Chapter 28
Multiple Campaign Management

Abstract Many database marketing programs are constructed as one-shot efforts – they determine the best campaign to implement *now*. However, more recently, academics and companies have recognized that the actions we take now influence what actions we will be compelled to take in the future, and if the current actions are not managed correctly, these future actions will not be successful. The key is to manage the series of communications holistically, taking into account the future as we design the current campaign, and to do so at the customer level. This chapter discusses "optimal contact models" for managing a series of campaigns. Many of the examples we draw on involve the catalog industry, although we also discuss examples involving e-mails, product magazines, promotional discounts, and even online survey panel management.

28.1 Overview

Managing multiple campaigns is an emerging issue in database marketing due to the confluence of two forces: First, optimizing a single contact using a predictive model has become commonplace – we know how to do this and can do it well. Second, there is concern that customers are becoming cluttered with direct marketing communications such as catalogs, e-mails, online advertising, etc.

Multiple campaigns can be managed by "optimal contact models." These models specify the number and/or schedule of communications including catalogs, e-mails, online advertising, even requests for online panelists to participate in surveys, at the customer level. Two factors make this a challenging task. First is dynamic customer response. Customer response to a contact changes over time, depending on the customer's previous contact and response history. Second, because of response dynamics, a natural way to optimize the *schedule* of contacts over time is to be "forward looking," i.e., recommend decisions for period t taking into account the impact this has on customer response and hence future profits in period $t + 1$, $t + 2$, etc. In Section 28.2 we discuss dynamic response. In Section 28.3 we discuss optimal contact models.

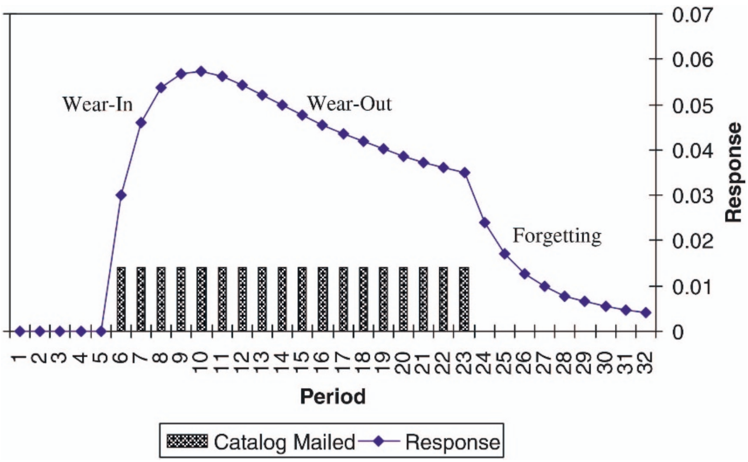

Fig. 28.1 Wear-in, wear-out, and forgetting.

28.2 Dynamic Response Phenomena

28.2.1 Wear-in, Wear-out, and Forgetting

These phenomena were first identified in the advertising literature (see Little 1979) but we explain them in the context of mailing catalogs to customers. Wear-in means it takes several mailed catalogs before the customer responds – it takes several mailings before we fully capture the customer's attention. Wear-out means that once a critical number of mailings is reached, subsequent catalogs produce lower response rates. The customer may no longer be paying attention to the catalog. Forgetting means that once mailings are halted, response does not instantly go to zero but decays gradually. This is because the customer still has the previous catalog(s) on hand, although eventually discards them. Figure 28.1 shows the three phenomena.

While wear-out is explained most easily by a lack of attention, there are three additional possibilities. First is that too much information confuses customers to the point that they make incorrect decisions (Assael 1995, pp. 231–232; Jacoby et al. 1974), or to make more mistakes in evaluating brands (Hutchinson and Alba 1991). A second explanation is that the customer gets angry at the constant "harassment" and refuses to buy. Third, the customer's needs could have been satiated by earlier catalogs.

Wear-in and forgetting can be modeled with a stock variable:

$$Stock_t = \lambda Stock_{t-1} + \beta C_t \tag{28.1}$$

C_t is an indicator variable (0–1) of whether a catalog was mailed to the customer in period t. The parameters $\lambda(0 < \lambda < 1)$ and $\beta(\beta > 0)$ refer to

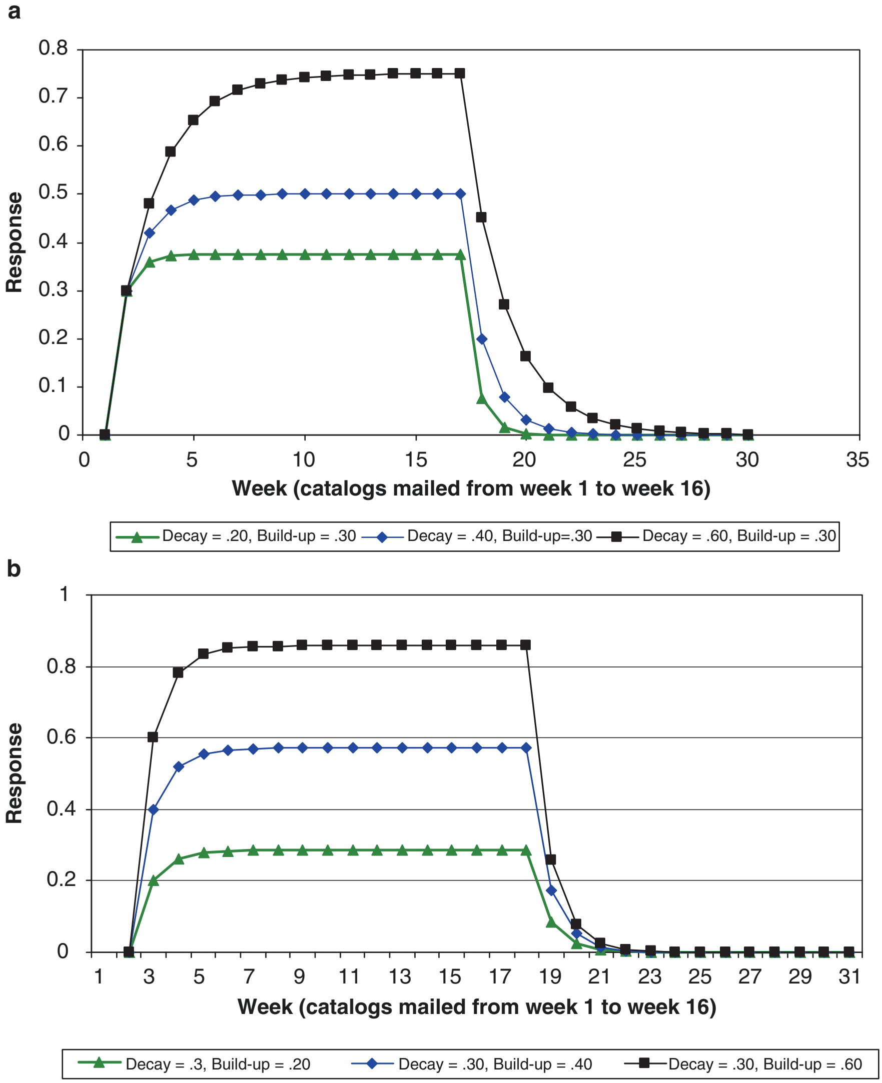

Fig. 28.2 Dynamic wear-in and forgetting generated from a stock model (Equation 28.1).

decay and build-up of catalog stock (see Fig. 28.2). λ controls the rate of wear-in and forgetting – large λ means faster wear-in and slower forgetting. Both β and λ determine the peak stock level:

$$\text{Peak Stock Level} = \frac{\beta}{(1-\lambda)} \tag{28.2}$$

The stock model includes wear-in and forgetting but not wear-out. Simon (1982) devised a simple way to accommodate wear-out and forgetting, as

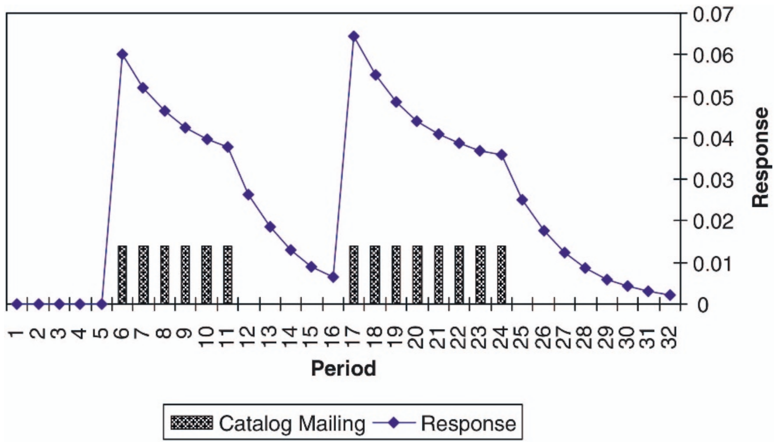

Fig. 28.3 Simon's (1982) model of wear-out and forgetting (Equation 28.3.)*
*$\mathrm{Prob(Buy)_t} = 0.7\mathrm{Stock_{t-1}} + 0.01\mathrm{C_t} + 0.05\max(0, \mathrm{C_t} - \mathrm{C_{t-1}})$

follows:

$$Stock_t = \lambda Stock_{t-1} + \beta C_t + \delta max(0, C_t - C_{t-1}) \quad (28.3)$$

The term $C_t - C_{t-1}$ creates an immediate increase ("shock") in stock whenever a catalog is mailed in the current period after not being mailed in the previous period. High values for δ boost stock higher than its maximum $(\beta/(1-\lambda))$. As more catalogs are mailed, stock declines (wears out) to this value. Figure 28.3 shows Equation 28.3 for specific values of the parameters. Wear-out and forgetting are clearly visible in the figure.

An even more flexible model that includes wear-in, wear-out, and forgetting is:

$$Stock_t = \lambda Stock_{t-1} + \beta_t C_t + \delta ShockStock_t \quad (28.4a)$$

$$ShockStock_t = \lambda' ShockStock_{t-1} + \beta' Max(0, C_t - C_{t-1}) \quad (28.4b)$$

$$\beta_t = \lambda'' \beta_{t-1} + \beta'' C_t \quad (28.4c)$$

Equation 28.4b smoothes the $Max(0, C_t - C_{t-1})$ shock variable. It peaks immediately at β' and then decays by a factor λ'. Equation 28.4c grows the immediate impact of catalog mailings up to a maximum $\beta^* = (\beta''/(1-\lambda''))$. The net result is that the maximum induced by the $C_t - C_{t-1}$ shock is not realized until some delay, creating a wear-in effect. Once we reach that maximum, we have wear-out down to the level $\beta^*/(1-\lambda)$, and then forgetting by the factor λ once the mailing stops (see Fig. 28.4).

One can graph the number of communications made over several periods versus response aggregated over several customers. This "aggregate response function" can take on several shapes depending on the degree of wear-in, wear-out, and forgetting occurring at the micro level (see Fig. 28.5).

