smoothing and Mixed Models," *Computational Statistics*, 18, 223-249.

Wedel, Michel and Jie Zhang (2004), "Analyzing Brand Competition Across Subcategories," *Journal of Marketing Research*, 41(4), 448-456.

West, Mike, and Jeff Harrison (1997), *Bayesian Forecasting and Dynamic Models*, Springer: New York, NY.

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

Yang, Sha and Greg M. Allenby (2003), "Modeling Interdependent Consumer Preferences," *Journal of Marketing Research*, 40 (3), 282-294.

FOOTNOTES

1. We also estimated our model using non-smoothed broadband data and obtained qualitatively similar results. Full details are available from the authors upon request.
2. Household level Internet usage data were collected as supplementary data in the Current Population Survey (CPS) from 8,162 national zip codes in October 1997, December 1998, August 2000, and September 2001. Although the CPS data match nicely with the time period for the Netgrocer.com data, they include only 670 (46%) of the zip codes in Pennsylvania. Therefore, we utilize the spatially-smoothed "Broadband Access" variable derived from the FCC data as this measure can be constructed for all 1,459 zip codes in Pennsylvania. The average zip code-level correlations between the CPS data and smoothed "Broadband Access" are 0.95 for the total US sample of 8,162 zip codes and 0.97 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.
3. Manski (1993, p. 532-537) 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, access to the Internet and to local retailers is controlled for in our model. Also, regional and temporal baselines account for region and time specific shocks. While 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.

4. This assumption is made for reasons of computational tractability and consistency with the previous literature (e.g., Claude 2002; LeSage and Pace 2005; Yang and Allenby 2003). In order to demonstrate that our empirical findings are robust to this assumption, we defined two additional proximity measures with an inverse function of *half* (=0.5) and *twice* (=2) the geographic distance, and re-estimated the models. Both measures provide consistent model estimates and thus the same qualitative insights. We thank an anonymous reviewer for suggesting this check.
5. One can specify that these random effects are spatially correlated, e.g., using a CAR (Conditional AutoRegressive) formulation (Cressie 1993). Albuquerque, Bronnenberg, and Corbett (2007), however, found that incorporating spatially-correlated errors do not improve their model's performance. Thus, we retain the i.i.d. specification. One can also specify a more general model with demographically correlated random effects, e.g., as a joint distribution across zip codes with correlation in demographic space. We thank the AE for this observation.
6. Giving these temporal parameters independent diffuse normal distributions (i.e., $N(0,100^2)$) is undesirable for two reasons. First, since these parameters measure the strength of imitation over time, one would expect them 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 time periods when estimating these parameters, and hence reduce 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).

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

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 paper should be interpreted as the best-case scenario.

Table 1. Summary Statistics for the Number of New Buyers

	Mean	**Std Dev**	**Min**	**Max**
May, 1997	0.001	0.037	0.000	1.000
Oct, 1997	0.008	0.090	0.000	1.000
Mar, 1998	0.055	0.282	0.000	3.000
Aug, 1998	0.198	0.631	0.000	8.000
Jan, 1999	0.065	0.322	0.000	5.000
June, 1999	0.083	0.350	0.000	6.000
Nov, 1999	0.212	0.602	0.000	7.000
Apr, 2000	0.220	0.607	0.000	5.000
Sep, 2000	0.219	0.823	0.000	22.000
Jan, 2001	0.235	0.802	0.000	19.000

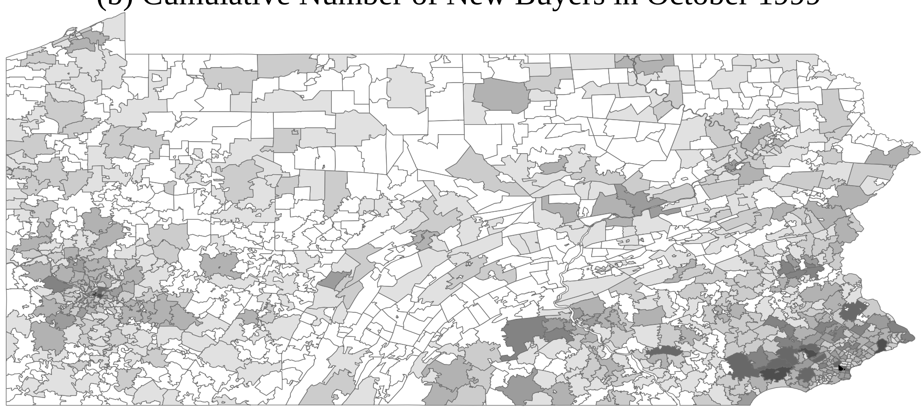
Table 2. Variable Descriptions and Summary Statistics for Zip Code Characteristics

Variable	Description	Mean	Std Dev
Local Environment			
Population	Total population	8391.600	11149.400
Population Density	Population density	1298.799	3117.592
Population Growth	Annual population growth rate from 2000 to 2004	0.004	0.011
Home Value	% of homes valued at $250,000 or more	0.060	0.108
Urban Housing	% of houses with 50 units or more	0.018	0.056
Land Area	Area in square miles	30.607	35.511
Household Characteristics			
Asian	% of Asians	0.008	0.016
Black	% of Blacks	0.038	0.112
White	% of Whites	0.938	0.130
College	% with bachelors and/or graduate degree	0.370	0.144
Elderly	% aged 65 and above	0.156	0.041
Wealthy	% of households earning $75,000+	0.165	0.118
Access to Retail Services			
Density General	Density of general stores within the second order neighboring zip codes	0.107	0.251
Density Supermarket	Density of supermarkets within the second order neighboring zip codes	0.224	0.393
Presence Warehouse	Presence of warehouse clubs within the second order neighboring zip codes	0.245	0.430
Access to the Internet			
Broadband Access	Number of high-speed Internet service providers		
Dec, 1999		1.784	1.320
June, 2000		2.060	1.749
Dec, 2000		2.940	2.665
June, 2001		2.840	2.773

