Management Science

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Traditional and IS-Enabled Customer Acquisition on the Internet

Jeonghye Choi, David R. Bell, Leonard M. Lodish,

To cite this article:
Jeonghye Choi, David R. Bell, Leonard M. Lodish, (2012) Traditional and IS-Enabled Customer Acquisition on the Internet. Management Science 58(4):754-769. http://dx.doi.org/10.1287/mnsc.1110.1447

MANAGEMENT SCIENCE
Vol. 58, No. 4, April 2012, pp. 754–769
ISSN 0025-1909 (print) | ISSN 1526-5501 (online)

http://dx.doi.org/10.1287/mnsc.1110.1447

Traditional and IS-Enabled Customer Acquisition on the Internet

Jeonghye Choi
Yonsei School of Business, Yonsei University, Seoul 120-749, Korea, jeonghye@yonsei.ac.kr

David R. Bell, Leonard M. Lodish
The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
{davidb@wharton.upenn.edu, lodish@wharton.upenn.edu}

Geographic variation in consumer use of Internet retailers is partly explained by variation in offline shopping costs. Explanations for geographic variation in the efficacy of different customer acquisition methods including traditional methods of offline word-of-mouth (WOM) and magazine advertising and information systems (IS)-enabled methods of online WOM and online search remain unexplored. We estimate a multivariate negative binomial distribution (NBD) model on zip code–level customer counts from a leading Internet retailer and provide new insights into factors explaining geographic variation in the success of these methods. First, we show that target customer density explains geographic variation over and above the impact due to the number of potential customers. Moreover, the effect of density is greatest for offline and online WOM acquisitions; this suggests that density contributes to contagion, connectivity, and a hypothesized "social multiplier." Second, when senders and recipients of WOM share consumption benefits, WOM is more powerful and compelling. We find that location-based convenience benefits have stronger effects on location-dependent offline WOM acquisitions than on location-independent online WOM acquisitions. Third, acquisition channels contribute differently to the total customer pool—offline WOM acquisitions are clustered, whereas magazine acquisitions are dispersed. Finally, separate click-to-conversion data from Coremetrics.com indicates that using the model-based predictions to target specific markets delivers a twofold improvement in actual click-to-order rates.

Key words: count model; Internet retailing; search; spatial analysis; word-of-mouth
History: Received July 14, 2009; accepted June 28, 2011, by Sandra Slaughter, information systems. Published online in *Articles in Advance* December 22, 2011.

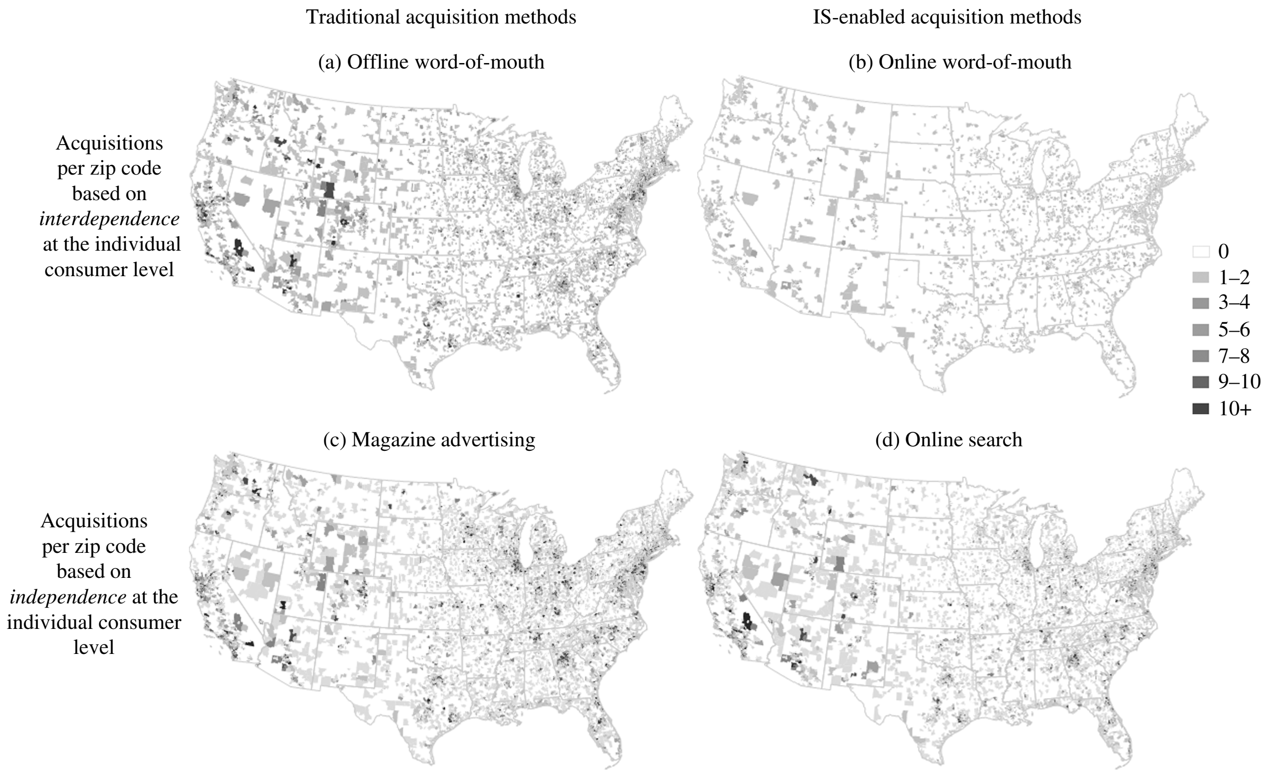
Introduction

Physical stores have relatively small trading areas (Fotheringham 1988, Huff 1964, Reilly 1931); this "downside" is, however, counterbalanced by the fact that customer acquisition efforts can be focused in a few neighborhoods. Internet retailers, on the other hand, have the "upside" of access to large geographic markets (Bell and Song 2007); however, this means it is unclear a priori which of many possible locations will yield the most online customers. Thus, the two institutional arrangements—traditional and online retailing—pose distinct and opposing advantages and disadvantages for sellers. At the same time, consumers deciding whether to shop offline or online also face contrasting cost–benefit trade-offs. Although they can easily discover and visit local offline stores, travel costs preclude inspection of too many geographically distant offline alternatives. Conversely, Internet retail alternatives for many consumer products are plentiful, but shoppers may not know how to initially "find" the site that best suits their needs.

We focus on managerially important research questions: which locations will generate the most online demand, by which acquisition methods, and why? To answer this question, we explain geographic variation in online demand in terms of variation in offline shopping costs, and in the propensity of Internet retail buyers to arrive through different acquisition channels. Recent studies (e.g., Anderson et al. 2010, Brynjolfsson et al. 2009, Choi and Bell 2011, Forman et al. 2009) showed that online retailer demand varies substantially across local markets as a function of the relative price, assortment, and convenience of local *offline* options. In other words, variation in offline shopping costs explains variation in online demand. Other research shows that proximity among target customers—which facilitates social influence—also plays a key role in buyer acquisition (Choi et al. 2010). In this paper we incorporate and build on these prior findings by showing how and why geographic variation in physical characteristics of local markets explains geographic variation in the number of new buyers acquired through four different modes—offline word-of-mouth (WOM), online WOM, online search, and magazine advertising.

The empirical analysis examines customer acquisitions at a leading U.S. Internet retailer, Childcorp

Figure 1 Geographic Variation in New Buyers per Acquisition Method



.com.[1] Childcorp.com focuses primarily on one product category—a ubiquitous category with individual items that are bulky, storable, and purchased and consumed repeatedly over time. The four methods used by Childcorp.com to attract customers differ conceptually and substantively. Figure 1 shows geographic variation in the number of new buyers by acquisition mode. The left two panels depict buyers arriving through "traditional" acquisition methods (offline WOM and magazine advertising) long used by marketers, and the right two panels show those arriving through new "information systems (IS)-enabled" acquisition methods (online WOM and online search). Figure 1 distinguishes acquisitions that are *interdependent* at the individual level (top panels), i.e., from WOM, from acquisitions that are *independent* at the individual level (bottom panels), i.e., from individual shoppers' online search and response to magazine advertising. Zip code–level variation in the number of target customers for Childcorp.com, i.e., the number of households with children under six years old, is shown in Figure 2. It is clear that this geographic variation in market potential alone will be insufficient to explain geographic variation in the success of different acquisition methods (Figure 1).

This research contributes three new findings. First, target customer density delivers online demand over and above that created through the total number of target customers alone. Moreover, target customer density induces significantly higher numbers of buyers acquired via *interdependent* methods versus *independent* acquisition methods at the same location. This finding is consistent with the notion that interdependent methods create a synergistic effect from positive social influence among buyers, i.e., a social multiplier. Second, when senders and recipients of WOM share consumption benefits, WOM is more powerful and compelling. We illustrate this principle by showing that location-based convenience benefits have stronger effects on location-dependent offline WOM acquisitions than on location-independent online WOM acquisitions. Third, there are systematic differences among the four acquisition modes in the way each contributes to the total customer base. WOM generates many geographically clustered buyers in a relatively small number of zip codes. Magazine advertising is more effective in generating geographically dispersed buyers over a large number of zip codes; online search contributes a relatively constant proportion of buyers, independent of location.[2]

[1] For reasons of confidentiality, we refer to this leading Internet retailer by the pseudonym "Childcorp.com." Acquisition information is collected during customer registration. We provide more details in Data and Measures.

[2] Because total acquisitions are decomposed by mode in the model, we can show how the local efficacy of each interacts with the characteristics of the local environment.

