ASIM ANSARI, CARL F. MELA, and SCOTT A. NESLIN*

The authors develop a model of customer channel migration and apply it to a retailer that markets over the Web and through catalogs. The model identifies the key phenomena required to analyze customer migration, shows how these phenomena can be modeled, and develops an approach for estimating the model. The methodology is unique in its ability to accommodate heterogeneous customer responses to a large number of distinct marketing communications in a dynamic context. The results indicate that (1) Web purchasing is associated with lower subsequent purchase volumes than when buying from other outlets; (2) marketing efforts are associated with channel usage and purchase incidence, offsetting negative Web experience effects; and (3) negative interactions occur between like communications (catalog × catalog or e-mail × e-mail) and between different types of communications (catalog × e-mail). The authors find that over the four-year period of their data, a Web-oriented "migration" segment emerged, and this group had higher sales volume. Their post hoc analysis suggests that marketing efforts and exogenous customer-level trends played key roles in forming these segments. The authors rule out alternative explanations, such as that the Web attracted customers who were already heavy users or that the Web developed these customers into heavier users. They conclude with a discussion of implications for both academics and practitioners.

Keywords: customer relationship management, multichannel marketing, Internet marketing, catalogs, e-mail

Customer Channel Migration

As multichannel distribution becomes increasingly prevalent, customers face an expanding array of purchase and communication options. For example, online sales are expected to increase 20% in 2006 to $211.4 billion, doubling the total revenue in 2003 (*The Wall Street Journal* 2006). As such, multichannel customer management is becoming a pivotal component in firms' marketing strategy. Despite this trend, we are aware of no empirical research that details (1) how customers migrate between channels in a multichannel environment and (2) the role of marketers in shaping migration through their communications strategy.

Some prior work has shown that customer preferences differ by channel (Liang and Huang 1998; Morrison and Roberts 1998; Shankar, Smith, and Rangaswamy 2003). However, this research does not investigate how preferences vary in the long run as customers gain experience with different channels or how marketing influences this evolution. Other studies have explored channel cannibalization (Biyalogorsky and Naik 2003; Deleersnyder et al. 2002). However, they do not model customer heterogeneity, which is central to the task of customer management.[1] In this vein, Fox, Montgomery, and Lodish (2004) develop an individual-level model of retail choice. However, our focus is on channel formats (Web versus catalog) rather than store variety within a particular retail format (grocery, mass merchandise, or drug).

The foregoing discussion suggests that researchers are beginning to recognize the considerable economic and behavioral ramifications of customer channel migration. However, many important questions remain:

- What determines whether customers migrate to the Internet, and what is the overall effect of this channel on demand in the long run?
- What are the short- and long-term effects of channel usage on channel selection and demand? For example, do customers

*Asim Ansari is Professor of Marketing, Columbia University Graduate School of Business (e-mail: maa48@columbia.edu). Carl F. Mela is Professor of Marketing, Fuqua School of Business, Duke University (e-mail: mela@duke.edu). Scott A. Neslin is Albert Wesley Frey Professor of Marketing, Tuck School of Business, Dartmouth College (e-mail: scott.neslin@dartmouth.edu). Russell Winer served as guest editor for this article.

[1]In a separate analysis not reported here, we found that dynamic effects were considerably larger when we did not include customer heterogeneity.

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

develop "channel loyalty" according to their channel usage experience?
•What is the role of marketing communications in channel migration? Does marketing affect channel selection, demand, or both?
•Do customer differences affect the channel migration process, and if so, how?

We develop and estimate a model of customer channel migration to investigate these substantive questions. Our contribution is twofold: First, we (1) propose a set of key phenomena that are related to channel migration behavior, (2) show how these phenomena can be modeled, and (3) develop an estimation approach for such a model. The migration model captures the effects of large numbers of marketing communications in the face of dynamics and customer heterogeneity. Second, we contribute to the substantive knowledge base on customer channel migration. One key finding is that Web use, when we control for marketing and other factors, is associated with a permanent decrease in the likelihood of buying from a firm, perhaps because the Internet can expand consideration sets and lower customer service levels.

We proceed as follows: We first describe the modeling framework and use it to identify key phenomena to be incorporated into the model. Next, we describe our model. Subsequently, we describe our data and report the results. Finally, we summarize key findings and conclude by offering managerial and research implications.

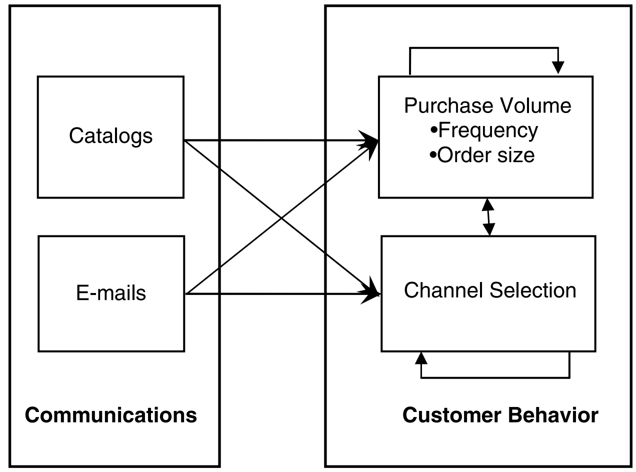
CHANNEL MIGRATION FRAMEWORK

Channel migration affects firm profit through its influence on cost and revenue. For example, it has been claimed that the Internet is more cost efficient than traditional channels. Although this might suggest that companies should migrate customers to the Web, the efficacy of this strategy depends on how migration affects overall demand. Thus, understanding how marketing actions are associated with demand is crucial in grasping how customer channel migration affects firm profitability.[2] In Figure 1, we provide an overview of the demand-side characteristics of channel migration.

We assume that the customer jointly decides how much to purchase from the firm and what channel to use. Both behaviors entail experience or learning effects, whereby previous purchases and channel selections can affect subsequent behavior. In addition, purchase volume and channel selection may be linked contemporaneously. For example, heavy purchasers may prefer certain channels. Finally, marketing communications can affect purchase volumes and channel share.[3]

To illustrate the ramifications of this framework, we consider its implications for research findings reinforced in the popular press that multichannel customers produce more sales than single channel shoppers (*Chain Store Age* 2001; *Infoworld* 2001; *Inter@ctive Week* 2000; Kumar and Venkatesan 2005; Kuswaha and Shankar 2005; Thomas and Sullivan 2005; *The Wall Street Journal* 2004; Yulinsky 2000). As a result, there is speculation that "multichannel customers are the best customers for a retailer, because they buy more and provide retailers with incremental gains over their lifetime" (*Inter@active Week* 2000, p. 50). It might be conjectured that firms should cultivate multichannel buying. However, the framework suggests several other possibilities regarding why the multichannel and Internet-loyal customers have high sales levels:

Figure 1
CONCEPTUAL FRAMEWORK



•Heavy users naturally migrate to the Internet (purchase volume → channel selection). Heavy usage might be correlated with various demographic factors.
•The Internet cultivates heavy buying; that is, customers buy more from a firm in the long run when they buy on the Internet (channel selection → purchase volume).
•Customers respond differently to marketing (communications → purchase volume and channel share). For example, customers who migrate might also respond more strongly to marketing.

Each explanation has a different implication for the profitability of channel migration, so it is desirable to disentangle them. For example, Internet buyers might not be prone to buy more; rather, they might receive more marketing, leading to the appearance of a greater proclivity to buy. Should a switch to the Internet be countervailed by lesser future demand (by encouraging consumers to shop other Web sites or by lowering service levels), the prescription to migrate people to the Web could be counterproductive. Thus, it is desirable to take a more systematic view of channel migration. In the next section, we formalize the model in Figure 1.

MODEL

We model purchase volume and channel selection as suggested by our framework. We model the purchase incidence and order-size components of purchase volume using a Type II Tobit specification and channel selection using a

[2]Channel migration can also affect revenues when prices differ across channels. In our data, prices are identical across channels.
[3]In Figure 1, we consider two forms of communication—catalogs and e-mails (because these are the instruments used by the firm in our data).

probit framework. Specifically, we assume that the customer jointly decides each month whether to purchase and, if so, how much to spend and what channel to select.[4] Let q_{it} indicate the dollar sales volume of purchases by customer i in month t conditioned on a decision to buy, and let b_{it} be an indicator variable of whether the customer buys or does not buy. Let q^*_{it} be a partially latent variable that is related to the observed order sizes (in \$), and let b^*_{it} be a latent variable related to the decision regarding whether to buy. We can write the Type II Tobit specification as

$$(1) \qquad b_{it} = \begin{cases} \text{Buy} & \text{if } b^*_{it} > 0 \\ \text{No buy} & \text{if otherwise} \end{cases},$$

$$(2) \qquad q_{it} = \begin{cases} \exp(q^*_{it}) & \text{if } b^*_{it} > 0 \\ 0 & \text{if otherwise} \end{cases}.$$

We exponentiate q^*_{it} to ensure that predicted actual quantities are positive.

For the channel selection decision, let w_{it} be an indicator variable that records channel choice, and let w^*_{it} be a latent variable that represents the difference between the customer's latent utilities for purchasing through the catalog and the Internet. Then, the binary probit model conditional on purchase is as follows:[5]

$$(3) \qquad w_{it} = \begin{cases} \text{Buy on catalog} & \text{if } w^*_{it} > 0 \text{ and } b^*_{it} > 0 \\ \text{Buy on Internet} & \text{if } w^*_{it} \le 0 \text{ and } b^*_{it} > 0 \end{cases}.$$

Four groups of variables accommodate the various substantive issues identified in our framework: customer characteristics, experience effects, communications effects, and time effects. We can write the relationship between the latent variables and these groups as

$$(4) \qquad \begin{aligned} b^*_{it} &= \text{Customer characteristics}_{bi} + \text{Experience}_{bit} \\ &\quad + \text{Communication}_{bit} + \text{Time effects}_{bit} + e_{bit}, \\ q^*_{it} &= \text{Customer characteristics}_{qi} + \text{Experience}_{qit} \\ &\quad + \text{Communication}_{qit} + \text{Time effects}_{qit} + e_{qit}, \text{ and} \\ w^*_{it} &= \text{Customer characteristics}_{wi} + \text{Experience}_{wit} \\ &\quad + \text{Communication}_{wit} + \text{Time effects}_{wit} + e_{wit}, \end{aligned}$$

where b subscripts purchase incidence, q labels order size, and w labels channel selection. The terms e_{bit}, e_{qit}, and e_{wit} represent unobserved factors that influence incidence, order size, and channel selection, respectively. The errors are assumed to be distributed multivariate normal, $N(\mathbf{0}, \boldsymbol{\Sigma})$. The off-diagonal elements in the covariance matrix $\boldsymbol{\Sigma}$ accommodate the contemporaneous correlations among the three customer decisions. For identification, we set the variances of e_{bit} and e_{wit} to 1 because the scales of b^*_{it} and w^*_{it} cannot be inferred from the observed binary choices. We now describe the specification of the effects within each group of variables.[6]