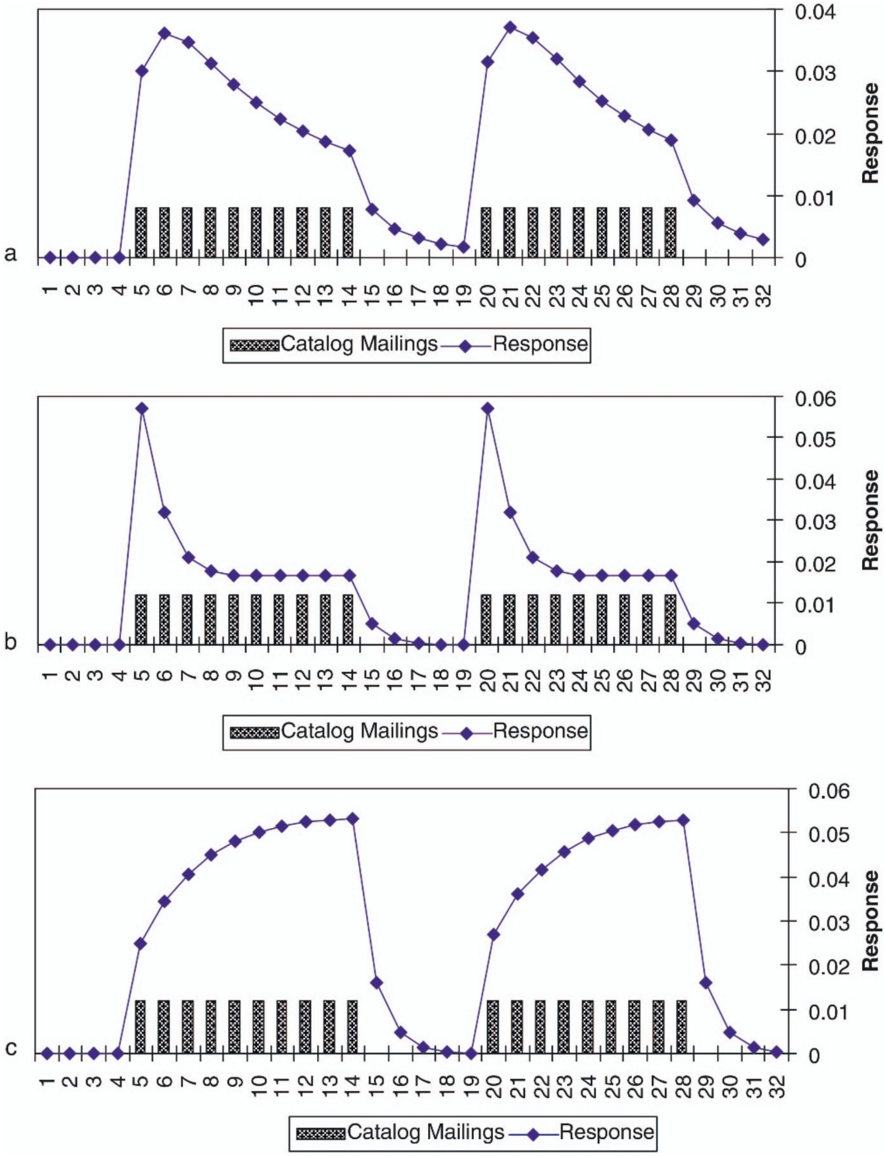

Fig. 28.4 A stock model including wear-in, wear-out, and forgetting (Equation 28.4)*. (a) Wear-in, wear-out, forgetting; (b) No wear-in, wear-out, forgetting; (c) Wear-in, no wear-out, forgetting;

***Parameter values (see Equations 28.4a–c)**

	Scenario 1	**Scenario 2**	**Scenario 3**
λ	0.3	0.3	0.3
δ	0.05	0.05	0.01
λ'	0.8	0.1	0.3
β'	0.5	1	1
λ''	0.4	0.4	0.6
β''	0.005	0.007	0.015

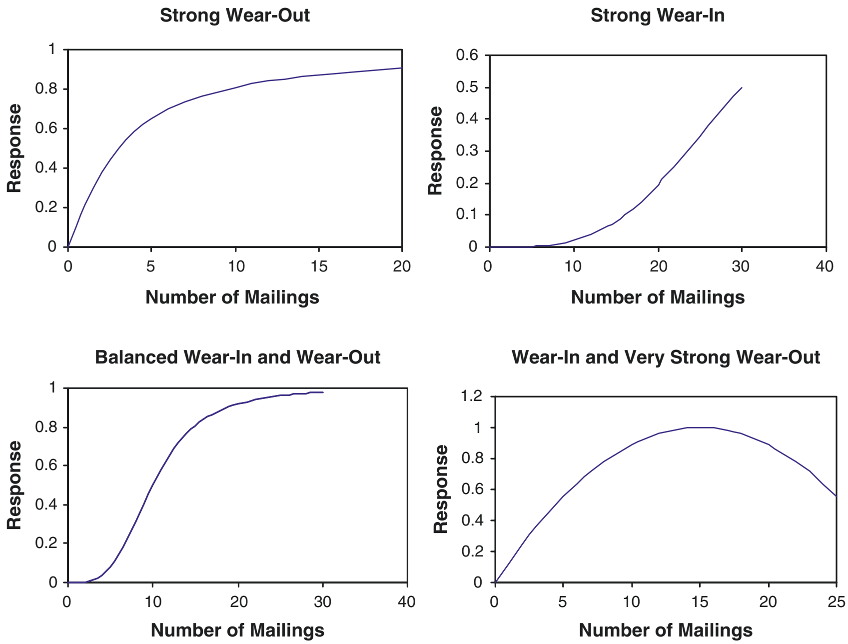

Fig. 28.5 Aggregate response functions depending on degree of wear-in and wear-out.

Wear-out is an especially crucial phenomenon because it provides a non-cost reason to limit the number of contacts. Fewer mailings may actually produce more sales (see Fig. 28.6). With 20 consecutive weeks of catalogs, response peaks relatively soon and then starts to wear out. However, with catalog "pulsing," the wear-out effect is mitigated and the total expected response increases. This is why aggregate response as a function of contacts can be inverse U-shaped as in Fig. 28.5**d**.

Ansari et al. (2008) estimate λ's ranging from 0.04 to 0.14 for e-mails and catalogs using weekly data, implying fast wear-in as well as forgetting.[1] Gönül et al. (2000) include time since the last catalog was received in a catalog response model. Its estimated coefficient was negative, consistent with forgetting. The authors also include the cumulative number of catalogs received after the last purchase and find a negative relationship. This supports wear-out. Campbell et al. (2001) report wear-out regarding net returns as a function of "advertising expenditure."

Eastlick et al. (1993) asked catalog buyers how many catalogs they received in the past 12 months, and how much they purchased during that time. The researchers fit a regression across buyers and find an inverse U-shaped relationship between catalogs and expenditures. This is consistent with very

[1] Note that estimates of λ can be biased upward when data are temporally aggregated, e.g., to the monthly or quarterly levels (see Leeflang et al. 2000, pp. 85–91. Estimating on the weekly level is therefore advisable.

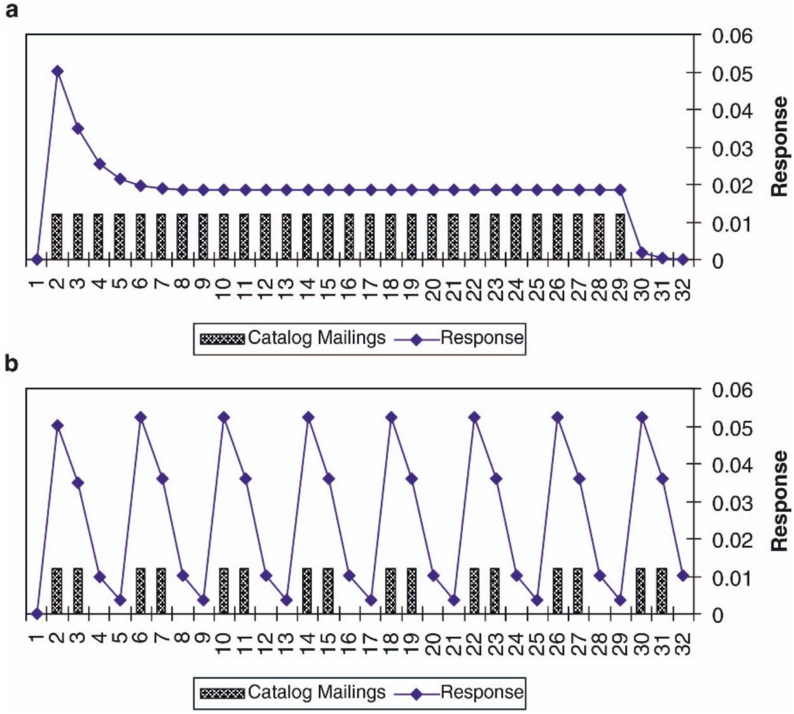

Fig. 28.6 With high wear-out, pulsing produces more responses than constant mailing, even with fewer total mailings. (a) Constant mailing strategy (total response = 0.580); (b) Pulsing mailing strategy (total response = 0.810)
Note: Model used to generate graphs based on Equations 28.4a–28.4c parameterized as follows:

$$\text{Stock}_t = 0.1^*\text{Stock}_{t-1} + \beta_t{}^*\text{C}_t + 0.05^*\text{ShockStock}_t$$
$$\text{ShockStock}_t = 0.4^*\text{ShockStock}_{t-1} + 0.8^*\text{Max}(0, \text{C}_t - \text{C}_{t-1})$$
$$\beta_t = 0.4^*\beta_{t-1} + 0.01^*\text{C}_t$$

We also assume that stock translates linearly into response, whether it be response rate, revenues, etc. One could in general assume Response = f(Stock_t), where "f" is a nonlinear function.

strong wear-out as in Fig. 28.5. Ganzach and Ben-Or (1996) note the authors did not distinguish between strong wear-out and very strong wear-out. Feinberg et al. (1996) reply that the wear-out is visible in the data. However, they note that only a "low number of subjects" received more than the estimated overload point.

28.2.2 Overlap

The database marketer contacts its customers with different communications. For example, L.L. Bean has a male clothing catalog, a women's catalog, etc.

The degree of content overlap between communications might moderate the wear-in, wear-out, and forgetting phenomena. For example, Equation 28.1 might be extended as follows:

$$Stock_t = (\lambda + \lambda_s Sim)^* Stock_{t-1} + (\beta + \beta_s Sim)^* C_t \quad (28.5)$$

where *Sim* is the similarity between the current and previous catalogs. Campbell et al. (2001) include overlap in a model of catalog profits as follows[2]:

$$Profit = R_p(1 - S_{qp}) + R_q(1 - S_{pq}) \quad (28.6)$$

where $R_{p(q)}$ is the profit from catalog $p(q)$, and S_{AB} is the "saturative" effect of catalog A on catalog B. $S_{pq} = 0.05$ means that catalog q's profits are reduced by 5% when catalog p is also mailed. The authors model the saturative effect as:

$$S_{AB} = [time\ index\ (A, B)]^* [similarity\ index\ (A, B)] \quad (28.7)$$

Both indices are between 0 and 1. The similarity index captures content overlap – if A and B are the same catalogs, similarity equals 1. The time index captures proximity of mail dates. If the catalogs are mailed at exactly the same time, the time index equals 1. So the greater the content overlap between closely mailed catalogs, the more saturation. The authors indeed find cannibalization between catalogs mailed close to each other.