Figure 1. Spatio-Temporal Evolution of Netgrocer.com Buyers in Pennsylvania

(a) Cumulative Number of New Buyers in July 1998

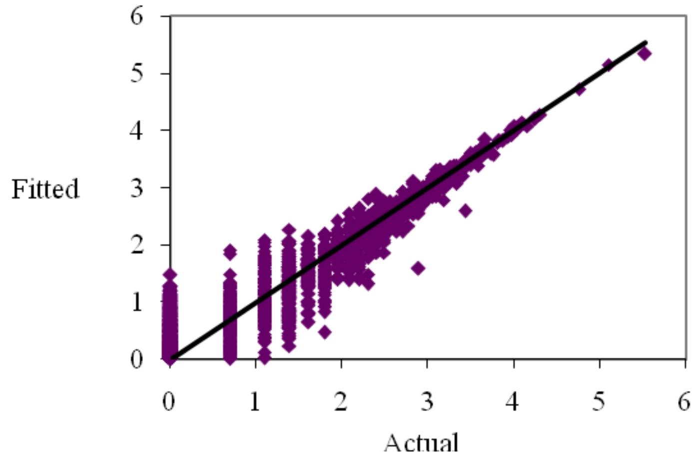
(b) Cumulative Number of New Buyers in October 1999



(c) Cumulative Number of New Buyers in January 2001

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))

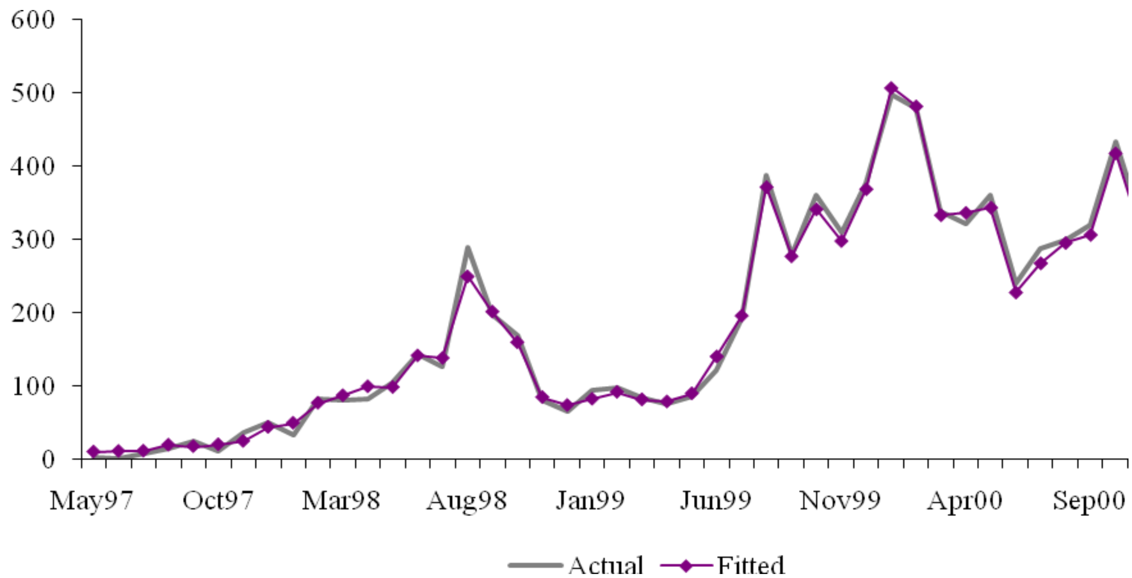


Figure 3. Posterior Means and 95% Posterior Intervals for the Temporal Baseline (ζ_t) and the Time-Varying Imitation Parameters (β_t^W, β_t^G and β_t^D)

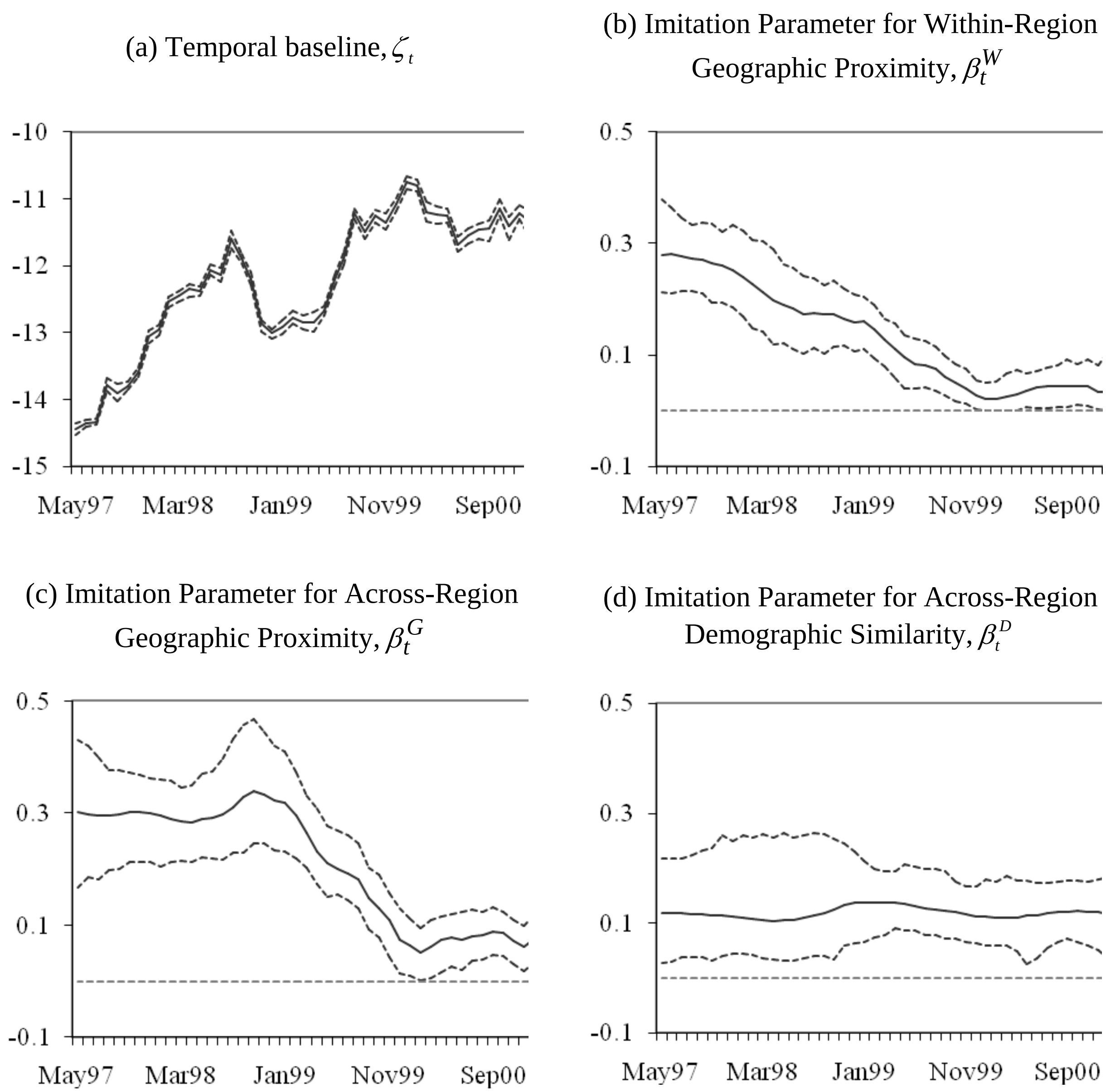


Figure 4. Expected Number of Buyers in Each County (aggregated over time) Under our Full Model, Compared to the Expected Number by Setting Imitation Coefficients to 0 (the gap is the marginal effect of imitation under our model framework)