Customer Characteristics

Customer characteristics include observed (e.g., demographics) and unobserved factors that vary cross-sectionally. Observed factors in our model include age, income, and whether the household has children (Appendix A defines all variables used in this study). We capture the direct impact of unobserved customer-specific variables through individual-level random intercepts in the three equations. Denoting the random intercepts as Int, we have the following:

$$(5) \qquad \begin{aligned} \text{Customer characteristic}_{bi} &= \text{Int}_{bi} + \xi^{cc}_{b1}\text{Age}_i + \xi^{cc}_{b2}\text{Income}_i \\ &\quad + \xi^{cc}_{b3}\text{Children}_i, \\ \text{Customer characteristic}_{qi} &= \text{Int}_{qi} + \xi^{cc}_{q1}\text{Age}_i + \xi^{cc}_{q2}\text{Income}_i \\ &\quad + \xi^{cc}_{q3}\text{Children}_i, \text{ and} \\ \text{Customer characteristic}_{wi} &= \text{Int}_{wi} + \xi^{cc}_{w1}\text{Age}_i + \xi^{cc}_{w2}\text{Income}_i \\ &\quad + \xi^{cc}_{w3}\text{Children}_i. \end{aligned}$$

Experience Effects

Experience variables vary across customers and over time. We incorporate transient effects through purchase recency (Since) and lagged variables for Web and catalog incidence and order sizes (Lweb and Lcat). These variables correspond to the RFM (recency, frequency, monetary value) variables typically used in database marketing applications. Permanent effects can arise when prior usage generates enduring changes in behavior. We capture this by Wuse, which is a function of the number of previous purchases on the Internet. Equation 6 captures the experience effects in our model:

$$(6) \qquad \begin{aligned} \text{Experience}_{bit} &= \xi^{e}_{b1i}\text{Lweb}_{bit} + \xi^{e}_{b2i}\text{Lcat}_{bit} + \xi^{e}_{b3i}\text{Wuse}_{it} \\ &\quad + \xi^{e}_{b4i}\text{Since}_{it}, \\ \text{Experience}_{qit} &= \xi^{e}_{q1i}\text{Lweb}_{qit} + \xi^{e}_{q2i}\text{Lcat}_{qit} + \xi^{e}_{q3i}\text{Wuse}_{it} \\ &\quad + \xi^{e}_{q4i}\text{Since}_{it}, \text{ and} \\ \text{Experience}_{wit} &= \xi^{e}_{w1i}\text{Lweb}_{wit} + \xi^{e}_{w2i}\text{Lcat}_{wit} + \xi^{e}_{w3i}\text{Wuse}_{it} \\ &\quad + \xi^{e}_{w4i}\text{Since}_{it} + \xi^{e}_{w5i}\text{Diff}_{it}, \end{aligned}$$

where the ξ terms represent random coefficients.

We define the variables in Equation 6 differently in each equation. In the purchase incidence equation (Experience_{bit}), Lcat and Lweb are binary variables that indicate whether the customer purchased from the firm in the pre-

[4]Our use of disaggregate customer-level data is consistent with emphasis on customer heterogeneity. An alternative approach is to use aggregate time-series models to focus more on lag structure. We thank an anonymous reviewer for this perspective.

[5]We use a binary probit instead of a bivariate probit because nearly 100% of household-month observations entail purchase on only one channel.

[6]We investigated nonstationarity in the form of a random walk by estimating a model with a random-walk intercept. This model yielded no appreciable change in the posterior distribution of the parameters and a lower log-marginal likelihood.

ceding month through either the catalog or the Web. In the order-size equation ($Experience_{qit}$), we define Lcat and Lweb as the order size of the previous purchase on the catalog or on the Web. The use of lagged incidence in the incidence equation and lagged order size in the order-size equation mimics the definitions of the dependent variables. The channel selection model ($Experience_{wit}$) incorporates both lagged volume and lagged channel selection effects.[7] In this equation, we define Lcat and Lweb as the previous month's purchase volume from the catalog or the Web, respectively. We also include Diff, which represents state dependence in channel selection. We set Diff to 1 if the previous purchase for the customer is a catalog purchase and to –1 if it is through the Internet.[8] We define Since as the number of months elapsed since the previous purchase. We include this recency measure in all three experience equations.

The function $Wuse_{it}$, defined as Log(1 + Web purchases to date), captures the permanent effect due to Web usage.[9] We use $Wuse_{it}$ in all three equations. By definition, this variable is independent of the duration between Web purchases because we attempt to capture forgetting and other transient effects due to previous channel usage through the Lweb variables. We specify this to have diminishing marginal returns because, consistent with Bayesian learning (Roberts and Urban 1988), we expect the first usage to have a greater effect on behavior than the last usage. Finally, we allow for individual-specific slopes for all the experience effects.

Communications Effects

We define communication c as a particular communication sent by the firm at a particular time (these can be e-mails or catalogs). Therefore, two different catalogs mailed at the same time are considered different communications, and the same catalog sent at two different times is considered two different communications. For this reason, the number of communications in our data is considerable, totaling C = 723. Although each customer could have received 723 communications, in practice, no customer received this many, and the number received varies across customers. Rather than model the effect of each communication separately, we decompose these effects (Campbell et al. 2001) into (1) the characteristics of the communication (e.g., communications of like kind have the same effect) and (2) the time since the communication was sent (i.e., the effect of the communication decays over time). In addition, we allow for the direct effect of a communication and its interaction with other communications because there is likely to be decreasing marginal returns to these communications, reflected in a negative interaction.

Direct communication effects. We define the direct effect of communication c on customer i at time t as[10]

$$\text{Direct\_Effect}_{ict} = \beta_{ic}\lambda_c^{r_{ict}}d_{ict}. \tag{7}$$

The variable d_{ict} indicates whether customer i has received communication c on or before time t. It equals 0 until the customer receives communication c, and it equals 1 each period thereafter. This ensures that the communication does not begin to have an impact until the customer receives it. The variable r_{ict} is the number of periods elapsed since customer i received communication c. The λ_c is the "decay" parameter and reflects dynamics. We expect that λ_c is between 0 and 1; a large λ_c means that the communication exerts an impact well into the future. The parameter β_{ic} is the magnitude of the direct effect of communication c and is household specific. Communications that are more effective have higher values of β_{ic}.

We can sum Equation 7 across all communications to compute the total direct effect of communications received by customer i as of time t:

$$\text{Total\_Direct\_Effect}_{it} = \sum_{c \in C} \beta_{ic}\lambda_c^{r_{ict}}d_{ict}. \tag{8}$$

Equation 8 implies that it is necessary to estimate C direct communication effect parameters, β_{ic}, for incidence, order size, and channel selection. Because C = 723, this implies 2169 parameters for each customer. To model these effects parsimoniously, we describe each communication by a set of M attributes. We define a_{cm} as being equal to 1 when communication c has attribute m and 0 when otherwise (m = 1, …, M). Then, we can express the communication effect for communication c as

$$\beta_{ic} = \sum_{m=1}^{M} \psi_{im}a_{cm}. \tag{9a}$$

We can write the decay parameter as[11]

$$\lambda_c = \frac{\exp\left(\sum_{m=1}^{M}\zeta_m a_{cm}\right)}{1 + \exp\left(\sum_{m=1}^{M}\zeta_m a_{cm}\right)}. \tag{9b}$$

Because M is small compared with C, we achieve great parsimony while allowing different communication types to have different effects. In our application, we use M = 2 and distinguish between catalogs and e-mails.[12] Accordingly, we have the following:

a_{c1} = 1 if communication c is a catalog and 0 if otherwise, and
a_{c2} = 1 if communication c is an e-mail and 0 if otherwise.

We use the label λ_{cat} to denote the decay parameter for a communication c ∈ Catalogs (the set of all communications that are catalogs), and this equals $\exp(\zeta_1)/[1 + \exp(\zeta_1)]$ for all catalogs. Similarly, we use the label $\lambda_{e\text{-}mail}$ to denote the decay parameter for communications c ∈ E-mails (the set

[7]We thank an anonymous reviewer for this suggestion.

[8]We derive this variable by differencing the latent utilities for Internet choice and catalog choice; therefore, it has a value of ±1. We add ξ_{w5i} to w* when a catalog is used and subtract it from w* when the Internet is used (see Equation 4).

[9]It is impossible to include a corresponding variable for catalog purchases to date because there is no information on the number of catalog purchases before the data. In contrast, the Web channel is new, so the amount of purchasing before our data is negligible for all households.

[10]Although we suppress the subscripts for the equation (b, q, or w) to simplify the presentation, all parameters are equation specific.

[11]We use this functional form to ensure $0 \le \lambda \le 1$.

[12]It is possible to consider additional attributes (e.g., men's catalogs versus women's), but the catalog versus e-mail distinction is fundamental and enables us to investigate the propositions we stated previously (i.e., our specification is theoretically driven).

of all e-mails), and this equals $\exp(\zeta_2)/[1 + \exp(\zeta_2)]$ for all e-mails. Partitioning the communications into e-mail and catalog, using the definitions for a_{c1}, a_{c2}, λ_{cat}, and $\lambda_{e\text{-}mail}$, and substituting Equations 9a and 9b into Equation 8, we obtain the following:

$$(10)\qquad \text{Total\_Direct\_Effect}_{it} = \sum_{c\,\in\,\text{Catalogs}} \psi_{i1}\lambda_{cat}^{r_{ict}}d_{ict} + \sum_{c\,\in\,\text{E-mails}} \psi_{i2}\lambda_{e\text{-}mail}^{r_{ict}}d_{ict}.$$

We call the first sum the catalog direct effect and the second sum the e-mail direct effect:

$$(11a)\qquad \text{Cat}_{it} = \sum_{c\,\in\,\text{Catalogs}} \psi_{i1}\lambda_{cat}^{r_{ict}}d_{ict}, \text{ and}$$

$$(11b)\qquad \text{E-mail}_{it} = \sum_{c\,\in\,\text{E-mails}} \psi_{i2}\lambda_{e\text{-}mail}^{r_{ict}}d_{ict}.$$

Communication interaction effects. We define the interaction between communications c and c′ as

$$(12)\qquad \text{Interaction\_Effect}_{icc't} = \beta_{cc'}\lambda_c^{r_{ict}}\lambda_{c'}^{r_{ic't}}d_{ict}d_{ic't}.$$

The parameter $\beta_{cc'}$ reflects the interaction between two communications. Its magnitude is modified by the temporal proximity of the communications. This modification is reflected in the $\lambda_c^{r_{ict}}\lambda_{c'}^{r_{ic't}}$ term. If one or both communications were received a long time ago, the interaction will be negligible because λ is raised to the power r (the number of periods elapsed since the communications were received). Because $\lambda_c \neq \lambda_{c'}$, Equation 12 allows for order effects in that the interaction between communication c and c′ can be different when c precedes c′ versus when c′ precedes c. For example, in Equation 12, the λ_c term dominates if c is received recently, whereas $\lambda_{c'}$ dominates if c′ is received recently.

