WEB APPENDIX

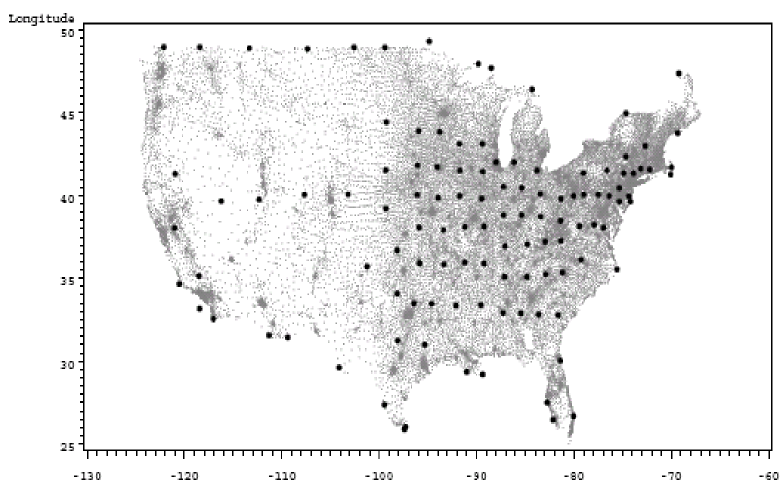
Bayesian Spatio-Temporal Analysis of Imitation Behavior Across New Buyers at an Online Grocery Retailer

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

Web Appendix A. 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.

Web Appendix B. 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.

Web Appendix C. 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:

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

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.

Web Appendix D. 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):

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

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 $\rho^{|x_i - x_j|}$, ρ < 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).

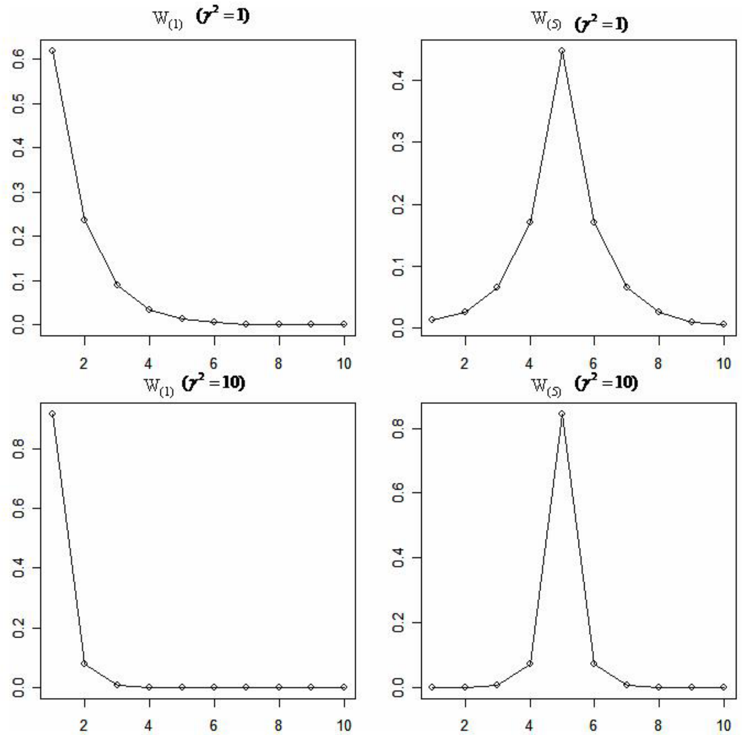


Figure IV.1. Numerical example of the estimator in Equation [IV-9] ($T = 10, \sigma = 100$).
Upper panel: γ^2=1; lower panel: γ^2=10.

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.

Web Appendix E. 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:

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

(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:

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

(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:

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

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.

Web Appendix F. Effects of Control Variables ($\vec{\tau}$)

The posterior means of the coefficients of the control variables are shown in Table WA-F. 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).

REFERENCES

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

Casella, George and Edward I. George (1992), "Explaining the Gibbs Sampler," *American Statistician*, 46 (3), 167–74.

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), 1285–88.

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

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

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

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

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

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

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.

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

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

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

Table WA-F. 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.