Figure 2 **Geographic Variation in Target Customers**

Note. Target customers for Childcorp.com are households with children aged less than six years old.

This paper is organized as follows. The next section summarizes key findings from the literature, and the two subsequent sections describe the data and the empirical model, respectively. The next reports the empirical findings and new implications for managers (including evidence for possible gains from geotargeting). This paper concludes with a summary of key findings and suggested directions for future research.

Background Literature

Prior studies analyze factors in the local offline environment that affect shoppers' trade-offs in deciding to shop online instead of offline; we briefly review some of these key drivers of online shopping as they relate to our research. Next, we explore the rationale for the conjectured positive effects of target customer density on both overall and WOM-induced demand, and the idea that shared benefits between senders and receivers will make WOM more effective.

Location-Based Drivers of Online Shopping

Customer benefits from using online retailers include lower prices (Anderson et al. 2010, Goolsbee 2000) and greater convenience (Brynjolfsson and Smith 2000, Forman et al. 2009). The difference between online sales tax rates (often zero) and offline sales tax rates create economic incentives for shopping online.[3] As evidence, Goolsbee (2000) finds that if Internet retail transactions were taxed at average offline rates (approximately 8%), online demand would decline by more than 20%. Similarly, Anderson et al. (2010) show that when an Internet retailer opens physical stores and collects sales tax in locations where it previously did not, Internet sales in those locations suffer. The importance of convenience is highlighted in a *Wall Street Journal* study (Gunn 2007) that evaluated several competing online sites (including Childcorp.com): "Getting an online discount doesn't matter much if you have to pinch-hit with pricier (products) from the grocery store while you wait for your order to arrive." For an Internet retailer, "convenience" is determined by (1) shipping time proximity of an Internet purchase to a customer ("time distance") and (2) physical travel distance from the customer's location to the nearest offline stores ("travel distance"). Internet retailers benefit when time distance is lower and travel distance is higher. Brynjolfsson and Smith (2000) find that some customers pay a premium for faster shipping (reduced time distance), and Forman et al. (2009) find that reduced travel distance to offline stores makes online retailers less attractive. Finally, the online demand potential at a particular location is also affected by demographics, access to the Internet, and so on (all such factors serve as controls in our empirical analysis).

[3] Because Internet retailers collect no sales tax in locations where they have no offline presence, most locations enjoy tax-free shopping from online retailers. The sensitivity of online sales to local offline sales tax rates mirrors cross-border shopping observed for traditional stores; consumers arbitrage tax rate disparities between online stores and local offline stores.

Target Customer Density, WOM, and the Social Multiplier

In economics, medicine, and sociology it is well known that the physical density of a target group has a positive effect on the areal spread of innovation, ideas, disease, and so forth (Choldin 1978; Fox et al. 1980; Glaeser et al. 1996, 2003). In retailing, high target customer density proxies for higher offline shopping costs for bulky products requiring transport and storage (e.g., Bell and Hilber 2006).[4] Thus, locations with greater target customer density should generate higher online demand in aggregate.

Physical proximity among target customers amplifies their propensity to communicate or observe each other's behavior (e.g., Yang and Allenby 2003); emulation among physically close customers has been reported for adoption of an Internet retailer (Bell and Song 2007, Choi et al. 2010). Thus, these studies imply that target customer density will have an additional effect on *offline* WOM acquisitions in particular beyond the general positive effect on acquisitions overall. A likely parallel effect on *online* WOM acquisitions has had recent support as well. Sinai and Waldfogel (2004) found that households in more densely populated urban areas are more likely to peruse the Internet for "content," which likely includes blogs and other sources of online WOM. Katona et al. (2011) found a persistent significant and positive effect of population location density on the propensity of individuals to join an online social network. These studies show that offline density correlates with online connectivity. Just as density creates opportunities for offline social contagion, density also facilitates online social contagion through online connectivity.

In summary, interdependency among target customers induced through density creates a synergistic effect, i.e., a social multiplier (Becker and Murphy 2000). This means that any factor generating positive social influence at the *individual* level delivers a larger demand coefficient in an aggregate model (Glaeser et al. 1996, 2003). Thus, we expect that the effect of target customer density on the count of buyers acquired via WOM (*interdependent* processes) will be significantly greater than the effect on counts of buyers acquired via online search and magazine advertising (*independent* processes).

Shared Benefits and the Effectiveness of WOM

In a classic study, Katz and Lazarsfeld (1955) found WOM is seven times as effective as magazine advertising and twice as effective as radio advertising. Recent research implies that the superiority of WOM over other acquisition methods holds for online retailers as well. Villanueva et al. (2008) found that buyers acquired via WOM have long-term equity twice that of buyers acquired by marketing.[5] To our knowledge, no prior study examines whether or not shared benefits among senders and recipients further enhance the effectiveness of WOM.

Research in sociology (e.g., Fernandez et al. 2000) implies that WOM is engendered by "benefit matching," i.e., when the recipient of a WOM recommendation experiences a positive fit with the information conveyed and the product or service recommended. Our empirical setting allows us to study a related idea, i.e., how shared benefits among senders and recipients can promote acquisitions through WOM. First, consider that acquisitions through *offline* WOM most likely involve co-located senders and recipients and that, on average, *online* WOM acquisitions will involve more geographically diffuse senders and recipients. Second, note that many of the costs and benefits of using an Internet retailer such as Childcorp .com are location dependent. Shipping time is an obvious location-based convenience benefit; access to offline retail stores reflects relative offline shopping costs and is largely location based. We therefore expect that these location based benefits of shopping convenience will have stronger effects on acquisitions through offline WOM that occurs among senders and recipients that are most likely co-located.[6]

Data and Measures

Zip Code–Level Cumulative Numbers of New Buyers at Childcorp.com

Childcorp.com is a pseudonym for a leading Internet retailer selling a large selection of brand name children's necessities that are distributed nationally through various offline stores (all supermarkets, discount stores, and warehouse clubs). The quality of items sold at Childcorp.com can be determined ex ante, i.e., the products possess few if any nondigital attributes (Lal and Sarvary 1999), and prices are comparable to those at Walmart. Shipping is free with orders over $49 (approximately 90% of orders are shipped free), and UPS ships from company warehouses located in both the eastern and western United States. Key to our study, when individual shoppers register at Childcorp.com they are asked, "How did you hear about our website?" Multiple responses are prevented through the use of a drop-down list, and all the answers are classified into the four mutually

[4] This is especially true when target customer density is correlated with higher population density (in our data, the correlation is 0.923). Furthermore, Steenburgh et al. (2003) also note that higher population density leads to higher inventory holding costs.

[5] The authors' definition of WOM is broad because it includes links from search engines and referrals from friends and colleagues. In our study, we distinguish between online search and offline WOM. More details are given in Data and Measures.

[6] It need not be the case that shared benefits are rooted in location. Information technology helps geographically dispersed individuals share information (Dellarocas 2003), and senders and recipients of WOM could exactly share other kinds of benefits. A recipient of an online WOM recommendation for a romantic comedy could share the same tastes as the sender; this would amplify the power of the WOM recommendation, independent of the location of the individuals involved.

Table 1 Numbers of New Buyers per Acquisition Mode per Zip Code

Acquisition process[a]	Mean	Standard deviation	Sum	Correlations[b] Offline WOM	Online WOM	Online search
Buyers from offline word-of-mouth	1.829	7.945	54,246	—		
Buyers from online word-of-mouth	0.347	1.132	10,300	0.786	—	
Buyers from online search	1.422	3.328	42,170	0.818	0.757	—
Buyers from magazine advertising	1.743	3.386	51,681	0.686	0.685	0.802
Total buyers	5.342	14.501	158,397			

[a]Zip code penetrations by acquisition mode are as follows: buyers from offline word-of-mouth (11,689 zip codes), buyers from online word-of-mouth (5,716 zip codes), buyers from online search (12,261 zip codes), buyers from magazine advertising (13,978 zip codes). There are 29,652 residential zip codes in the database; 18,244 of these zip codes (about 62%) have at least one buyer.
[b]All the correlations are significantly different from zero ($p < 0.01$).

exclusive and collectively exhaustive categories introduced earlier—offline WOM, online WOM, online search, and magazine advertising.[7]

Offline WOM includes personal referrals from friends, colleagues, or acquaintances and accidental referrals from unacquainted people in local regions. *Online WOM* includes referrals through online message boards, blogs, and online communities. *Online search* includes paid and organic keyword search from search engines and connections from sponsored price comparison sites. *Magazine advertising* includes ads in an affiliated magazine targeted at the customer group.

We model zip code–level counts of new buyers acquired through each of the four processes, from the inception of Childcorp.com in January 2005 through March 2008. Table 1 presents corresponding summary statistics. The coefficient of variation is higher for the WOM processes than for online search and magazine advertising, suggesting that WOM acquisitions are relatively more "concentrated" (some visual evidence is also seen in Figure 1).[8] The high geographic correlations underscore the need to control for regional baseline effects by each mode as well as their intercorrelation.