Because there are 723 × 722/2, or 261,003, potential communication interaction terms in our model, we model the $\beta_{cc'}$ as functions of the communications' underlying attributes. Accordingly, we assign pairs of communications to three unique categories: Both communications are catalogs, both are e-mails, or one is a catalog and the other an e-mail. In Appendix B, we show that this implies the following:

$$(13)\qquad \text{Total\_Interaction\_Effect}_{it} = \text{Cat\_Cat}_{it} + \text{E-mail\_E-mail}_{it} + \text{Cat\_E-mail}_{it},$$

where

$$(14a)\qquad \text{Cat\_Cat}_{it} = \sum_{\substack{c,c'\\ \text{both catalogs}}} \theta_1\lambda_{cat}^{r_{ict}}\lambda_{cat}^{r_{ic't}}d_{ict}d_{ic't},$$

$$(14b)\qquad \text{E-mail\_E-mail}_{it} = \sum_{\substack{c,c'\\ \text{both e-mails}}} \theta_2\lambda_{e\text{-}mail}^{r_{ict}}\lambda_{e\text{-}mail}^{r_{ic't}}d_{ict}d_{ic't}, \text{ and}$$

$$(14c)\qquad \text{Cat\_E-mail}_{it} = \sum_{\substack{c\text{ is a catalog}\\ c'\text{ is an e-mail}}} \theta_3\lambda_{cat}^{r_{ict}}\lambda_{e\text{-}mail}^{r_{ic't}}d_{ict}d_{ic't}.$$

Combining Equations 11 and 14, we can write the total effect of communications on customer i at time t ($\text{Communication}_{it}$) as

$$(15)\qquad \text{Communication}_{it} = \text{Cat}_{it} + \text{E-mail}_{it} + \text{Cat\_Cat}_{it} + \text{E-mail\_E-mail}_{it} + \text{Cat\_E-mail}_{it}.$$

Equation 15 accommodates direct and interaction effects of communications, differential effects across forms or types of communications, and dynamics. Despite the relative parsimony of our approach, the computational burden associated with the summation across all pairs is considerable. We develop a recursive scheme, outlined in Appendix C, that reduces the computational complexity considerably.

Time Effects

Time effects include time trend and seasonality. For trend, we include Time_t, a monthly trend variable. To reduce seasonal indicators to a more parsimonious set, we first regressed total sales on monthly dummies and determined which were significant at $p < .05$. We included only "significant" seasonal dummies into our volume and selection models and further combined months whose parameters did not significantly differ from each other. This led us to include the following seasonality indicators: July/February (JlF_t), to account for months with low sales, and October (Oct_t), November (Nov_t), and December (Dec_t).[13] Therefore,

$$(16)\qquad \text{Time effects}_{bit} = \xi_{b1i}^{te}\text{Trend}_t + \xi_{b2}^{te}\text{Oct}_t + \xi_{b3}^{te}\text{Nov}_t + \xi_{b4}^{te}\text{Dec}_t + \xi_{b5}^{te}\text{JlF}_t,$$

$$\text{Time effects}_{qit} = \xi_{q1i}^{te}\text{Trend}_t + \xi_{q2}^{te}\text{Oct}_t + \xi_{q3}^{te}\text{Nov}_t + \xi_{q4}^{te}\text{Dec}_t + \xi_{q5}^{te}\text{JlF}_t, \text{ and}$$

$$\text{Time effects}_{wit} = \xi_{w1i}^{te}\text{Trend}_t + \xi_{w2}^{te}\text{Oct}_t + \xi_{w3}^{te}\text{Nov}_t + \xi_{w4}^{te}\text{Dec}_t + \xi_{w5}^{te}\text{JlF}_t,$$

where ξ are parameters to be estimated. Note that the seasonality variables vary over time and not across customers. However, we include cross-sectional heterogeneity in trend to reflect the possibility that, for example, different customers adopted the Internet at different rates.

Unobserved Heterogeneity

We specify customer-specific random effects for model intercepts, experience, direct communication, and trend parameters within each equation. Our initial efforts to accommodate unobserved heterogeneity in all communications parameters were thwarted by collinearity between the direct and interaction effects, so we do not specify random effects for the interactions. We specify the random effects to be correlated both within and across equations. We investigate both a multivariate normal and a multivariate t popu-

[13]Because this procedure does not account for other variables in the model, we also estimated the full model (Equation 4) with seasonal dummies for each month. The log-marginal likelihood of this model was lower, and the remaining posterior parameter distributions were essentially identical.

Table 1
DESCRIPTIVE STATISTICS

Variable	*M*	*Mdn*	*SD*	*Minimum*	*Maximum*
Average purchase ($)	131	111	75	26	564
Purchases (per year)	2.2	1.7	1.3	0	8.3
Income ($)	84,487	81,500	40,765	1,000	150,000
Age (years)	50	48	12	24	97
E-mails (per month)	.04	0	.05	0	3.4
Catalogs (per month)	3.4	3.2	1.6	.3	7.9
Catalog share (of orders)	89%	100%	.22	22%	100%

lation distribution for modeling unobserved heterogeneity. The t is a robust alternative to the normal because it has fatter tails. We use Bayesian methods for inference regarding the parameters. Because the posterior distribution is not completely known, we use Markov chain Monte Carlo (MCMC) techniques to obtain draws from the posterior distribution of the unknowns. We describe the priors and the full conditional distributions for the unknown parameters in Appendix D.

DATA

Data were provided by a retailer that sells consumer durable and apparel products in mature categories over the Internet and through a catalog. The data span four years, from February 1998 to February 2002. We restrict attention to active customers who bought at least three times in at least one of the years during this period. This restriction allows for changes in behavior over time. In our data, 37% of the customers used both the Internet and catalog for purchases, 1% exclusively used the Internet, and 62% used only catalogs. The entire data set consists of 40,000 customers; we randomly select 500 customers for our analysis. This suggests that 500 households × 48 months = 24,000 observations are available for estimation; however, the initialization period necessary to create lagged variables reduces our estimation sample to 19,064 observations.

The data consist of several files. A catalog purchase file includes information on how much was spent by whom and when, and an Internet purchase file provides the same information for Internet purchases. Catalog and e-mail data files indicate who received which communications when. In addition, a demographics file includes the age, income, and number of children for each household. Demographic data were purchased by the firm from companies that use either publicly available data sources or surveys. We aggregate data to the monthly level because the median purchase frequency is approximately 1.7 purchases per year. Finer gradations yield an excess of observations with zero sales, and coarser gradations result in multiple purchases within a single interval.[14] That is, the monthly sampling rate corresponds largely to the decision processes we model.

Table 1 presents the means of some of the key variables in the raw data. Collinearity in the data is modest because the condition indexes for the regressor sets in each of our three equations are all below 30 (Belsley, Kuh, and Welsch 1980, p. 105).[15] An item of note in Table 1 is the high level of catalog mailing.

Substantial channel migration is evident in our data, and multichannel and Internet buyers purchase more. In 1998, a large proportion (96% of customers) made more than 95% of their purchases through the catalog. By 2001, this share had fallen to 77%. Moreover, people who purchased through the Internet and the catalog tended to buy more; the mean purchase level of those who made more than 95% of their purchases on the catalog was $267, in contrast to $444 for those who made less than 95% of purchases through the catalog.

RESULTS

Model Comparisons

We estimated four models. The first, M1, is the full model specified in Equation 4; it incorporates heterogeneity using a multivariate t population distribution. The second model, M2, is also the full model, but it assumes that the random effects are distributed multivariate normal. This model is useful in assessing the relative merit of using the t-distribution for the random effects. The third model, M3, assumes no marketing effects. The fourth, M4, does not account for communication dynamics (by assuming that the communications have only an immediate impact and do not have a delayed effect). Both M3 and M4 use t-heterogeneity. M3 enables us to ascertain whether marketing contributes to model fit, whereas M4 enables us to test the role of marketing dynamics. Table 2 displays the log-marginal likelihoods based on the MCMC draws.

The best model is M1, which includes t-heterogeneity as well as marketing and experience dynamics. The superiority of M1 over M2 suggests that the t-distribution, owing to its fatter tails, is better able to capture heterogeneity than the normal distribution. The superiority of M1 over M3

Table 2
MODEL COMPARISON

Model	*Description*	*Log-Marginal Likelihood*
M1	Full model	–16,896.6
M2	Normal heterogeneity	–17,095.5
M3	No marketing	–18,965.3
M4	No communications dynamics	–17,077.4

[14]With monthly aggregation, multiple purchases are negligible, amounting to .29% of total observations and 1.61% of choice occasions. When there are multiple purchases, we classify the channel with the higher order size as the channel of choice.

[15]The most noticeable source of collinearity in our data is among the communications variables and their interactions; correlations between these variables range up to .94. Because collinearity increases standard errors, our data afford a conservative test of our hypotheses.

indicates that including marketing variables results in model improvement. Finally, the superiority of M1 over M4 suggests that the inclusion of dynamics is also desirable.

Model Prediction

To check the predictive validity of the models, we hold out the last three months and reestimate the model for the first 45 months (parameter estimates from this shorter period are comparable to the full period estimates).[16] We then use these estimates to predict aggregate purchase volume and channel share per month, using the first 45 months as "in-sample" or "calibration" and the last 3 months as "out-of-sample" or "holdout." Panels A and B of Figure 2 show the results. This figure indicates that calibration and holdout sample predictions are roughly equal among M1, M3, and M4, and all models track the data well. Computing the mean absolute errors, we find no statistically significant difference across models for either calibration or holdout.

That the full model does not forecast better is unsurprising because even the simpler models are fairly rich in their accommodation of customer heterogeneity, which is often crucial for good predictive performance. We also note that in the calibration data, the log-marginal likelihood is substantially superior for M1, even though the mean absolute errors are indistinguishable across models. These are different measures. The marginal likelihood is a measure of the probability of observing the data, given that the model is true. It is composed from the individual customer-level data. It also accounts for the uncertainty in the unknown parameters by integrating them out over their priors. In contrast, the mean absolute error is based on point estimates of the parameters and computes an aggregate prediction for a month and then averages across the months. In addition, these predictions rely on simulated lagged experience variables. This adds "noise" to the predictions that might mask differences in model performance.

Parameter Estimates

The parameter estimates for the best model (M1) appear in Table 3. We discuss the results for incidence, order size, and channel selection separately.

Purchase volume. Notably, few effects are significant (by significant, we mean that the 95% posterior confidence interval excludes zero) in the order-size model, whereas several estimates are significant in the incidence model. This suggests that most variation in purchase volume arises from incidence.