28.2.3 Purchase Acceleration, Loyalty, and Price Sensitivity Effects

Database marketing communications often entail promotions. Catalogs contain descriptive information, but they are similar to weekly store circulars in that they list products and prices. E-mails as well as online advertising are also similar to feature advertising. In addition, all these communications often offer price discounts.

Database marketing communications therefore can produce the same long-term effects as do promotions. These include accelerating forward in time sales that would have occurred anyway (Blattberg et al. 1981; Neslin et al. 1985; Macé and Neslin 2004), changing brand loyalty (Guadagni and Little 1983; Gedenk and Neslin 1999; Seetharaman 2004), and increasing price sensitivity (Mela et al. 1997; Jedidi et al. 1999).

[2] Note this differs a little bit from equation 1 in Campbell et al. (2001, p. 89), but is consistent with their definition of saturation on page 81.

Findings of a negative impact of recency (the more recently the customer has purchased, the less likely the customer is to purchase now) suggest purchase acceleration, assuming the recent purchase was stimulated by communication. Ansari et al. (2008) and Gönül et al. (2000) find evidence of negative recency effects.

Anderson and Simester (2004) studied the long-term impact of promotion depth. Each of three experiments included a "control" and "promotion" catalog. Both catalogs had the same number of promotional prices, clearly communicated as "Regular Price $X, Sale $Y," but Y was smaller for the promotion catalog. Purchasers were followed for at least 24 months, and all purchasers, whether in the control or promotion group, received the same catalogs during this period. In all three tests, the promotions clearly increased sales. But the key question was, what happened in the long term?

The authors investigate acceleration, repeat purchasing, and price-sensitivity effects. They also determine whether promotions draw a different group of customers. They find evidence for all these effects. For example, promotion purchasers bought fewer units in months 1–12 after purchasing from a promotional catalog than they did in months 13–24. This is suggestive of acceleration.

They also found that promotion purchasers bought fewer units after the purchase compared to non-promotion purchasers. This effect however vanished when the authors controlled for selection, i.e., they found that the promotion catalog drew a lower RFM customer and after controlling for this, the number of units purchased in the future was unaffected by promotion (see Neslin and Shoemaker 1989). The authors found that customers who were infrequent, not recent purchasers bought additional units in the future, whereas the number of units subsequently purchased by higher RF groups was unaffected. This suggests that inexperienced, low RF customers learned about the positive aspects of the product due to the purchase experience induced by the promotion. Finally, the authors found that customers who had historically paid high prices purchased at lower prices after purchasing on a steep promotion, whereas customers who had historically paid low prices continued to do so. This is consistent with sensitizing heretofore not-price-sensitive customers to buying on deal.

Overall, there is some evidence, mostly through Anderson and Simester's study, that database marketing communications can act like promotions and induce the same long-term effects observed in the promotions literature (see Neslin 2002). More empirical work is needed to measure these effects in other settings (e.g., e-mail communications, non-promotional catalogs, etc.), but the evidence suggests that optimal contact models need to consider these issues. For example, a communication may accelerate a purchase, so it would not make sense to communicate again until sufficient time had elapsed for the customer to need the product again.

28.2.4 Including Wear-in, Wear-out, Forgetting, Overlap, Acceleration, and Loyalty

Ansari et al. (2008) develop a model to study customer channel migration that includes wear-in, wear-out, forgetting, acceleration, and loyalty. They refer to the first four phenomena as "communication effects," while the last two they call "experience effects." The communications model is:

$$\textit{Communication Effect}_{it} = \textit{Direct Effect}_{it} + \textit{Interaction Effects}_{it} \tag{28.8}$$

The communication effect includes the impact of all communications currently and previously received by customer i on that consumer's decisions at time t.

The direct effect is each communication's impact in isolation and allows the model to capture wear-in and forgetting. The interactions are between communications and allow the model to capture wear-out and overlap. Specifically,

$$\textit{Direct Effect}_{it} = \sum_{c \in C} \beta_{ic} \lambda_c^{\tau_{ict}} d_{ict} \tag{28.9a}$$

$$\textit{Interaction Effects}_{it} = \sum_{c,c' \in C} \delta_{icc'} \lambda_c^{\tau_{ict}} \lambda_{c'}^{\tau_{ic't}} d_{ict} d_{ic't} \tag{28.9b}$$

where

C = Set of all communications distributed by the firm. Particular communications are denoted by c or c'.
β_{ic} = Direct response of customer i to communication c.
λ_c = Decay parameter for communication c.
τ_{ict} = Time since customer i received communication c, as of period t.
d_{ict} = Step indicator equal to 1 if customer i received communication c on or before period t; 0 otherwise. The indicator turns on once the customer has received the communication, and remains on thereafter.
$\delta_{icc'}$ = Interaction response of customer i between communications c and c'.

The model does not measure overlap explicitly, but the authors model the δ's, as well as the β's and λ's, as functions of communication attributes. The attributes could include content (e.g., men vs. women's catalogs, etc.) as well as vehicle types (catalog vs. e-mails). Indeed they find that the interaction terms between like vehicles (catalogs and catalogs; e-mails and e-mails) are generally stronger than between different vehicles (catalogs and e-mails). This suggests an overlap effect due to vehicle overlap.

That the direct effects determine wear-in and forgetting is shown in Fig. 28.7**a**. In that figure, all the interactions (δ) are set equal to zero. We see that over the course of four communications, response builds toward a maximum, then declines to zero, although not instantly, once the communications end. Figure 28.7**b** demonstrates that with negative interactions, the model produces wear-out. In this example, the first two communications have a stronger interaction than any other pair. The effect is so strong that response declines when the second communication is delivered. This could be due to strong overlap between the first two communications.

Equation 28.9b captures the timing effects suggested by Campbell et al. (2001). Wear-out/overlap effects will be strongest when communications are delivered consecutively (because the λ terms for both communications will be relatively large).

Ansari et al. capture repeat purchase and acceleration effects through lagged incidence, order size, or channel choice variables. For example, they include a variable called "Wuse" equal to log(1+ the cumulative number of purchases made on the Internet). They find this variable has a negative average coefficient in their purchase incidence model, suggesting that purchasing on the Internet decreases future purchase incidence. They include a recency variable, "since," equal to the time since the previous purchase. The coefficient for this variable was positive on average, meaning that it is less likely the customer will purchase in the current period if the customer has purchased last period. Since the marketing variables being studied – catalogs and e-mails – made it more likely that a purchase took place, this suggests purchase acceleration.

28.3 Optimal Contact Models

Optimal contact models determine the number and/or schedule of communications to be delivered to each customer in a given time frame. These models consist of a response model and an optimization. The response model predicts how the customer will respond to a particular "contact." This depends on the "state" the customer is in at a given point in time, for example, how long it has been since the customer was last contacted. Response is probabilistic – there is uncertainty as to whether the customer will respond. The optimization is often forward looking because actions the firm takes in the current period may influence the actions it should take in future periods. For example, if we contact the customer now, we may need to wait a few periods before it becomes worthwhile to contact the customer again. The probabilistic and forward looking aspects suggest that optimal contact models be formulated as a stochastic dynamic program (Ross 1983; Bertsekas 1995). This technique is designed to handle optimizations where the outcomes of firm decisions are probabilistic and there are dynamics in the response to these decisions. For example, the first optimal

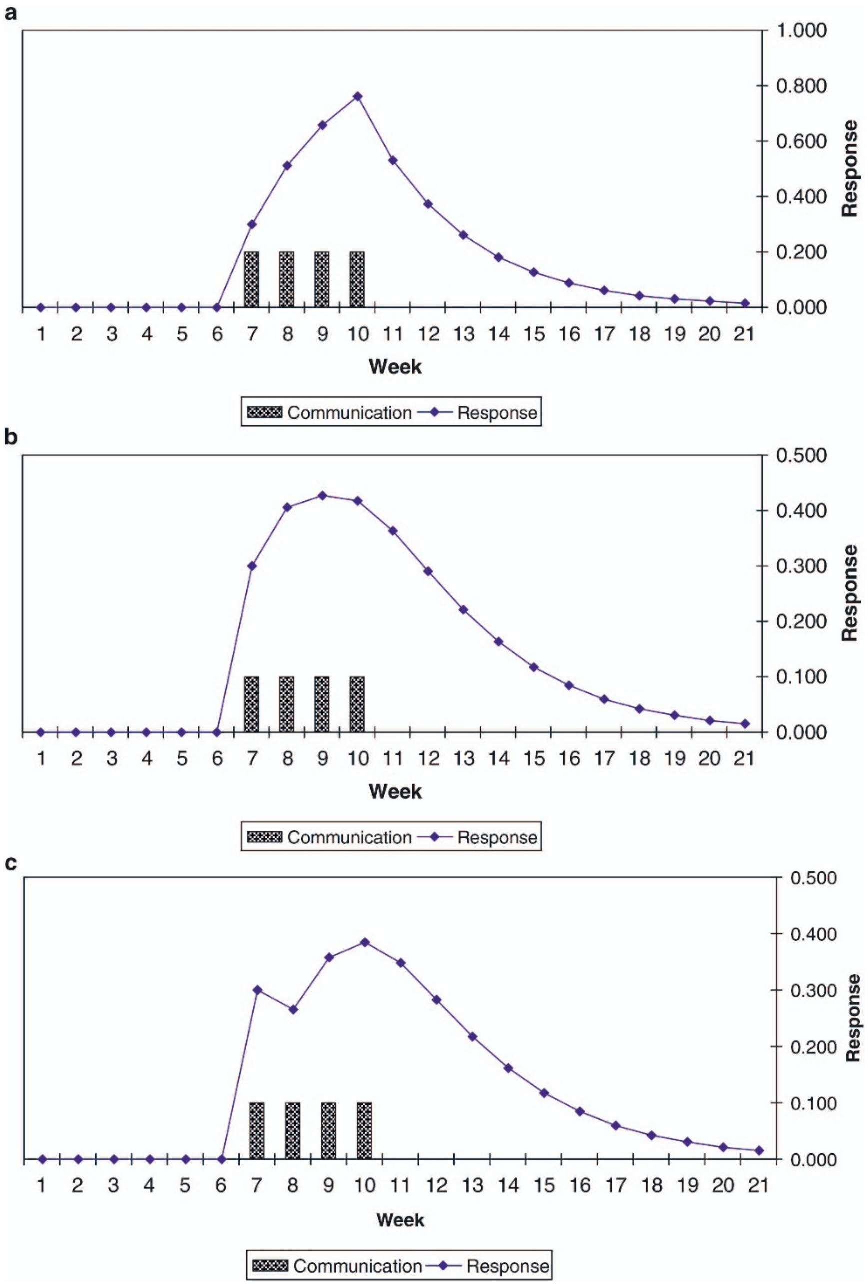

Fig. 28.7 Phenomena captured by Ansari et al. (2008) model Equations 28.8–28.9. (a)* Wear-in and forgetting: Communications in Periods 7–10; (b)+ Wear-in, wear-out, forgetting, and overlap: Communications in Periods 7–10; (c)^ Wear-in, wear-out, forgetting, and overlap: Communications in Periods 7–10 – stronger overlap between communications 1 and 2 than between others (From Ansari et al. 2008)

* $\beta = 0.3, \lambda = 0.7, \delta = 0$

\+ $\beta = 0.3, \lambda = 0.7, \delta = -0.15$ for all pairs of communications

^ $\beta = 0.3, \lambda = 0.7, \delta = -0.35$ between 1st and 2nd communications, $\delta = -0.15$ for other communication pairs.

contact model we discuss (Ching et al. 2004) makes direct use of this methodology.