The calibration data set is created by using the zip code indicator to match Childcorp.com data with four other data sources: (1) the 2000 U.S. Census, (2) UPS shipping times, (3) local sales tax rate schedules, and (4) the 2007 U.S. Census of Business and Industry. Table 2 provides a description and summary statistics for all model variables, which are elaborated on in more detail below. Working at the zip code level is both practical in terms of data requirements (detailed individual-level information is not collected and unavailable) and managerially useful as many retailers collect sales information at the zip code level (see, for example, Steenburgh et al. 2003).[9] During the data period, Childcorp.com did not engage in locally targeted marketing.[10]

"Pick One" vs. "Pick Any" Data. Shoppers chose one (and only one) of the four acquisition modes during registration at Childcorp.com. This practice is consistent with that used by many other Internet retailers and confers both advantages and challenges for empirical analysis. Asking customers to assign weights to multiple sources of influence is taxing and potentially increases nonresponse rates; allowing customers to "pick any" modes that were relevant yields 15 ($2^4 - 1$) possible response combinations. These "pick any" data could be analyzed with a multinomial choice model, but absent information on the individual-level weights for each mode, the

[7] About 70% of all buyers answered this question. The ordering behavior of the remaining 30% who make up the nonrespondent group does not differ significantly from that of the respondent group. Specifically, total spending averages approximately $250 for the respondent group and $247 for the nonrespondent group ($p > 0.10$). The average amount spent by buyers in the respondent group does however differ significantly across the four acquisition modes as follows: $315 (offline WOM), $176 (online WOM), $240 (search), and $204 (magazine advertising). All p-values for the pairwise differences are < 0.05. Thus, we believe the data are relatively free of nonresponse bias and, moreover, that there is no reason to believe individuals systematically distort their self-reported acquisition mode. Finally, the model specification error at the zip code level helps to account for "imperfect memory" of individual consumers as well as the possibility that an individual was influenced in multiple ways (see the Empirical Model section for a detailed discussion). We are very grateful to an anonymous reviewer for suggesting these checks.

[8] To examine this more formally, we compute the Getis–Ord G^* statistic (Getis and Ord 1992) for each process. G^* statistics are higher for both types of WOM acquisitions than they are for online search and magazine advertising, supporting the observation that WOM acquisitions are more locally concentrated.

[9] Numerous common data sources including the United States census capture geographic variation at the zip code level; furthermore, Childcorp.com faces local offline competitors in nearly all zip codes in the United States.

[10] There was no locally targeted spending for online search. Magazine subscriptions and circulation are exogenous to the firm's control; our model controls for magazine advertising exposure using circulation information.

Table 2 Summary Statistics for Model Covariates

Variable	Mean	Standard deviation	Min	Max
Target customer density				
Density of Households with Children ≤ 6 *Years Old*	65.890	256.242	0	7,398.909
Convenience benefit: *Time Distance*				
Shipping Days to Zip Code	2.624	0.967	1	4
Convenience benefit: *Travel Distance*				
Distance to Nearest Supermarket	4.064	4.366	0	65.029
Distance to Nearest Discount Store	13.105	13.921	0	180.283
Distance to Nearest Warehouse Club	31.760	33.316	0.044	332.644
Control variables				
Online price benefit				
No Tax = 1 *if No Tax is Levied in Zip Code*	0.171	0.377	0	1
Local Sales Tax Rate (%)[a]	6.655	1.186	2.900	9.750
Magazine circulations				
Magazine Circulations (in thousands)[b]	33.214	35.382	3.272	195.867
High-speed Internet access				
High-Speed Internet Connections[c]	2.733	0.899	0	5
Geodemographic characteristics				
Number of Households with Children ≤ 6 *Years Old*	562.525	850.750	0	9,705
Growth Rate in Number of Households (2000–2004)	0.013	0.018	−0.126	0.337
Percentage Population Aged 20 to 39 Years Old	0.258	0.068	0	0.868
Percentage Households with Working Female	0.032	0.051	0	1
Percentage of Whites	0.850	0.198	0	1
Percentage of Blacks	0.076	0.157	0	0.985
Percentage with College Education	0.452	0.163	0	1
Percentage Households Earning $50,000–$75,000	0.188	0.059	0	1
Percentage Households Earning $75,000–$150,000	0.188	0.059	0	1
Percentage Households Earning $150,000 *or more*	0.142	0.093	0	1

[a]Summary statistics for the local sales tax rate are computed across 24,573 residential zip codes that have local sales taxes on Childcorp.com products.
[b]Summary statistics for the magazine circulations are computed across 48 contiguous states.
[c]The high-speed Internet connections are coded from 0 to 5 depending on penetration rates. Its summary statistics are computed across 3,089 counties.

analysis is problematic. A shopper with unobserved weights of 80–20 on search and offline WOM would be counted in the {search, offline WOM} category, but so would a shopper with 20–80 weight on these two modes. Alternatively, a "pick one" approach gets it approximately right, assigning one customer to the search count and one to the offline WOM count. Conversely, if the weights are more evenly distributed over modes, e.g., 60–40 on search and online WOM, then the "pick any" approach might work better.

Focal Variables: Target Customer Density and Location-Based Convenience Benefits

Target customers for Childcorp.com are households with children aged less than six years old; hence, target customer density is the number of these households per square mile in each zip code. We are interested in the main effect of target customer density via offline shopping costs and in the secondary effect via a social multiplier, namely, that customer density is a facilitator of contact and observation and thereby a factor contributing to additional acquisitions through *offline* WOM; furthermore, because density is correlated with individuals' connectivity over and above that explained by their social networks alone (e.g., Katona et al. 2011) and their use of content on the Internet (Sinai and Waldfogel 2004), it should drive *online* WOM acquisitions as well.

We follow Brynjolfsson and Smith (2000) and measure the time convenience benefit through exogenously determined shipping times between buyers' zip codes and Childcorp.com warehouses (shoppers learn "days to ship" when they place orders). We also follow prior literature (e.g., Bell and Song 2007, Forman et al. 2009) for our measures of travel convenience benefit. We use eight-digit North American Industry Classification System (NAICS) codes to obtain location information on three major local offline competitors—supermarkets, discount stores (Walmart and Target), and warehouse clubs—and calculate the expected travel distance from each zip code to the nearest store of each format.[11] Convenience

[11] Although six-digit NAICS codes are often used in research, greater accuracy is achieved with our approach. For example, six-digit NAICS codes for supermarkets include candy stores and other

based on time distance and travel distance are both location-based benefits that co-located senders and recipients of WOM will share; hence, we expected they will have more pronounced effects on acquisitions through *offline* WOM than through online WOM.

Control Variables and Spatial Clustering of Zip Codes

Online Price Benefit. Childcorp.com prices are the same in every zip code, but the relative online price advantage varies across zip codes with variation in offline sales tax rates. We cannot measure price levels at all of Childcorp.com's competitors, but we can proxy for the price benefit of shopping online for particular products by using tax rates (see, for example, Anderson et al. 2010, Choi and Bell 2011, Goolsbee 2000). Zip code–level sales tax rates were compiled from public information from the Department of Revenue in each state. We called over 1,000 randomly selected stores in an exhaustive manual check to verify the tax status of Childcorp.com products because local areas may have tax rates that are different from those in their states.

Magazine Circulations. Because we model geographic variation in the number of customers acquired via magazine advertising, we must control for observed heterogeneity in magazine circulation. To do so, we collected data on magazine circulations for the key magazine used by Childcorp.com, by state and for six months ending on June 30, 2009 (the data include both paid subscriptions and single copy sales), although for reasons of confidentiality we were unable to secure zip code–level circulation data. Our model also controls for spatial variation via model random effects and the specification error (see Empirical Model).

High-Speed Internet Access. The Federal Communications Commission collects, by location, Internet access services faster than 200 kbps in at least one direction between ISPs (Internet service providers) and households. Connections are coded from zero to five depending on the penetration: 0 for 0% and 1–5 for each 20% incremental range. We use such data collected as of June 30, 2009, for individual counties in the United States to proxy for geographic variation in Internet penetration.

smaller retail formats that differ from what is typically thought of as a supermarket. These NAICS codes have exact correspondence with SIC codes. The physical distance a shopper must travel to an offline store parallels transportation costs in spatial differentiation models (e.g., Balasubramanian 1998, Bhatnagar and Ratchford 2004).

Geodemographic Characteristics.[12] Potential market size in a zip code is measured by the number of households with children less than six years of age and serves as an offset variable in the multivariate negative binomial distribution (NBD) model (Agresti 2002, Greene 2008). Standard zip code–level control variables that are expected to affect online demand include measures of age, income, ethnicity, and education. Following Dhar and Hoch (1997), these variables are expressed as percentages and skewed away from simple averages to generate more geographic variation, e.g., we use "percentage of households with a college degree" rather than "average years in school."

Spatial Clustering of Zip Codes. The U.S. Census Bureau groups zip codes into metropolitan statistical areas (MSAs) and micropolitan statistical areas (μSAs) on the basis of strong social and economic ties.[13] Zip codes in the same MSA or μSA share average characteristics, so we define regional clusters of zip codes using these designations; zip codes that do not belong to MSAs or μSAs are grouped by states (there are 358 MSAs and 567 μSAs in the 48 contiguous states). In the model, regional-cluster random effects efficiently capture the difference in baseline acquisition rates across regional clusters.

Empirical Model

Zip code–level buyer acquisition numbers are nonnegative integers, so we model them in a Poisson framework. We assume that $y_{k,\,z(m)}$, the number of new buyers acquired by process k in zip code z in regional cluster m, is Poisson distributed:

$$y_{k,\,z(m)} \sim \text{Poisson}(\lambda_{k,\,z(m)}), \tag{1}$$

where $k =$ offline WOM, online WOM, online search, and magazine advertising. We justify our modeling choice on both theoretical and empirical grounds. First, the Poisson is widely applied in spatial models when the occurrence of an event is rare in comparison with the target population (Wikle and Hooten 2006, Knorr-Held and Besag 1998), as is the case here. Second, in Online Appendix I

[12] There is no significant multicollinearity among these variables. The largest pairwise correlation is 0.48, and most pairwise correlations are less than 0.30. Also, the largest variance inflation factor (VIF) is 4.06 in the regression model of count data in log form. We thank an anonymous reviewer for suggesting this check.