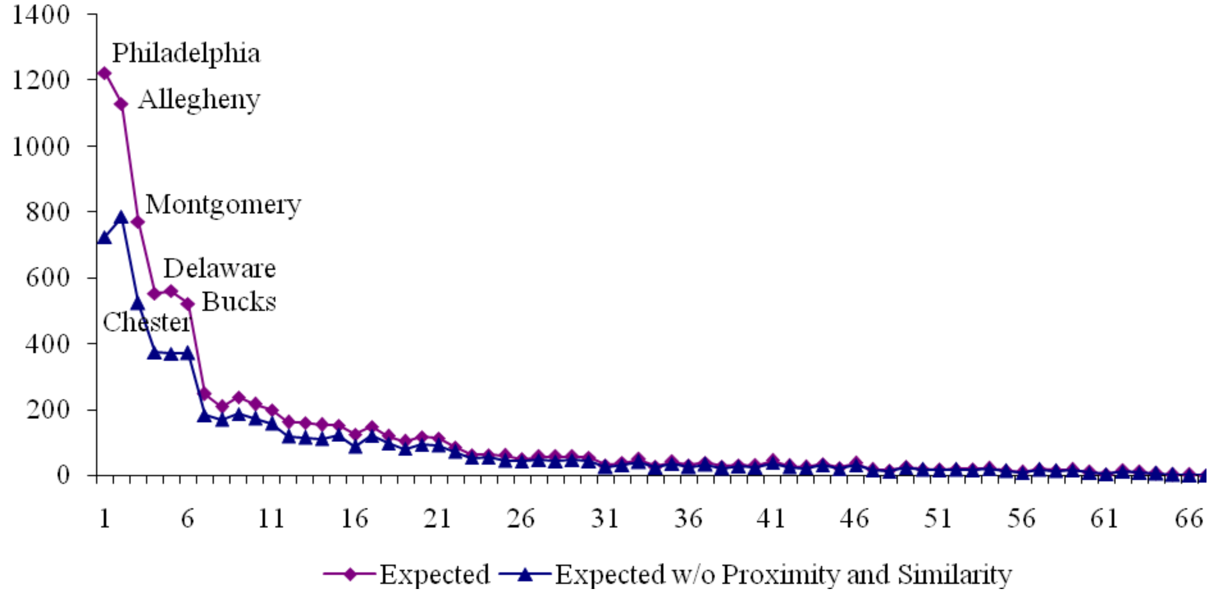


Figure 5. Expected Number of New Buyers in Each Month (aggregated over zip codes) Under our Full Model, Compared to the Expected Number by Setting Imitation Coefficients to 0 (the gap is the marginal effect of imitation under our model framework)