Many experience variables significantly influence purchase incidence. Internet usage is negatively associated with long-term purchase incidence (Wuse < 0).[17] Several explanations for this result may exist. First, when people migrate to the Internet, search costs are lowered, resulting in a greater likelihood of purchase elsewhere. Second, the lack of a human interface may loosen the psychological bonds between customer and firm (Ariely, Lynch, and Moon 2002). Third, the lack of contact with a sales agent may limit opportunities for cross-selling. The consequent attenuation in service levels implies that customer relationships with the firm may weaken over time as customers shop over the Internet, resulting in lower sales.

With respect to transient effects, Lweb is not significant in either model. In contrast, Lcat has a positive association with subsequent purchase volumes in the incidence model and a negative association with order size given incidence. This implies that increased catalog usage is associated with consumers buying more often but less on each occasion (possibly spreading their purchases over more catalogs).[18] The finding of little inertia for Internet purchases, coupled with the permanent negative association with purchase incidence, is provocative. It raises the possibility that Internet usage can have a long-term deleterious effect on demand.

The coefficient for Since is positive, indicating that recent purchasers are less likely to buy this period. This suggests that Since represents an inventory effect on average rather than a preference effect.

The marketing variables all have a positive direct effect in the incidence model. In addition, the interactions between communications are negative, implying cannibalization and decreasing return effects. The decay parameter estimates are significant but small, suggesting that these communications operate more as a call to action than by creating changes in attitudes over time.

Several of the control variables are significant. In particular, there is significant seasonality in sales, with a peak around Christmas. The effect of income is positive, suggesting that people with higher incomes buy more often, consistent with intuition.

Channel selection. The effect of Lcat is positive, suggesting that people who purchased high quantities on the catalog in the prior month are more likely to use the catalog if they purchase in the current month. The effect of Since is positive, suggesting that people who have not purchased in a while are more likely to use the catalog. E-mail has a negative association with catalog selection, suggesting that e-mails are associated with increased use of the Internet.[19] The positive interaction between catalogs and catalogs suggests that increasing numbers of catalogs solidifies the customer as a catalog user. The positive interaction between e-mails and e-mails, when coupled with the negative direct effect on catalogs, suggests a diminishing marginal return to e-mails in terms of driving people to the Internet.

As with the purchase volume model, some of the control variables are significant. The trend effect is negative, reflecting increased Internet use over time, possibly as a result of macroeconomic trends, such as increased computer penetration. The coefficient for age is positive, suggesting that older people are less likely to use the Internet, consistent with previous research during the period of our data (Jupiter Communications 2000).

[16]Note that our holdout period is relatively short because e-mails were used only in the latter part of the data. By using the last three months, we ensure that we have a sufficient number of observations in the calibration period to gauge their effectiveness reliably.

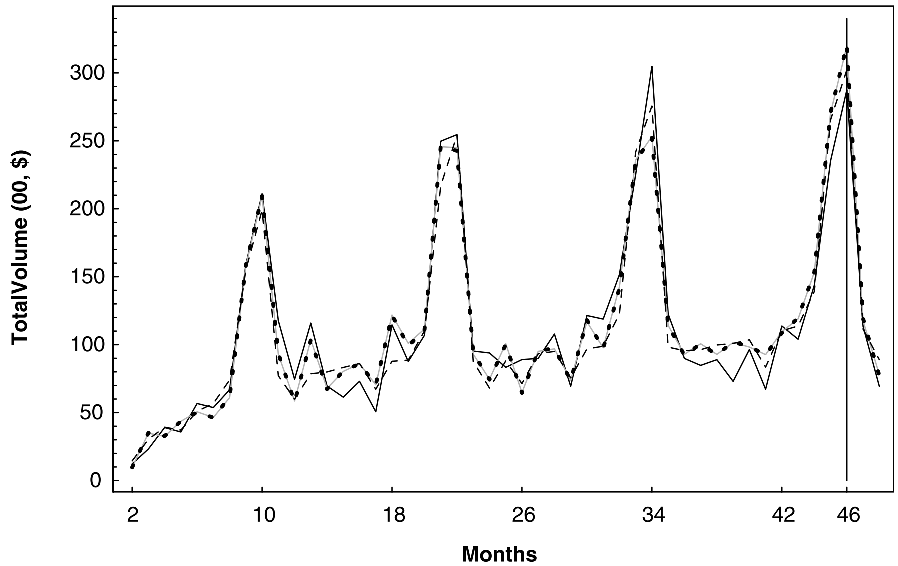
[17]This result could also be interpreted to mean that those with lower incidence rates are more inclined to shop online. However, we control for this possibility by allowing the random intercepts for purchase incidence and channel choice to be correlated and by allowing the error terms in these two models to be correlated (i.e., we "sweep out" the fixed over-time effects).

[18]Note that the null result for Lweb could be the result of poor statistical power (because there are fewer Web purchases), though there are 19,064 observations in total.

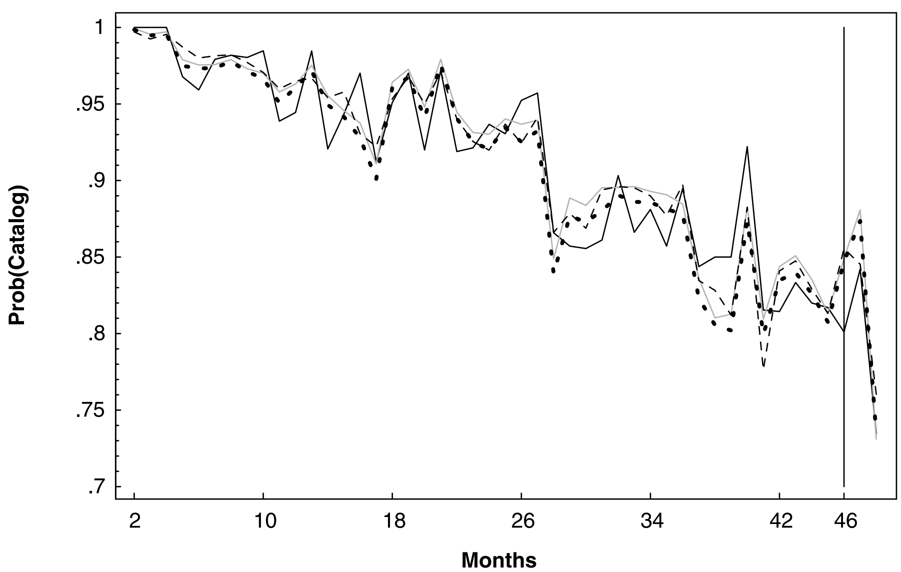
[19]It is possible that this result implies that Web users get more e-mails, though our data supplier did not target e-mails. It might also be that e-mail is a proxy for Internet access.

Figure 2
PREDICTION RESULTS

A: Purchase Volume



B: Channel Share



Notes: In-sample from Periods 2–45, and out-of-sample from Periods 46–48. The solid dark line is actual data. The gray line is for the full model (M1), the dashed line is for the no-marketing model (M3), and the dotted line is for the no-dynamics model (M4).

Finally, the decay parameter estimates are small but nonzero. The average (.14 + .11)/2 = .125 (or 1/8), which implies an infinite horizon effect for the communications on choice of 1/(1 – 1/8), or 8/7.[20] Thus, 1/7 of the communications effect occurs in periods after the communication is received. Although this appears small, this firm sends 45 catalogs a year. As such, the communication decay is tantamount to increasing the effect of these catalogs by 45/7 = 6.5 additional catalogs compared with the case in which communications were immediately perishable. Summed across people, the revenue effects of these lagged factors are considerable. The leads to the question whether these decay effects would be slower if fewer communications were sent.

[20]The total effect is calculated as $1 + .125 + .125^2 + .125^3 + ... = 1/(1 – .125) = 8/7$.

Table 3
PARAMETER ESTIMATES

		Incidence		*Order Size|Incidence*		*Choice|Incidence*	
Parameter		*M*	*SD*	*M*	*SD*	*M*	*SD*
Customer effects	Int	**–1.84**	**.24**	–.09	.34	–.01	1.08
	Age	.02	.04	–.01	.04	**.80**	**.16**
	Income	**.02**	**.01**	.00	.01	.03	.04
	Child	–.06	.10	.01	.10	.26	.34
Experience effects	Wuse	**–.47**	**.14**	.01	.15	.17	.35
	Since	**.59**	**.06**	.07	.06	–.03	.23
	Diff					.02	.17
	Lcat	**.18**	**.05**	**–.10**	**.02**	**.64**	**.19**
	Lweb	.13	.16	–.09	.10	–.06	.27
Marketing effects	Cat	**.13**	**.02**	.04	.03	.04	.12
	E-mail	**.15**	**.05**	.01	.05	**–.46**	**.14**
	Cat_Cat	**–.02**	.00	–.01	.00	**.05**	**.02**
	E-mail_E-mail	**–.02**	**.01**	.01	.01	**.09**	**.03**
	E-mail_Cat	**–.01**	**.01**	.00	.01	.02	.02
Time effects	Trend	**–.08**	**.02**	.01	.02	**–.26**	**.11**
	JIF	**–.18**	**.04**	–.02	.05	**–.48**	**.18**
	Oct	.03	.04	.10	.05	.12	.25
	Nov	**.54**	**.04**	**.25**	**.06**	.29	.21
	Dec	**.93**	**.04**	**.17**	**.08**	–.10	.22
	λ_{cat}	**.04**	**.02**	**.03**	**.01**	**.11**	**.04**
	$\lambda_{e\text{-mail}}$	**.04**	**.02**	**.04**	**.02**	**.14**	**.07**

Notes: Bold indicates that the 95% posterior interval excludes zero.

Communications Impulse Response Functions

The foregoing results suggest that the effects of communications are manyfold (e.g., e-mails increase sales in the short run but also switch demand to the Internet, which decreases sales in the long run). The combined effect of these dynamics is not apparent but can be determined by simulating the effect of an additional e-mail or catalog communication on purchase behavior. To do this, we increment the number of e-mails by one unit in Month 2 for all consumers and measure the impact on response in the subsequent 10 months. We repeat this procedure for Months 3–35, after which we averaged these response functions over time.[21]

From these simulations, we find that a catalog generates an incremental $.57 in revenue and that an e-mail generates an additional $.79 in the current period. It is possible that the smaller incremental effect for the catalog arises from the large number of catalogs sent per year (40), which may lead to decreasing returns. In contrast, few e-mails were sent during the interval of our data, suggesting that an incremental increase in e-mails could be more effective. Summing across all ten periods, the total effect of a catalog is $.48, whereas for e-mail, it is $.68.[22] Thus, the negative long-term effects are approximately 16%–17% the magnitude of the short-term effects. This is smaller than negative effects of promotion observed in scanner data, which are approximately 40% (Jedidi, Mela, and Gupta 1999; Macé and Neslin 2004).

