Customer Channel Migration

Asim Ansari

Carl F. Mela

Scott A. Neslin[1]

July 4, 2006

[1] Asim Ansari is Professor of Marketing, Columbia University Graduate School of Business. Carl Mela is Professor of Marketing, The Fuqua School of Business, Duke University. Scott Neslin is Albert Wesley Frey Professor of Marketing, Amos Tuck School of Business, Dartmouth College.

Customer Channel Migration

Abstract

We develop a model of customer channel migration and apply it to a retailer that markets over the Web and through catalogs. The model (1) identifies the key phenomena required to analyze customer migration, (2) shows how these phenomena can be modeled, and (3) 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. Results indicate that (1) Web purchasing is associated with lower subsequent purchase volumes relative to 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 email/email) as well as between different types of communications (catalog/email). We further find that, over the four-year period of our data, a Web-oriented "migration" segment emerged, and this group had higher sales volume. Our *post hoc* analysis suggests that marketing efforts and exogenous customer-level trends played key roles in forming these segments. We 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. We conclude with a discussion of implications both for academics and practitioners.

INTRODUCTION

Overview

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 BB, doubling the total revenue in 2003 (Wall Street Journal 2006). As such, multichannel customer management is becoming a pivotal component in firms' marketing strategy. In spite of this trend, we are aware of no empirical research detailing (1) how customers migrate between channels in a multi-channel environment, and (2) the role marketers play 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-term as customers gain experience with different channels or how marketing influences this evolution. Other studies have explored channel cannibalization (Biyalogorsky and Naik (2003); Deleersnyder, Geyskens, Gielens, and Dekimpe (2002). However, they do not model customer heterogeneity which is central to the task of customer management.[2] In this vein, Fox, Montgomery and Lodish (2002) develop an individual-level model of retail choice. However, our focus is across channel formats (Web vs. catalog) as opposed to store variety within a particular retail format (grocery, mass merchandise, or drug).

The forgoing discussion suggests that researchers are beginning to recognize the considerable economic and behavioral ramifications of customer channel migration. Yet 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?

---

[2] In a separate analysis not reported here, we found dynamic effects to be considerably larger when we did not include customer heterogeneity.

- What are the short- and long-term effects of channel usage on channel selection and demand? For example, do customers develop "channel loyalty" based on their channel usage experience?

- What role do marketing communications play 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 the substantive questions posed above. Our contribution is twofold: First, we (a) propose a set of key phenomena that are related to channel migration behavior, (b) show how these phenomena can be modeled, and (c) 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 regarding customer channel migration. One key finding is that Web use, controlling 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. First we describe the modeling framework and use it to identify key phenomena to be incorporated in the model. Next, we describe our model. Subsequently, we describe our data and report our results. Finally, we summarize key findings and conclude by offering managerial and research implications.

CHANNEL MIGRATION FRAMEWORK

Channel migration affects firm profit via its influence on cost and revenue. For instance, the Internet channel is said to be more cost efficient compared to traditional channels. While this might suggest that companies should migrate customers to the Web, the efficacy of this strategy depends upon how migration affects overall demand. Thus, understanding how marketing actions

are associated with demand is crucial in grasping how customer channel migration impacts firm profitability.[3] So, in Figure 1, we overview the demand-side characteristics of channel migration:

[Insert Figure 1 About Here]

We assume 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. Also, 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.[4]

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 (Kumar and Venkatesan 2005; Kuswaha and Shankar 2005; Thomas and Sullivan 2005; Chain Store Age 2001; Infoworld 2001; Inter@ctive Week, 2000; 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). One might conjecture that firms should therefore cultivate multichannel buying. However, the framework suggests several other possibilities regarding why the multichannel and Internet-loyal customers have high sales levels:

- 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, i.e., customers buy more from a firm in the long term when they buy on the Internet (Channel Selection => Purchase Volume).

---

[3] Channel migration can also affect revenues when prices differ across channels. In our data, prices are identical across channels.

[4] In Figure 1, we consider two forms of communication – catalogs and emails (because these are the instruments used by the firm in our data).

- 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, but 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), then the prescription to migrate persons to the Web could be counter-productive. Hence 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. The purchase incidence and order-size components of purchase volume are modeled using a type-II tobit specification and channel selection is modeled using a 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.[5] 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 b_{it}^* be a latent variable related to the decision regarding whether to buy. The type-II tobit specification can be written as follows:

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

---

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

$$(2) \qquad q_{it} = \begin{cases} Exp(q_{it}^*) & \text{if } b_{it}^* > 0 \\ 0 & \text{otherwise} \end{cases}$$

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.

Let w_{it}^* be a latent variable that represents the difference between the customer's latent utilities

for purchasing via the catalog and via the Internet. The binary probit model conditional on

purchase is then[6]:

$$(3) \qquad w_{it} = \begin{cases} Buy\ on\ Catalog, & \text{if } w_{it}^* > 0\ and\ b_{it}^* > 0 \\ Buy\ on\ Internet & \text{if } w_{it}^* \leq 0\ and\ b_{it}^* > 0 \end{cases}$$

Four groups of variables accommodate the various substantive issues identified in our framework:

customer characteristics, experience effects, communications effects and time effects. The

relationship between the latent variables and these groups can be written as:

$$b_{it}^* = Customer\ Characteristics_{bi} + Experience_{bit} + Communication_{bit} + Time\ Effects_{bt} + e_{bit}$$
$$(4) \quad q_{it}^* = Customer\ Characteristics_{qi} + Experience_{qit} + Communication_{qit} + Time\ Effects_{qi} + e_{qit}$$
$$w_{it}^* = Customer\ Characteristics_{wi} + Experience_{wit} + Communication_{wit} + Time\ Effects_{wi} + e_{wit},$$

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 distributed multivariate normal, $N(\mathbf{0}, \mathbf{\Sigma})$.

The off-diagonal elements in the covariance matrix $\mathbf{\Sigma}$ accommodate the contemporaneous

correlations among the three customer decisions. For identification, the variances of e_{bit} and e_{wit}

---

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

are set to 1, as 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.[7]

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 1 defines all variables used in this study). The direct impact of unobserved customer-specific variables is captured via individual-level random intercepts in the three equations. Letting the random intercepts be denoted by *Int*, we have,

$$
\begin{aligned}
Customer\ Characteristic_{bi} &= Int_{bi} + \xi_{b1}^{cc} Age_i + \xi_{b2}^{cc} Income_i + \xi_{b3}^{cc} Children \\
Customer\ Characteristic_{qi} &= Int_{qi} + \xi_{q1}^{cc} Age_i + \xi_{q2}^{cc} Income_i + \xi_{q3}^{cc} Children \\
Customer\ Characteristic_{wi} &= Int_{wi} + \xi_{w1}^{cc} Age_i + \xi_{w2}^{cc} Income_i + \xi_{w3}^{cc} Children.
\end{aligned}
$$

(5)

Experience Effects

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

$$
\begin{aligned}
Experience_{bit} &= \xi_{b1i}^{e} Lweb_{bit} + \xi_{b2i}^{e} Lcat_{bit} + \xi_{b3i}^{e} Wuse_{it} + \xi_{b4i}^{e} Since_{it} \\
Experience_{qit} &= \xi_{q1i}^{e} Lweb_{qit} + \xi_{q2i}^{e} Lcat_{qit} + \xi_{q3i}^{e} Wuse_{it} + \xi_{q4i}^{e} Since_{it} \\
Experience_{wit} &= \xi_{w1i}^{e} Lweb_{wit} + \xi_{w2i}^{e} Lcat_{wit} + \xi_{w3i}^{e} Wuse_{it} + \xi_{w4i}^{e} Since_{it} + \xi_{w5i}^{e} Diff_{it},
\end{aligned}
$$

(6)

where, the ξ terms represent random coefficients.