[13] MSAs are formed around a central urbanized area, i.e., a contiguous area of relatively high population density, and surrounding areas that have "strong ties" (as measured by commuting and employment) to the central area. Likewise, μSAs consist of adjacent areas that have at least one urban cluster. This spatial demarcation is more comprehensive than one based on geographical boundaries alone. Delaware Valley, for example, is a metropolitan area comprising several counties in Delaware, Maryland, New Jersey, and Pennsylvania.

(available at http://marketing.wharton.upenn.edu/people/faculty.cfm?id=227), we outline a mathematical argument (adapted from Berry 1994 and Blum and Goldfarb 2006) that the Poisson approximation for zip code–level counts can be motivated from individual-level utility maximizing choices between an online retailer and an outside offline option. Third, the Poisson model is flexible enough to accommodate the geographic variation in baseline acquisition rates, correlations among the four acquisition processes, and specification error in the dependent variable (details follow below). The inclusion of cross-sectional heterogeneity in the Poisson model leads naturally to the negative binomial model (see Equation (AI.5) and Online Appendix I). In the next section we show that our proposed model provides an excellent fit to the data and very good predictive accuracy in holdout samples.

The rate parameter $\lambda_{k,z(m)}$ is modeled as a function of (1) target customer density, (2) location-based benefits, (3) the number of target customers, (4) a set of control variables that capture observed heterogeneity, (5) unobserved baseline by regional cluster, and (6) zip code–level measurement error:

$$\log(\lambda_{k,z(m)}) = x'_{k,z(m)}\beta_k + \varepsilon_{k,z(m)} \quad \text{and} \tag{2}$$

$$\begin{aligned} x'_{k,z(m)}\beta_k = {} & \varphi_k \cdot \mathit{Target\ Customer\ Density}_{z(m)} \\ & + \Delta_k \cdot \mathit{Location\text{-}Based\ Benefits}_{z(m)} \\ & + \log(n_{z(m)}) + \Psi_k \cdot \mathit{Controls}_{z(m)} \\ & + \alpha_{k,0} + \alpha_{k,m}, \end{aligned} \tag{3}$$

where φ_k is a scalar parameter that varies by acquisition mode k. A vector of six dummies for one-, two-, and three-day shipping on the East Coast and West Coast, relative to the four-day benchmark, plus expected travel distance to three different types of offline stores, is captured by *Location-Based Benefits*$_{z(m)}$ and Δ_k is the corresponding parameter vector.[14] The number of target customers, $n_{z(m)}$, enters the model in log form to serve as an offset variable (Agresti 2002, Greene 2008).[15] A vector containing all the other measures for observed heterogeneity summarized in Table 2 (e.g., offline sales tax rates, Internet penetration, magazine circulations, etc.) is captured by *Controls*$_{z(m)}$ and Ψ_k is the corresponding parameter vector.

The geographic variation in the raw data (Figure 1 and Table 1) dictates that we control for unobserved heterogeneity in the regional baselines by acquisition mode. Hence, the baseline for regional cluster m consists of the overall baseline, $\alpha_{k,0}$, and the random deviation of regional cluster m from the overall baseline, $\alpha_{k,m}$. Because all four demand processes emerge from the same regional cluster m, the four random effects follow a multivariate normal distribution (MVN) (Gueorguieva 2001, Thum 1997):

$$\begin{pmatrix} \alpha_{\text{offlineWOM},m} \\ \alpha_{\text{onlineWOM},m} \\ \alpha_{\text{Search},m} \\ \alpha_{\text{Magazine},m} \end{pmatrix} \sim \text{i.i.d. MVN}\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \tau_1^2 & r_{21}\tau_2\tau_1 & r_{31}\tau_3\tau_1 & r_{41}\tau_4\tau_1 \\ r_{21}\tau_2\tau_1 & \tau_2^2 & r_{32}\tau_3\tau_2 & r_{42}\tau_4\tau_2 \\ r_{31}\tau_3\tau_1 & r_{32}\tau_3\tau_2 & \tau_3^2 & r_{43}\tau_4\tau_3 \\ r_{41}\tau_4\tau_1 & r_{42}\tau_4\tau_2 & r_{43}\tau_4\tau_3 & \tau_4^2 \end{pmatrix} \right). \tag{4}$$

Our multivariate random effects approach delivers several estimation and interpretation benefits: (1) the four acquisition modes are modeled simultaneously and accommodate a variety of nested cases; (2) the four acquisition modes are modeled as a function of the same variables, and the parameters are jointly estimated, so direct comparison of the separate effects of one specific variable, e.g., target customer density, across modes is straightforward; and (3) the multivariate model offers good control over the Type I error rates in multiple tests and generates more efficient parameter estimates.

Our model also accounts for the possibility that the mode-specific numbers of new buyers per zip code could be an imperfect reflection of the true acquisition process at the individual level. Some buyers could, for example, fail to indicate their true acquisition modes because of imperfect memory or be exposed to multiple sources of influence but answer with the one mode that is most salient or most relevant as they are forced to respond in a "pick one"

[14] From January 2005 through December 2005, orders were shipped from one warehouse on the East Coast. From January 2006 onward, orders shipped from two warehouses, one on each coast. Under the two-warehouse regime, orders ship from whichever warehouse is closer to the zip code receiving the order, and zip codes along the West Coast saw improvements in shipping times from five to six days to one to three days.

[15] The parameter for the offset variable is constrained to one, which allows the numbers of new buyers per each acquisition mode to be interpreted as the rate relative to the number of target buyers (i.e., the number of new buyers divided by the number of households with children). Using the number of target customers as an offset variable is justified in two ways. First, this approach is standard when the number of buyers is very small compared to the size of the potential customers (as is the case in our data, see Tables 1 and 2). Second, the offset can be derived mathematically from individual-level utility maximization decisions made by these same households residing in a common zip code (see Online Appendix I). The natural log form for the offset variable is also justified as the canonical link for the Poisson distribution.

Table 3 Model Fit Comparisons of Proposed Model and Nested Models

Model	Specification	Log-likelihood	Mean absolute error[a]	
			In sample	Out of sample
Proposed model	NBD model with multivariate random effects	−115,800	0.810	0.821
Nested model 1	NBD model with univariate random effects ($r_{kk'}=0$ for all k and k', $k \neq k'$ in Equation (4))	−115,999	0.848	0.871
Nested model 2	NBD model with no random effects ($a_{k,m}=0$ for all k and m in Equation (3))	−117,218	0.911	0.917
Nested model 3	NBD model with no random effects ($a_{k,m}=0$), holding the parameter vector for control variables (Ψ_k) constant across four modes (k's) in Equation (3)	−118,296	0.931	0.934
Nested model 4	NBD with model no random effects ($a_{k,m}=0$), holding all parameters (φ_k, Γ_k, Δ_k, and Ψ_k) constant across four modes (k's) in Equation (3)	−118,785	0.945	0.952

[a]We conduct holdout tests by performing 10-fold cross validation on each partition of the estimation and validation data sets (Breiman and Spector 1992, Kim et al. 2005). Estimation and validation data sets include 26,687 and 2,965 residential zip codes, respectively.

format. These potential measurement errors average over consumers within a zip code, and we account for this and additional specification error in the *zip code–level* dependent variable by the disturbance term $\varepsilon_{k,z(m)}$.[16] We assume that $\exp(\varepsilon_{k,z(m)})$ is independently and identically Gamma distributed with shape and scale parameter, θ_k (equal scale and shape parameters are needed for identification; see Cameron and Trivedi 1986, Greene 2008), so that the density for $y_{k,z(m)}$ after integrating out over $\exp(\varepsilon_{k,z(m)})$ becomes one form of the NBD with mean $\mu_{k,z(m)}$ and variance $\mu_{k,z(m)}(1+\theta_k^{-1}\mu_{k,z(m)})$, and is given by

$$f(y_{k,z(m)}|x_{k,z(m)}) = \frac{\Gamma(\theta_k + y_{k,z(m)})}{\Gamma(y_{k,z(m)}+1)\Gamma(\theta_k)} r_{k,z(m)}^{y_{k,z(m)}} (1-r_{k,z(m)})^{\theta_k}, \quad (5)$$

where $\mu_{k,z(m)} = \exp(x'_{k,z(m)}\beta_{k,z(m)})$ and $r_{k,z(m)} = (\mu_{k,z(m)})/(\mu_{k,z(m)}+\theta_k)$ (see Equation (AI.7) in Online Appendix I for the derivation). The specification error $\varepsilon_{k,z(m)}$ also allows the variance of the dependent data to be larger than the mean, and a test of the Poisson assumption is given by $\theta_k^{-1}=0$.

Equation (5) has a closed form up to the random effects, so the likelihood is evaluated via numerical integration over the random effects. Computational demands increase with the dimensionality of the random effects, so we follow Fieuws and Verbeke (2006) and Fieuws et al. (2006) and fit all pairwise bivariate models separately. We then calculate the parameter estimates and their sampling variation for the full multivariate model (see Online Appendix II, available at http://marketing.wharton.upenn.edu/people/faculty.cfm?id=227) and obtain the multivariate model likelihood through Monte Carlo sampling.

[16] We thank an anonymous reviewer for the following observation—if a zip code contains a reasonable number of new customers and potential customers, then individual-level imperfect memory, if present, will "average out" so that the recorded counts will reliably reflect mode and geographic variation in actual counts. (The average zip code has five customers and approximately 563 potential customers, and the average MSA has 391 customers and 39,134 potential customers.)