28.3.1 A Promotions Model (Ching et al. 2004)

28.3.1.1 Response Model

Ching et al. (2004) develop a stochastic dynamic program to decide when to offer a promotion to customers of a computer services company. Customers are assigned to one of four states depending on their usage in the previous week: (1) 0 min, (2) 1–20 min, (3) 21–40 min, (4) $>$ 40 min. Using historical data, the authors estimate ${P_{ik}}^j$, the probability the customer moves from state i to k, if the customer receives promotion j. Both i and k take on values 1, 2, 3, or 4; j can equal 1 (promotion), or 2 (no promotion). The authors calculate c_i^j, the expected revenue from a customer in state i who receives promotion j. The state definitions, "transition probabilities" P_{ik}^j, and revenues c_i^j are the ingredients for a stochastic dynamic program.

28.3.1.2 Optimization Model

In each period, the firm can observe what state the customer is in. The question is whether to promote or not to promote to this customer. This problem can be formulated as the following recursion:

$$v_i(t) = \underset{j=1,2}{Max} \left\{ c_i^j - d_j + \delta \sum_{k=1}^{4} p_{ik}^j v_k(t-1) \right\} \tag{28.10}$$

where:

$v_i(t)$ = The expected optimal revenue given the customer is in state i and there are t periods remaining in the planning horizon.
d_j = Cost of implementing promotion j ($j = 1$ or 2).
δ = Discount rate applied to future profits.

The $c_i^j - d_j$ term represents expected profits in period t depending on whether the firm promotes ($j = 1$) or not ($j = 2$). The $\delta \sum_{k=1}^{4} p_{ik}^j v_k(t-1)$ term represents the discounted expected profit for the remaining $t - 1$ periods after the current period. Depending on whether the firm promotes or not in the current period, the customer progresses with probability p_{ik}^j to state k in the next period, and if so, expected optimal revenues are $v_k(t-1)$. The expression of the optimization in this recursive form is a fundamental "principle of optimality" in dynamic programming. It says that the optimal solution can be determined by deciding what to do in the current period, taking into account the repercussions for future periods. Appropriately, $v_i(t)$ is called the "value function."

Equation 28.10 assumes a finite time horizon. In this case the optimal solution will prescribe what action to take if the customer is found to be in state i at time t. The authors also consider the infinite horizon problem, which assumes the firm is maximizing over an infinite period. This may sound a bit unrealistic (will the firm be in existence forever!), but the discount factor effectively limits the time horizon, and makes the model an optimization of lifetime value. When considering an infinite horizon, the stochastic dynamic program prescribes what action should be taken in "steady state," i.e., for any arbitrary period, the optimal action is determined solely by what state the customer is in.

Steady state solutions can be obtained using various techniques, including successive approximations, policy improvement, and linear programming (Ross 1983, pp. 35–42). Finite horizon solutions can be obtained by solving Equation 28.10 using backward induction. Ching et al. solve the infinite horizon version of their model using linear programming and the finite horizon version using backward induction. They include computer programs in Excel that illustrate. Their application finds, intuitively, that if promotion costs (d) are large, they should only be used for inactive customers (state 1). However, if the cost is small, the promotion should be used for light users (state 2). In the finite horizon case, the authors also impose the constraint that a maximum of four promotions can be administered. They find (for low promotion costs), that promotions should be administered to inactive customers as soon as possible, while generally speaking the light users should receive their promotions as late as possible. The first result is intuitive, while the latter may be due to the particular p_{ik}^{j}'s.

The Ching et al. model is a straightforward yet powerful example of deriving an optimal contact strategy from a stochastic dynamic program. Many optimal contact models either embellish the state definitions, the response function, or the optimization requirements. However, these can be nontrivial improvements, both managerially and technically.

28.3.2 Using a Decision Tree Response Model (Simester et al. 2006)

28.3.2.1 Response Model

Simester et al. (2006) design an optimal contact model for a single catalog. Their approach is particularly rich in its use of a decision tree for the response model (Chapter 17). Customer states are defined based on which end node of the tree the customer is classified in at a given point in time. The authors' decision tree method is distinctive in two respects: (1) the dependent variable is a measure of long-term potential, not response to a single catalog;

(2) they develop a decision tree algorithm for handling this continuous dependent variable. The end nodes of the decision tree define customer states. This is an innovative approach. The alternative would be to use a decision tree with 0–1 response as the dependent variable. However, the authors use both response and expenditure, and the long term rather than the short term, to define their dependent variable and derive their customer states.

28.3.2.2 Optimization Model

The stochastic dynamic program is set up as follows:

$$V^{\pi}(s) = E_{r,T,s'}[r_{s,\pi(s)} + \delta^T V^{\pi}(s')|s, \pi(s)] \tag{28.11}$$

where:

$\pi(s)$ = Mailing policy; a decision rule of whether or not to mail a catalog to the customer who is in state s.
$V^{\pi}(s)$ = Expected long-term profits for the customer in state s under mailing policy π.
$r_{s,\pi(s)}$ = Immediate period profits for the customer in state s under mailing policy π.
δ^T = Discount factor given time T between catalog mailings.

The random elements taken into account in the expected value are the short-term response r, the time T between mailings, and the future states s' that customer may enter as a result of the mailing policy. Equation 28.11 is a value function. It says that the expected long-term profits for the customer in state s under policy π is the sum of the current period optimal profits plus expected future optimal profits.

The authors calculate that their method significantly increases profits over the current policy used by the firm they study, especially if the future is not highly discounted. This is sensible because the states are defined in terms using long-term potential as the dependent variable. Their policy mails more catalogs to customers who have not recently received catalogs, and few catalogs to customers who have recently received catalogs. The authors find that profits increase with more states. However, this may simply be taking advantage of chance variation. Indeed, in a holdout test, profits are independent of the number of states (although still greater than under the current policy).

The authors field test their approach over a 6-month period. They test three customer groups (low, moderate, and high value), and two methods (model versus current). They find the model improves profit for low and moderate value customers, but decreases it for high value customers. The model under-mails these customers relative to the current method, but by the end of the test, the gap tightens. The authors diagnose the problem was that the historical data were skewed toward mailing many catalogs to high value customers. In prescribing fewer catalogs for these customers, the model was going out of the range of the data. The lesson is that the historical data for optimal contact models need to include ample variation in contacts. Elsner et al. (2003, 2004) use field tests to generate the data to estimate their response models.

28.3.3 Using a Hazard Response Model (Gönül et al. 2000)

28.3.3.1 Response Model

Gönül et al. (2000) develop a catalog optimal contact model based on a proportional hazard response model (Cox 1972). The response model is:

$$h_i(t|X) = h_{0i}(t)\psi_i(X) \tag{28.12}$$

where:

t = Time since the last purchase.
$h_i(t|X)$ = The likelihood that customer i purchases at time t.
$h_{0i}(t)$ = The baseline hazard for customer i, due only to the passage of time t.
ψ_i(X) = The proportional covariate adjustment for customer i – due to covariates that vary across customers or over time.

The authors operationalize Equation 28.12 as follows:

$$h_i(t|X) = exp(\gamma_{0i} + \gamma_{1i}t + \gamma_{2i}ln(t) + \gamma_{3i}t^2) \tag{28.13a}$$

$$\psi_i(X) = exp(\beta_1 MALE_i + \beta_2 AVG\_CONSUMP_i + \alpha_{1i} PROM\_REST_i + \alpha_{2i} WEAROUT_i) \tag{28.13b}$$

The baseline hazard model captures recency, since t is the time since the last purchase. Equation 28.13a can capture many monotonic and non-monotonic relationships. The covariate adjustment consists of several multipliers. The MALE variable allows for gender to influence response. AVG_CONSUMP is defined as the average daily expenditure of the household. It combines frequency and monetary value.

PROM_REST is the number of periods since the last catalog was mailed to the customer. The authors hypothesize that α_{1i} should be negative to reflect forgetting. The impact of a catalog in the current period is implied by setting PROM_REST to 0. WEAROUT is defined as the number of catalogs since the last response. If α_{2i} is negative the likelihood of responding in the current period decreases if a large number of catalogs have been mailed since the last purchase. This can be interpreted as wear-out. Note however that if α_{2i} is positive, that could be interpreted as wear-in.

The model includes heterogeneous response in the baseline hazard (γ's) and mailing response parameters (α's). The authors use a latent class approach (Kamakura and Russell 1989) to model heterogeneity, yielding different parameters for each segment ($i = 1, \ldots, S$). The authors account for endogeneity of the catalog mailing variables (PROM_REST and WEAROUT) using instrumental variables. They use a logistic regression of catalogs mailed as a function of RFM variables and use the predictions from this model to calculate PROM_REST and WEAROUT. Interestingly, they do not find much impact of this procedure on their estimated parameters.

The authors estimate their model for 979 customers. They find a two-segment model fits best ($S = 2$). The product category is a durable good, so the customer typically would not need to re-order until a long time had elapsed since purchase. Accordingly, the authors find the baseline hazard is monotonically increasing for Segment 1. However, it is U-shaped for the second. Perhaps these customers order another product to augment the first immediately after purchase, but baseline hazard then decreases and rises again as the customer needs to replace the product(s).

The authors find that males are less likely to respond ($\beta_1 < 0$) and that heavy users are more likely to respond ($\beta_2 > 0$). They also find in both segments that a response is less likely if there has been a longer the time since the last catalog was mailed ($\alpha_{i1} < 0$). This indicates forgetting is a real phenomenon in catalog mailing. They find a significantly negative WEAROUT coefficient in one segment ($\alpha_{i2} < 0$), suggesting wear-out.

28.3.3.2 Optimization Model

The optimal policy considers each customer at time t and recommends a mailing at that time if expected profit over the period $t + x$ is greater with a mailing than without, where x is the time horizon. They find that the qualitative findings for different x's do not vary for $x \in [1, 12]$ and use $x = 3$. Profit for customer i is:

$$\pi_i(D_i) = mE(A_i)[D_i P_i^c - (1 - D_i)P_i^n] - cD_i \tag{28.14}$$

where:

$D_i = 1$ if mail to customer i at time t; 0 if not.
$\pi_i(D_i)$ = Profit earned on customer i over next x months depending on whether or not the customer is mailed a catalog.
m = Profit margin per response.
$E(A_i)$ = Expected expenditure level for customer i over next x months if the customer purchases.
P_i^c, *or* P_i^n = Probability that customer i purchases ("responds") over the next x months depending on whether he or she receives a catalog at time t ("c") or does not receive a catalog ("n").
c = Cost of mailing catalog (production plus mail cost).