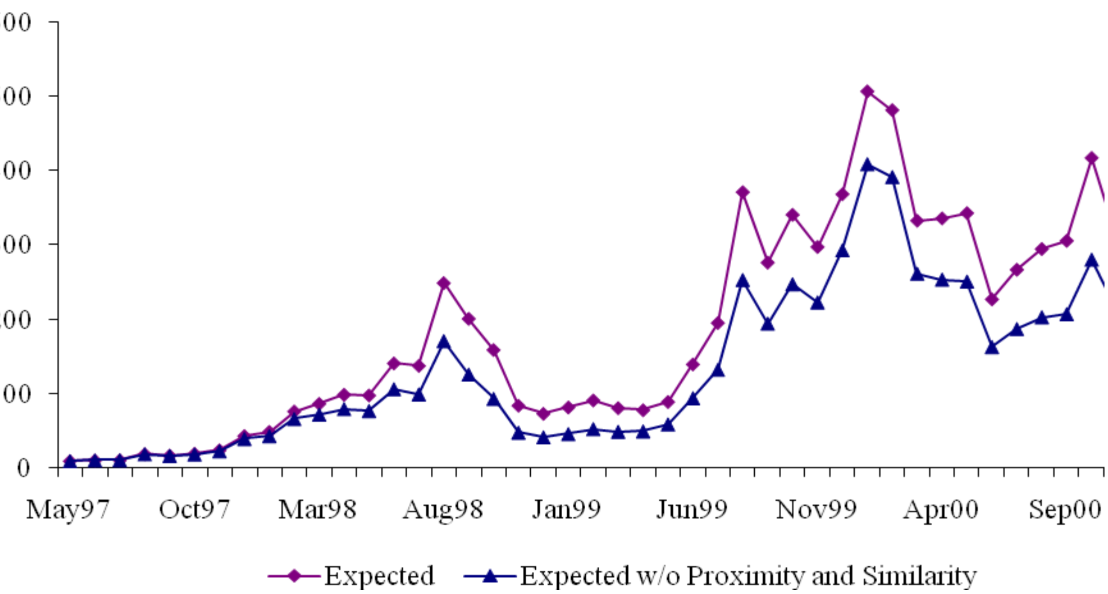


Figure 6. Relative Magnitudes of the Proximity and Similarity Effects Over Time

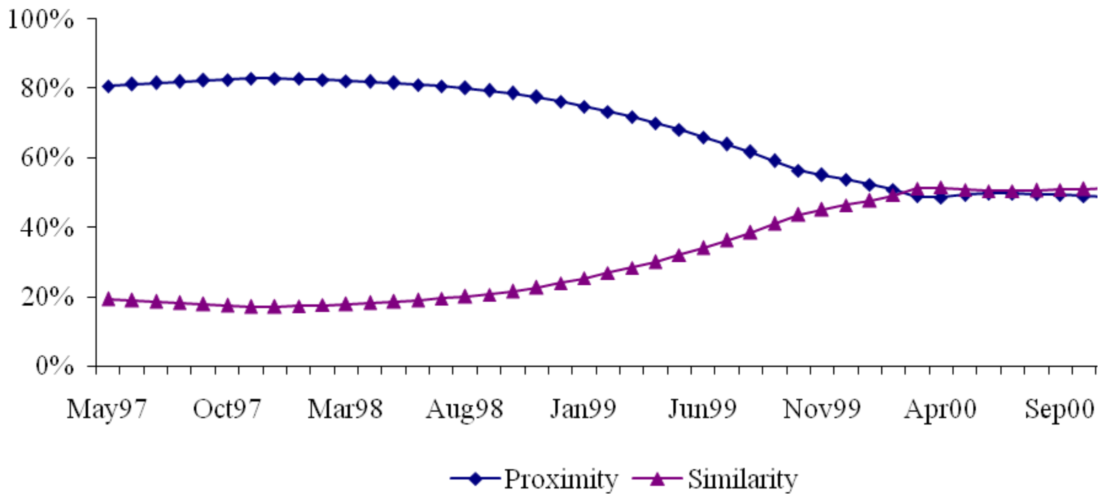
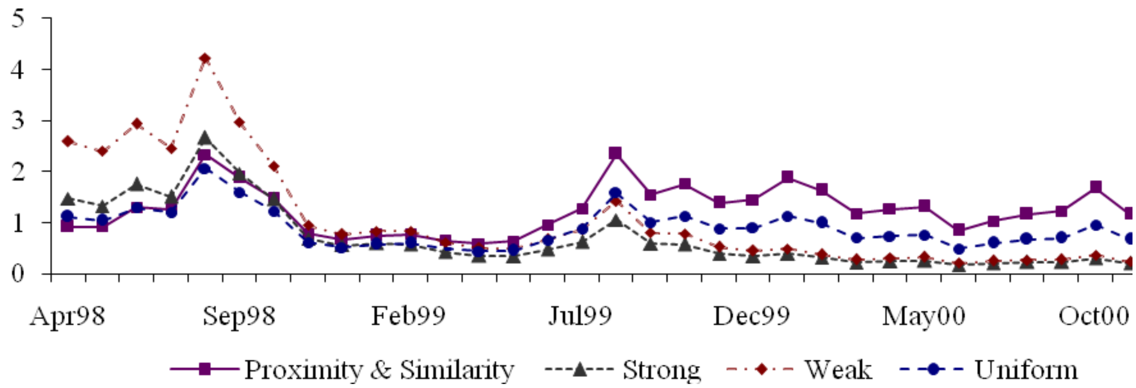
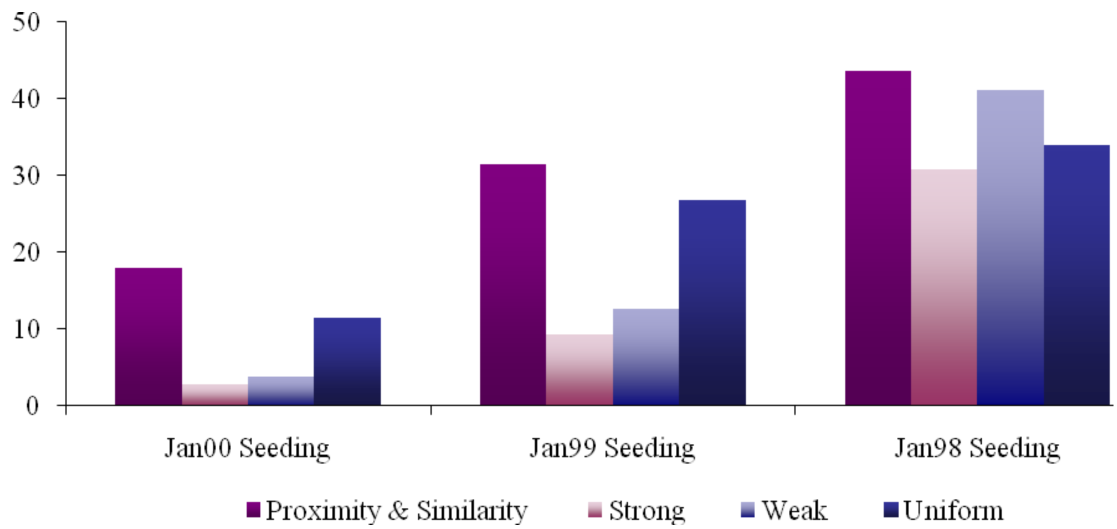


Figure 7. Hypothetical Seeding Experiments

(a) Temporal Trajectory of the Number of Incremental New Buyers from the One-Time Seeding in April 1998 through January 2001.
(50 new buyers were seeded in April 1998)



(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.
(50 new buyers were seeded in these seeding events)



WEB APPENDIX

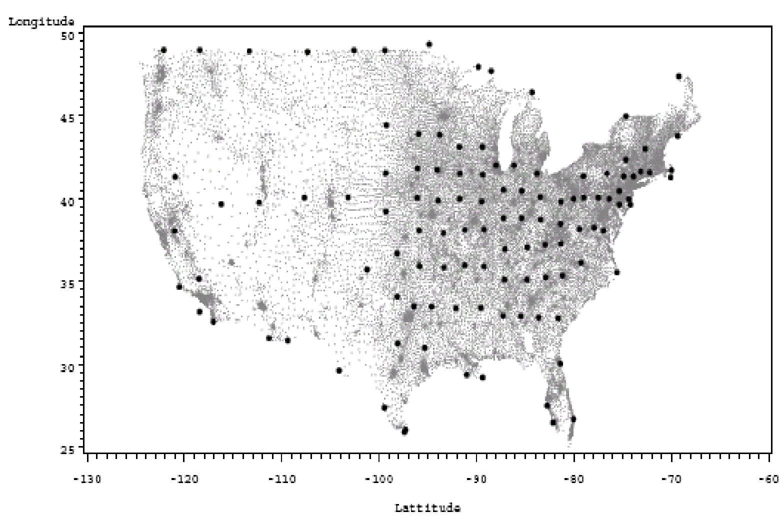
Bayesian Spatio-Temporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer

Jeonghye Choi, Sam K. Hui, David R. Bell

I. Low-rank spatial smoothing of the broadband access variable

Our measure of the Internet access availability has a few known imperfections: there are some missing data for some zip codes, and in some cases services are under-reported. To improve the quality of this variable, we implement a low-rank thin plate spline smoother (Wand 2003) to correct for measurement errors. Here, we provide an outline for our implementation; readers are encouraged to see Wand (2003) for more details.