---

[7] We investigated non-stationarity 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.

The variables in (6) are defined differently in each equation. In the purchase incidence equation (*Experience*$_{bit}$), *Lcat* and *Lweb* are binary variables indicating whether the customer purchased from the firm in the preceding month via either the catalog or the Web. In the order-size equation (*Experience*$_{qit}$), *Lcat* and *Lweb* are defined 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 mimic the definitions of the dependent variables. The channel selection model (*Experience*$_{wit}$) incorporates both lagged volume and lagged channel selection effects.[8] In this equation, *Lcat* and *Lweb* are defined as the previous month's purchase volume from the catalog or the Web respectively. We also include *Diff*, which represents state dependence in selection. *Diff* is set to 1 if the previous purchase for the customer is a catalog purchase and to -1 if it is via the Internet.[9] *Since* is defined as the number of months elapsed since the previous purchase. This recency measure is included in all three experience equations.

Wuse$_{it}$, defined as Log(1+Web purchases to date), captures the permanent effect due to Web usage.[10] *Wuse*$_{it}$ is used in all three equations. By definition, this variable is independent of the duration between Web purchases because we seek to capture forgetting and other transient effects due to previous channel usage via the *Lweb* variables. We specify this to have diminishing marginal returns because, consistent with Bayesian Learning (Roberts and Urban 1988), we expect the very first usage to have a greater effect on behavior than the last usage. Finally, note that 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 emails or catalogs). Therefore, two different catalogs mailed at the same time

---

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

[9] This variable is derived from differencing the latent utilities for Internet choice and catalog choice, hence its value of ± 1. Thus ξ_{w5i} is added to w^* when a catalog is used, and subtracted from w^* when the Internet is used (see equation 4).

[10] Note that it is impossible to include a corresponding variable for catalog purchases to date, as there exists no information on the number of catalog purchases prior to the data. In contrast, the Web channel is new, so the amount of purchasing prior to our data is negligible for all households.

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 are considerable, totaling $C = 723$. Though 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 a) the characteristics of the communication (e.g., communications of like kind have the same effect, and b) 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 as well as its interaction with other communications, as there is likely to be decreasing marginal returns to these communications, reflected in a negative interaction.

Direct Communication Effects. The direct effect of communication c on customer i at time t is defined as:[11]

$$(7) \qquad\qquad Direct\_Effect_{ict} \;=\; \beta_{ic}\lambda_c^{r_{ict}} d_{ict}$$

The variable d_{ict} indicates whether customer i has received communication c on or before time t. It equals zero until the customer receives communication c, and equals one 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 time 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 zero and one; a large λ_c means that 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 will have higher values of β_{ic}.

One can sum equation (7) across all communications to compute the total direct effect of communications received by customer i as of time t:

---

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

$$(8) \qquad\qquad Total\_Direct\_Effect_{it} \; = \; \sum_{c \in C} \beta_{ic} \lambda_c^{r_{ict}} d_{ict}$$

Equation (8) implies one needs to estimate C direct communication effect parameters, β_{ic}, for incidence, order-size, and channel selection. As $C = 723$, 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} equal to 1 when communication c has attribute m, else 0 ($m=1,…,M$). Then, the communication effect for communication c can be expressed as:

$$(9a) \qquad\qquad \beta_{ic} = \sum_{m=1}^{M} \psi_{im} a_{cm},$$

and the decay parameter can be written as[12]:

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

As M is small relative to C, we achieve great parsimony while allowing different communication types to have different effects. In our application, we will use $M=2$ and distinguish between catalogs and emails.[13] Accordingly, we have,

$\quad a_{c1} \;=\; 1$, if communication c is a catalog; 0 if not.

$\quad a_{c2} \;=\; 1$, if communication c is an email; 0 if not.

We label as λ_{cat} the decay parameter for a communication $c \in$ *Catalogs* (the set of all communications that are catalogs) and this equals $\exp(\zeta_1)/(1+\exp(\zeta_1))$ for all catalogs. Similarly, we label λ_{email} to be the decay parameter for communications $c \in$ Emails (the set of all emails) and this equals $\exp(\zeta_2)/(1+\exp(\zeta_2))$ for all emails. Partitioning the communications into email and

---

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

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

catalog, using the above definitions for a_{c1}, a_{c2}, λ_{cat}, λ_{email} and substituting equations (9) into (8), it

follows that:

$$(10) \qquad Total\_Direct\_Effect_{it} = \sum_{c \in Catalogs} \psi_{i1} \lambda_{cat}^{r_{ict}} d_{ict} + \sum_{c \in Emails} \psi_{i2} \lambda_{email}^{r_{ict}} d_{ict}$$

We call the first sum the catalog direct effect, and the second sum the email direct effect:

$$(11a) \qquad Cat_{it} = \sum_{c \in Catalogs} \psi_{i1} \lambda_{cat}^{r_{ict}} d_{ict}$$

$$(11b) \qquad Email_{it} = \sum_{c \in Emails} \psi_{i2} \lambda_{email}^{r_{ict}} d_{ict}$$

Communication Interaction Effects. We define the interaction between communications c and

c' as:

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

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). As $\lambda_c \neq \lambda_{c'}$, (12) allows for order effects in that the interaction

between communication c and c' to 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.