[21]We commence with Month 2 because the preceding month is used for initialization. We conclude with Month 35 because this leaves several months to observe the post–impulse response.

[22]Note that the impulse response functions are calculated over 485 customers. We lose 15 customers because they did not have any purchases during the first 34 months, so we could not initialize their experience variables. On average, we have 38 months of data per household.

Heterogeneity in the Effects of Marketing on Channel Migration

The previous discussion centered on population-level effects, but there are also differences across customers. Consider the effect of direct communications in the channel selection model. The population-level effect of catalogs in the channel selection model is .04 and is not significantly different from zero (see Table 3). However, the random effects are highly dispersed and left-skewed (suggesting that it is misleading to conclude that catalogs have little effect). The 95% interval for the random effects ranges from –.60 to .28 (i.e., the interval that excludes the 2.5% of the highest random effects and the 2.5% of the lowest random effects). Therefore, for most people, the effect of catalog is to induce them to buy on catalog, but for some people, the highly negative direct effect of catalogs migrates them to the Internet (the lower 95% effect in the population is .04 + –.60 = –.56). A similar calculation for e-mails shows that it is possible for e-mails to move people to the catalog; the highest household-level effect in the 95% interval is .21. Compared with the –.56 effect for a catalog migrating a household to the Web, this .21 is small. As such, there is a greater potential to move some people to the Web with a catalog than it is to move people to a catalog with e-mail.

DIAGNOSING THE EMERGENCE OF THE INTERNET CHANNEL SEGMENT

Recall that our data provide evidence for (1) customers migrating to the Internet and (2) Internet and multichannel users buying more. Although this result would seem to suggest that consumers should be migrated to the Internet to increase sales, this argument ignores marketing and experience effects. In the next section, we try to disentangle these effects.

Internet Migration

To ascertain more precisely why customers migrated to the Internet, we calculate changes in the experience effects, marketing effects, and time effects in the channel selection equation (w* in Equation 4) that occurred between 1998 and 2001. The strategy is to observe how these factors changed for customers who migrated between 1998 and 2001 (n = 69) versus those who did not (n = 312) and to interpret these changes to determine why the migrations occurred.[23]

Equation 4 implies that latent utility in the channel selection can be written as follows:

$$E(w^*_{it}) = \text{Customer characteristics}_{wi} + \text{Experience}_{wit} + \text{Communications}_{wit} + \text{Time effects}_{wit}.$$

Because only one survey record exists for each customer, the measured customer characteristics do not change between 1998 and 2001. Thus, these factors cannot explain an increase in channel latent utility over time. However, experience, communications (marketing), and time contributions to latent utility do change. Thus, we calculate the averages of these utilities for both years, difference them, and then compare these differences between customers who migrated from the catalog to the Internet between Year 1 (1998) and Year 4 (2001) and those who did not. Table 4 displays the results.

Negative signs in Table 4 suggest that the corresponding factor facilitates a migration to the Internet. First, experience effects are equal for both groups. Therefore, the change in experience utility is the same for both groups. Ergo, experience effects were not associated with migration behavior for the two groups. Second, the time factor is negative for both segments, capturing a trend toward the Internet, and it is greater for customers who migrated. It is tempting to argue that this larger effect arises from our definition of a "migrater" as a household that used the Internet in later periods. However, note that we measured the trend effect while controlling for experience and marketing effects, and it is possible that these effects alone would have predicted the migration toward the Internet. So, the finding that those who migrated had a stronger trend toward the Internet is not an artifact of the model. Third, the change in marketing utility is positive for the no-migration group but negative for the migration group. This suggests that marketing both enhanced the likelihood that some customers would migrate and inhibited the likelihood that other customers would migrate.

Further inspection reveals that these differences in marketing utility arise from both changes in the levels of marketing and differences in marketing response across groups. In Table 5, we consider these factors. Customers who migrated were exposed to more marketing and switched more in response to it.[24] Because the migration group's response parameter for e-mail is more negative than that for catalogs and because the absolute level of change is higher, it appears that e-mail played the greater role.[25]

Changes in Purchase Volumes

Previously, we offered three explanations regarding why consumers who migrate to the Internet purchase greater quantities:

1. Migrating households were heavier users to begin with,
2. A positive experience on the Web encouraged higher purchase volumes, and
3. Migrating households simply reacted to marketing.

To disentangle these explanations, we decomposed the incidence model latent utility (b*) into experience, marketing, customer characteristics, and time factors, similar to the previous analysis. (We focus on incidence, rather than order size conditioned on incidence, because most of the variance in purchase volume is explained by the incidence model.) We report changes in these factors in Table 6. A positive entry means that changes in this factor contributed to higher sales in Year 4 than in Year 1. We also show the average customer characteristic utility (which does not change over time) and the average sales response to catalog and e-mail.

Table 6 eliminates the "heavier-users" explanation for increased use because the average customer characteristic

[23]These two groups do not sum to 500 households because (1) they include only those that bought in both 1998 and 2001 (we consider changes from Year 1 to Year 4) and (2) they do not include those that did not have sufficient data in Year 1 for initialization of experience variables.

[24]Again, a negative sign means that channel selection utility decreases, which, given our coding of w*, means that the factor is associated with customers using the Web.

[25]Note that additional e-mail by itself is not sufficient to migrate customers to the Web. They must also respond to that e-mail by moving to the Web. Of the 36.2% of nonmigrating customers who received e-mail, on average, they received 2.6 additional e-mails per month, and their channel selection response parameter averaged –.30. Of the 76.7% of migrating customers who received e-mail, they received on average 2.3 additional e-mails, and their average channel selection response parameter was –.54.

Table 4

CHANGES IN CONTRIBUTION TO LATENT UTILITY: YEAR 1 VERSUS 4

	No-Migration Group		*Migration Group*	
	Mean Change	*SE*	*Mean Change*	*SE*
Experience	–.15	.01	–.12	.09
Marketing	.51	.06	–.55	.17
Time	–.56	.04	–2.05	.12

Table 5

CHANGES IN MARKETING LEVELS AND RESPONSE PARAMETERS: YEAR 1 VERSUS 4

	No-Migration Group		*Migration Group*	
	M	*SE*	*M*	*SE*
Change in Marketing				
Catalogs/month	1.02	.09	1.54	.16
E-mails/month	.95	.08	1.76	.21
Response Parameter				
Catalog	.10	.01	–.16	.03
E-mail	–.43	.01	–.52	.05

Table 6

CHANGES IN LATENT INCIDENCE UTILITY, CUSTOMER CHARACTERISTICS, AND MARKETING RESPONSE BY MIGRATION GROUP

	No-Migration Group		*Migration Group*	
	M	*SE*	*M*	*SE*
Change in experience utility	.03	.03	–.19	.04
Change in marketing utility	.07	.02	.17	.05
Change in time utility	–.46	.05	.04	.08
Customer characteristic utility	–1.53	.03	–1.81	.07
Catalog response	.12	.01	.14	.01
E-mail response	.14	.01	.16	.03

utility is similar for both the no-migration and migration groups; this utility is slightly lower for the migration group. The data also refute the "positive-experience" explanation because experience utility shows a downward trend for the migration group. However, Table 6 supports the "marketing" explanation. The increase in the marketing utility offsets the decrease in experience utility. The result suggests that marketing obscures the negative association between Internet usage and demand. This is similar to the findings of Kopalle, Mela, and Marsh (1999) and Abraham and Lodish (1993), who show that increased sales promotions can mask a receding brand baseline. Table 6 also shows that time effects are more negative for customers who did not migrate, suggesting that these users are buying diminished amounts over time.

Although we do not find much difference in latent customer characteristic utility between customers who migrated and those who did not, the explanation that certain users are more likely to migrate has considerable intuitive appeal. Table 7 considers this explanation in more detail. Table 7 indicates that customers in the no-migration group are somewhat older, have lower income, and are less likely to have children than customers in the migration group. However, the lower half of Table 7 shows that these factors had little impact on differences in purchase incidence utility. Thus, these characteristics do not explain why the migration group purchased higher volumes.

CONCLUSION

Modeling Migration

We develop a model of customer channel migration and use it to (1) investigate the general nature of channel migration and (2) diagnose the migration process that occurred for a major, multichannel retailer between 1998 and 2001. We integrate several factors into our model, including the following:

- A model of purchase volume and channel selection;
- The effect of marketing and channel experience on purchase volume and channel selection;
- Customer heterogeneity in marketing, experience effects, base purchase volumes, and channel preference; and
- Controls for observed customer characteristics, seasonality, and market trends.

Moreover, our approach can accommodate large numbers of communications with dynamic effects in the presence of heterogeneity in consumer response. We find the interplay of these factors to be important because the estimated effects of dynamics are affected considerably by failure to capture unobserved heterogeneity in response.

By integrating all the foregoing factors into a single model, we can examine their effects on Web migration and purchase volumes. In our data, marketing is associated with both a migration to the Web and increased sales volumes. In contrast, we find that migrating customers were not inherently heavy users who were attracted to the Web, because there is little difference in baseline sales between those who migrated and those who did not. Moreover, Web experience does not increase sales volume; on the contrary, it is associated with lower future sales volume. This latter result contrasts with the common wisdom reported in the trade press (*Inter@active Week* 2000).

Summary of Results and Managerial Implications

We find (1) a negative long-term association between Internet usage and sales and (2) limited loyalty effects for purchases made on the Internet. Our conjecture is that migration to the Internet lowers switching costs (Brynjolfsson and Smith 2000), making it easier to compare products across firms. In addition, Internet interactions decrease the likelihood that users interact with people on the telephone

Table 7

CUSTOMER CHARACTERISTIC DIFFERENCES BETWEEN MIGRATION AND NO-MIGRATION GROUPS AND THEIR CONTRIBUTION TO PURCHASE INCIDENCE

		No-Migration Group		*Migration Group*	
		M	*SE*	*M*	*SE*
Demographics	Age (years)	50.95	.71	45.80	1.28
	Income ($10K)	8.49	.24	9.14	.47
	Children (%)	24.36	2.43	27.54	5.42
Contribution to customer characteristic utility	Intercept	–1.81	.03	–2.09	.06
	Age	.09	.001	.08	.002
	Income	.20	.01	.21	.01
	Children	–.02	.002	–.02	.003
	Overall	–1.53	.03	–1.81	.07

or in the store. The attendant reduction in personal service could lead to lower loyalty (Ariely, Lynch, and Moon 2002; Kacen, Hess, and Chiang 2003). Thus, the notion that migration is unqualifiedly positive because it lowers costs and increases demand should be tempered by the admonition that it can be negatively associated with long-term purchase patterns.