Empirical Findings

Model Fit, Validation, and Spatial Autocorrelation Test

Model fits and validation results for the multivariate NBD model and the four nested models are given in Table 3. The multivariate model has the largest log-likelihood, but to ensure that it is not overfitting we conduct predictive validation using holdout tests. The data are cross-sectional with no natural ordering, so we perform 10-fold cross validation on each combination of the estimation and validation data sets (Breiman and Spector 1992, Kim et al. 2005). As shown in Table 3 the multivariate model has the smallest mean absolute error in the estimation and validation data sets. To check that there is no remaining spatial autocorrelation in the residuals of the multivariate model, we compute Moran's I statistics using a spatial weighting matrix based on an exponential distance decay function (Moran 1950).[17] The Moran's I values are very small and statistically insignificant, which indicates that conditional upon the observed covariates and control for unobserved heterogeneity, there is no remaining unaccounted for spatial autocorrelation.

Target Customer Density and WOM Acquisitions

Table 4 reports the estimation results from the multivariate NBD model. Note that the parameter estimates for a single covariate are directly comparable

[17] The pairwise weight between zip code i and zip code j is an exponential function of the inverse distance in miles, d_{ij}, and equal to $\exp(-\Delta d_{ij})$. We further assume Δ is one. The latter assumption is made for computational tractability and consistency with prior work (e.g., Claude 2002, LeSage and Pace 2005, Yang and Allenby 2003).

Table 4 Parameter Estimates from the Multivariate NBD Model

	Multivariate NBD Model[a]									
	Offline WOM		Online WOM		Online search		Magazine ads		Total buyers[b]	
Variable	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE	Estimate	SE
Target customer density										
φ, *Density, HH with Children Aged* $\leq$ *6 Yrs*	0.071*	0.010	0.064*	0.007	0.034*	0.009	0.033*	0.006	0.048*	0.005
Convenience benefit: *Time Distance*										
Δ_1, *One-Day Shipping, Eastern US*	1.189*	0.126	0.733*	0.194	0.853*	0.105	0.746*	0.102	0.889*	0.062
Δ_2, *Two-Day Shipping, Eastern US*	0.555*	0.084	0.308*	0.103	0.377*	0.061	0.411*	0.058	0.431*	0.045
Δ_3, *Three-Day Shipping, Eastern US*	0.326*	0.059	0.218*	0.063	0.254*	0.046	0.297*	0.042	0.290*	0.039
Δ_4, *One-Day Shipping, Western US*	0.662*	0.164	0.442*	0.135	0.451*	0.112	0.285*	0.099	0.460*	0.092
Δ_5, *Two-Day Shipping, Western US*	0.285*	0.081	0.136+	0.073	0.202*	0.060	0.035	0.054	0.150*	0.055
Δ_6, *Three-Day Shipping, Western US*	0.026	0.094	−0.138	0.100	−0.014	0.068	−0.046	0.055	−0.095	0.061
Convenience benefit: *Travel Distance*										
Δ_7, *Distance to Nearest Supermarket*	−0.076*	0.020	−0.044	0.033	−0.061*	0.017	−0.061*	0.015	−0.074*	0.011
Δ_8, *Distance to Nearest Discount Store*	0.268*	0.030	0.179*	0.029	0.231*	0.019	0.192*	0.019	0.230*	0.012
Δ_9, *Distance to Nearest Warehouse Club*	0.126*	0.021	0.073*	0.031	0.060*	0.017	0.143*	0.017	0.098*	0.013
Control variables										
α_0, *Model Intercept*	−7.045*	0.209	−8.328*	0.184	−6.832*	0.147	−6.288*	0.117	−5.449*	0.087
Online price benefit										
Ψ_1, *No Tax Dummy*	0.150	0.199	0.141	0.157	0.186	0.137	0.048	0.110	0.126	0.086
Ψ_2, *Local Sales Tax Rate* (%)	0.048*	0.023	0.043*	0.021	0.042*	0.019	0.016	0.015	0.026*	0.012
Magazine circulations										
Ψ_3, *Magazine Circulations*	0.058	0.046	0.013	0.036	0.060*	0.026	0.054*	0.024	0.046*	0.017
High-speed Internet access										
Ψ_4, *High-Speed Internet Connections*	0.005	0.046	−0.002	0.071	0.017	0.032	−0.015	0.032	0.009	0.010
Geodemographic characteristics										
Ψ_5, *Growth Rate in Number of HH*	0.181*	0.026	0.145*	0.031	0.188*	0.021	0.194*	0.019	0.200*	0.006
Ψ_6, *Percent Population Aged 20 to 39 Years*	0.155*	0.030	0.164*	0.032	0.101*	0.026	0.059*	0.020	0.097*	0.009
Ψ_7, *Percent HH with Working Female*	0.010	0.041	0.005	0.045	−0.027	0.024	0.018	0.021	0.001	0.013
Ψ_8, *Percent with College Education*	0.596*	0.037	0.494*	0.050	0.478*	0.030	0.359*	0.025	0.458*	0.012
Ψ_9, *Percent of Whites*	0.339*	0.060	0.276*	0.056	0.236*	0.034	0.321*	0.040	0.245*	0.017
Ψ_{10}, *Percent of Blacks*	0.087	0.058	0.050	0.043	0.069*	0.034	0.060+	0.036	0.031*	0.014
Ψ_{11}, *Percent HH Earning* \$50K–\$75K	−0.029	0.030	0.016	0.030	0.028	0.020	0.048*	0.021	0.008	0.011
Ψ_{12}, *Percent HH Earning* \$75K–\$150K	−0.149*	0.035	−0.129*	0.040	−0.167*	0.028	−0.079*	0.023	−0.114*	0.013
Ψ_{13}, *Percent HH Earning* \$150K or more	0.078*	0.015	0.065*	0.017	0.008	0.010	0.025*	0.012	0.059*	0.009
Variances										
τ	0.378*	0.028	0.258*	0.027	0.257*	0.021	0.218*	0.017	0.323*	0.017
θ	2.481*	0.119	2.935*	0.216	4.715*	0.244	5.406*	0.285	2.737*	0.042
r_{21} (Online WOM, Offline WOM)							0.986*	0.039		
r_{31} (Online search, Offline WOM)							0.959*	0.018		
r_{32} (Online search, Online WOM)							0.963*	0.011		
r_{41} (Magazine ads, Offline WOM)							0.787*	0.215		
r_{42} (Magazine ads, Online WOM)							0.707*	0.244		
r_{43} (Magazine ads, Online search)							0.818*	0.073		

Note. For each estimate, we test the null hypothesis that the parameter is equal to zero.
[a]The dependent variable is the number of new buyers acquired through each process in each zip code, and all the variables except those for local sales tax and time distance are standardized (see Equations (1)–(4)).
[b]The dependent variable is the total number of new buyers aggregated over the four processes in each zip code, and all the variables except those for local sales tax and time distance are standardized.
$^*p < 0.05$; $^+p < 0.10$.

across the four outcome variables. The final column of Table 4 reports the estimates from a model in which the dependent variable is the *total* buyer count per zip code, $y_{z(m)} = \sum_k y_{k,z(m)}$, i.e., no distinction is made as to the acquisition mode.

Target customer density has the expected positive and significant effect on total new buyer acquisitions ($\varphi = 0.048$, $p < 0.05$) and on all four acquisition modes individually—this is consistent with the conjecture that density is a proxy for offline shopping costs. Furthermore, the mode-specific estimates of φ_k show differences. The largest incremental effects are on *interdependent* acquisitions via WOM compared to independent acquisitions via online search and advertising. Estimates for offline WOM and online WOM are not different from each other, but both are significantly greater than the estimates for search and advertising ($p < 0.01$), which are not different from each other. As argued previously, the larger estimates for WOM acquisitions are consistent with the social multiplier effect: the offline presence of positive social contagion is enabled by physical proximity among target customers (Yang and Allenby 2003), and online connectivity is positively correlated with physical population density (Katona et al. 2011, Sinai and Waldfogel 2004).

Quantitative effects of target customer density show important implications for geotargeting—the firm cannot affect density but it can use readily available secondary data to identify locations with dense populations of target customers. If we select 100 zip codes that have values of all the model variables in Table 2 at their means, this yields 121.8 expected new buyers in total. The expected total breaks down into 32.2 offline WOM buyers, 7.3 online WOM buyers, 34.4 search buyers, and 47.9 magazine buyers. Increasing customer density by one standard deviation brings 5.7 additional buyers: 2.4 from offline WOM, 0.5 from online WOM, 1.2 from search, and 1.6 from magazine advertising. In other words, WOM buyers account for approximately one-fourth of the pool of buyers in average markets, but they account for half of the lift that comes from a change in target customer density. As we show later in Figures 3 and 4 this density effect indicates that as the firm penetrates locations with a higher target density, WOM will be the most effective acquisition mode. Conversely, as the firm penetrates into rather sparse areas of lower target density, online search and magazine advertising acquisitions will be more effective.

Location-Based Convenience Benefits and Offline vs. Online WOM Acquisitions

Location-based convenience benefits in our study are measured by time distance and travel distance. Fast shipping has an obvious positive and significant effect on online demand. Each of one-, two-, and three-day shipping speeds produces statistically significantly more customers than their corresponding slower shipping speeds (i.e., $\Delta_1 > \Delta_2 > \Delta_3$ on the East Coast and $\Delta_4 > \Delta_5 > \Delta_6 = 0$ on the West Coast), and this rank ordering is preserved in all customer acquisition modes. Of more substantive interest is the fact that fast shipping—a key location-based benefit—is more effective in generating new buyers through *offline* WOM, where senders and recipients of WOM are likely to share locations, than through *online* WOM, where this is less likely. Table 4 shows that Δ_1(offline WOM) is greater than Δ_1(online WOM) and this pattern repeats for $\Delta_2 - \Delta_6$. The difference is statistically significant for $\Delta_1 - \Delta_5$ ($p < 0.01$), and the Δ_6 estimates are not different from zero. Thus, although the same benefit—fast shipping—could be part of both offline and online WOM conversations, it is significantly more powerful when senders and recipients are more likely to be physically co-located.