Step 1. Choose knots: We obtain "knots" based on centroids of a *k-d* tree (Molenberghs and Verbeke 2006). Starting with the entire set of zip codes, the *k-d* tree partitions the space until all partitions contain at most 300 regions. The region nearest to the centroid of each partition is chosen as a knot which creates 117 knots in our application, as shown in the figure below.



Step 2. Bivariate radial smoothing: We then apply the low-rank thin plate spline smoothing with a radial basis function. ISPs in region *i* at time *t*, x_{it}, are specified to follow a negative binomial distribution with parameters r and κ_{it} (French, Kammann, and Wand. 2001; Molenberghs and Verbeke 2006). κ_{it} is then spatially-smoothed based on the Euclidean distances from the set of knots, $z_1, z_2, ..., z_K$, and a proper covariance function.

II. Alternative measures for *G* and *D* matrices

The matrix *G* can also be specified from information on the shared boundaries among zip codes. Two alternatives are: (1) the shared boundary approach, and (2) the contiguity approach. The shared boundary weighting matrix is

$$G_{ij} = \begin{cases} l_{ij} / l_i, & l_{ij} > 0 \\ 0, & otherwise \end{cases} \qquad \text{[II-1]}$$

where l_{ij} is the length of zip code *i*'s boundary shared with zip code *j* and l_i is the total length of *i*'s boundary shared with all its contiguous zip codes, i.e., $l_i = \sum_j l_{ij}$. This weighting system is appropriate when two regions with a longer shared boundary might be expected to exert greater influence on each other. The shared boundary weighting matrix can be simplified to a case where two neighboring regions have equal influence on the focal region as long as they share boundaries with focal region, and this simpler form is called a contiguity weighting matrix,

$$G_{ij} = \begin{cases} 1, & l_{ij} > 0 \\ 0, & otherwise. \end{cases} \qquad \text{[II-2]}$$

Two alternative measures for *D* are: (1) Inverse Exponential Mahalanobis Distance, and (2) "Affiliation" (Van Alstyne and Brynjolfsson 2005).

The first measure, Inverse Exponential Mahalanobis Distance, is based on Mahalanobis distance as suggested by Van Alstyne and Brynjolfsson (2005). It measures scale-free *dissimilarity* between regions *i* and *j* and takes into account correlations in the data

$$d_{ij} = \sqrt{(v_i - v_j)'\Sigma^{-1}(v_i - v_j)}, \qquad \text{[II-3]}$$

where v_i is a vector of socio-demographic characteristics of region *i* and Σ^{-1} is the corresponding covariance matrix. As equation [II-3] is a measure of dissimilarity, similarity can be specified as an inverse function of the exponentiated socio-demographic distance (Yang and Allenby 2003):

$$D_{ij} = \exp(-d_{ij}) \qquad \text{[II-4]}$$

Affiliation is derived to be directly consistent with analytical work in Van Alstyne and Brynjolfsson (2005). Instead of using regional profile vectors directly, we define regional "vectors of types" in the following way. We compute the empirical distribution of each individual element of the fifteen elements of the profile vectors described in the paper. That is, we look across all 1,459 regions in the sample and compute the first quartile, median, and third quartile of the distribution of a particular characteristic. As a result, for each region and each characteristic, we can assign the region to one of four mutually exclusive and collectively exhaustive "types" along each element: "high" (top quartile and above) "moderate" (between median and top quartile), "low" (between bottom quartile and median), and "very low" (below bottom quartile). For example, imagine that the first quartile of the distribution of the ethnic subcategory "Black" is 10% (i.e., one quarter of the regions in the sample have a population which contains 10% or fewer Blacks). If the Black proportion of a region is 5%, then its type is defined as a region with a very low proportion of Black residents (compared to the overall population). If another region also has a small portion of Black residents, say 7%, then these two regions are assumed to be implicitly affiliated on the Black dimension. The extent of affiliation comparing two regions is as

$$D_{ij} = \sum_k I(e_{ik}, e_{jk}) \qquad \text{[II-5]}$$

where e_{ik} is an element of the vector of socio-demographic types of region i, and $I(\cdot)$ is an indicator function which takes one if two elements are equal, and zero otherwise.

III. Justification of Poisson distribution in Equation (3)

We denote the number of individiuals in zip code i as N_i, and the number of buyers as Y_i. Let y_{ij} (j = 1, 2, … N_i) be an indicator variable which takes value 1 if the j-th individual in zip code i adopts, and 0 otherwise. In other words, we have $Y_i = \sum_{j=1}^{N_i} y_{ij}$. The Poisson distribution has been shown to be an adequate limiting distribution under the following three assumptions:

(i) The adoption probabilities are equal across individuals,

(ii) the adoption probabilities are low, and

(iii) adoption behaviors across individuals, *during the same time period*, are independent.

As discussed in Section 3, assumption (ii) holds in our dataset; assumption (i) and (iii), however, are fairly strong assumptions that may not hold in reality. In the following argument, adapted from Knorr-Held and Besag (1998) and Ross (1996), we show that under a reasonable relaxation of assumptions (i) and (iii), the Poisson distribution is still a valid approximation.

Heterogeneous adoption probabilities (relaxing assumption (i))

We begin by assuming that the j-th individual in the i-th zip code adopts with probability θ_{ij} , and θ_{ij} is beta-distributed across the different individuals, i.e., $\theta_{ij} \sim \text{Beta}\,(a_i, b_i)$.