Another novel result is our finding of decreasing returns for communications in the purchase volume model. Because e-mails are virtually costless, it could be tempting to think that the optimal e-mail strategy is to e-mail customers daily. However, decreasing returns imply that a pulsing strategy might be more effective. That is, total response can be higher by sending e-mails intermittently and getting their full impact than by continually e-mailing and diminishing their effectiveness (see Blattberg, Kim, and Neslin 2008). Note also that customers react differently to the same marketing stimuli. For example, firms need to learn which types of customers will be unreceptive to the Internet and to the company if they are channeled to the Internet.

Limitations and Extensions

Although the firm we analyze can be characterized as a "typical" retailer, channel migration can be affected by industry, product line, marketing policy, customer base, and time. For example, Zhang and Wedel (2004) find high Internet loyalty for grocery goods, suggesting positive state dependence in that industry. This can be explained by the list feature offered by online grocers, in which consumers invest considerable effort setting up a shopping list to facilitate subsequent shopping. In addition, we consider data from 1998 to 2001. It is possible that firms have since addressed the lack of a human interface and improved cross-selling on the Internet.

Note also that we analyze secondary data and not a controlled experiment. Accordingly, it is not possible to make strong causal claims or rule out all alternative explanations. An alternative explanation is selection bias. For example, selectivity could become a problem in our channel choice model if unobserved variables governing the receipt of e-mails are correlated with unobserved variables determining channel choice (Manchanda, Rossi, and Chintagunta 2004). For example, an unobserved factor could be customer propensity to favor the Internet. This would make customers more likely to reveal to the company their e-mail addresses and more likely to use the Internet to make purchases. This would induce a spurious correlation between receipt of e-mails and channel selection. Although we cannot definitively rule out this possibility, we believe that selectivity is not a severe problem in our case for three reasons.

First, we modeled selectivity with observed variables by incorporating surrogates for Internet propensity explicitly in the choice model. Foremost among these is an individual-specific trend term (ξ^{te}_{w1i} in Equation 16). This term controls for changes in customers' Internet propensity over time. We also include demographics, such as age and income, that could be correlated with Internet propensity. Indeed, we find that younger people are more likely to choose the Internet. Although the company in this case did not explicitly target e-mails on the basis of previous behavior, previous RFM factors, including lagged volume and incidence, lagged channel usage, and time since last purchase, might be related to Internet propensity. For example, people who recently bought on the Internet may evidence a greater propensity to that channel. Despite the inclusion of heterogeneous trend, demographics, and prior behaviors, we still find a significant e-mail coefficient with regard to channel choice (Table 3).

Second, several additional factors mitigate the potential correlation between unobserved factors affecting channel choice and the receipt of e-mails. For example, many customers who received e-mails did not migrate to the Internet (66%), and many customers who migrated to the Internet did not receive e-mails (23%). If there were a strong correlation between unobserved factors, these patterns would be difficult to obtain. In addition, there is considerable variation in the number of e-mails received from week to week (some weeks the customer receives no e-mails, some weeks the customer receives one, some weeks the customer receives two, and so on). This pattern of variation differs from what might be expected in terms of changes in Internet propensity over time. This also suggests a weak link between unobserved factors governing e-mail receipt and Internet choice. Furthermore, e-mails were not sent to anyone until well after some people commenced their Web purchases, further suggesting that the contemporaneous correlation between unobserved factors governing receipt and choice is likely not high. Finally, the firm indicated that it did not target its e-mails. Rather, e-mail addresses are collected from prior purchases (both Web and catalog), and then communications are sent to all names on the lists. Together, these factors suggest a low correlation between unobserved factors affecting e-mail receipt and Internet choice.

Third, to the extent these arguments indicate that we have included effective surrogates for Internet propensity in the model and that the factors we do not observe that influence receipt of e-mails and channel selection behaviors are not highly correlated, selectivity effects will be minimal. Perhaps for similar reasons, Gönül and Shi (1998) find selectivity (albeit in a catalog context) to be insignificant.

We also note that our finding that e-mails could influence choice of the Internet has both face and convergent validity. In terms of face validity, e-mails typically contain links to the company Web site, so it is easy for the customer to transfer to the Web site and make a purchase there. In terms of convergent validity, other researchers have found associations between e-mails and Internet usage (Knox 2005; Lohse, Bellman, and Johnson 2000).

It is theoretically possible to control for selectivity bias by appending two selection equations (one for e-mails and one for catalogs) to our three-equation model of channel selection, purchase volume, and incidence. However, this may be infeasible in practice because the three-equation model is already heavily parameterized. In light of these considerations and the discussion in the foregoing paragraphs, the cost of selectivity controls (poor reliability and convergence) exceeds the value of additional insights that might accrue. However, although our previous discussion would suggest that selectivity is not problematic in our context, we cannot empirically rule out the potential for selectivity bias. Accordingly, further research on this problem is warranted.

Several additional research extensions exist. Perhaps the most pressing are to (1) consider a richer array of communications attributes and (2) design an optimal contact strategy predicated on this model. With respect to the latter, it may be possible to develop an optimization algorithm in which a marketer selects from among a set of communications to optimize demand and overtouching (due to saturation). It would also be of interest to analyze the effect of e-mails on the decision to unsubscribe. Finally, Morrison and Roberts (1998) note that industry characteristics moderate the relationship between migration and sales, so this is also an area of interest. We hope that our work helps managers understand the role of multiple channels and marketing communications on demand and sparks additional research into the phenomenon of channel migration.

APPENDIX A: LIST OF VARIABLES IN THE MODEL

Customer Variables

- Int_i = Intercept: household-specific random effect.
- Age_i = Age: age of customer i in years.
- Inc_i = Income: income of customer i.
- $Child_i$ = Children: equals 1 if the household has children and 0 if otherwise.

Experience Variables

- $Wuse_{it}$ = Web usage: Log(1 + number of Web purchases made by customer i up to period t – 1). We use a log rather than a quadratic function to capture diminishing marginal effects of this variable because the log function has fewer parameters.
- $Since_{it}$ = Recency: the number of periods since customer i last made a purchase before period t.
- $Diff_{it}$ = State dependence: equals 1 if customer i's last purchase was with a catalog and –1 if customer i's last purchase was on the Web.
- $Lweb_{it}$ = Lagged Web sales: In the incidence model, this is an indicator variable that represents whether a household bought on the Web in month t – 1. In the conditional order-size model, this represents the dollar volume purchased from the Web on the last purchase occasion. In the selection model, this represents the purchase volume, in dollars, from the Web in month t – 1.
- $Lcat_{it}$ = Lagged catalog sales: In the incidence model, this is an indicator variable that represents whether a household bought from the catalog in month t – 1. In the conditional order-size model, this represents the dollar volume purchased from the catalog on the last purchase occasion. In the selection model, this represents the purchase volume, in dollars, from the catalog in month t – 1.

Note that the effect of Wuse does not decay with time because time-varying usage behavior is captured with Lweb and Lcat. Allowing both to vary over time would confound the permanent and transient effects.

Marketing Variables

- Cat_{it} = Catalog stock: weighted summation of previous catalogs for customer i in period t (see Appendix C).
- $E\text{-}mail_{it}$ = E-mail stock: weighted summation of previous e-mails for customer i in period t (see Appendix C).
- Cat_Cat_{it} = Catalog saturation stock: weighted summation of previous catalog × catalog interactions for customer i in period t (see Appendix C).
- $E\text{-}mail\_E\text{-}mail_{it}$ = E-mail saturation stock: weighted summation of previous e-mail × e-mail interactions for customer i in period t (see Appendix C).
- $Cat\_E\text{-}mail_{it}$ = Catalog × e-mail interaction stock: weighted summation of previous e-mail × catalog interactions for customer i in period t (see Appendix C).

Time Variables

- $Trend_t$ = Time trend: month index, t = 1, ..., 48.
- JlF = February/July seasonal: equals 1 if period t is July or February and 0 if otherwise.
- Oct_t = October seasonal: equals 1 if period t is October and 0 if otherwise.
- Nov_t = November seasonal: equals 1 if period t is November and 0 if otherwise.
- Dec_t = December seasonal = equals 1 if period t is December and 0 if otherwise.

APPENDIX B: DERIVATION OF INTERACTIONS MODEL

We can write the total interaction effect across all communications by summing Equation 12:

$$\text{(B1)}\quad \text{Total\_Interaction\_Effect}_{it} = \sum_{c=1}^{C}\sum_{c'=c}^{C} \beta_{cc'}\lambda_c^{r_{ict}}\lambda_{c'}^{r_{ic't}} d_{ict}d_{ic't}.$$

Note that this sum is over all combinations, not permutations. This is because the interaction between communication c and c′ is the same as that between c′ and c.

We model the $\beta_{cc'}$ as functions of the communications attributes:

$$\text{(B2)}\qquad \beta_{cc'} = \sum_{m=1}^{M}\sum_{m'=1}^{M} \pi_{mm'}a_{cm}a_{c'm'},$$

where $\pi_{mm'}$ measures how the interaction between communications c and c′ is influenced by communication c's indicator variable for attribute m and c′'s indicator variable for attribute m′. In our application, we have M = 2, so Equation B2 can be written as

$$\text{(B3)}\quad \beta_{cc'} = \pi_{11}a_{c1}a_{c'1} + \pi_{12}a_{c1}a_{c'2} + \pi_{21}a_{c2}a_{c'1} + \pi_{22}a_{c2}a_{c'2}.$$

Now, let

- $\theta_1 = \pi_{11}$ = interaction if both communications are catalogs,
- $\theta_2 = \pi_{22}$ = interaction if both communications are e-mails, and
- $\theta_3 = \pi_{12} = \pi_{21}$ = interaction if one communication is an e-mail and the other is a catalog.

We set $\pi_{12} = \pi_{21}$ because the interaction between a catalog and an e-mail is the same as the interaction between an e-mail and a catalog. Substituting Equation B3 into Equation B1 and using the definitions of θ to simplify, we find the following expression for the total interaction effect:

(B4) $\text{Total\_Interaction\_Effect}_{icc't}$

$$= \sum_{c=1}^{C}\sum_{c'=c}^{C} \beta_{cc'}\lambda_c^{r_{ict}}\lambda_{c'}^{r_{ic't}} d_{ict} d_{ic't}$$

$$= \sum_{c,c'\text{both catalogs}} \theta_1 \lambda_{cat}^{r_{ict}}\lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't}$$

$$+ \sum_{c,c'\text{both e-mails}} \theta_2 \lambda_{e\text{-mail}}^{r_{ict}}\lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't}$$

$$+ \sum_{\substack{c\text{ is a catalog}\\ c'\text{ is an e-mail}}} \theta_3 \lambda_{cat}^{r_{ict}}\lambda_{e\text{-mail}}^{r_{ic't}} d_{ict} d_{ic't}.$$

This corresponds to Equations 13 and 14 in the text. The first term in Equation B4 is the catalog × catalog interaction effect (Cat_Cat_{it}), the second term is the e-mail × e-mail interaction ($\text{E-mail\_E-mail}_{it}$), and the last term is the catalog × e-mail interaction (Cat_E-mail_{it}).