Co-located senders and recipients of WOM have the same access to offline stores, and Table 4 indicates that offline WOM acquisitions (where senders and recipients are more likely co-located) are indeed more sensitive to offline travel distance, the second location-based convenience benefit. Table 4 shows that for travel distance to discount stores Δ_8(offline WOM) is greater than Δ_8(online WOM), and this difference is significant ($p < 0.01$). The same is true for travel distance to warehouse clubs as Δ_9(offline WOM) is greater than Δ_9(online WOM); again this is significant ($p < 0.05$). For discount stores and warehouse clubs, the coefficients have intuitive positive signs—the greater the expected distance a shopper in a given location must travel to an offline store, the greater the online demand.

Somewhat less initially intuitive are the negative estimates for Δ_9 "distance to the nearest supermarket," implying that when shoppers are *closer* to supermarkets they are *more likely* to shop online at Childcorp.com and when they are *further away* they are *less likely* to shop there.[18] (Although Δ_9(offline WOM) is greater in absolute value than Δ_9(online WOM), the estimates are not significantly different.) Our explanation for the negative sign is as follows. First, Childcorp.com prices are lower than typical supermarket prices, and Childcorp.com shoppers spend, on average, approximately \$1,500 per year on the products in question. Shoppers living *closer* to supermarkets shop more frequently (Bell and Lattin 1998) and have superior price knowledge for product categories (Dinesh et al. 2008). Every time they see the higher supermarket prices, the wisdom of their Childcorp.com purchases is reinforced. This implies a negative sign: Shorter travel distances to supermarkets make for more frequent and price-informed supermarket shoppers, which drives demand online. Second, shoppers who travel *further* to supermarkets buy larger baskets of items (including Childcorp.com products) to amortize fixed travel costs (Tang et al. 2001). Because fixed cost amortization implies a greater likelihood of more categories being in the average (supermarket) shopping basket of these households, this makes them have less need for a "single category" online retailer. Hence, this

[18] This does not result from multicollinearity among the three expected distance variables. The pairwise correlations are 0.55 (supermarkets and discount stores) and 0.49 (supermarkets and warehouse clubs), and the VIF values are small: 1.60, 2.23, and 2.30 for the distances to supermarket, department stores, and warehouse clubs, respectively, in the regression model of count data in log form.

also implies a negative sign: Longer travel distances to supermarkets make for large-basket shoppers who therefore have less need for a single-category online retailer.

Control Variables

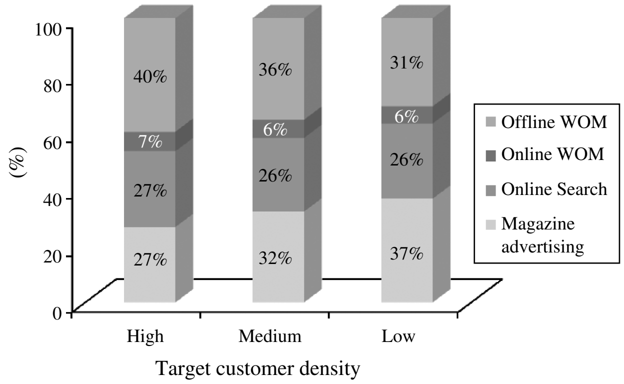
Control variable estimates either replicate findings from prior research or have intuitive signs if the variables are unique to our study. Controlling for the presence of sales tax, the effect of saving on sales tax is positive and statistically significant for WOM buyers and for search buyers. Prior research shows that search buyers are motivated by price (Bakos 1997, Lal and Sarvary 1999), and we find WOM buyers are also sensitive to price benefits.[19] Magazine advertising by Childcorp.com did not stress an online price advantage, so it's perhaps not surprising that there is no effect of tax savings on acquisitions through this mode.

Higher magazine circulation increases the total number of acquisitions in a location, but the decomposition shows that this effect is driven solely by increases in buyers via online search and magazine advertising. Conditional on the other controls, high-speed Internet penetration is not significant for any acquisition mode, and estimates for geodemographic control variables typically have intuitive signs (online demand is higher in zip codes with higher population growth rates, more college educated and wealthy individuals, etc.).

New Managerial Insights

Which Method Works Where. The findings deliver new managerial insights into the geographically complementary nature of different customer acquisition modes. To demonstrate, we use the estimates to compute the expected number of buyers per acquisition mode per zip code. To illustrate a key distinction across modes, all zip codes contained within MSAs are assigned to one of three groups. The groups are constructed so that they have approximately equal numbers of target customers and new buyers, but differ significantly on the dimension of target customer density.[20] Specifically, each group has approximately 4.7 million target customers and 50,000 new buyers; however, the average density of households

Figure 3 **Decomposition of New Buyers by Acquisition Mode and by Target Customer Density**



Note. The high, medium, and low groups of zip codes are defined so that each group has roughly equal numbers of target customers and new buyers, but differs substantially by target population density.

with children declines from 590.5 (high) to 121.9 (medium) to 11.0 (low). Group 1 (high density) contains 2,319 zip codes, Group 2 (medium density) contains 3,541 zip codes, and Group 3 (low density) contains 10,064 zip codes.

Figure 3 shows the percentage decomposition of buyers by acquisition mode (y-axis) plotted against high, medium, and low target customer density groups (x-axis). Offline WOM acquisitions account for 40% of total buyers in the high-density group but only 30% in the low-density group. Magazine advertising acquisitions show a reverse pattern—they start at around 27% and increase to 37% of the total buyers. Offline WOM is especially effective in high potential locations that are also fertile for interaction, whereas magazine advertising has more reach into many regions with relatively low potential individually, but that collectively account for a sizable portion of the customer base (see Table 1).

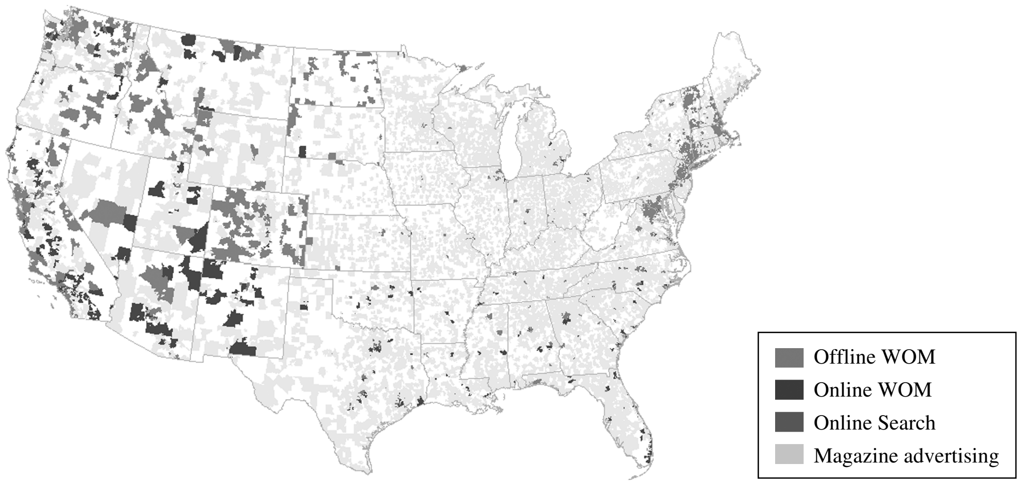
Figure 4, (a) and (b), complements Figure 3. Expected acquisitions are placed on a physical map of the United States, and each zip code is colored according to which the of the four acquisition modes is most effective at that location. Figure 4(a) shows this information for zip codes with at least one expected buyer; Figure 4(b) shows it for zip codes with at least 10 expected buyers. In Figure 4(a) there are many light gray zip codes, i.e., zip codes where magazine advertising generates the most expected new buyers. However, among "high-performing" zip codes in Figure 4(b), there are relatively few light gray regions, and many more gray regions where offline WOM is most effective. Offline WOM dominates in a small number of very high-performing spatially clustered zip codes, whereas traditional magazine advertising is effective in spatially dispersed (and individually low-performing) zip codes. This reinforces a key finding: IS-enabled methods of acquisition are important in

[19] The "shared benefit" argument may explain why the two types of WOM buyers have the same sales tax estimates. When a potential buyer hears about the benefit of "saving on sales tax" via WOM, he/she can easily understand the size of saving independent of whether that WOM arrived offline or online.