Table 28.1 Optimal versus actual mailing policy for Gönül et al. model (From Gönül et al. 2000)

		Actual		
		Send	Do not send	Total
Optimal	Send	16	92	108
	Do not send	92	779	871
	Total	108	871	979

The hazard response model provides the response probability estimates. Note P_i^n does not equal zero because the customer has been mailed catalogs before and could order from those catalogs. The variables that change when the customer receives a catalog are WEAROUT (increases by 1) and PROM_REST (resets to 0). Gönül et al. (2000) decision rule is to mail to customer i if:

$$\Delta\pi_i = \pi_i(D_i = 1) - \pi_i(D_i = 0) > 0 \tag{28.15}$$

The authors apply their method to a durable household products catalog. Table 28.1 compares the optimal and actual policies. Out of 108 customers actually sent catalogs, 16 should have been sent the catalog. However 92 of them (close to 90%) should not have been sent catalogs. The total expected profit is \$6,327 under the optimal policy compared to \$5,968 under the actual policy.

These findings suggest the cataloger is mis-targeting catalogs. The authors speculate this may be due to management not understanding the wear-out and forgetting phenomena captured by the hazard model, or are not considering heterogeneity.

The Gönül et al. (2000) approach is a rigorous, practical approach to catalog mailing. The model is not a dynamic program in that it only optimizes one mailing at a time. Gönül and Ter Hofstede (2006) extend the approach to address this by considering a finite decision period of length P periods, and evaluate 2^P possible mailing schedules. They evaluate each schedule in terms of the utility of the firm, using risk-neutral as well as risk-neutral profit functions. They also use simulation to "integrate out" the uncertainty in customer parameter values. The method evaluates each of the 2^P schedules separately. For a 52-week schedule, this could get prohibitive in terms of computer resources. However, the authors show that $P = 6$ improves over a myopic ($P = 1$) optimization, so practically speaking, the model promises improvements over non-forward looking mail decisions.

28.3.4 Using a Hierarchical Bayes Model (Rust and Verhoef 2005)

28.3.4.1 Response Model

Rust and Verhoef (2005) use a hierarchical Bayes model to estimate customer response to two marketing mix interventions – direct mail and a relationship magazine. The setting is a Dutch insurance company that must decide how many direct mail pieces and how many "relationship magazines" to send to each customer in the coming year. The response model is as follows:

$$\Delta R_{i,(t-1)\to t} = [ln(\overrightarrow{M}_i + 1)]\beta_i + \varepsilon_i \tag{28.16a}$$

$$\beta_i = \overrightarrow{Z}_i \alpha + \delta_i \tag{28.16b}$$

where:

$\Delta R_{i,(t-1)\to t}$ = Change in profits (gross of marketing costs) for customer i between previous and current year.

$\overrightarrow{M}_i = \{M_{i1}, M_{i2}\}$, where M_{i1} is the number of direct mail pieces sent to customer i, and M_{i2} is the number of relationship magazines sent to customer i, in the current year.

$\beta_i = \{\beta_{i1}, \beta_{i2}\}$, customer i's responsiveness to direct mail and relationship magazines respectively.

$\overrightarrow{Z}_i$ = Vector of behavioral and demographic variables for customer i, such as lifetime duration, number of products purchased, gender, etc.

$\alpha = \{\alpha_1, \alpha_2\}$ Impact of behavioral and demographic variables on customer i's responsiveness to direct mail and relationship magazines, respectively.

ε_i, δ_i = Unobserved factors influencing customer i's change in profits in the current year and responsiveness to marketing, respectively.

Equation 28.16a reflects diminishing returns to marketing efforts through the log transformation (the "1" is to avoid having to take the log of zero). Equation 28.16a represents the impact of behavioral and demographic variables on customer response to marketing. For example, the authors hypothesized that in general, loyal customers would be more receptive to the relationship magazine and less receptive to direct mail.

The model was estimated for 1,580 customers using MCMC methods implemented in WinBugs. The dependent variable is *change* in profits over a 1-year horizon. The authors found several interesting results, generally supportive of their hypothesis regarding loyalty. For example, cumulative number of purchases had a negative impact on responsiveness to direct mail, whereas had no impact on responsiveness to the relationship magazine. Membership in the company's loyalty program had a stronger effect on response to the magazine than it did to response to direct mail.

28.3.4.2 Optimization Model

The objective is to maximize each customer's change in profits in the coming year:

$$\Pi_{i,(t-1)\to t} = \Delta R_{i,(t-1)\to t} - \overrightarrow{M}_i \overrightarrow{C} \tag{28.17}$$

where $\overrightarrow{C} = \{C_1, C_2\}$ is the per unit cost per customer of direct mail and relationship magazines, respectively. Given this formulation and the response function, the optimal level of marketing instrument k (k = direct mail, relationship magazine) for customer i can be obtained using simple calculus:

$$M^*_{ik} = \frac{\beta_{ik}}{C_k} - 1 \tag{28.18}$$

Equation 28.18 says that more of marketing instrument k should be allocated to customer i if customer i is more responsive to that instrument, and marketing instrument k is less expensive. Since responsiveness is the only factor that varies across customers, it is the key measure, and it is provided by the estimation of Equations 28.16.

The authors calculate the optimal level of marketing for each customer and find it is quite heterogeneous due to heterogeneity in the response measures (β). They compare their model to three others: segmentation based on demographics, segmentation based on RFM variables, and latent structure segmentation. They find that their model fits better than the other two, and that the predicted profits generated by their model are higher than those generated by the competitive models. In particular, they find:

Model	Mean square Error (Fit)	Projected average Profit (Guilders)
Demographic	12.98	14.46
RFM	13.44	8.61
Latent class	23.49	3.12
Hierarchical	12.42	23.12

Profits under the marketing plan currently used by the company generated 10.57 guilders, so the hierarchical model outperformed both the other models and current practice.[3]

The Rust and Verhoef model is a very practical yet rigorous approach to deciding customer-specific investments in the intermediate term. Like Gönül et al., it is not a dynamic optimization – it does not take into account the investments made in the coming year have on long-term retention rates and lifetime value. It does not explicitly model wear-in, wear-out, and forgetting, and so could not be used to *schedule* marketing activities within the year. However, the model does include decreasing returns on an aggregate basis, so it implicitly accounts for these factors for the 1-year time horizon. The model depends heavily on the customer-specific response parameters estimated on the calibration sample of 1,580 customers. A challenge would be to infer the coefficients for the rest of the firm's customers.

[3] Note the authors use the hierarchical model to project profits for both their model and the other models. The justification is that the hierarchical model predicted best, so would make the most accurate projection.

28.3.5 Incorporating Customer and Firm Dynamic Rationality (Gönül and Shi 1998)

28.3.5.1 Response Model

Many optimal contact models assume that the firm is forward looking, i.e., "dynamically rational." However, there is growing evidence that customers are also dynamically rational – they consider the impact of their current purchase on future costs and benefits. For example, consumers have been shown to take into account the likelihood of future promotions in deciding whether to purchase in period t (Gönül and Srinivasan 1996; Sun et al. 2003).[4]

Gönül and Shi's (1998) optimal contact model takes into account that *both* the customer and the firm may be forward looking. The customer's utility function is:

$$u_{it} = \alpha + \beta_m m_{it} + \beta_{1r} r_{it} + \beta_{2r} r_{it}^2 + \beta_{1f} f_{it} + \beta_{2f} f_{it}^2 + \varepsilon_{it} \qquad (28.19)$$

where:

u_{it} = Utility for customer i of making a purchase in period t; = 0 if the customer does not make a purchase.
m_{it} = 1 if customer i receives a catalog in period t; 0 if not.
r_{it} = Recency, the number of periods since the last purchase.
f_{it} = Frequency, the number of purchases made by the customer since the beginning of the data.

Utility is considered quadratic functions of both recency and frequency for flexibility.

The customer is assumed to maximize his or her long-term utility of making a purchase in period t, taking into account the future impact of a current purchase on his or her recency and frequency variables. Recency and frequency are the state variables in the customer's dynamic program, summarized by $S_{it} = \{r_{it}, f_{it}\}$. Each period, the customer decides whether to buy ($d_{it} = 1$) or not buy ($d_{it} = 0$) by considering the following:

$$V_{it}(S_{it}) = \begin{cases} u_{it} + \delta_c E[V_{i,t+1}(S_{i,t+1}|d_{it} = 1)] & \textit{if } d_{it} = 1 \\ 0 + \delta_c E[V_{i,t+1}(S_{i,t+1}|d_{it} = 0)] & \textit{if } d_{it} = 0 \end{cases} \qquad (28.20)$$

Customers realize that purchasing or not purchasing changes recency and frequency, and that will affect future utility depending on the parameter values in Equation 28.19. Gönül and Shi (1998) estimate Equation 28.19 by maximum likelihood (see also Keane and Wolpin 1994).

[4] Note that competitive economic models in the database marketing literatures have considered the case that both firms and customers are forward looking. See Chapter 2 for discussion.

Gönül and Shi (1998) find the dynamic model fits better than a static model. This suggests customers consider the future when deciding whether to buy now. The mail variable has a positive coefficient as expected. Both recency and frequency have U-shaped impacts. The recency result means that the customer is most likely to buy right after the previous purchase or after a significant lapse of time. The frequency result implies that customers who have bought the product very frequently or very infrequently have more need for the product in the current period.

Gönül and Shi's response model does not take into account wear-in, wear-out, and forgetting. However, this could be done through lagged mailing variables as in Gönül et al. (2000). The model includes a "structural model" of acceleration in the sense that the customer will purchase earlier if he or she realizes that this will increase his or her future utility. See Li et al. (2005) for an extension of the Gönül and Shi model optimizing two elements of the marketing mix (messages and price).

28.3.5.2 Optimization Model

The firm's problem is to decide whether to mail to a customer each period depending on which state the customer is in. The current period profit for the firm is:

$$\pi_{it}(S_{it}, m_{it}) = R\ Prob_{it}(d_{it} = 1|S_{it}, m_{it}) - cm_{it} \tag{28.21}$$

where:

$\pi_{it}(S_{it}, m_{it})$ = Profit for customer i in period t, given the customer is in state S_{it} and a decision to mail or not mail.
$m_{it} = 1$ if mail to customer i in period t; 0 otherwise.
R = Revenues from customer i if the customer purchases.
$d_{it} = 1$ if customer purchases; 0 otherwise.
c = Cost to mail to customer i.

The recency/frequency state (S_{it}) is the state variable for the dynamic program. The firm decides on the mailing policy that maximizes long-term profits:

$$P_{it}(S_{it}) = \sum_{j=t}^{\infty} \delta_f^{j-t} \pi_{it}(S_{it}, m_{it}^*(S_{it})) \tag{28.22}$$

where:

$P_{it}(S_{it})$ = Maximum expected profits to be gained through an optimal mailing policy for customer i who starts period t in state S_{it}.

By the principle of optimality, long-term profits equal the profits from maximizing current period profits plus the maximal profits to be earned from period $t + 1$ onward, i.e.,:

$$\begin{aligned} P_{it}(S_{it}) &= max\,\{\pi_{it}(S_{it}, m_{it}) + P_{it+1}(S_{it+1})\} \\ &= \max_{m_{it}}\,\{\pi_{it}(S_{it}, m_{it}) + \delta_f[Prob_{it}(d_{it}=1|S_{it}, m_{it})P_{it+1}(S_{it+1}|d_{it}=1) \\ &\quad + Prob_{it}(d_{it}=0|S_{it}, m_{it})P_{it+1}(S_{it+1}|d_{it}=0)]\} \end{aligned} \tag{28.23}$$

Gönül and Shi (1998) calculate the steady state optimal mail decision using successive approximation (Ross 1983). The authors maximize the firm's future profits taking into account the customer's forward looking response to the mailing decision. They apply their model to a durable product. The authors find that if recency is low, the firm does not mail, but if recency is medium or high, the firm does mail. The reason for this is that if recency is low, the customer is likely to buy anyway (recall the U-shaped finding for recency) so mailing is unnecessary. If recency is medium, response probability is low without a mailing so the mailing is needed. When recency is high, customers are likely to buy on their own, but it pays to mail to these customers anyway to make sure they purchase and push up frequency to a higher level where the customer will buy on his/her own.