APPENDIX C: DERIVATION OF STOCK VARIABLES

The computation of the direct effects in Equations 11a and 11b and the interaction effects in Equations 14a–14c is difficult because of the large number of catalogs and e-mails in our data. The computational complexity arises because we need to compute aggregates involving a large number of communication dummies for each observation in the data and for each sampled value of the discount terms within the MCMC iterations. However, considering that all communications of a given type (i.e., catalogs or e-mails) a person receives within a month are exchangeable (i.e., they have the same effect on all subsequent observations), we can use a much simpler representation that involves a recursive definition of the direct and interaction effects.

For the recursive definitions, we do not need the communications dummies. Because of the exchangeability of communications received in the same period, we need to know only two variables, C_{ij} and E_{ij}, which contain the number of catalogs and the number of e-mails, respectively, received by customer i in month t. We now show how these variables can be used to compute recursively the direct and interactions effects.

Direct Effects

According to Equation 11a, the direct effect of the catalogs can be written in terms of the communications dummies as follows:

(C1) $$\text{Cat}_{it} = \psi_{i1} \sum_{c\,\in\,\text{Catalogs}} \lambda_{cat}^{r_{ict}} d_{ict}.$$

Because ψ_{i1} is a coefficient that is common to all terms within the summation, we can ignore this coefficient and define RCat_{it} to be the raw catalog effect:

(C2) $$\text{RCat}_{it} = \sum_{c\,\in\,\text{Catalogs}} \lambda_{cat}^{r_{ict}} d_{ict}.$$

Because of the exchangeability of catalogs received in a given month, all catalogs in a month will have the same τ_{ict} variable. Thus, the direct effect of catalogs in Month 1 can alternatively be written as

(C3) $$\text{RCat}_{i1} = \sum_{c\,\in\,\text{Catalogs}} d_{ic1}.$$

This is the number of catalogs received in Month 1 by customer i, as $r_{ic1} = 0$ for all catalogs in Period 1. Thus, we have the following:

(C4) $$\text{RCat}_{i1} = C_{i1}.$$

The direct effect of the catalogs in Month 2 is given by

(C5) $$\text{RCat}_{i2} = C_{i2} + \lambda_{cat}C_{i1} = C_{i2} + \lambda_{cat}\text{RCat}_{i1}.$$

Thus, we can compute the direct effect recursively by discounting the previous period's direct effect and by adding the direct effect due to all catalogs received in the current month. Generically, the recursive scheme results in the following representation:

(C6) $$\text{RCat}_{it} = C_{it} + \lambda_{cat}\text{RCat}_{it-1}.$$

The computational complexity is reduced considerably because on any given observation, only a single term is added to an already computed value obtained from the previous period.

The direct effect for e-mails can analogously be written as

(C7) $$\text{RE-mail}_{it} = E_{it} + \lambda_{e\text{-mail}}\text{RE-mail}_{it-1},$$

where RE-mail_{it} refers to the raw e-mail direct effect. We can compute the total e-mail direct effect, E-mail_{it}, by multiplying RE-mail_{it} with the coefficient ψ_2.

Interaction Effects

We begin by focusing on the Cat_Cat_{it} interaction effect. According to Equation 14a, we can write this as

(C8) $$\text{Cat\_Cat}_{it} = \theta_1 \sum_{cc'\,\in\,\text{Catalogs}} \lambda_{cat}^{r_{ict}}\lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't}.$$

Again, ignoring the θ_1 coefficient, we define

(C9) $$\text{RCat\_Cat}_{it} = \sum_{cc'\,\in\,\text{Catalogs}} \lambda_{cat}^{r_{ict}}\lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't}.$$

Given the exchangeability of catalogs received within the same period, considerable simplifications result. For example, for the first period, we can write

(C10) $$\text{RCat\_Cat}_{i1} = \frac{C_{i1}(C_{i1}-1)}{2}$$

to represent the interaction effects between all catalog pairs received in Month 1. Similarly, for Month 2, we can write the interaction effect as

(C11) $$\text{RCat\_Cat}_{i2} = \frac{C_{i2}(C_{i2}-1)}{2} + C_{i2}\lambda_{cat}C_{i1} + \lambda_{cat}^2 \frac{C_{i1}(C_{i1}-1)}{2}.$$

The first term in Equation C11 represents the interaction among the catalogs received in the current month. The second term represents the interaction between the current catalogs and the catalogs received in Month 1. The third term represents the discounted impact of the interaction effects between catalog pairs received in the previous month. We can also write this equation as

$$\text{(C12)} \qquad \text{RCat\_Cat}_{i2} = \frac{C_{i2}(C_{i2}-1)}{2} + C_{i2}\lambda_{cat}\text{RCat}_{i1} + \lambda_{cat}^2\text{RCat\_Cat}_{i1}.$$

Thus, for an arbitrary period t, we can write the overall interaction effect as

$$\text{(C13)} \qquad \text{RCat\_Cat}_{it} = \frac{C_{it}(C_{it}-1)}{2} + C_{it}\lambda_{cat}\text{RCat}_{it-1} + \lambda_{cat}^2\text{RCat\_Cat}_{it-1}.$$

It is clear that this recursive scheme considerably reduces the computational burden. Rather than computing products of terms involving all catalog pairs in the data, we now simply add a single term to previously computed quantities to obtain the requisite interaction effect.

Using similar logic, we can show that the e-mail interaction effects can be recursively computed using the scheme

$$\text{(C14)} \qquad \text{RE-mail\_E-mail}_{it} = \frac{E_{it}(E_{it}-1)}{2} + E_{it}\lambda_{e\text{-mail}}\text{RE-mail}_{it-1} + \lambda_{e\text{-mail}}^2\text{RE-mail\_E-mail}_{it-1},$$

and the catalog-e-mail interaction effects can be computed using the formula

$$\text{(C15)} \qquad \text{RCat\_E-mail}_{it} = \text{RCat}_{it} \times \text{RE-mail}_{it}.$$

Note that the recursive schemes for the interaction effects involve the computed quantities for the direct effects, and thus the direct effect terms must be computed before beginning the computation of the interaction effects.

APPENDIX D: PRIORS AND FULL CONDITIONAL DISTRIBUTIONS

Let $\mathbf{u}_{it} = \{b_{it}^*, q_{it}^*, w_{it}^*\}$, $\mathbf{e}_{it} = \{e_{bit}, e_{qit}, e_{wit}\}$, and the vector $\boldsymbol{\lambda} = \{\lambda_{cat}^b, \lambda_{e\text{-mail}}^b, \lambda_{cat}^q, \lambda_{e\text{-mail}}^q, \lambda_{cat}^w, \lambda_{e\text{-mail}}^w\}$. Then, conditional on $\mathbf{u}_{it}$, we have a nonlinear mixed model specification given by

$$\mathbf{u}_{it} = \mathbf{X}_{it}(\lambda)\boldsymbol{\mu} + \mathbf{Z}_{it}(\lambda)\gamma_i + \mathbf{e}_{it},$$

where the matrix $\mathbf{X}_{it}(\boldsymbol{\lambda})$ contains all the variables in the three equations and $\mathbf{Z}_{it}(\boldsymbol{\lambda})$ contains a subset of the variables in $\mathbf{X}_{it}(\boldsymbol{\lambda})$, whose coefficients are assumed to vary across individuals. The conditioning on $\boldsymbol{\lambda}$ highlights that the communication variables are composed from the decay parameters and thus are "random." The errors are $\mathbf{e}_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\gamma}_i \sim t_\omega(\mathbf{0}, \boldsymbol{\Gamma})$, where ω is the degree of freedom for the t distribution.

Priors

We place diffuse but proper priors on the unknown parameters. The prior for $\boldsymbol{\mu}$ is multivariate normal, $N(\boldsymbol{\eta}, \mathbf{C})$. The covariance matrix $\mathbf{C}$ is diagonal with large values (1000) for the variances to reflect a lack of precise knowledge regarding the population mean, and $\boldsymbol{\eta} = \mathbf{0}$. We assume a Wishart prior $W[\rho,(\rho\boldsymbol{\Omega})^{-1}]$ for the precision matrix $\boldsymbol{\Gamma}^{-1}$. In our parametrization of the Wishart, the matrix $\boldsymbol{\Omega}$ can be considered the expected prior variance of the random effects $\boldsymbol{\gamma}_i$. Smaller values for ρ correspond to more diffuse prior distributions. We set ρ to be the number of random effects across the three equations and set $\boldsymbol{\Omega}$ as identity. For the covariance matrix of the errors, $\boldsymbol{\Sigma}$, we use the decomposition $\boldsymbol{\Sigma} = \mathbf{DRD}$, where $\mathbf{D}$ is a diagonal matrix containing the standard deviations in $\boldsymbol{\Sigma}$ and $\mathbf{R}$ is the corresponding correlation matrix. For identification, we set $\sigma_{11} = 1$ and $\sigma_{33} = 1$. Let $\chi = \log(\sigma_{22})$. We assume that $\chi \sim N(0, 1)$. The correlation matrix $\mathbf{R}$ has three nonredundant parameters. We assume that the prior for these correlations is the product of truncated univariate normal tn(0, 1) distributions, where the truncation is over the interval [–1, 1], together with the joint restriction that the resulting $\mathbf{R}$ matrix is a proper correlation matrix (i.e., it is positive definite). Finally, note that each element of $\boldsymbol{\lambda}$ lies in the interval [0, 1]. Let $\boldsymbol{\varphi}$ be the vector obtained by applying the logit transform on each element of $\boldsymbol{\lambda}$. We assume independent univariate normal priors over each of the elements in $\boldsymbol{\varphi}$. For example, we first transform λ_{cat}^q into $\phi_{q1} = \ln[\lambda_{cat}^q/(1 - \lambda_{cat}^q)]$ and then specify a N(0, 1) prior for ϕ_{q1}.

Full Conditional Distributions

First, we use a mini-Gibbs sampler for generating the latent and partially latent variables in $\mathbf{u}_{it} = \{b_{it}^*, q_{it}^*, w_{it}^*\}$ as part of a data augmentation step of the MCMC algorithm. We draw each latent variable from a univariate conditional normal distribution obtained from the joint trivariate normal distribution of $\mathbf{u}_{it}$. We obtain the full conditional for b_{it}^* conditioned on the values of $\{q_{it}^*, w_{it}^*\}$ and draw b_{it}^* for each observation from its conditional normal distribution that is right truncated at 0 if $b_{it} = 0$ and left truncated at 0 if $b_{it} = 1$. The full conditional for q_{it}^* is conditioned on the values of $\{b_{it}^*, w_{it}^*\}$, and we draw q_{it}^* from its conditional normal distribution if the observed quantity is 0 and set it equal to $\log(q_{it})$ if otherwise. Finally, we draw $w_{it}^*|\{b_{it}^*, q_{it}^*\}$ from its conditional normal distribution that is right truncated at 0 if $w_{it} = 0$ and left truncated at 0 if $w_{it} = 1$, and we draw it from its conditional normal without truncation for observations on which no purchase is observed.