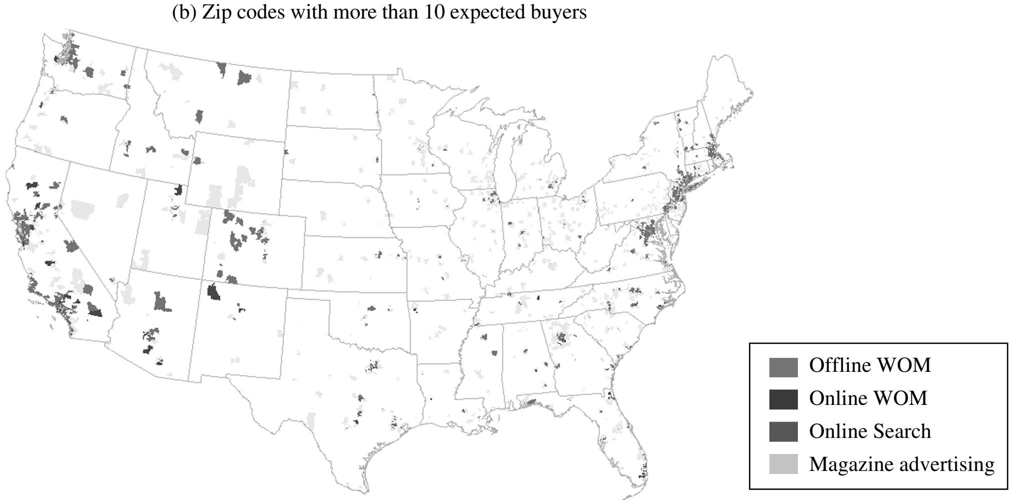
[20] We limit the analysis to zip codes within MSAs to ensure shoppers face reasonably comparable local environments, and this is also consistent with prior research (e.g., Forman et al. 2009, Sinai and Waldfogel 2004). Moreover, we obtain qualitatively identical results when we include all the zip codes.

Figure 4 Geographic Variation in the Most Effective Acquisition Mode

(a) Zip codes with more than one expected buyer



(b) Zip codes with more than 10 expected buyers



Note. The shades of gray indicate which mode is most effective in each location, i.e., which mode generates the greatest expected number of new buyers.

the new Internet retail economy, but traditional methods remain vital in a complementary manner.[21]

Preliminary Evidence for Gains from Geotargeting. Figures 3 and 4 raise an important question: What decisions should the firm make differently in light of our findings? We answer by showing how the firm might think about the locally customized purchase of search keywords. Search engines charge for sponsored links on a cost-per-click basis, and although it is possible to purchase search keywords on a geographical basis, Childcorp.com has never done this. To explore the potential of this option, we examine improvements that could result from locally targeted search keywords, where promising local targets are identified by the model.

We obtained conversion rates from "first click" to "first order" among first-time visitors at Childcorp .com for approximately 1,200 major cities in the United States, from October 2007 through March 2008, from Coremetrics.com.[22] We then compared actual

[21] As noted in the Data section, Childcorp.com did no locally targeted marketing with any acquisition method during the period of our data. Childcorp.com or other Internet retailers could however employ locally adjusted acquisition strategies. Out of the four modes, online search is directly under the firm's control, and search spending could be tailored by location. Magazine subscriptions are beyond the firm's control, but measurable. Online WOM can be promoted through bloggers and online brand communities established via social networking sites. Finally, Childcorp.com can facilitate offline WOM by supporting local moms' communities (see also Godes and Mayzlin 2009 for a discussion of firm-initiated WOM). We thank an anonymous reviewer for these suggestions.

[22] Our data are at the zip code level, whereas Coremetrics.com data are at the city level. Coremetrics.com specializes in tracking visitor browsing and purchasing behavior at online sites, for visitors coming from major U.S. cities. It started collecting data for Childcorp.com management from October 2007. The number of new buyers in these major cities accounts for 52% of the total new buyers, despite the relatively small number of cities included.

Table 5 A Comparison of Model Predictions and Click-to-Order Conversions

Cities per group	HHs w/children (1)	Expected buyers (2)	First orders (3)	First clicks (4)	Expected buyers per HHs w/children (5) = (2)/(1)	Conversion rates (6) = (4)/(3)
Top two groups						
1	67,098	5,194	9,924	54,119	0.077	0.183
21	80,260	2,405	2,260	11,904	0.030	0.190
Middle two groups						
46	228,172	2,154	1,133	10,673	0.009	0.106
16	208,026	1,914	1,013	11,207	0.009	0.090
Bottom two groups						
42	394,773	1,816	886	10,942	0.005	0.081
44	252,416	976	905	10,946	0.004	0.083

Notes. Each group of cities has about 11,000 of clicks (i.e., roughly equal marketing costs), and all cities in a group have approximately equal predictions for the expected number of new buyers per household (HH). The best-performing group contains one city, New York City. The number of cities in the other groups is variable. In the interests of space, we show only six groups of cities and indicate the differences between the "best" (top two), "average" (middle two), and "worst" (bottom two) groups of cities. Full information for all 50 groups is available from the authors upon request.

conversion i.e., click to order, in a city with the model-based predictions of potential for that city. To do this, we used the model estimates to generate an overall prediction for the total number of new buyers for each zip code. We used predictions for the total number of buyers because (1) new buyers are likely to access Childcorp.com via search engines regardless of their initial acquisition mode, and (2) there were no acquisition-mode-specific conversion rate data available (Coremetrics.com does not provide this information).

Next, we aggregated zip code predictions in each of the 1,200 major cities in the Coremetrics.com database and sorted the cities from highest to lowest according to the expected number of new buyers per household in the target population. After this sorting, we formed 50 separate groups of cities from the initial pool of the major cities. The 50 groups of cities are defined so that each group has approximately equal numbers of new clicks, i.e., approximately equal marketing costs, and cities in each group have similar "predicted performance," i.e., model predicted numbers of new buyers per number of households with children aged less than six years old.

For the sake of brevity, Table 5 shows results for only 6 (of 50) groups of cities: the top 2, middle 2, and bottom 2 groups. Column (5) gives the model-based prediction of new buyers per household with children, and column (6) gives the actual click-to-order conversion rates captured by Coremetrics.com. Top groups of cities have conversion rates of about 18%–19% and need, on average, 5.5 new clicks to obtain one new buyer. This increases to 10 and 12 clicks for the middle and bottom groups, respectively. Table 5 implies that targeting groups of cities with good model-based expected performance could improve efficiency in click-through rates by a factor of about 2. This preliminary evidence from completely separate conversion information suggests that predictions leveraged from our geographic model based on "old economy" geodemographic data could deliver meaningful improvements in (roughly doubling) the effectiveness of marketing expenditures on keywords.

Finally, the Coremetrics.com data also shows that shoppers in cities with good model-based expected performance (1) click more pages per session and (2) stay longer at Childcorp.com per session. Both observations suggest these shoppers are more engaged with Childcorp.com than are buyers in lower quality locations. Thus, our findings represent an interesting complement to those in recent studies of conversion efficacy. Ghose and Yang (2009) find that an improvement in landing page quality increases conversion rates, and Yang and Ghose (2010) report that conversion rates are higher when both paid and organic search results are present than when paid search is paused. These studies clearly show that specific improvements in information quality at the site aids conversion—our research highlights the fact that conversion rates respond positively to an improved ability to identify *locations* with receptive customers.

Conclusion and Future Research Directions

An online retailer is by definition ubiquitous because shoppers almost anywhere have the *potential* to use it. It is, however, becoming well established that the propensity for shoppers to buy online varies significantly by geography in accordance with the physical characteristics of shoppers' locations (e.g., Brynjolfsson et al. 2009, Choi and Bell 2011, Forman et al. 2009). Relatively unexplored are explanations for geographic variation in the success of different customer acquisition methods (see Figure 1) that are unrelated to just variation in market potential alone (see Figure 2). This is a key area for research because Internet retailers have vast trading areas and potentially face quite different cost–benefit trade-offs for different acquisition methods in different locations. Our main empirical findings are as follows.

• *Acquisitions in general and WOM acquisitions in particular benefit from physical proximity among target customers*. Target customer density explains geographic variation in total online demand through all modes of acquisition even after controlling for the total number of potential customers as well as observed and unobserved heterogeneity. In the case of the bulky, repeat-purchase consumables sold by Childcorp.com, density is likely to be a proxy for higher offline shopping costs. Target customer density also heightens the possibility for social observation and social interaction both offline *and* online. It therefore has a further positive incremental effect on acquisitions through offline and online WOM.

• *Location-based benefits enhance offline WOM acquisitions more than they enhance online WOM acquisitions*. Not surprisingly, online demand responds positively to time convenience (faster shipping speeds) and travel convenience (longer distances to direct offline competitors). More interestingly, the effects of these benefits are amplified when senders and recipients of WOM are more likely to be co-located, i.e., when acquisitions are through *offline* WOM. This suggests that the effectiveness of the WOM channel interacts with the type of benefit and with the locations of senders and receivers of WOM.

• *Acquisition modes are complementary and gains from geotargeting are possible.* Offline WOM acquisitions are geographically clustered, whereas magazine advertising acquisitions are geographically dispersed. IS-enabled acquisitions are relatively location independent and generate a roughly constant proportion of new customers in each location. This mode-based variation coupled with likely differences in the cost of acquiring customers through different modes suggests opportunities for geotargeting. Our model validation exercise on a separate data set from Coremetrics.com found that high-performing cities identified by the model have actual click-to-conversion rates approximately double those of low-performing cities.

Limitations and Directions for Future Research

The limitations of this article suggest a number of avenues for future work. First, it would be helpful to identify a comprehensive set of "geographic factors" that make some locations more viable than others for online retailers. Some considered thus far include access to offline stores (e.g., Forman et al. 2009), preference isolation (e.g., Choi and Bell 2011), and offline tax rates (e.g., Anderson et al. 2010). Second, we should learn more about what leads to WOM conversations, whom they are among, and what is discussed. Findings to date are that product characteristics influence WOM volume (e.g., Berger and Schwartz 2011) and that observational learning and WOM conversations have distinct as well as interactive effects (e.g., Chen et al. 2011). Third, it would be useful to develop more comprehensive modeling approaches that can handle slope heterogeneity over locations—even with the very large data sets typical of Internet retail businesses. In conclusion, Internet retailing is the fastest growing retail sector both in the United States and in many other international markets, including China, where sales reached $40 billion in 2010. It is therefore vital that researchers and practitioners alike build new theories and analyses to understand why consumers choose online stores over offline stores and how the fixed geography of consumer locations shapes consumer behavior online.