Gönül and Shi (1998) calculate that profits using their model would have been 16% higher than what the firm actually earned during the data period. They also note that if one just considers the short term, the firm would probably not do any mailing. This is because the break-even incremental response probability to justify a single period mailing is 9.37%, and the incremental response probability from a mailing is typically less than that. However, from a long-term perspective, a mailing in the current period boosts consumers to profitable recency and frequency states.

In summary, the Gönül and Shi approach embeds a dynamic rational customer response model within a dynamic firm optimization. While no other mail policy will increase firm profits, other mail policies might increase customer utility. This is similar to a Stackelberg game where the leader is the firm and the follower is the customer. The customer is assumed to know the mailing schedule. Indeed, many customers probably do learn how often they receive catalogs from a given company. However, it would be interesting to include customer catalog mailing expectations in the model.

28.3.6 Incorporating Inventory Management (Bitran and Mondschein 1996)

28.3.6.1 Response Model

Bitran and Mondschein (1996) use an RFM model (Chapter 12) as follows:

S_i = Market segment or "state" i, defined by specific values of recency, frequency, and monetary value.

$p_{S_i,S_j,k}$ = Probability that a customer moves from state S_i to S_j when he or she receives k mailings in a given time period.

Recency is defined as the number of periods since the last purchase. Frequency is defined as 1 if the customer has bought once and 2 if more than once. Monetary value is defined as two values, \$55 and \$80. This makes for a total of $7 \times 2 \times 2 = 28$ RFM states. The authors estimate the response probabilities using historical data available from the catalog company with whom they applied their model.

28.3.6.2 Optimization Model

Bitran and Mondschein's (1996) optimization considers (1) how many catalogs to mail to house list *and* rental list customers each season, and (2) how much the firm should invest in inventory, subject to a budget. The model takes into account firm-level constraints; it does not "simply" optimize individual customers. The optimization is a stochastic dynamic program, but because the model is at the firm level, the states are numerous and continuous. They include the *number* of customers in each RFM state, plus the budget available and inventory levels. The model is not easily solved because of the "curse of dimensionality" – there are too many state variables and each of them takes on too many values. That is, if there are 27 RFM states, there are 27 number-of-customers variables plus a budget variable plus an inventory variable, resulting in 29 continuous state variables. This compares to just four states for the Ching et al. (2004) model.

To simplify, the authors first calculate the number of catalogs to mail to a customer to maximize lifetime value, assuming no budget restrictions or inventory costs. Second, they calculate optimal inventory re-ordering to maximize one-period profit. Third, they incorporate their lifetime value and inventory calculations to derive the optimal mailing policy across all customers. The lifetime value optimization is:

$$LF(s_i) = \underset{k}{Max} \begin{cases} \bar{b}_{s_i k} + \beta \sum\limits_{s_j} p_{s_i s_j k} LF(s_j) & k = 1 \ldots K \\ \beta \sum\limits_{s_j} p_{s_i s_j 0} LF(s_j) & k = 0 \end{cases} \tag{28.24}$$

where:

$s_i =$ RFM state i, defined by particular values of RFM.
$LF(s_i) =$ Lifetime value of a customer in state s_i.
$k =$ Number of catalogs mailed under optimal policy (this will differ for each s_i). K is the maximum number of catalogs to be mailed.
$p_{s_i s_j k} =$ Probability customer in state s_i migrates to state s_j if mailed k catalogs.
$\beta =$ Discount factor.
$\bar{b}_{s_i k} =$ Current period profit if customer in state s_i mailed k catalogs.

$$= \sum (d_{s_i s_j}(1-g) - c_1) p_{s_i s_j k} - k c_h \tag{28.25}$$

where:

$d_{s_i s_j}$ = Amount of money spent by customer who migrates from s_i to s_j.
g = Cost of goods sold as percentage of revenue.
c_1 = Cost of filling order.
c_h = Cost of mailing to member of house list.

The inventory optimization calculates the amount of inventory to order to maximize one-period profit, assuming the firm enters period t with a certain level of inventory. The authors find the optimal amount to invest in inventory is:

$$Z_t = \sigma_t a + \mu_t - I_t \tag{28.26}$$

where:

Z_t = Amount to invest in inventory at the beginning of period t.
μ_t = Expected demand during period t.
σ_t = Standard deviation of demand in period t.
I_t = Inventory value at beginning of period t.
a = Solution to the following equation:

$$F_y(a) = \frac{1 + c_2 - g}{1 + c_2 + c_3 - \beta g} \tag{28.27}$$

where:

$F_y(a)$ = The cumulative standard normal distribution.
c_2 = Penalty cost for unfulfilled demand.
c_3 = Inventory holding cost.

From Equation 28.26, it follows immediately that the amount to order is increasing in average demand and decreasing in inventory. From Equation 28.27, it can be shown that the firm should order more if the penalty costs for unfulfilled demand are higher, less if it costs more to hold inventory, and order less if COGS is high.

Note that the inventory investment depends on demand, which depends on the mailing policy and response. Likewise, the amount spent on inventory influences the number of mailings because it limits cash availability. Bitran and Mondschein make a major contribution by combining marketing and operations decisions. The optimal firm-level mailing policy is derived by the following linear program:

$$max \sum_{s_i} \sum_{k} LF(s_i k) d(s_i, k) + \sum_{j} LF(j) d(j) \tag{28.28}$$

subject to:

$$\sum_{s_i} \sum_{k} k c_h d(s_i, k) + \sum_{j} c_m d(j) \leq Y_t + a_t \tag{28.29a}$$

$$\sum_{s_i} d(s_i, k) \le N_{s_i t} \qquad \forall s_i \tag{28.29b}$$

$$d(j) \le L_{jt} \qquad \forall j \tag{28.29c}$$

where:

$LF(s_i k)$ = Lifetime value of customer in state s_i assuming mail k catalogs in current time period, and the optimal number thereafter according to Equation 28.24. j refers to customers on rental list j.
$d(s_i, k)$ = Number of customers in RFM state s_i who receive k catalogs in current period.
$d(j)$ = Number of catalogs mailed to rental list j in current time period.
Y_t = Available funds after mailing and inventory investment.
a_t = Exogenously provided funding from corporate level in time t.
$N_{s_i t}$ = Number of customers in state s_i in period t.
L_{jt} = Number of customers available from rental list j in period t.

The decision variables are $d(s_i, k)$ *and* $d(j)$. The first constraint says that the firm cannot spend more on mailings than available funds. Equations 28.29b–c ensure that the total number of customers who receive catalogs does not exceed the sum of individuals who are in each RFM group or on each rental list.

The authors implement the optimization by calculating the optimal mailing plan via Equations 28.28–28.29, calculating the optimal re-order implied by that plan (via Equation 28.26), then checking to make sure the re-order plus mailing costs are within the cash constraint. If not, the lowest lifetime value segment (determined by Equation 28.24) is dropped. The process iterates until the cash constraint is satisfied.

The authors apply their model using data from a catalog company, and compare simulated profits generated from their method to a theoretical upper bound. They find their method does well compared to the upper bound, usually capturing more than 95% of the upper bound profits. The authors generate several insights based on their simulations. For example, when constrained by cash availability, it is better mail to more customers than mail more often to a smaller set of customers. This is to prevent the not-mailed-to customer's recency from becoming so high that they effectively exit the house list. Also, start-up catalogs should use multiple mailings early to build up frequency so that the customer becomes firmly entrenched in the house list. This is the intuitive notion that new companies should emphasize acquisition over retention (see Chapter 26).

28.3.7 Incorporating a Variety of Catalogs (Campbell et al. 2001)

28.3.7.1 Response Model

Campbell et al. (2001) describe the methodology implemented by Fingerhut, a cataloger that was mailing more than 340,000,000 catalogs to 7,000,000

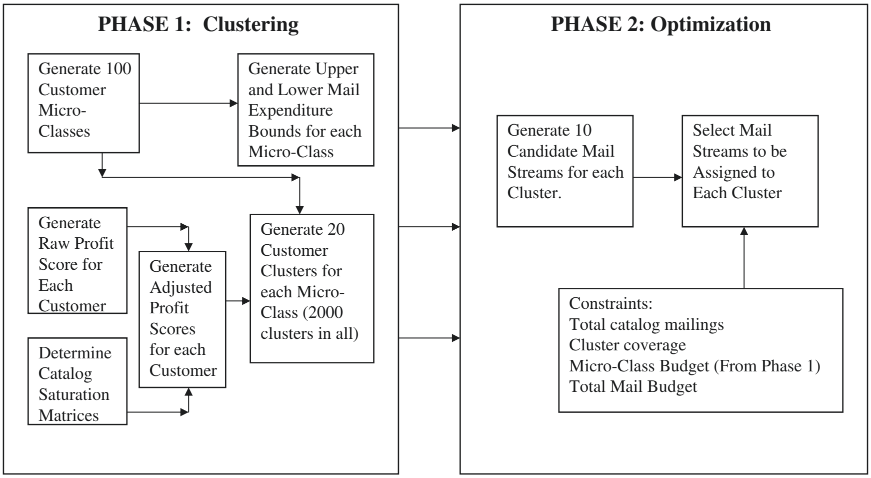

Fig. 28.8 Catalog mail stream optimization system (From Campbell et al. 2001).

customers annually. The problem is to generate a *set* of catalogs (a "mail stream") to send to each customer over the planning horizon. Catalogs differ in terms of their content and when they are to be sent. The method consists of two steps, "Clustering" and "Optimization," as shown in Fig. 28.8. The Clustering phase measures response, while the Optimization phase determines which mail stream to send to which cluster.

The Clustering phase first produces 100 micro classes homogeneous with respect to customer value, lifetime duration, and catalog productivity. This is used to set mailing budgets per class. A second stage produces 20 clusters within each micro class that are homogeneous with respect to predicted response to the catalogs under consideration. Mail streams are then tailored to each of the resulting 2,000 (100 × 20) clusters.

28.3.7.2 Optimization Model

The customer-level optimization problem can be stated as:

$$MAX\ \ Profit = \Pi = \sum_i \sum_p [G_p^i - F_p^i] X_p^i - \sum_i \sum_p \sum_{p'} G_p^i S_{p,p'} X_p^i X_{p'}^i \tag{28.30a}$$

such that:

$$\sum_i \sum_p F_p^i X_p^i \leq B \tag{28.30b}$$

where:

G_p^i = Gross profit from mailing catalog p to customer i.
F_p^i = Cost to mail catalog p to customer i.