Second, the full conditional for $\boldsymbol{\mu}$ is multivariate normal, given the conjugacy of the priors. We define the adjusted vector of latent variables, $\mathbf{u}_{it\mu} = \mathbf{u}_{it} - \mathbf{Z}_{it}\boldsymbol{\gamma}_i$. The posterior distribution for $\boldsymbol{\mu}$ is given by $\boldsymbol{\mu} \sim N(\mathbf{V}_\mu\Sigma_{i=1}^{I}\Sigma_{t=1}^{n_i}\mathbf{X}_{it}'\Sigma^{-1}\mathbf{u}_{it\mu}, \mathbf{V}_\mu)$, where $\mathbf{V}_\mu^{-1} = \Sigma_{i=1}^{I}\Sigma_{t=1}^{n_i}\mathbf{X}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{X}_{it} + \mathbf{C}^{-1}$, n_i indicates the number of observations for individual i, and I denotes the number of customers.

Third, we write the full conditional distribution for the random effects $\boldsymbol{\gamma}_i$ using the scale mixtures of normal representation for the t_ω distribution. This involves introducing a random variable $\kappa_i \sim \text{gamma}(\omega/2, \omega/2)$ and letting $\gamma_i \sim N(\mathbf{0}, \kappa_i^{-1}\boldsymbol{\Gamma})$. We define $\mathbf{u}_{it\gamma_i} = \mathbf{u}_{it} - \mathbf{X}_{it}\boldsymbol{\mu}$. The posterior distribution is $\gamma_i \sim N(\mathbf{V}_{\gamma_i}\Sigma_{t=1}^{n_i}\mathbf{Z}_{it}\Sigma^{-1}\mathbf{u}_{it\gamma_i}, V_{\gamma_i})$, where $\mathbf{V}_{\gamma i}^{-1} = \Sigma_{t=1}^{n_i} \mathbf{Z}'_{it}\boldsymbol{\Sigma}^{-1}\mathbf{Z}_{it} + \kappa_i\boldsymbol{\Gamma}^{-1}$.

Fourth, we use the Metropolis method to make independent draws for the elements in $\boldsymbol{\varphi}$, the vector of the transformed decay parameters. The likelihood is

$$L(\boldsymbol{\varphi}) \propto \prod_{i,t} \exp\left\{-\frac{1}{2}[\mathbf{u}_{it}^* - \mathbf{X}_{it}(\boldsymbol{\varphi})\boldsymbol{\mu} - \mathbf{Z}_{it}(\boldsymbol{\varphi})\boldsymbol{\lambda}_i]' \right.$$
$$\left. \sum{}^{-1}(\chi, \mathbf{R})[\mathbf{u}_{it}^* - \mathbf{X}_{it}(\boldsymbol{\varphi})\boldsymbol{\mu} - \mathbf{Z}_{it}(\boldsymbol{\varphi})\boldsymbol{\lambda}_i]\right\}.$$

Given the normal prior for $p(\boldsymbol{\varphi})$, the posterior is proportional to $L(\boldsymbol{\varphi})p(\boldsymbol{\varphi})$. Because the prior is not conjugate to the likelihood, we use a random-walk Metropolis algorithm to draw each element independently. For generating candidate draws, we use a normal proposal distribution centered on the previous draw and with a variance of .02.

Fifth, we use a Metropolis step to generate the log standard deviation, χ. Given the likelihood in the previous step and the normal prior for χ, we use a random-walk Metropolis step with the proposal distribution centered on the previous draw and with a proposal variance of .1 that is tuned to obtain rapid mixing.

Sixth, the correlations are bounded and constrained because of positive definiteness requirements. The full conditional for any correlation is not completely known because of the positive definiteness constraint. Therefore, we use the guided-walk Metropolis algorithm (Gustafson 1998) to generate each correlation separately. In generating the candidate correlation from a normal proposal distribution, we ensured that each correlation was obtained from an interval that kept the correlation matrix $\mathbf{R}$ positive definite (see Barnard, McCulloch, and Meng 2000).

Seventh, the full conditional for $\boldsymbol{\Gamma}^{-1}$ is Wishart and is given by $\boldsymbol{\Gamma}^{-1} \sim W(\rho + I, \mathbf{S})$, where ρ is the prior degrees of freedom, I is the number of customers in the data, and $\mathbf{S} = (\rho\boldsymbol{\Omega} + \Sigma_{i=1}^{I}\kappa_i\gamma_i\gamma'_i)^{-1}$.

REFERENCES

Abraham, Magid M. and Leonard M. Lodish (1993), "An Implemented System for Improving Promotion Productivity Using Store Scanner Data," *Marketing Science*, 12 (3), 248–69.

Ariely, Dan, John Lynch, and Youngme Moon (2002), "Taking Advice from Smart Agents: The Advice Likelihood Model," paper presented at Marketing Science Institute/Duke Joint Conference on Customer Relationship Management, Durham, NC (January 31–February 1).

Barnard, John, Robert McCulloch, and Xiao-Li Meng (2000), "Modeling Covariance Matrices in Terms of Standard Deviations and Correlations with Applications to Shrinkage," *Statistical Sinica*, 10 (4), 1281–1311.

Belsley, David A., Edwin Kuh, and Roy E. Welsch (1980), *Regression Diagnostics*. New York: John Wiley & Sons.

Biyalogorsky, Eyal and Prasad Naik (2003), "Clicks and Mortar: The Effect of Online Activities on Offline Sales," *Marketing Letters*, 14 (1), 21–32.

Blattberg, Robert C., Byung-Do Kim, and Scott A. Neslin (2008), *Database Marketing: Analyzing and Managing Customers*. New York: Springer, forthcoming.

Brynjolfsson, Erik and Michael Smith (2000), "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 46 (4), 563–85.

Campbell, Deb, Randy Erdahl, Dough Johnson, Eric Bibelnieks, Michael Haydock, Mark Bullock, and Harlan Crowder (2001), "Optimizing Customer Mail Streams at Fingerhut," *Interfaces*, 31 (1), 77–90.

Chain Store Age (2001), "And Now, for the Main Event," (April), 32.

Deleersnyder, Barbara, Geyskens Inge, Katrijn Gielens, and Marnik G. Dekimpe (2002), "How Cannibalistic Is the Internet Channel? A Study of the Newspaper Industry in the United Kingdom and the Netherlands," *International Journal of Research in Marketing*, 19 (4), 337–48.

Fox, Edward J., Alan L. Montgomery, and Leonard M. Lodish (2004), "Consumer Shopping and Spending Across Retail Formats," *Journal of Business*, 77 (2), S25.

Gönül, Füsun and Meng Ze Shi (1998), "Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models," *Management Science*, 44 (9), 1249–62.

Gustafson, Paul (1998), "A Guided Walk Metropolis Algorithm," *Statistics and Computing*, 8 (4), 357–64.

Infoworld (2001), "Going Multichannel," (May 14), 46.

Inter@active Week (2000), "The Promise of Multichannel Retailing," (October 9), 50.

Jedidi, Kamel, Carl F. Mela, and Sunil Gupta (1999), "Managing Promotions and Advertising for Long-Term Profitability," *Marketing Science*, 18 (1), 1–22.

Jupiter Communications (2000), "Assessing the Digital Divide," (May 20), (accessed October 10, 2007), [available at http://www.jupiterresearch.com/bin/item.pl/research:concept/63/id=84367,pos=2/].

Kacen, Jacqueline J., James D. Hess, and Wei-yu Kevin Chiang (2003), "Bricks or Clicks? Consumer Attitudes Toward Traditional Stores and Online Stores," working paper, Bauer College of Business, University of Houston.

Knox, George (2005), "Modeling and Managing Customers in a Multichannel Setting," working paper, The Wharton School, University of Pennsylvania.

Kopalle, Praveen, Carl F. Mela, and Lawrence Marsh (1999), "The Dynamic Effect of Discounting on Sales: Empirical Analysis and Normative Pricing Implications," *Marketing Science*, 18 (3), 317–32.

Kumar, V. and Rajkumar Venkatesan (2005), "Who Are Multichannel Shoppers and How Do They Perform? Correlates of Multichannel Shopping Behavior," *Journal of Interactive Marketing*, 19 (Spring), 44–61.

Kushwaha, Tarun L. and Venkatesh Shankar (2005), "Multichannel Shopping Behavior: Antecedents and Implications for Channel and Customer Equity," working paper, Mays Business School, Texas A&M University.

Liang,Ting-Peng and Jin-Shiang Huang (1998), "An Empirical Study on Consumer Acceptance of Products on Electronic Market: A Transaction Cost Model," *Decision Support Systems*, 24 (1), 29–43.

Lohse, Gerald L., Steven Bellman, and Eric J. Johnson (2000), "Consumer Buying Behavior on the Internet: Findings from Panel Data," *Journal of Interactive Marketing*, 14 (1), 15–29.

Macé, Sandrine and Scott A. Neslin (2004), "The Determinants of Pre- and Postpromotion Dips in Scanner Data," *Journal of Marketing Research*, 41 (August), 339–50.

Manchanda, Puneet, Peter E. Rossi, and Pradeep Chintagunta (2004), "Response Modeling with Nonrandom Marketing-Mix Variables," *Journal of Marketing Research*, 41 (November), 467–78.

Morrison, Pamela D. and John H. Roberts (1998), "Matching Electronic Distribution Channels to Product Characteristics: The Role of Congruence in Consideration Set Formation," *Journal of Business Research*, 41 (3), 223–29.

Roberts, John H. and Glen L. Urban (1988), "Modeling Multiattribute Utility and Risk and Belief Dynamics for New Consumer Durable Brand Choice," *Management Science*, 34 (2), 167–85.

Shankar, Venkatesh, Amy Smith, and Arvind Rangaswamy (2003), "The Relationship Between Customer Satisfaction and Loyalty in Online and Offline Environments," *International Journal of Research in Marketing*, 20 (2), 153–75.

Thomas, Jacquelyn S. and Ursula Y. Sullivan (2005), "Managing Marketing Communications with Multichannel Customers," *Journal of Marketing*, 69 (October), 239–51.

The Wall Street Journal (2004), "Shoppers Who Blend Store, Catalog, Web Spend More," (September 3), A7.

——— (2006), "Web Sales to Surpass $200 Billion," (May 23).

Yulinsky, Corey (2000), "Multi-Channel Marketing," *McKinsey Marketing Practice*, (August), 1–9.

Zhang, Jie and Michel Wedel (2004), "The Effectiveness of Customized Promotions in Online and Offline Stores," working paper, Robert H. Smith School of Business, University of Maryland.