We can now derive the marginal distribution of y_{ij} as follows. Since y_{ij} can take only the value 0 or 1, we consider the marginal probability that y_{ij} takes value 1:

$$P(y_{ij}=1)=\int_0^1 P(y_{ij}=1\,|\,\theta_{ij})\text{Beta}(\theta_{ij}\,|\,a_i,b_i)d\theta_{ij}=\int_0^1 \theta_{ij}\text{Beta}(\theta_{ij}\,|\,a_i,b_i)d\theta_{ij}=\frac{a_i}{a_i+b_i} \qquad \text{[III-1]}$$

By writing $p_i=\dfrac{a_i}{a_i+b_i}$, we can write the marginal distribution of y_{ij} as:

$$y_{ij} \sim \text{Bernoulli}\,(p_i) \qquad \text{[III-2]}$$

Thus, the marginal distribution of Y_i is $\text{Binomial}(N_i, p_i)$. When p_i is small, we can use the classical Poisson approximation (Ross 1996) to obtain:

$$Y_i \sim \text{Poisson}(N_i p_i) = \text{Poisson}\,(\lambda_i) \qquad \text{where } \lambda_i = N_i p_i\,. \qquad \text{[III-3]}$$

Weakly-correlated adoption (relaxing assumption (iii))

To relax the assumption that *same-period* adoptions across individuals are independent, we need to consider the possibility of positive correlations across individual adoption behaviors. We assume that imitation behavior takes time to develop, and hence same-period imitation is weak; thus, individuals' adoption behavior during the same period is assumed to be at most weakly correlated. (Again, this is reasonable given the sparseness of our data; see Figure 2.)

Researchers have developed error bounds for the Poisson approximation when correlations across individuals are present. The bound, derived using the Stein-Chen method (Ross 1996), is as follows:

$$\left|P\{Y\in A\}-\sum_{i\in A} e^{-\lambda}\lambda^i/i!\right| \le \min(1,1/\lambda)\sum_{i=1}^{n}\lambda_i E[|\,Y-V_i\,|] \qquad \text{[III-4]}$$

where V_i is such that $P\{V_i=k\}=P\left\{\sum_{j\neq i} X_j = k \mid X_i = 1\right\}\ \ \forall k$.

For a detailed derivation of [III-4], we encourage readers to refer to Ross (1996) or Barbour, Holst, and Janson (1992). Here, we note that the errors of the Poisson approximation are proportional to the quantity $E[|Y - V_i|]$, which is assumed to be arbitrarily small given our assumption that imitation takes time (thus same-period imitation is limited). Empirically, the validity of the Poisson approximation is also supported by the empirical evidence presented in Figure 2; the predicted distribution of adopters under our model closely resembles that of the actual empirical distribution.

IV. Embedding a Frequentist polynomial smoother within a Bayesian model

In this appendix, we discuss how we embed a polynomial smoother within our Bayesian model by exploiting the parallel between the Gaussian random walk prior specification and polynomial smoothing. We begin by providing a brief introduction of smoothing techniques commonly used in Frequentist nonparametric statistics, and then explain how we embed such techniques into our model using Gaussian random walk priors.

Smoothing techniques

In non-parametric statistics, a smoother is often used to produce a smooth curve of *y* against *x*, given a scatterplot of (*x*,*y*) values (Simonoff 1986). The underlying model is of the form $y_i = f(x_i) + \varepsilon_i$, and interest is typically centered on estimating the function *f*(.). Since *y* is measured with error ε_i, smoothing helps the estimation of $f(x_i)$ by considering not only the observations at x_i, but also observations that have *x* values "close" to x_i. When estimating $f(x_i)$, these "neighboring" observations are down-weighted by their distance from x_i. For instance, a kernel smoother is of the form (Hastie, Tibshirani, and Friedman 2001):

$$\hat{y}_i = \sum_{j=1}^{n} y_j \, K\left(\frac{x_i - x_j}{b}\right) \Bigg/ \sum_{j=1}^{n} K\left(\frac{x_i - x_j}{b}\right) = \sum_{j=1}^{n} w(x_{i,} x_j) y_j \qquad \text{[IV-1]}$$

where $w(x_i, x_j)$ denotes the "weight" of the *j*-th observation on the estimation of y_i, which is governed by the distance of x_j to x_i. Different types of smoothers are defined based on the functional form of $w(x_i, x_j)$. In particular, for a *polynomial* smoother, the weights are defined to be proportional to $\rho^{|x_i - x_j|}$, $\rho < 1$.

Gaussian random walk prior

A Gaussian random walk prior, as defined in Equation (7)-(10), allows us to embed a polynomial smoother within our Bayesian model. In the discussion below, we explore the parallel between a random walk prior and polynomial smoothing using a simplified set-up as follows (t = 1,2,…,T):

$$y_t \mid \theta_t \sim N(\theta_t, 1) \qquad \text{[IV-2]}$$

$$\theta_t \mid \theta_{t-1} \sim N(\theta_{t-1}, \gamma^2) \qquad \text{[IV-3]}$$

$$\pi(\theta_0) \sim N(0, \sigma^2) \qquad \text{[IV-4]}$$

Equation [IV-2] states that y_t is θ_t observed with error, in the same way that the adoption number y_{it} is a noisy observation based on the time-varying coefficients (and other controls) in Equation (3). Equation [IV-3] is the Gaussian random walk prior similar to that in Equation (7)-(10). Equation [IV-4] is a conjugate prior for the first period parameter; the variance term σ^2 can be set to a large number (e.g., 100^2) to obtain a diffuse prior.

We now explore the properties of the random walk prior in the simplified setting in Equations [IV-2]-[IV-4]. First, we show that the posterior mean estimate of θ_t (and hence $\hat{y}_t$) is a linear function of $\vec{y}$. We then proceed to show more concretely, using a numerical example, that the properties of these estimators mirror that of a polynomial smoother.

Since the conditional distribution for each $\theta_t \mid \theta_{t-1}$ is normal, it follows that their joint prior distribution is also normal (Ravishanker and Dey 2002). Thus, we can write

$$\pi(\vec{\theta}) = MVN(\vec{\mu}_0, \Lambda_0) \qquad \text{[IV-5]}$$

where (after algebraic manipulations)

$$\Lambda_0(i, j) = \sigma^2 + \gamma^2[\min(i, j) + 1] \, . \qquad \text{[IV-6]}$$

Clearly, given Equation [IV-4] and the structure of Equation [IV-3], the marginal expectation of $\vec{\theta}$ is a zero vector. Thus,

$$\pi(\vec{\theta}) = MVN(0, \Lambda_0) \qquad \text{[IV-7]}$$

Equation [IV-2] implies that

$$\vec{y} \mid \vec{\theta} \sim N(\vec{\theta}, I) \qquad \text{[IV-8]}$$

From Equation [IV-7] and Equation [IV-8], we obtain (after some algebraic simplifications),

$$E(\vec{\theta} \mid \vec{y}) = \left(\Lambda_0^{-1} + I\right)^{-1} \vec{y} = W\vec{y} \qquad \text{[IV-9]}$$

which is linear in $\vec{y}$, as desired.