Acknowledgments

The authors are grateful to Eric Bradlow, Christophe Van den Bulte, three anonymous *Management Science* reviewers, the area editor, and the department editor for valuable and detailed suggestions. They also thank seminar participants at Carlson School of Management, Goizeta School of Business, Google, Harvard Business School, Kellogg School of Management, Tel Aviv University, University of Iowa (Department of Geography), Rotman School of Management, the Wharton School, Yonsei University School of Business, and the INFORMS 2009 annual meeting (San Diego) for comments. This work was supported by the National Research Foundation of Korea Grant funded by the Korean Government [NRF-2011-330-B00076]. Correspondence about this paper should be addressed to jeonghye@yonsei.ac.kr.

References

Agresti, A. 2002. *Categorical Data Analysis*. Wiley, New York.

Anderson, E. T., N. M. Fong, D. I. Simester, C. E. Tucker. 2010. How sales taxes affect customer and firm behavior: The role of search on the Internet. *J. Marketing Res.* **47**(2) 229–239.

Bakos, Y. 1997. Reducing buyer search costs: Implications for electronic marketplaces. *Management Sci.* **43**(12) 1676–1692.

Balasubramanian, S. 1998. Mail versus mall: A strategic analysis of competition between direct marketers and conventional retailers. *Marketing Sci.* **17**(3) 181–195.

Becker, G., K. M. Murphy. 2000. *Social Economics*. Harvard University Press, Cambridge, MA.

Bell, D. R., C. A. L. Hilber. 2006. An empirical test of the theory of sales: Do household storage constraints affect consumer and store behavior? *Quant. Marketing Econom.* **4**(2) 87–117.

Bell, D. R., J. M. Lattin. 1998. Shopping behavior and consumer preference for store price format: Why "large basket" shoppers prefer EDLP. *Marketing Sci.* **17**(1) 66–88.

Bell, D. R., S. Song. 2007. Neighborhood effects and trial on the Internet: Evidence from online grocery retailing. *Quant. Marketing Econom.* **5**(4) 361–400.

Berger, J., E. M. Schwartz. 2011. What drives immediate and ongoing word of mouth? *J. Marketing Res.* **48**(5) 869–880.

Berry, S. T. 1994. Estimating discrete-choice models of product differentiation. *RAND J. Econom.* **25**(2) 242–262.

Bhatnagar, A., B. T. Ratchford. 2004. A model of retail format competition for non-durable goods. *Internat. J. Res. Marketing* **21**(1) 39–59.

Blum, B. S., A. Goldfarb. 2006. Does the Internet defy the law of gravity? *J. Internat. Econom.* **70**(2) 384–405.

Breiman, L., P. Spector. 1992. Submodel selection and evaluation in regression: The X-random case. *Internat. Statist. Rev.* **60**(3) 291–319.

Brynjolfsson, E., M. D. Smith. 2000. Frictionless commerce? A comparison of Internet and conventional retailers. *Management Sci.* **46**(4) 563–585.

Brynjolfsson, E., Y. Hu, M. S. Rahman. 2009. Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Sci.* **55**(11) 1755–1765.

Cameron, A. C., P. K. Trivedi. 1986. Econometric models based on count data: Comparisons and applications of some estimators and tests. *J. Appl. Econometrics* **1**(1) 29–53.

Chen, Y., Q. Wang, J. Xie. 2011. Online social interactions: A natural experiment on word of mouth versus observational learning. *J. Marketing Res.* **48**(2) 238–254.

Choi, J., D. R. Bell. 2011. Preference minorities and the Internet. *J. Marketing Res.* **48**(4) 670–682.

Choi, J., S. K. Hui, D. R. Bell. 2010. Spatio-temporal analysis of imitation behavior across new buyers at online grocery retailer. *J. Marketing Res.* **47**(1) 75–89.

Choldin, H. M. 1978. Urban density and pathology. *Annual Rev. Sociol.* **4** 91–113.

Claude, B. 2002. A spatial autoregressive specification with a comparable sales weighting scheme. *J. Real Estate Res.* **24**(2) 193–212.

Dellarocas, C. 2003. The digitization of word of mouth: Promise and challenges of online feedback mechanisms. *Management Sci.* **49**(10) 1407–1424.

Dhar, S., S. J. Hoch. 1997. Why store brand penetration varies by retailer. *Marketing Sci.* **16**(3) 208–227.

Dinesh, K. G., K. Sudhir, D. Talukdar. 2008. The temporal and spatial dimensions of price search: Insights from matching household survey and purchase data. *J. Marketing Res.* **45**(2) 226–240.

Fernandez, R. M., E. J. Castilla, P. Moore. 2000. Social capital at work: Networks and employment at a phone center. *Amer. J. Sociol.* **105**(5) 1288–1356.

Fieuws, S., G. Verbeke. 2006. Pairwise fitting of mixed models for the joint modeling of multivariate longitudinal profiles. *Biometrics* **62**(2) 424–431.

Fieuws, S., F. Boen, C. Delecluse. 2006. High dimensional multivariate mixed models for binary questionnaire data. *J. Royal Statist. Soc.: Ser. C (Appl. Statist.)* **55**(4) 449–460.

Forman, C., A. Ghose, A. Goldfarb. 2009. Competition between local and electronic markets: How the benefit of buying online depends on where you live. *Management Sci.* **55**(1) 47–57.

Fotheringham, S. A. 1988. Consumer store choice and choice set definition. *Marketing Sci.* **7**(3) 299–310.

Fox, B. J., J. Fox, R. W. Marans. 1980. Residential density and neighbor interaction. *Sociol. Quart.* **21**(3) 349–359.

Getis, A., J. K. Ord. 1992. The analysis of spatial association by use of distance statistics. *Geographical Anal.* **24** 189–206.

Ghose, A., S. Yang. 2009. An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management Sci.* **55**(10) 1605–1622.

Glaeser, E. L., B. Sacerdote, J. A. Scheinkman. 1996. Crime and social interactions. *Quart. J. Econom.* **111**(2) 507–548.

Glaeser, E. L., B. Sacerdote, J. A. Scheinkman. 2003. The social multiplier. *J. Eur. Econom. Assoc.* **1**(2–3) 345–353.

Godes, D., D. Mayzlin. 2009. Firm-created word-of-mouth communication: Evidence from a field test. *Marketing Sci.* **28**(4) 721–739.

Goolsbee, A. 2000. In a world without borders: The impact of taxes on Internet commerce. *Quart. J. Econom.* **115**(2) 561–576.

Greene, W. 2008. *Econometric Analysis*. Prentice Hall, Upper Saddle River, NJ.

Gueorguieva, R. 2001. A multivariate generalized linear mixed model for joint modeling of clustered outcomes in the exponential family. *Statist. Modeling* **1**(3) 177–193.

Gunn, E. 2007. Keeping baby dry: We buy bulk diapers. *Wall Street Journal* (October 2), http://online.wsj.com/article/SB119146035058248489.html.

Huff, D. L. 1964. Defining and estimating a trading area. *J. Marketing* **28**(3) 34–38.

Katona, Z., P. P. Zubcsek, M. Sarvary. 2011. Network effects and personal influences: The diffusion of an online social network. *J. Marketing Res.* **48**(3) 425–443.

Katz, E., P. F. Lazarsfeld. 1955. *Personal Influence*. Free Press, Glencoe, IL.

Kim, Y. S., W. N. Street, G. J. Russell, F. Menczer. 2005. Customer targeting: A neural network approach guided by genetic algorithms. *Management Sci.* **51**(2) 264–276.

Knorr-Held, L., J. Besag. 1998. Modeling risk from a disease in time and space. *Statist. Medicine* **17**(18) 2045–2060.

Lal, R., M. Sarvary. 1999. When and how is the Internet likely to decrease price competition? *Marketing Sci.* **18**(4) 485–503.

LeSage, J. P., R. K. Pace. 2005. A matrix exponential spatial specification. *J. Econometrics* **140**(1) 190–214.

Moran, P. A. P. 1950. Notes on continuous stochastic phenomena. *Biometrika* **37**(1–2) 17–23.

Reilly, W. J. 1931. *The Law of Retail Gravitation*. Knickerbocker Press, New York.

Sinai, T., J. Waldfogel. 2004. Geography and the Internet: Is the Internet a substitute or a complement for cities? *J. Urban Econom.* **56**(1) 1–24.

Steenburgh, T. J., A. Ainslie, P. H. Engebretson. 2003. Massively categorical variables: Revealing the information in zip codes. *Marketing Sci.* **22**(1) 40–57.

Tang, C. S., D. R. Bell, T.-H. Ho. 2001. Store choice and shopping behavior: How price format works. *California Management Rev.* **43**(2) 56–65.

Thum, Y. M. 1997. Hierarchical linear models for multivariate outcomes. *J. Educational Behavioral Statist.* **22**(1) 77–108.

Villanueva, J., S. Yoo, D. M. Hanssens. 2008. The impact of marketing-induced vs. word-of-mouth customer acquisition on customer equity. *J. Marketing Res.* **45**(1) 48–59.

Wikle, C. K., M. B. Hooten. 2006. Hierarchical Bayesian spatio-temporal models for population spread. J. S. Clark, A. Gelfand, eds. *Applications of Computational Statistics in the Environmental Sciences: Hierarchical Bayes and MCMC Methods*. Oxford University Press, New York.

Yang, S., G. M. Allenby. 2003. Modeling interdependent consumer preferences. *J. Marketing Res.* **40**(3) 282–294.

Yang, S., A. Ghose. 2010. Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Sci.* **29**(4) 602–623.