As there are 723*722/2, or 261,003 potential communication interaction terms in our model,

we again model the $\beta_{cc'}$ as functions of the communications' underlying attributes. Acordingly,

we assign *pairs* of communications into three unique categories: both communications are

catalogs, both are emails, or one is a catalog and the other an email. In Appendix 2, we show that

this implies:

$$(13) \qquad Total\_Interaction\_Effect_{it} = Cat\_Cat_{it} + Email\_Email_{it} + Cat\_Email_{it}$$

11

where,

$$(14a) \qquad Cat\_Cat_{it} = \sum_{\substack{c,c' \\ both\ cata\log s}} \theta_1 \lambda_{cat}^{r_{ict}} \lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't}$$

$$(14b) \qquad Email\_Email_{it} = \sum_{\substack{c,c' \\ both\ emails}} \theta_2 \lambda_{email}^{r_{ict}} \lambda_{email}^{r_{ic't}} d_{ict} d_{ic't}$$

$$(14c) \qquad Cat\_Email_{it} = \sum_{\substack{c\ is\ a\ cata\log \\ c'\ is\ an\ Email}} \theta_3 \lambda_{cat}^{r_{ict}} \lambda_{email}^{r_{ic't}} d_{ict} d_{ic't}$$

Combining equations (11) and (14), we can write the total effect of communications on customer i at time t ($Communication_{it}$) as:

$$(15) \qquad Communication_{it} = Cat_{it} + Email_{it} + Cat\_Cat_{it} + Email\_Email_{it} + Cat\_Email_{it}$$

Equation (15) accommodates direct and interaction effects of communications, differential effects across forms or types of communications, and dynamics. In spite of 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 3 which 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 $p < 0.05$. 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).[14] Therefore:

---

[14] As 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.

$$Time\ Effects_{bit} = \xi_{b1i}^{te}Trend_t + \xi_{b2}^{te}Oct_t + \xi_{b3}^{te}Nov_t + \xi_{b4}^{te}Dec_t + \xi_{b5}^{te}JlF_t$$

(16) $$Time\ Effects_{qit} = \xi_{q1i}^{te}Trend_t + \xi_{q2}^{te}Oct_t + \xi_{q3}^{te}Nov_t + \xi_{q4}^{te}Dec_t + \xi_{q5}^{te}JlF_t$$

$$Time\ Effects_{wit} = \xi_{w1i}^{te}Trend_t + \xi_{w2}^{te}Oct_t + \xi_{w3}^{te}Nov_t + \xi_{w4}^{te}Dec_t + \xi_{w5}^{te}JlF_t,$$

where the ξ's are parameters to be estimated. Note 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 were adopting 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* population distribution for modeling

unobserved heterogeneity. The *t* is a robust alternative to the normal as it has fatter tails. We use

Bayesian methods for inference regarding the parameters. As the posterior distribution is not

completely known, we use MCMC techniques to obtain draws from the posterior distribution of

the unknowns. The priors and the full conditional distributions for the unknown parameters are

described in Appendix 4.

DATA

Data are provided by a retailer who sells consumer durable and apparel products in

mature categories over the Internet and via 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 while 1%

exclusively used the Internet and 62% used only catalogs. The entire dataset consists of 40,000

customers; we randomly select 500 customers for our analysis. This suggests 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 email 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, as the median purchase

frequency is about 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.[15] 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, as the condition indices for the regressor sets in each of our three equations are all below

30 (Belsley, Kuh and Welsch 1980, p. 105).[16] One item of note in Table 1 is the high level of

catalog mailing.

[Insert Table 1 About Here]

There is substantial channel migration evidenced in our data, and multi-channel

and Internet buyers purchase more. In 1998, a large proportion (96% of customers) made

more than 95% of their purchases on the catalog. By 2001 this share had fallen to 77%.

Moreover, those that purchased on the Internet as well as the catalog tended to buy more;

the mean purchase level of those that made more than 95% of their purchases on the

---

[15] Using monthly aggregation, multiple purchases are negligible, amounting to 0.29% 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.
[16] The most noticeable source of collinearity in our data is among the communications variables and their
interactions; correlations between these variables range up to 0.94. As collinearity increases standard
errors, our data afford a conservative test of our hypotheses.

catalog was \$267, in contrast to \$444 for those that made less than 95% of purchases via catalog.

<div align="center"><i>RESULTS</i></div>

Model Comparisons

We estimated four models. The first, M1, is the full model specified in equation (4) and incorporates heterogeneity using a multivariate t population distribution. The second model, M2, is again the full model, but 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 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 allows us to test the role of marketing dynamics. Table 2 displays the Log-marginal likelihoods based on the MCMC draws.

<div align="center">[Insert Table 2 About Here]</div>

The best model is M1, which includes t-heterogeneity as well as marketing and experience dynamics. M1's superiority to M2 suggests that the t-distribution, owing to its "fatter" tails, is better able to capture heterogeneity than the normal distribution. M1's superiority to M3 indicates that including marketing variables results in model improvement. Finally, M1's superiority to 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 re-estimate the model for the first 45 months (parameter estimates from this shorter period are

comparable to the full period estimates).[17] We then use these estimates to make both in-sample predictions for the first 45 months and out of sample predictions for the final 3 months. Figures 2a and 2b show the results for purchase volume and channel share respectively. The out-of-sample period begins in period 46. The figures indicate that calibration and holdout sample predictions are roughly equal among M1, M3, and M4. Computing the mean absolute errors we find no statistically significant difference across models for either calibration or holdout. Though the full model better explains the data, it forecasts no better than the null models. This is not surprising as even the simpler models are fairly rich in their accommodation of customer heterogeneity which is often crucial for good predictive performance.[18]

[Insert Figure 2 About Here]

Parameter Estimates

The parameter estimates for the best model (M1) are presented in Table 3 (bold indicates that the 95% posterior interval excludes zero). We discuss the results for incidence, order size, and channel selection separately.

[Insert Table 3 About Here]

Purchase Volume. Interestingly, few effects are significant (by significant, we mean 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 ($Wuse < 0$).[19] Several explanations for

---

[17] Note that our hold-out period is relatively short because emails 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 reliably gauge their effectiveness.

[18] We note that the log-marginal likelihood is substantially superior for M1 even though the MAE's are indistinguishable across models. This might reflect the difference in how the measures are computed.

[19] One could also interpret this result 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).

this result may exist. First, when persons 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, lack of contact with a sales agent may limit opportunities for cross-selling. The consequent attenuation in service-levels implies 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).[20] 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. Additionally, 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 those with higher incomes buy more often, consistent with intuition.

---

[20] It is important to note that the null result for *Lweb* could be the result of poor statistical power (since there are fewer Web purchases) although there are 19,064 observations in total.

Channel Selection. The effect of *Lcat* is positive, suggesting that persons 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 persons who have not purchased in a while are more likely to use the catalog. Email has a negative association with catalog selection, suggesting that emails are associated with increased use of the Internet.[21] The positive interaction between catalogs and catalogs suggests that increasing numbers of catalogs solidifies the customer as a catalog user. The positive interaction between emails and emails, when coupled with the negative direct effect on catalogs, suggests a diminishing marginal return to emails in terms of driving persons 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 older persons are less likely to use the Internet, consistent with previous research during the period of our data (Jupiter Communications 2000).

Finally, the decay parameter estimates are small, but non-zero. The average $(0.14+0.11)/2 = 0.125$ (or $1/8$), which implies an infinite horizon effect for the communications on choice of $1/(1-1/8)$, or $8/7$.[22] Thus, $1/7^{th}$ of the communications effect occurs in periods after the communication is received. While this appears small, this firm sends 45 catalogs per year. As such, the communication decay is tantamount to increasing the effect of these catalogs by $45/7 = 6.5$ additional catalogs compared to the case where communications were immediately perishable. Summed across persons, the revenue effects of these lagged factors are considerable. One wonders, further, if these decay effects would be slower if fewer communications were sent.