To further explore the properties of the estimator in Equation [IV-9], we conduct a numerical experiment. In the numerical results below, we set $T = 10, \sigma = 100,$ and explore two values for γ^2, γ^2 = 1 (more smoothing) and γ^2 = 10 (less smoothing). Note that in the actual implementation, the random walk variances (i.e., $\sigma_\zeta^2, \sigma_W^2, \sigma_G^2, \sigma_D^2$) are all sampled along with other parameters, and hence the degree of smoothing is also governed by the data.

Figure IV.1 plots the 1st (to estimate $\hat{\theta}_1$) and 5th row (to estimate $\hat{\theta}_5$) of *W*, for both values of γ^2. The figure shows that the estimator induced by a random walk prior in Equation [IV-9] mirrors that of polynomial smoothing; for example, when estimating $\hat{\theta}_1$, a (polynomially) decreasing function is applied to the y_t's based on the distance between *t* and 1 (see the upper left panel of the figure). The same holds for the estimation of $\hat{\theta}_5$ (see the upper right panel). Further, by comparing the lower panels with the upper panels, we see that the amount of smoothing is controlled by the value of γ^2; a higher γ^2 leads to less smoothing (more weight on the observation at *t*).

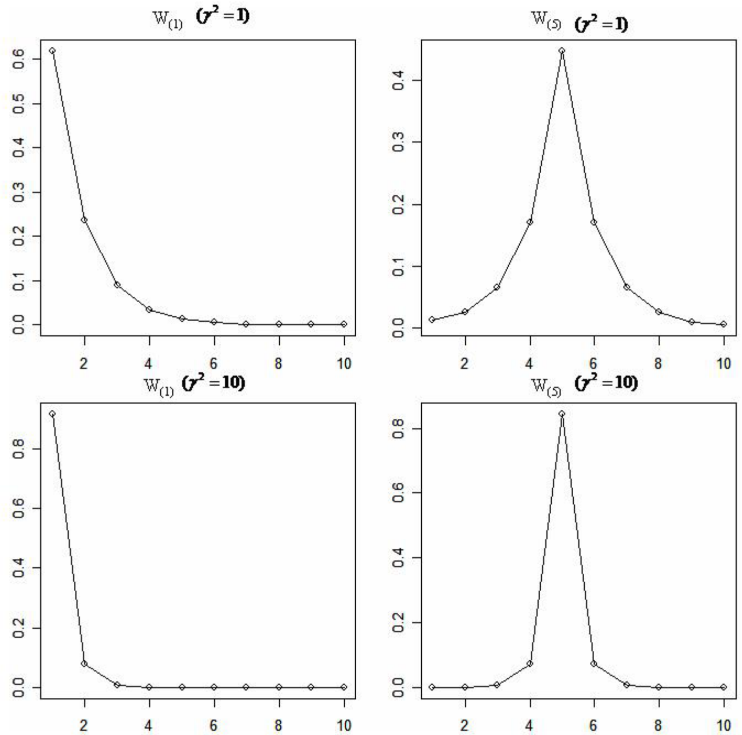


Figure IV.1. Numerical example of the estimator in Equation [IV-9] ($T = 10, \sigma = 100$).
Upper panel: γ^2=1; lower panel: γ^2=10.

The Gaussian random walk prior offers several statistical advantages. First, it allows for smooth variation in the behavior of coefficients over time without a need to pre-define a parametric form (thus, data rather than model assumptions drive inferences about the temporal evolution of coefficients). Second, it links coefficients at different time points together, allowing our estimation procedure to “borrow strength” across all data observations and thereby yield more accurate estimates. Third, it is a special case of a Bayesian dynamic linear model (West and Harrison 1997), and behaves as a conjugate prior when we draw coefficients. Consequently, posterior samples are drawn very efficiently using the Gibbs sampler.

V. MCMC procedure

As we discussed in the *Prior Specification and Smoothing* section, proper conjugate priors are specified for each of the parameters. That is, $\zeta_1, \beta_1^W, \beta_1^G, \beta_1^D \sim N(0, \sigma_1^2)$, $\vec{\tau} \sim N(0, \sigma_\tau^2 I)$, and the variance parameters are given $Inv-\chi^2(\kappa_0, \xi_0^2)$ conjugate priors. With these conjugate priors, the full conditional distribution for all parameters except λ_{it} are of standard forms. The Gibbs sampler (Casella and George 1992) is used to sample from them. Samples of λ_{it} are generated from a random walk Metropolis-Hastings algorithm (Hastings 1970).

In the discussion below, we outline the full conditional distributions for the other parameters. The following three expressions are used in this process.

(i) First, we consider a vector *y* of *n* i.i.d. observations from $N(\mu, \sigma^2)$ with known variance. Given the conjugate prior on μ ($\mu \sim N(\mu_0, \sigma_0^2)$), the conditional posterior distribution is:

$$\mu \mid y \sim N\left(\frac{\left(\sigma_0^2\right)^{-1}\mu_0 + \left(\sigma^2/n\right)^{-1}\bar{y}}{\left(\sigma_0^2\right)^{-1} + \left(\sigma^2/n\right)^{-1}}, \frac{1}{\left(\sigma_0^2\right)^{-1} + \left(\sigma^2/n\right)^{-1}} \right) \qquad \text{[V-1]}$$

(ii) Second, we consider a linear model of $y_i \mid \vec{\beta}, \sigma^2, \vec{x} \sim N(\vec{x}'\vec{\beta}, \sigma^2)$ with known variance. Given the conjugate prior on $\vec{\beta}$ ($\vec{\beta} \sim N(\beta_0, \sigma_\beta^2 I)$), the conditional posterior distribution is:

$$\vec{\beta} \mid y \sim N\left(\left(\left(\sigma^2\right)^{-1} X'X + \left(\sigma_\beta^2\right)^{-1} I\right)^{-1} \left(\left(\sigma^2\right)^{-1} X'y + \left(\sigma_\beta^2\right)^{-1}\beta_0\right), \left(\left(\sigma^2\right)^{-1} X'X + \left(\sigma_\beta^2\right)^{-1} I\right)^{-1} \right) \qquad \text{[V-2]}$$

(iii) Next, we consider a vector *y* of *n* i.i.d. observations from $N(\mu, \sigma^2)$ with known mean. Given a conjugate prior on σ^2 ($\sigma^2 \sim Inv-\chi^2(v_0, s_0^2)$), the conditional posterior distribution is:

$$\sigma^2 \mid y \sim Inv-\chi^2\left(v_0 + n, \frac{v_0 s_0^2 + ns^2}{v_0 + n}\right) \text{ where } s^2 = \frac{\sum (y_i - \mu_i)^2}{n} \qquad \text{[V-3]}$$

We now outline how we sample each individual model parameter.