$X_p^i = 1$ if mail catalog p to customer i; 0 otherwise.
$S_{p,p'}$ = Saturative effect between catalogs p and p′ (Sect. 28.2.2)
B = Mailing budget constraint.

The saturative effects provide interesting dynamics. However, the authors cannot maximize the above because there are 7,000,000 customers times 40 potential catalogs, or 280,000,000 decision variables. That is simply too large to solve directly.

Accordingly, Campbell et al. simplify by dividing customers into the 2,000 clusters described above. They then conduct the optimization in two steps. First they generate 10 candidate mail streams for each cluster. Next they decide which mail streams to use for each cluster. The individual cluster optimization is virtually the same as Equations 28.30a–b, and is described by Campbell et al. (2001) as follows:

$$MAX\ Z = \sum_p [R_p - E_p]Y_p - \sum_p \sum_{p'} R_p S_{p,p'} Y_p Y_{p'} \tag{28.31a}$$

$$\text{such that} : \sum_p E_p Y_p \leq B \tag{28.31b}$$

where:

R_p = Gross profit from mailing catalog p to the cluster.
E_p = Cost to mail catalog p to the cluster.
$Y_p = 1$ if mail catalog p to the cluster; 0 otherwise.

The difference between Equations 28.31a–b and Equations 28.30a–b is dropping the i subscript for individuals. Optimization 28.31a–b is solved for 2,000 different customer clusters. For each optimization, there are 40 decision variables corresponding to 40 catalogs, and a budget constraint. The optimization is solved using a range of 10 values for the budget constraint, so 10 candidate mail streams are generated for each cluster.

A crucial output from this stage is the expected profit per customer for mailing mail stream m to cluster j within micro-class k, G_m^{kj}. This quantity drives the objective function for the global optimization that maximizes profits across clusters:

$$Max \sum_k \sum_j \sum_m G_m^{kj} X_m^{kj} \tag{28.32}$$

subject to:

$$Q_p \leq \sum_k \sum_j \sum_m C_{pm}^{kj} X_m^{kj} \leq \bar{Q}_p \quad \forall p \tag{28.33a}$$

$$A^k \leq \sum_m \sum_j F_m^{kj} X_m^{kj} \leq \bar{A}^k \qquad \forall k \tag{28.33b}$$

$$\sum_{m} X_m^{kj} = V^{kj} \qquad \forall j, k \tag{28.33c}$$

$$T \leq \sum_{k} \sum_{j} \sum_{m} F_m^{kj} X_m^{kj} \leq \bar{T} \tag{28.33d}$$

where:

G_m^{kj} = Profit per customer for mailing mail stream m to cluster j within micro-class k (from stage 1 optimization).
X_m^{kj} = Number of customers in cluster j, micro-class k, who receive mail stream m.
$Q_p, \bar{Q}_p$ = Lower and upper bounds for number of catalogs of type p that can be distributed.
C_{pm}^{kj} = 1 if catalog p is included in mail stream m, for cluster j, micro-class k.
$A^k, \bar{A}^k$ = Lower and upper bounds for catalog mailing budget for micro-class k.
F_m^{kj} = Cost of mailing mail stream m to cluster j in micro-class k.
V^{kj} = Total number of customers in cluster j in micro-class k.
$T, \bar{T}$ = Lower and Upper bounds for total mailing costs.

The decision variable is the number of customers within a given cluster who receive a particular mail stream. There are 20,000 decision variables since there are 2,000 clusters and 10 potential mail streams per cluster. Equations 28.33a–d represent 2,141 constraints. The first (Equation 28.33a) is that each catalog has an upper and lower bound for the total number of mailings. These cover catalog development costs and maintain firm positioning. There are 40 such constraints, one for each catalog.

The second constraint (Equation 28.33b) is that each of the 100 micro-classes has a minimum and maximum level of catalog mailing investment. There are 100 such constraints. The bounds are generated in the first phase of the system. This ensures that customer segments receive minimum levels of investment while avoiding wear-out. The third constraint (Equation 28.33c) is that all customers in each cluster must receive a mail stream. There are thus 2,000 such constraints. The final constraint (Equation 28.33d) requires that the total mailing investment must be between upper and lower bounds.

In summary, Campbell et al. (2001) replace an optimization over 280,000,000 decision variables with one constraint by two optimizations – one with 40 decision variables and one constraint that is solved 2,000 times, and another that has 20,000 decision variables and 2,140 constraints.

Campbell et al. (2001) report a field test consisting of 700,000 test and 700,000 control customers. The goal was to see if the system could generate incremental profit by lowering mailing costs. Indeed, the system reduced mailing costs by 6%, and as a result, revenues fell by 1.5%. However, the net impact was a profit gain of 2%. The effects were particularly strong for customers who had not bought recently from Fingerhut. Campbell et al. (2001)

report that the system is "directly responsible for a \$3.5 million annual profit gain," and that the "project paid for itself within the first year" (p. 86).

The system is very innovative in its use of saturation interactions (Sect. 28.2.2). The formation of clusters assumes customers within cluster have homogeneous response to mail streams, but the clustering is based not on response to mail streams, but to individual catalogs and aggregate measures of mailing response. On the optimization side, it is not clear how much is lost by optimizing in two stages rather than one. In conclusion, Campbell et al's (2001) method is innovative in its response function, practical in its optimization, has demonstrated value in the real world, and provides ample opportunities for future research.

28.3.8 Multiple Catalog Mailings (Elsner et al. 2003, 2004)

28.3.8.1 Response Model

Elsner et al. (2003, 2004) develop a "dynamic multilevel model" (DMLM) for optimizing the targeting of a single catalog, and a "dynamic multidimensional model" (DMDM) to target different catalogs. We focus on the DMLM model and then discuss how it is extended to DMDM. DMLM consists of three steps or levels, each of which requires its own response function analysis:

1. Determine how many catalog campaigns to conduct during the next 12 months, what should be the timing between campaigns, and on what day to mail the catalogs for a given campaign. Let n_{opt} be the optimal number of campaigns.
2. Determine which customer segments should receive the n_{opt} campaigns.
3. Conduct an additional segmentation analysis that determines which customers are "inactive" and hence should receive a "reactivation package" and which if any should receive the normal catalog mailing

In the first step, the authors conduct field tests that provide data for regressions that relate response rates and order sizes to the number of catalogs distributed, the day of which customers received catalogs, and the time between mailings. An interesting finding is that Saturday is the optimal day to deliver a catalog. This makes sense in that Saturday begins the weekend, when customers have more time to read through catalogs.

In Step 2, the authors divide customers into three segments based on recency. They then estimate the response rate for each segment per catalog. The authors assume that for a given segment, the response rates do not change from campaign to campaign. They can then forecast how customers will migrate between recency segments, similar to Bitran and Mondschein

(1996). They then calculate profits if a segment participates in n_{opt} campaigns and hence whether it is profitable receive n_{opt} campaigns. An important output of this step is a breakeven cut-off s^*. If a given segment's expected sales rate is less then s^*, that segment does not receive the n_{opt} campaigns.

Step 3 looks at a complete array of RFM and other variables and determines whether a customer segment will achieve the critical breakeven point or not. If not, further analysis is conducted to determine if it is worthwhile to send the customers in that segment a special "reactivation" package.

Note that the authors assume in step 2 that response rates for a given segment are constant over time and do not depend on the frequency or timing of catalogs. It thus appears that this model does not take into account wear-in, wear-out, and forgetting. However, as in the Rust and Verhoef (2005) model, step 1 implicitly does, since it regresses at an aggregate level total response and order size as a function of frequency and timing.

28.3.8.2 Optimization Model

They authors find in Step 1 that 25 bimonthly catalog campaigns, spaced 14 days apart, and delivered on Saturdays is optimal (Elsner et al. 2003; Fig. 28.4). The authors then divide their customer base into three recency segments (e.g., recency $<$ 12 months, 12 $<$ recency $<$ 24, and recency $>$ 24). Given the response rate and order size for each segment, as determined in Step 2, they derive expressions for the expected number of customers in each segment and hence its profitability, if that segment receives n_{opt} catalogs. They take into account that customers may be acquired or leave the database entirely by moving without a forwarding address, etc. In summary, these expressions calculate the total profit as a function of the customer migrations that occur between recency states depending on whether a customer receives and responds to a given catalog. Using these expressions, they calculate s^*. If a given segment j's s_j = response rate $\times$ order size is greater than s^*, the segment receives the n_{opt} catalogs.

Step 3 provides a predictive model that identifies more specifically (on the basis of more than just recency variables) which customers will have $s_j < s^*$. For those customers, the authors conduct additional analysis, scrutinizing their response rates, etc., to determine if it is worthwhile to send them a "reactivation package."

The authors apply their procedure to a German catalog company, Rhenania, and report improvements in sales and the size of the customer base. Profit starts increasing a year later. The company does so well that they acquire another catalog company, Akzente, and apply the model to that company. Similar to the results for Rhenania, the number of active customers, sales growth, and even profit immediately start to increase after 2–3 years of decline.

The authors attribute the success to: (1) a forward-looking optimization rather than optimizing one mailing at a time; (2) the use of segmentation to help decide what minimum number of expected sales was required in order to mail to the segment, and (3) further segmentation to identify active versus inactive customers, and using a reactivation campaign selectively on customers considered most likely to respond profitably.

The authors found that after acquiring additional direct mail companies, a new model was needed, DMDM, to optimize customer contacts across three different types of catalogs. The authors follow generally the same three steps as in DMLM, however, for example in Step 1, they also consider response to the total number of mailings, across the three catalogs. Implicitly included is the cross-correlation between response to catalogs for the different brands. In applying this model (Elsner et al. 2004), they find for example there is more cross-buying of products from different catalogs.

28.3.9 Increasing Response to Online Panel Surveys (Neslin et al. 2007)

28.3.9.1 Response Model

Neslin et al. (2007) develop a model to increase response rates for online survey panels. Online survey panels have become an important source of survey data. Nearly 80% of consumer goods and 74% of B2B companies use online panels (Thornton 2005). Online panels provide fast turnaround, lower operations costs (compared to mail surveys or personal interviews), and more specialized sample frames. To increase response rates, the online panel manager might increase participation incentives or recruit more panelists. Either way this increases costs. Another alternative is to use an optimal contact model. This identifies panelists who are likely to respond and uses them judiciously over time to maximize response rates.

Neslin et al. use a decision tree to model response to previous survey solicitations. They consider several potential predictors; their final model contains the following four:

- Days between the previous invitation or joining the panel and the current invitation (INVJOIN): Lower INVJOIN was associated with higher response.
- E-mail response confirmation (CONFIRM): The firm had sent an e-mail to panelists asking whether they were still interested in participating. Possible responses were "Yes," "No," and "No response." Those who said yes were most likely to respond to mailings, while those who said no were unlikely to response. No-response customers fell in the middle.
- Response to previous invitation (PREVRESP): Respondents might have responded to the previous invitation, not responded, or never been invited

before. Responders were obviously more likely to respond to subsequent invitations, non-responders were least likely, and never-invited were in the middle.
- Gender: Females were somewhat more likely to respond than males.

The decision tree includes two variables that change over time, INVJOIN and PREVRESP. The authors found no evidence of wear-out. This may be due to the range of the data. Very few panelists in the data had been invited to more than two studies.