---

[21] It is possible that this result implies Web users get more emails, though our data supplier did not target emails. It might also be that email is a proxy for Internet access.
[22] The total effect is calculated as $1+0.125+0.125^2+0.125^3+ \ldots = 1/(1-0.125))=8/7$.

Communications Impulse Response Functions

The foregoing results suggest the effects of communications are many-fold (e.g., emails increase sales in the short-term, but also switch demand to the Internet which decreases sales in the long-term). The combined effect of these dynamics is not apparent, but can be determined by simulating the effect of an additional email or catalog communication on purchase behavior. To do this, we increment the number of emails 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 through month 35 after which we averaged these response functions over time.[23]

From these simulations, we find that a catalog generates an incremental $0.57 in revenue, and an email generates an additional $0.79 in the current period. It is interesting to speculate that the smaller incremental effect for the catalog might arise from the large number of catalogs sent per year (40), which may lead to decreasing returns. In contrast, few emails were sent during the interval of our data, suggesting that an incremental increase in emails might be more effective. Summing across all 10 periods, the total effect of a catalog is $0.48 whereas, for email it is $0.68.[24] Thus, the negative long-term effects are about 16-17% the magnitude of the short-term effects. This is smaller than negative effects of promotion observed in scanner data, which are roughly 40% (Jedidi, Mela and Gupta 1999; Macé and Neslin 2004).

Heterogeneity in the Effects of Marketing on Channel Migration

While the previous discussion centered on population-level effects, there are 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 0.04 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

---

[23] 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.

[24] Note 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.

95% interval for the random effects ranges from -0.60 to 0.28 (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 persons the effect of catalog is to induce them to buy on catalog, but for some persons, the highly negative direct effect of catalogs migrates them to the Internet (the lower 95% effect in the population is 0.04 + -0.60 = -0.56). A similar calculation for emails suggests that it is possible for emails to move persons to the catalog; the highest household-level effect in the 95% interval is 0.21. Relative to the -0.56 effect for a catalog migrating a household to the Web, this 0.21 is small. As such, there is a greater potential to move some persons to the Web with a catalog than it is to move persons to a catalog with email.

DIAGNOSING THE EMERGENCE OF THE INTERNET CHANNEL SEGMENT

Recall that our data provides evidence for (1) customers migrating toward the Internet, and (2) Internet and multi-channel users buying more. Though 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 more precisely ascertain why customers migrated toward 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 see how these factors changed for those that migrated between 1998 and 2001 (n = 69) versus those who did not (n = 312), and interpret these changes to determine why the migrations occurred.[25]

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

$E(w_{it}^*) = $ *Customer Characteristics*$_{wi}$ + *Experience*$_{wit}$ + *Communications*$_{wit}$ + *Time Effects*$_{wit}$. As

only one survey record exists for the each customer, the measured customer characteristics do not

---

[25] These two groups do not sum to 500 households because a) they only include those who bought in both 1998 and 2001 (we consider changes from year 1 to year 4), and b) do not include those who did not have sufficient data in year one for initialization of experience variables.

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. We thus calculate the averages of these utilities for both years, difference them, and then compare these differences between those who migrated from the catalog to the Internet between year 1 (1998) and year 4 (2001), versus those who did not. Table 4 displays the results.

[Insert Table 4 About Here]

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 those that migrated. It is tempting to argue this larger effect arises from our definition of a "migrater" as a household that used the Internet in later periods. However, it is important to note that the trend effect is measured 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 artificat 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 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:

[Insert Table 5 About Here]

Those who migrated were exposed to more marketing and switched more in response to it.[26]

As the migration group's response parameter for email is more negative than that for catalogs,

and the absolute level of change is higher, it would appear that email played the greater role.[27]

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.
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, as opposed to order-size conditioned on incidence, as most of the

variance in purchase volume is explained by the incidence model). Changes in these factors are

reported in Table 6. A positive entry means that changes in this factor contributed to higher sales

in year 4 compared to year 1. We also show the average customer characteristic utility (which

doesn't change over time) as well as the average sales response to catalog and email.

[Insert Table 6 About Here]

The table eliminates the "heavier-users" explanation for increased use because the average

customer characteristic utility is similar for both the no migration and migration groups; in fact,

this utility is slightly *lower* for the migration group. The data in the table also refute the "positive

experience" explanation as experience utility is trending downward for the migration group.

However, the table does support the "marketing" explanation. The increase in the marketing

utility offsets the decrease in experience utility. The result suggests that *marketing obscures the*

---

[26] 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.

[27] Note that additional email by itself is not sufficient to migrate customers to the Web. They must also respond to that email by moving to the Web. Of the 36.2% of non-migrating customers who received email, they received on average 2.6 additional emails per month and their channel selection response parameter averaged -0.30. Of the 76.7% of migrating customers who received email, they received on average 2.3 additional emails and their average channel selection response parameter was -0.54.

negative association between Internet usage and demand. This is very similar to the findings of Kopalle, Mela and Marsh (1999), who find that increased sales promotions can mask a receding brand baseline. Table 6 also shows that time effects are more negative for those that did not migrate, suggesting these users are buying diminished amounts over time.

Though we do not find much difference in latent customer characteristic utility between those 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.

[Insert Table 7 About Here]

Table 7 indicates that customers in the no migration group are somewhat older, have lower income, and are less likely to have children when compared to 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. So 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, multi-channel retailer between the years 1998 and 2001. We integrate a number of factors into our model, including:

- 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.

- 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, as the estimated effects of dynamics is 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 that were attracted to the Web, because there is little difference in baseline sales between those that migrate and those that do not. Moreover, Web experience does not increase sales volume; in fact 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 (i) a negative long-term association between Internet usage and sales and (ii) 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 decreases the likelihood that users interact with persons on the phone 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 associated negatively with long-term purchase patterns.

Another novel result is our finding of decreasing returns for communications in the purchase volume model. As emails are virtually costless, one might be tempted to think that the optimal email strategy is to email customers daily. However, decreasing returns imply that a pulsing strategy might be more effective. That is, total response can be higher by sending emails intermittently and getting their full impact, than by continually emailing and diminishing their effectiveness (see Blattberg, Kim, and Neslin 2004). It is also important to note 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

While 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 on-line grocers, wherein consumers invest considerable effort setting up a shopping list to facilitate subsequent shopping. Also, we consider data from 1998-2001. 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. One alternative explanation is selection bias. For example, selectivity could become a problem in our channel choice model if unobserved variables governing the receipt of emails are correlated with unobserved variables determining channel choice (Manchanda, Rossi and Chintagunta 2004). One such unobserved factor could be customer propensity to favor the Internet. This would make customers more likely to reveal to the company their email addresses, and more likely to use the Internet to make purchases. This would induce a spurious correlation between receipt of emails and channel selection. While we cannot definitively rule out this possibility, we believe selectivity is not a severe problem in our case for the following reasons:

- We modeled selectivity via observed variables by incorporating surrogates for Internet propensity explicitly in the choice model. Foremost among these is an individual-specific trend term (ξ_{w1i}^{te} 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

25

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 emails based on previous behavior, previous RFM factors including lagged volume and incidence, lagged channel usage, and time since last purchase might relate to Internet propensity. For example, those who recently bought on the Internet may evidence a greater propensity to that channel. Despite the inclusion of heterogeneous trend, demographics, and past behaviors, we still find a significant email coefficient with regard to channel choice (Table 3).