Regional random effects, γ_i

Let $\phi_{it} = \log(\lambda_{it}) - \left(\log(n_{it}) + \zeta_t + \vec{x}_i'\vec{\tau} + \beta_t^W z_{it} + \beta_t^G G_{(i)}\vec{z}_t + \beta_t^D D_{(i)}\vec{z}_t\right)$. Then,

$\phi_{it} \sim N(\gamma_i, \sigma_\varepsilon^2)$. With the prior $\gamma_i \sim N(0, \sigma_\gamma^2)$, we apply equation [V-1].

Parameters of control variables $\vec{\tau}$

Let $\phi_{it} = \log(\lambda_{it}) - \left(\log(n_{it}) + \zeta_t + \gamma_i + \beta_t^W z_{it} + \beta_t^G G_{(i)}\vec{z}_t + \beta_t^D D_{(i)}\vec{z}_t\right)$. Then,

$\phi_{it} \sim N(\vec{x}_i'\vec{\tau}, \sigma_\varepsilon^2)$. With the prior $\vec{\tau} \sim N(0, \sigma_\tau^2 I)$, we apply equation [V-2].

Time-varying coefficients $\zeta_t, \beta_t^W, \beta_t^G, \beta_t^D$

Let $\phi_{it} = \log(\lambda_{it}) - \left(\log(n_{it}) + \gamma_i + \vec{x}_i'\vec{\tau}\right)$. Then, priors for $\vec{\zeta}$ are given below depending on time periods; those of $\beta_t^W, \beta_t^G, \beta_t^D$ are of the same form.

For $t = 1$, $$\zeta_t \sim N\left(\frac{\left(\sigma_\zeta^2\right)^{-1}\zeta_2}{\left(\sigma_1^2\right)^{-1} + \left(\sigma_\zeta^2\right)^{-1}}, \frac{1}{\left(\sigma_1^2\right)^{-1} + \left(\sigma_\zeta^2\right)^{-1}}\right)$$

For $1 < t < T$, $$\zeta_t \sim N\left(\frac{\zeta_{t-1} + \zeta_{t+1}}{2}, \frac{\sigma_\zeta^2}{2}\right) \qquad \text{[V-4]}$$

For $t = T$, $$\zeta_t \sim N\left(\frac{\zeta_{t-1}}{2}, \sigma_\zeta^2\right)$$

With these priors, we apply equation [V-2] to obtain posterior distributions.

Variance parameters (e.g., σ_ε^2)

Let $\phi_{it} = \log(\lambda_{it}) - \left(\log(n_{it}) + \gamma_i + \zeta_t + \vec{x}_i'\vec{\tau} + \beta_t^W z_{it} + \beta_t^G G_{(i)}\vec{z}_t + \beta_t^D D_{(i)}\vec{z}_t\right)$. Then,

$\phi_{it} \sim N(0, \sigma_\varepsilon^2)$. With the conjugate prior, $\sigma_\varepsilon^2 \sim Inv - \chi^2(\kappa_0, \xi_0^2)$, we can apply equation [V-3] to obtain posterior samples.

VI. Effects of Control Variables ($\vec{\tau}$)

The posterior means of the coefficients of the control variables are shown in Table VI.1. Here we simply note a few interesting observations that may warrant future studies. In general, Netgrocer.com has a higher rate of new buyers in zip codes that have higher population growth, more urban housing units, and higher levels of educational attainment. While new buyers are gained more rapidly in urban areas (e.g., Philadelphia and Pittsburgh) this is driven mostly by the larger population size, and *not* a higher underlying adoption rate. Since the overall number of new buyers is relatively low, and there is a large disparity between population sizes in the highly urban areas and more rural ones, it turns out that the adoption rate, i.e., the number of buyers relative to the population, is negatively correlated with population density. The adoption rate is negatively correlated with the density of general stores (e.g., Wal-Mart) and the presence of warehouse clubs, the two offline formats that compete directly with Netgrocer.com. It is positively related to the density of supermarkets, a complementary format (Netgrocer.com does not sell perishable products).

[Insert Table VI.1 about here]

Table VI.1 Posterior Means of the Coefficients of Control Variables and Variance Parameters

Variable	Description	Posterior Mean	Posterior Std Dev
Local Environment			
Population Density	Population density	-0.102	0.013
Population Growth	Annual population growth rate from 2000 to 2004	0.037	0.011
Home Value	% of homes valued at $250,000 or more	0.033	0.017
Urban Housing	% of houses with 50 units or more	0.138	0.010
Land Area	Area in square miles	-0.034	0.011
Household Characteristics			
Asian	% of Asians	-0.019	0.013
Black	% of Blacks	-0.155	0.016
White	% of Whites	-0.050	0.006
College	% with bachelors and/or graduate degree	0.360	0.019
Elderly	% aged 65 and above	-0.081	0.010
Wealthy	% of households earning $75,000+	-0.125	0.028
Access to Retail Services			
Density General	Density of general stores within the second order neighboring zip codes	-0.158	0.056
Density Supermarket	Density of supermarkets within the second order neighboring zip codes	0.256	0.056
Presence Warehouse	Presence of warehouse clubs within the second order neighboring zip codes	-0.042	0.016
Access to the Internet			
Broadband Access	Number of high-speed Internet service providers	0.026	0.004
Variances			
$\sigma_{\varepsilon}^{2} \times 10$	Variance of errors	2.065	0.108
$\sigma_{\gamma}^{2} \times 10$	Variance of regional random effects	1.159	0.070
$\sigma_{\zeta}^{2} \times 10^{2}$	Variance of baseline adoption	9.361	1.928
$\sigma_{W}^{2} \times 10^{4}$	Variance of within-region proximity effects	4.106	1.055
$\sigma_{G}^{2} \times 10^{4}$	Variance of across-region proximity effects	7.069	2.864
$\sigma_{D}^{2} \times 10^{4}$	Variance of across-region similarity effects	3.784	1.049

Notes
[1] The dependent variable is the number of new buyers in each zip code in each month.
[2] All the variables concerning the local environment, household characteristics, and access to retail services are cross-sectional and standardized, while the broadband access variable is time-varying and un-standardized.