As in Simester et al. (2006), the decision tree end nodes divide customers into states. Customers migrate from state to state. For example, if the cut-off for being in the low INVJOIN state is 61 days, then after 61 days, the customer migrates to the INVJOIN > 61 state, where response rates are generally lower. Similarly, since PREVRESP is a predictor, panelists migrate to different states depending on whether they respond to a given invitation. This customer migration plays a critical role in the optimization, as it does for several of the other models reviewed in this chapter.

28.3.9.2 Optimization Model

The optimization model is forward looking over a finite horizon the authors chose to be the next four studies. The decision is how many customers in state j to invite to participate in a given study. The optimization takes into account that new panelists may be added to the database over time. It also takes into account that given studies may require demographic balance, for example, an equal number of males and females.

Specifically, the authors formulate their optimization as a linear program:

$$\underset{X_{js}}{Minimize} \sum_{s=1}^{S} \sum_{j=1}^{N} X_{js} \tag{28.34}$$

subject to:

$$X_{js} \leq A_{js} \quad j = 1, \ldots, N; s = 1, \ldots, S \tag{28.35a}$$

$$\sum_{j \in M_g} r_j X_{jt} \geq Q_{sg} \quad g = 1, \ldots, G; s = 1, \ldots, S \tag{28.35b}$$

$$A_{ks} = \sum_{j=1}^{N} p_{jk} \ X_{j,s-1} + \sum_{j=1}^{N} q_{jk}(A_{j,s-1} - X_{j,s-1}) + R_{ks} \quad k = 1, \ldots, N; s = 2, \ldots, S \tag{28.35c}$$

where:

X_{js} = Number of panelists in state j invited to participate in study s.
A_{js} = Number of panelists in state j available to participate in study s.

r_j = Response rate for panelists in state j.
Q_{sg} = Number of respondents desired from demographic group g for study s.
p_{sj} = Probability panelist in state j migrates to state k if invited to participate in a given study.
q_{jk} = Probability panelist in state j migrates to state k if invited to participate in a given study.
R_{ks} = Number of newly recruited panelists joining state k in time for study s.

The objective is to minimize the number of invites over the horizon. Since the model includes constraints on the desired number of respondents for each study, this is equivalent to maximizing average response rate. Constraint 28.32a ensures the solution can not invite more panelists from a given state to participate in a given study than are available. Constraint 28.32b states the required number of respondents from each demographic group; M_g is the set of states that contain panelists from demographic group g. Constraint 28.32c keeps track of panelists available for each state, for each study. The number of panelists in state k equals the number of responding panelists who migrate to state k plus the number of non-responding panelists who migrate to state k, plus the number of newly recruited panelists who enter state k. The migration is governed by the migration probabilities p and q. These in turn are derived by the definition of states as determined by the end nodes of the decision tree, and to the schedule of studies.

The authors use a rolling horizon implementation. Rolling horizons are used frequently in operations management (Baker 1977; Chand et al. 2002) as a pragmatic way to implement models when uncertainty is involved. In the authors' context, the approach is: (1) Find the optimal solution for Studies 1, 2, 3, and 4. This solution is based on *expected* panelist migration calculated using Equation 28.35c. (2) Implement the solution for Study 1. (3) Observe who actually responds or does not, thus calculating the *actual* numbers of customers in each state as of Study 2. (4) Find the optimal solution for Studies 2, 3, 4, and 5. (5) Implement the solution for Study 2, etc.

The authors field test the model and compare it to random selection and the firm's current heuristic for selecting panelists. Figure 28.9 shows the model outperforms both alternatives. The reason the optimization distinguishes itself particularly for the last three studies is that for the first study, there were not many panelists available in what the predictive model identified as high-responding states. However, in the first study, the model solicits panelists to discern whether they were in the high-responding group or not. For example customers with low INVJOIN and PREVRESP = "respond" are high response panelists. However, there were not any of these available for Study 1. By inviting high INVJOIN previous responders, the model could create a pool of low INVJOIN previous responders for Study 2 who would be highly likely to respond to that study. This strategy evidently worked well.

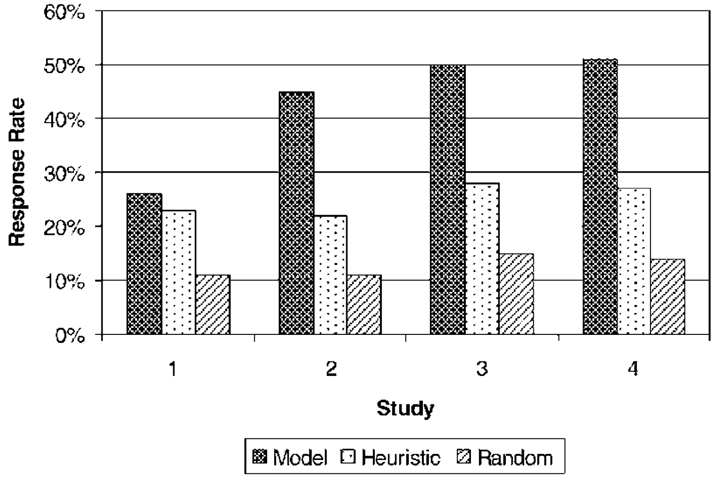

Fig. 28.9 Field test of optimal contact model for increasing online panel response rates (From Neslin et al. 2007).

28.4 Summary

Table 28.4 summarizes the various features contained in the optimal contact models discussed in this section. All but one of the models is applied to direct mailings or promotions. Neslin et al. (2007) show that the method is more broadly applicable (see also Sun and Li 2005, Chapter 25, for an application to call centers). Other potential applications include e-mails, online advertising, and multi-channel promotions. Most of the methods focus on one type of communication, e.g., a single catalog, rather than a selection of catalogs. Campbell et al. (2001), Rust and Verhoef (2005) and the extension of the basic model in Elsner et al. (2003, 2004) are important exceptions. Considering a selection of communications raises the issue of communication overlap, which Campbell et al. model in an innovative way.

Optimization methods range from simple profit cut-offs (Gönül et al. 2000) and linear programs (Neslin et al. 2007) to multi-stage optimizations (Campbell et al. 2001; Elsner et al. 2003, 2004). In both these multi-stage optimizations, the first consideration is the schedule of catalogs, while the second is which customer segment should receive which schedule. This approach may be a necessary simplification when there are different types of communications under consideration.

Most of the models assume the firm is forward looking. Rust and Verhoef (2005) are one exception. They focus on the aggregate level of marketing effort to expend on each customer within a year, without worrying about several years or the schedule within a year. The detailed scheduling of marketing effort is what creates a complex dynamic optimization. It would be very

Table 28.2 Comparing optimal catalog mailing procedures

		Bitran and Mondschein (1996)	Gönül and Shi (1998)	Gönül et al. (2000)	Campbell et al. (2001)	Elsner et al. (2003)	Ching et al. (2004)	Rust and Verhoef (2005)	Simester et al. (2006)	Neslin, Novak, Baker, and Hoffman (2007)
Decisions	Catalog mailing	√	√	√	√	√	("Promotions")	(Direct mail; magazine)	√	[a]
	Catalog diversity	–	–	–	√	[b]	–	√	–	–
Optimization	Method	Dynamic program	Dynamic Stackelberg	Short-term profit cut-off	Integer and linear programs	Calculus	Dynamic program	Calculus	Dynamic program	Linear Program
	Decision variable	# of catalogs to mail to segments each season[c]	Mail catalog to customer i at time t	Mail catalog to customer i at time t	Use mailstream m for segment j.	Timing and # of catalogs – which segment receives these catalogs	Whether to promote to a given customer	Number of direct mail pieces and magazines during 1 year	Mail to customer in segment s.	# of panelists in segment j to invite for study s
	Customer growth	√	–	–	–	√	–	–	–	√
	Budget	√	–	–	√	√	–	–	–	[d]
	Forward looking	√	√ (Firm and customer)	–	√	√	√	–	√	√
	Optimal solution	Steady state	Steady state	Steady state	Steady state	Steady state	Steady state and finite horizon	Single-period calculus-derived	Steady state	Rolling horizon

Costs	Mailing	√	√	√	√	√	√	√	√	–
	Inventory	√	–	–	–	–	–	–	–	NA
	Out-of-stock	√	–	–	–	–	–	–	–	NA
	COGS	√	√	√	√	√	√	√	√	NA
	Order processing	√	–	–	–	–	–	–	–	–
Response model	Type	RFM	Dynamic rational	Hazard	Regression	Regression, R and RFM	Response = f(customer volume)	Hierarchical Bayes' Regression (Profit)	Decision tree	Decision tree
Response phenomena in model	Response rate	√	√	√	√	√	√		√	√
	Order size	√			√	√	√	–	√	
	Wear-out	Implicit	–	√	√	Implicit		Implicit	√	√
	Wear-in	Implicit	–	–	–	Implicit	–	Implicit	–	–
	Forgetting	Implicit	√	√	–	Implicit	–	Implicit	–	√
	Overlap	–	–	–	√	–	–	–	–	–

[a] Determine the schedule of invitations to an online survey panel to maximize average response rate.
[b] Extension of base model considers different catalogs and correlation between responses to different catalogs.
[c] Also decide how much to invest in inventory each season.
[d] No explicit monetary budget; budget takes form of required number of respondents per study.

interesting to compare the aggregate approach with an *ad hoc* scheduling rule to a true dynamic optimization.

Gönül and Shi (1998) are unique in allowing the customer also to be forward looking. Gönül and Shi are also unique in considering potential endogeneity of the mailing decision in the data they use to estimate the predictive model. That they do not find this to be an issue is re-assuring, but there is need to investigate this more fully (see Ansari et al. 2008).

Most of the models devise a steady state decision rule, as in "if the customer is in this RFM state at time t, mail a catalog to the consumer" (e.g., Simester et al. 2006). One exception is Neslin et al. (2007), who use a rolling schedule approach so there is no general decision rule.

Bitran and Mondschein (1996) are unique in their treatment of a broader class of decisions, particularly dealing with the operations side of the business. They consider "back door" inventory and ordering costs, which are vital to a catalog organization.

The methods also use a variety of response models, including RFM, hazard models, and decision trees. The RFM and decision tree models divide customers into segments; membership in these segments changes over time, allowing for dynamics in the optimization. Another important, practical aspect considered by Bitran and Mondschein (1996), Campbell et al. (2001), and Elsner et al. (2003, 2004) is to consider both whether the customer responds, and if so, how much does the customer spend. Finally, a key issue uncovered by Simester et al. (2006) is that it is important that the data used to estimate the response functions represent a broad range of mailing histories. The point is very important. Optimization models do not explicitly consider the quality of the data that drive them. Simester et al. point out that in practice, ample data variation is crucial.

A final note is that while the collection of models summarized in this chapter collectively show that optimal contact models can be constructed and implemented, more work is needed to demonstrate they improve over current practice or simpler models (e.g., myopic models as mentioned above) in actual field tests. Campbell et al. (2001) and Neslin et al. (2007) demonstrate successes in controlled field tests, Elsner et al. (2003, 2004) provide quasi-experimental evidence of success, and Simester et al. 2006) demonstrate mixed results. More field testing is needed. Given the complexity of customer-specific multi-campaign scheduling, it is crucial to understand where the simplifications can be made, and which issues need to be confronted head on without any simplification.