- Several additional factors mitigate the potential correlation between unobserved factors affecting channel choice and the receipt of emails. First, many customers who received emails did not migrate to the Internet (66%), and many customers that migrated to the Internet did not receive emails (23%). Were there a strong correlation between unobserved factors, these patterns would be difficult to obtain. Second, there is considerable variation in the number of emails received from week to week (some weeks the customer receives no emails, some weeks one, some weeks two, etc.). This pattern of variation differs from what one might expect in terms of changes in Internet propensity over time. This also suggests a weak link between unobserved factors governing email receipt and Internet choice. Third, emails were not sent to any person until well after some individuals commenced their web purchases, further suggesting that the contemporaneous correlation between unobserved factors governing receipt and choice is likely not high. Fourth, the firm indicates it did not target its emails. Rather, email addresses are collected from past purchases (both Web and catalog) and then communications are sent to all names on the lists. Together, these four factors suggest a low correlation between unobserved factors affecting email receipt and Internet choice.

- To the extent these arguments indicate that we have included effective surrogates for Internet propensity in the model, and that the factors we don't observe that influence

receipt of emails 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 emails could influence choice of the Internet has both face

and convergent validity. In terms of face validity, emails typically contain links to the company

website so it is easy for the customer to transfer to the website and make a purchase there. In

terms of convergent validity, other researchers have found associations between emails and

Internet usage (Knox 2005; Lohse, Bellman, Johnson 2000).

It is theoretically possible to control for selectivity bias by appending two selection equations

(one for emails; another for catalogs) to our three-equation model of channel selection, purchase

volume and incidence. Yet we 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, though our previous discussion

would suggest selectivity not to be problematic in our context, we can not empirically rule out the

potential for selectivity bias. Accordingly, more research on this problem is warranted.

Several additional research extensions exist. Perhaps the most pressing are to a) to consider a

richer array of communications attributes, and b) design an optimal contact strategy predicated

upon this model. With respect to the latter, one could envision an optimization algorithm wherein

a marketer selects from among a set of communications to optimize demand and over-touching

(due to saturation). Also of interest is an analysis of the effect of emails on the decision to

unsubscribe. Finally, Morrison and Roberts (1998) note industry characteristics moderate the

relationship between migration and sales, so this is also an area of future interest. We hope our

work helps managers to understand the role of multiple channels and marketing communications

on demand, and sparks additional research into the phenomenon of channel migration.

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 household has children, else equals 0.

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 quadratic function to capture diminishing marginal effects of this variable as the log function has fewer parameters.

$Since_{it}$ — Recency – Number of time 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, equals -1 if customer i's last purchase was on the Web.

$Lweb_{it}$ — Lagged Web Sales – In the <u>incidence model</u>, this is an indicator variable that represents whether or not a household bought on the Web in month t-1. In the <u>conditional order-size</u> model, this represents the dollar volume purchased from the Web on the last purchase occasion. In the <u>selection model</u>, this represents the purchase volume, in dollars, from the Web in month t-1.

$Lcat_{it}$ — Lagged Catalog Sales – In the <u>incidence model</u>, this is an indicator variable that represents whether or not a household bought form the catalog in month t-1. In the <u>conditional order-size</u> model, this represents the dollar volume purchased from the catalog on the last purchase occasion. In the <u>selection model</u>, this represents the purchase volume, in dollars, from the catalog in month t-1.

Note: The effect of Wuse is not decaying with time because time varying usage behavior is captured via LWeb and LCat. Allowing both to time vary 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 3).

Email$_{it}$	Email Stock – Weighted summation of previous emails for customer *i* in period *t* (see Appendix 3).
Cat_Cat$_{it}$	Catalog Saturation Stock – Weighted summation of previous catalog-catalog interactions for customer *i* in period *t* (see Appendix 3).
Email_Email$_{it}$	Email Saturation Stock – Weighted summation of previous email – email interactions for customer *i* in period *t* (see Appendix 3).
Cat_Email$_{it}$	Catalog-Email Interaction Stock – Weighted summation of previous email/catalog interactions for customer *i* in period *t* (see Appendix 3).
Time Variables	
Trend$_t$	Time trend – Month index, $t = 1, \dots, 48$.
JlF	February/July Seasonal – Equals 1 if period t is July or February, else equals 0.
Oct$_t$	October Seasonal – Equals 1 if period *t* is October; else equals 0.
Nov$_t$	November Seasonal – Equals 1 if period *t* is November; else equals 0.
Dec$_t$	December Seasonal – Equals 1 if period *t* is December; else equals 0.

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

$$Total\_Interaction\_Effect_{it} = \sum_{c=1}^{C} \sum_{c'=c}^{C} \beta_{cc'} \lambda_c^{r_{ict}} \lambda_{c'}^{r_{ic't}} d_{ict} d_{ic't} \tag{A2.1}$$

Note 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.

The $\beta_{cc'}$'s are modeled as functions of the communications attributes, as follows:

$$\beta_{cc'} = \sum_{m=1}^{M} \sum_{m'=1}^{M} \pi_{mm'} a_{cm} a_{c'm'} \tag{A2.2}$$

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 (A2.2) can be written as:

$$\beta_{cc'} = \pi_{11} a_{c1} a_{c'1} + \pi_{12} a_{c1} a_{c'2} + \pi_{21} a_{c2} a_{c'1} + \pi_{22} a_{c2} a_{c'2} \tag{A2.3}$$

Now, let

$$\begin{aligned}
\theta_1 &= \pi_{11} = && \text{Interaction if both communications are catalogs.} \\
\theta_2 &= \pi_{22} = && \text{Interaction if both communications are emails.} \\
\theta_3 &= \pi_{12} = \pi_{21} = && \text{Interaction if one communication is an email and the other is a} \\
& && \text{catalog.}
\end{aligned}$$

We set $\pi_{12} = \pi_{21}$ since the interaction between a catalog and an email is the same as the interaction between an email and a catalog. Substituting equation (A2.3) into equation (A2.1) and using the definitions of θ to simplify, we find the following expression for the total interaction effect:

$$\begin{aligned}
Total\_Interaction\_Effect_{icc't} &= \sum_{c=1}^{C} \sum_{c'=c}^{C} \beta_{cc'} \lambda_c^{r_{ict}} \lambda_{c'}^{r_{ic't}} d_{ict} d_{ic't} \\
&= \sum_{c,c' both\ catalogs} \theta_1 \ \lambda_{cat}^{r_{ict}} \lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't} + \sum_{c,c' both\ emails} \theta_2 \ \lambda_{email}^{r_{ict}} \lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't} \\
&\quad + \sum_{\substack{c\ is\ a\ catalog \\ c'\ is\ an\ email}} \theta_3 \ \lambda_{cat}^{r_{ict}} \lambda_{email}^{r_{ic't}} d_{ict} d_{ic't}
\end{aligned} \tag{A2.4}$$

This corresponds to equations 13 and 14 in the text. The first term in equation (A2.4) is the catalog-catalog interaction effect (which we call *Cat_Cat_{it}*), the second term is the email-email interaction (*Email_Email_{it}*), and the last term is the catalog-email interaction (*Cat_Email_{it}*).

The computation of the direct effects in Equations 11a-11b and the interaction effects in equations 14a-14c is difficult because of the large number of catalogs and emails 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 the fact that all communications of a given type (i.e., catalogs or emails) that a person receives within a month are exchangeable (i.e., that 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 time period, we need to know only two variables, C_{ij} and E_{ij}, which contain respectively, the number of catalogs and number of emails received by customer i in month t. We now show how these variables can be used to recursively compute 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:

$$Cat_{it} = \psi_{i1} \sum_{c \in Catalogs} \lambda_{cat}^{r_{ict}} d_{ict} \tag{3.1}$$

As 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:

$$RCat_{it} = \sum_{c \in Catalogs} \lambda_{cat}^{r_{ict}} d_{ict} \tag{3.2}$$

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:

$$RCat_{i1} = \sum_{c \in Catalogs} d_{ic1} \tag{3.3}$$

This is nothing but the number of catalogs received in month 1 by customer i, as $r_{ic1} = 0$ for all catalogs. Thus we have:

$$RCat_{i1} = C_{i1} \tag{3.4}$$

The direct effect of the catalogs in month 2 is given by:

$$RCat_{i2} = C_{i2} + \lambda_{cat}C_{i1} = C_{i2} + \lambda_{cat}RCat_{i1} \tag{3.5}$$

Thus, the direct effect can be computed recursively by discounting the previous periods' 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:

$$RCat_{it} = C_{it} + \lambda_{cat}RCat_{it-1} \tag{3.6}$$

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 emails can analogously be written as:

$$REmail_{it} = E_{it} + \lambda_{email}REmail_{it-1} \tag{3.7}$$

where $REmail_{it}$ refers to the raw email direct effect. The total email direct effect, $Email_{it}$, can be computed by multiplying $REmail_{it}$ with the coefficient ψ_2.

Interaction Effects

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

$$Cat\_Cat_{it} = \theta_1 \sum_{cc' \in Catalog\, s} \lambda_{cat}^{r_{ict}} \lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't} \tag{3.8}$$

Again, ignoring the θ_1 coefficient, we define

$$RCat\_Cat_{it} = \sum_{cc' \in Catalog\, s} \lambda_{cat}^{r_{ict}} \lambda_{cat}^{r_{ic't}} d_{ict} d_{ic't} \tag{3.9}$$

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

$$RCat\_Cat_{i1} = \frac{C_{i1}(C_{i1}-1)}{2} \tag{3.10}$$

to represent the interaction effects between all catalog pairs received in month 1. Similarly, for month 2, the interaction effect can be written as:

$$RCat\_Cat_{i2} = \frac{C_{i2}(C_{i2}-1)}{2} + C_{i2}\lambda_{cat}C_{i1} + \lambda_{cat}^2 \frac{C_{i1}(C_{i1}-1)}{2} \tag{3.11}$$

The first term in equation (3.11) 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. The above equation can also be written as:

$$RCat\_Cat_{i2} = \frac{C_{i2}(C_{i2}-1)}{2} + C_{i2}\lambda_{cat}RCat_{i1} + \lambda_{cat}^2 RCat\_Cat_{i1} \tag{3.12}$$

Thus, for an arbitrary period t, the overall interaction effect can be written as:

$$RCat\_Cat_{it} = \frac{C_{it}(C_{it}-1)}{2} + C_{it}\lambda_{cat}RCat_{it-1} + \lambda_{cat}^2 RCat\_Cat_{it-1} \tag{3.13}$$

It is clear that the above recursive scheme considerably reduces the computational burden. Instead of 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, one can show that the email interaction effects can be recursively computed using the scheme:

$$REmail\_Email_{it} = \frac{E_{it}(E_{it}-1)}{2} + E_{it}\lambda_{email}REmail_{it-1} + \lambda_{email}^2 REmail\_Email_{it-1}$$

(3.14)

and the catalog-email interaction effects can be computed using the formula:

$$RCat\_Email_{it} = RCat_{it} * REmail_{it} \tag{3.15}$$

Notice that the recursive schemes for the interaction effects involve the computed quantities for the direct effects, and hence the direct effect terms need to be computed before beginning the computation of the interaction effects.

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 let the vector $\boldsymbol{\lambda} = \{\lambda_{cat}^b, \lambda_{e-mail}^b, \lambda_{cat}^q, \lambda_{e-mail}^q,$ $\lambda_{cat}^w, \lambda_{e-mail}^w\}$. Then conditional on $\mathbf{u}_{it}$, we have a non-linear mixed model specification given by

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

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 are thus "random". The errors $\mathbf{e}_{it} \sim N(\mathbf{0}, \boldsymbol{\Sigma})$ and $\boldsymbol{\gamma}_i \sim t_\omega(\mathbf{0}, \boldsymbol{\Gamma})$, where ω is the df 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 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 as the expected prior variance of the random effects $\boldsymbol{\gamma}_i$'s. 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 non-redundant parameters. We assume 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

1. 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. Each latent variable is drawn from a univariate conditional normal distribution obtained from the joint trivariate normal distribution of $\mathbf{u}_{it}$. The full conditional for b_{it}^* is obtained conditioned on the values of $\{q_{it}^*, w_{it}^*\}$, and b_{it}^* for each observation is drawn 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 q_{it}^* is drawn from its conditional normal distribution if the observed quantity is zero, and is set equal to $\log(q_{it})$,

otherwise. Finally, $w_{it}^* \mid \{b_{it}^*, q_{it}^*\}$ is drawn 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 is drawn from its conditional normal without truncation for observations on which no purchase is observed.

2. The full conditional for $\boldsymbol{\mu}$ is multivariate normal, given the conjugacy of the priors. 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 $\boldsymbol{\mu} \sim N(\sum_{i=1}^{I}\sum_{t=1}^{n_i} \mathbf{X}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{u}_{it\mu}, \mathbf{V}_{\mu})$, where, $\mathbf{V}_{\mu}^{-1} = \sum_{i=1}^{I}\sum_{t=1}^{n_i} \mathbf{X}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{X}_{it} + \mathbf{C}^{-1}$ and n_i indicates the number of observations for individual i, and I denotes the number of customers.

3. 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 $\boldsymbol{\gamma}_i \sim N(0, \kappa_i^{-1}\boldsymbol{\Gamma})$. Define $\mathbf{u}_{it\gamma_i} = \mathbf{u}_{it} - \mathbf{X}_{it}\boldsymbol{\mu}$. The posterior distribution is $\boldsymbol{\gamma}_i \sim N(\mathbf{V}_{\gamma_i}\sum_{t=1}^{n_i}\mathbf{Z}_{it}\boldsymbol{\Sigma}^{-1}\mathbf{u}_{it\gamma_i}, \mathbf{V}_{\gamma_i})$ where, $\mathbf{V}_{\gamma_i}^{-1} = \sum_{t=1}^{n_i}\mathbf{Z}_{it}'\boldsymbol{\Sigma}^{-1}\mathbf{Z}_{it} + \kappa_i\boldsymbol{\Gamma}^{-1}$.

4. 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
$$L(\boldsymbol{\varphi}) \propto \prod_{i,t} \exp\left(-\frac{1}{2}\left(\mathbf{u}_{it}^* - \mathbf{X}_{it}(\boldsymbol{\varphi})\boldsymbol{\mu} - \mathbf{Z}_{it}(\boldsymbol{\varphi})\boldsymbol{\lambda}_i\right)' \boldsymbol{\Sigma}^{-1}(\chi, \mathbf{R})\left(\mathbf{u}_{it}^* - \mathbf{X}_{it}(\boldsymbol{\varphi})\boldsymbol{\mu} - \mathbf{Z}_{it}(\boldsymbol{\varphi})\boldsymbol{\lambda}_i\right)\right).$$
Given the normal prior for $p(\boldsymbol{\varphi})$, the posterior is proportional to $L(\boldsymbol{\varphi})p(\boldsymbol{\varphi})$. As 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 0.02.

5. 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 0.1 that is tuned to obtain rapid mixing.

6. The correlations are bounded and constrained due to positive definiteness requirements. The full conditional for any correlation is not completely known because of the positive definiteness constraint. We therefore 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 et al., 2000).

7. 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 df, I is the number of customers in the data, and $\mathbf{S} = \left(\rho\boldsymbol{\Omega} + \sum_{i=1}^{I}\kappa_i\boldsymbol{\gamma}_i\boldsymbol{\gamma}'_i\right)^{-1}$.

Table 1
DESCRIPTIVE STATISTICS

Variable	Mean	Median	Standard Deviation	Minimum	Maximum
Average Purchase ($)	131	111	75	26	564
Purchases (per year)	2.2	1.7	1.3	0	8.3
Income ($)	84487	81500	40765	1000	150000
Age (years)	50	48	12	24	97
E-mails (per month)	0.4	0.0	0.5	0	3.4
Catalogs (per month)	3.4	3.2	1.6	0.3	7.9
Catalog Shr. (of orders	89%	100%	0.22	22%	100%

Table 2
MODEL COMPARISON

Model	Description	Log-marginal Likelihood
M1	Full Model	-16896.6
M2	Normal Heterogeneity	-17095.5
M3	No Marketing	-18965.3
M4	No Communications Dynamics	-17077.4

Table 3
PARAMETER ESTIMATES

Parameter		Incidence		Order Size|Incidence		Choice|Incidence	
		Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Customer Effects	Int	**-1.84**	**0.24**	-0.09	0.34	-0.01	1.08
	Age	0.02	0.04	-0.01	0.04	**0.80**	**0.16**
	Income	**0.02**	**0.01**	0.00	0.01	-0.03	0.04
	Child	-0.06	0.10	0.01	0.10	-0.26	0.34
Experience Effects	Wuse	**-0.47**	**0.14**	0.01	0.15	0.17	0.35
	Since	**0.59**	**0.06**	0.07	0.06	-0.03	0.23
	Diff					0.02	0.17
	Lcat	**0.18**	**0.05**	**-0.10**	**0.02**	**0.64**	**0.19**
	Lweb	0.13	0.16	-0.09	0.10	-0.06	0.27
Marketing Effects	Cat	**0.13**	**0.02**	0.04	0.03	0.04	0.12
	Email	**0.15**	**0.05**	0.01	0.05	**-0.46**	**0.14**
	Cat_Cat	**-0.02**	**0.00**	-0.01	0.00	**0.05**	**0.02**
	Email_Email	**-0.02**	**0.01**	0.01	0.01	**0.09**	**0.03**
	Email_Cat	**-0.01**	**0.01**	0.00	0.01	0.02	0.02
Time Effects	Trend	**-0.08**	**0.02**	0.01	0.02	**-0.26**	**0.11**
	JlF	**-0.18**	**0.04**	-0.02	0.05	**-0.48**	**0.18**
	Oct	0.03	0.04	0.10	0.05	0.12	0.25
	Nov	**0.54**	**0.04**	**0.25**	**0.06**	0.29	0.21
	Dec	**0.93**	**0.04**	**0.17**	**0.08**	-0.10	0.22
	λ_{cat}	**0.04**	**0.02**	**0.03**	**0.01**	**0.11**	**0.04**
	λ_{email}	**0.04**	**0.02**	**0.04**	**0.02**	**0.14**	**0.07**

Table 4
CHANGES IN CONTRIBUTION TO LATENT UTILITY: YEAR 1 VERSUS 4

	No Migration Group		Migration Group	
	Mean Change	Std. Error	Mean Change	Std. Error
Experience	-0.15	0.01	-0.12	0.09
Marketing	0.51	0.06	-0.55	0.17
Time	-0.56	0.04	-2.05	0.12

Table 5
CHANGES IN MARKETING LEVELS AND RESPONSE
PARAMETERS:YEAR 1 VS. 4

	No Migration Group		Migration Group	
	Mean	Std. Error	Mean	Std. Error
Change in Marketing				
Catalogs / month.	1.02	0.09	1.54	0.16
Emails / month.	0.95	0.08	1.76	0.21
Response Parameter				
Catalog	0.10	0.01	-0.16	0.03
Email	-0.43	0.01	-0.52	0.05

Table 6
CHANGES IN LATENT INCIDENCE UTILITY, CUSTOMER CHARACTERISTICS, AND
MARKETING RESPONSE BY MIGRATION GROUP

	No Migration Group		Migration Group	
	Mean	Std. Error	Mean	Std. Error
Change in Experience Utility	0.03	0.03	-0.19	0.04
Change in Marketing Utility	0.07	0.02	0.17	0.05
Change in Time Utility	-0.46	0.05	0.04	0.08
Customer Characteristic Utility	-1.53	0.03	-1.81	0.07
Catalog Response	0.12	0.01	0.14	0.01
Email Response	0.14	0.01	0.16	0.03

Table 7
CUSTOMER CHARACTERISTIC DIFFERENCES BETWEEN MIGRATION AND NO
MIGRATION GROUPS, AND THEIR CONTRIBUTION TO PURCHASE INCIDENCE

		No Migration Group		Migration Group	
		Mean	Std. Error	Mean	Std. Error
Demographics	Age (years)	50.95	0.71	45.80	1.28
	Income ($10K)	8.49	0.24	9.14	0.47
	Children (%)	24.36	2.43	27.54	5.42
Contribution	Intercept	-1.81	0.03	-2.09	0.06
to Customer	Age	0.09	0.001	0.08	0.002
Characteristic	Income	0.20	0.01	0.21	0.01
Utility	Children	-0.02	0.002	-0.02	0.003
	Overall	-1.53	0.03	-1.81	0.07

Figure 1
CONCEPTUAL FRAMEWORK

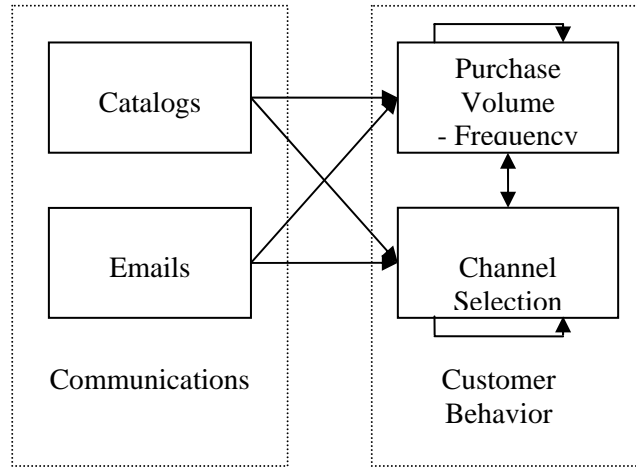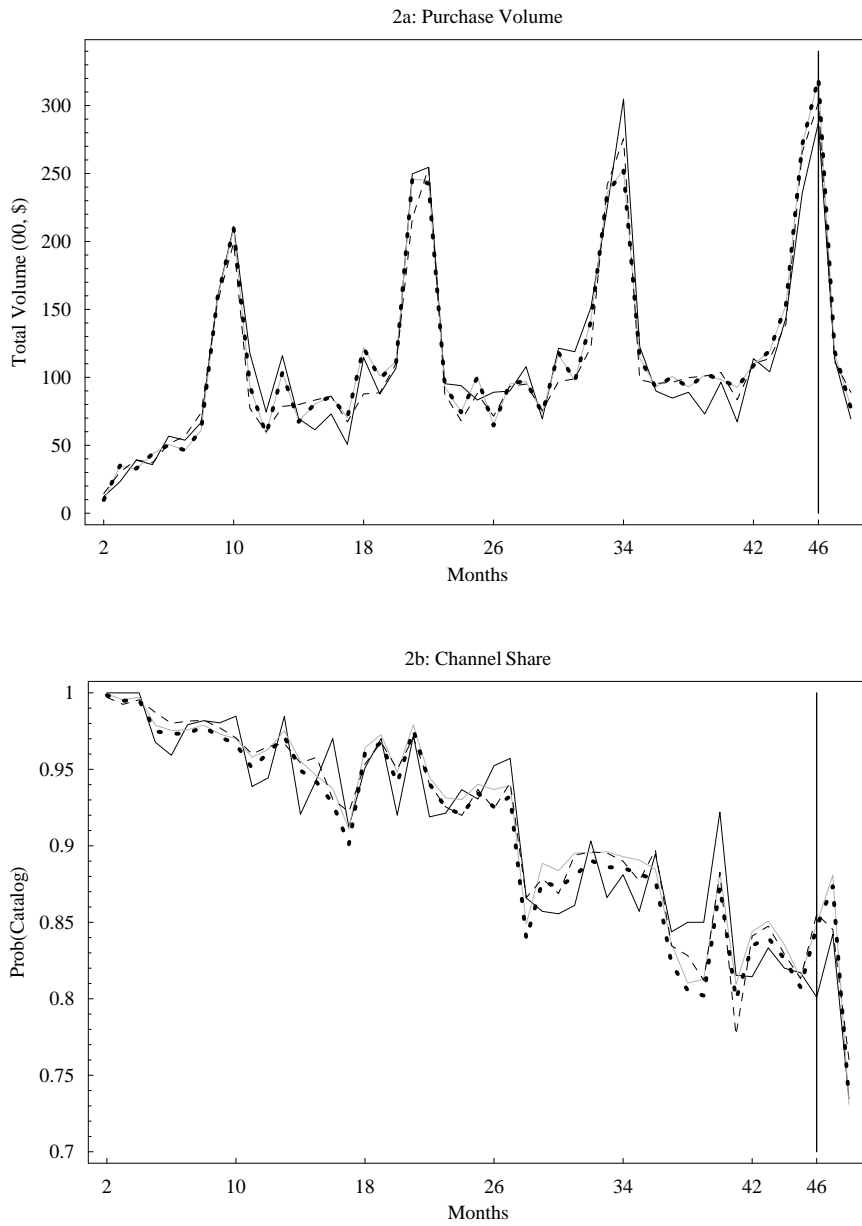## Figure 2
PREDICTION RESULTS

2a: Purchase Volume

2b: Channel Share

Note: In-sample from Periods 2-45 and out-of-sample 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).

REFERENCES

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 2002.

Belsley, D. A., E. Kuh and R. E. Welsch (1980). *Regression Diagnostics.* New York: Wiley.

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

Blattberg, Robert C., Byung-do Kim, and Scott A. Neslin (2004), *Database Marketing*, Book manuscript in preparation.

Brynjolfsson, E., and Smith, M. (2000), "Frictionless Commerce? A Comparison of Internet and Conventional Retailers," *Management Science*, 46, 4, 563-585.

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," *Chain Store Age*, April 2001, p. 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), 337-348.

Fox, Edward J., Alan L. Montgomery, and Leonard M. Lodish (2002), "Consumer Shopping and Spending Across Retail Formats," working paper, Carnegie-Melon University.

Gönül, Füsun, and Meng Ze Shi (1998), "Optimal Mailing of Catalogs: A New Methodology Using Estimable Structural Dynamic Programming Models," *Management Science*, 44, 9 (September), 1249-1262.

Gustafson, Paul (1998), "A Guided Walk Metropolis Algorithm," *Statistics and Computing*, 8, 4, 357 - 364.

Infoworld (2001) "Going Multichannel," May 14, 2001, 23 (20), p. 46.

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

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, 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, 317-332.

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 Shoppping Behavior: Antecedents and Implications for Channel and Customer Equity," Working Paper, Texas A&M University, College Station, TX.

Liang, T.P. and Huang, J.S. (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, 3 (August), 339-350.

Manchanda, Puneet, Peter E. Rossi and Pradeep Chintagunta (2004), "Response Modeling with Nonrandom Marketing-Mix Variables," *Journal of Marketing Research*, 41, 4 (November), 467-478.

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, 223-229.

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 (February), 167-185.

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 (June), 153-175.

Thomas, Jacquelyn S. and Ursula Y. Sullivan (2005), "Managing Marketing Communications with Multichannel Customers," *Journal of Marketing,* 69 (4), 239-251.

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

Wall Street Journal, "Web Sales to Surpass $200Billion" May 23, 2006.

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

Zhang, Jie, and Michel Wedel (2004), "The Effectiveness of Customized Promotions in Online and Offline Stores," working paper, University of Michigan.