Marketing Spending and Customer Lifetime Value for Firms with Limited Capacity

Anton Ovchinnikov and Phillip E. Pfeifer
Darden Graduate School of Business, University of Virginia, 100 Darden Blvd., Charlottesville, VA 22903 USA
aovchinnikov@darden.virginia.edu, pfeiferp@darden.virginia.edu

The concept of customer lifetime value (CLV) is widely used by marketing practitioners and academics in making decisions about customer acquisition and retention spending. The traditional view of CLV, however, assumes that the firm has the unlimited capacity to serve all its acquired and retained customers. In this paper we consider a firm with limited capacity and determine the role that CLV plays in its acquisition and retention spending decisions.

We find that the optimal spending for the firm with limited capacity increases in capacity and decreases in the number of customers. This stands in contrast to the firm with unlimited capacity that spends the CLV-maximizing amount regardless of the number of other customers. Surprisingly, however, the firm with limited capacity may spend more than the CLV-maximizing amount, which happens when the time value of money is large. When the time value of money is small, the two spending levels are nearly identical when the firm is well under capacity, but as the firm grows and approaches its capacity limit, optimal acquisition and retention spending decrease to substantially lower, but still non-zero levels. We characterize optimal spending for the firm at capacity in closed form and discuss the role CLV plays in optimizing customer mix.

We also examine customer selection decisions. We again find that CLV alone does not determine those decisions. In particular, we show the firm with limited capacity may prefer a customer with a lower CLV.

Lastly, we consider capacity expansion. We show that the expansion decision is of the threshold type: intuitively, the firm should expand when it becomes large enough. Interestingly, however, the value of expansion is largest when the firm still has significant unutilized capacity. CLV again offers little guidance in determining this value – per unit value of expanded capacity is much larger than the CLV of the customer who will occupy it. We also observe an interesting behavior of spending during the timelag between the decision to expand and the period when the additional capacity becomes operational: in that timelag the firm drastically reduces spending on the low-value customers, while increasing the spending on the high-value ones in order to rebalance its customer mix in anticipation of additional capacity.

1. Introduction

The suggestion to view customer relationships as assets that generate cash flows over time dates back to Bursk (1966). Bursk’s proposal to use the “investment value” of a customer to guide marketing spending decisions is the key idea behind the concept now commonly referred to as customer lifetime value (CLV). Defined for our purposes as the present value of the future cash flows attributed to the customer relationship, CLV is a measure that is commonly used to guide both acquisition and retention marketing spending. Attention directed toward CLV shifts the focus of marketing from transactions to relationships and underlines modern customer relationship management (CRM).

A typical CRM-focused firm will choose its marketing spending so as to maximize separately the CLV of each customer. Such CLV-maximizing spending (defined in more detail in Section 4.1) is optimal when the firm has independent relationships with its customers; i.e., when acquiring (retaining) one customer does not affect the cash flows from any of the firm’s other current and future customers/prospects.

It is rather obvious, however, that CLV-maximizing spending is suboptimal whenever there exist dependencies among customer relationships. Demand-side dependencies, such as customer network effects, are well recognized, e.g., Pfeifer (1999), Hogan et al. (2003), and Kumar et al. (2007). In contrast, supply side dependencies, have received only limited attention in the CRM literature.
Berger and Bechwati (2001) discuss the supply side dependency caused by the limited promotional budget per person in every time period. Liu et al. (2007) discuss the case in which a limited salesforce capacity is available to a single customer over multiple time periods. In general, there are a variety of supply-side dependencies that make CLV-maximizing spending suboptimal, Pfeifer and Ovchinnikov (2010). The focus of this paper is on a very fundamental form of supply-side dependency – the case when the number of customers that the firm can serve is limited by some $\bar{N}$, to which we refer as the *capacity constraint*.

Capacity constraints are a reality for many CRM-focused businesses. Capacity limit can be physical if the firm simply cannot serve more than some $\bar{N}$ customers. Performing arts companies, apartment buildings, sports venues, some colleges, and the like face constraints of this kind. A capacity limit can also be strategic, when firms for which quality of service depends on the number of customers served because of queueing or “crowding” (restaurants, health-clubs, medical practices, delivery companies, etc.) strategically limit the number of customers they serve. For example Kowalczyk (2006) reports the case of a hospital that “shuts doors to new patients.” In either case, capacity constrained CRM-focused firms face a major disconnect: the set of policies that are commonly applied to capacity constrained firms, such as the revenue management tools employed by airlines, focus on short-term profitability of sales transactions typically at the expense of customer relationships, while the “typical” CRM policies are designed to maximize separately the value of each relationship, and thus ignore capacity constraints. The goal of our paper is to address this disconnect.

A capacity constraint is fundamentally different from other types of limited resources, in particular those mentioned above, along at least three dimensions. First, capacity constraint is often practically impossible to remove, e.g., the Green Bay Packers, a professional football team in the U.S., has 81,000 names on its wait list for season tickets with an average wait of 30 years, but no apparent plans to move to a different stadium\(^1\). Similarly, several months prior to the start of its 2010 season Opera Company C (name disguised for confidentiality) was sold out in 9 of 14 seating areas, and expansion was not an option either\(^2\). We refer to such cases as the *hard* capacity constraints.

Second, even when capacity can be expanded, doing so is typically discrete and discontinuous (“lumpy”): expansion involves a one-time and usually large fixed cost, which brings the capacity up by some usually large number of units, possibly with some time lag. We refer to such cases as the *soft* capacity constraints; restaurants, medical clinics, sports clubs, and the like face constraints of this kind. This discreteness contrasts drastically with sales force or marketing budgets: both can be increased effectively continuously and involve no fixed costs. For example, it is reasonable that a firm might increase its marketing budget by 3%, but if a doctor’s office moves (to expand), then a 3% capacity increase is inconceivable; a far more likely scenario is that they will add, say 30% capacity, even though initially some of this newly added capacity can be idle\(^3\). In addition, it is not clear that such continuous forms of limited resources as budget constraints should even be treated as constraints in the first place: e.g., if the cost of funds is properly accounted for in the cost of servicing the customer, then the problem reduces to unconstrained CLV maximization. That is

\(^1\)http://www.packers.com/fan_zone/faq/

\(^2\)That is, theoretically and eventually the company may move to a different building, but that decision depends on many factors (availability of physical space for construction - a major issue in a large metropolis, major (tens of millions of dollars) donations, similar size government support, artistic needs, etc.), which by far outweigh CRM in making the expansion decision. Thus, when C makes its CRM decisions, it treats the number of seats in the building as a constraint now and in the foreseeable future.

\(^3\)This adds another, less critical distinction: somewhat unfortunately budgets rarely get underspent, managers are afraid to end up with unused funds, thinking that this will jeopardize future budget allocations; temporarily under-utilized capacity, however, is rather typical.
not the case with capacity constraint: even if the variable costs of capacity is accounted for in the contribution margin, the fixed cost cannot be generally allocated to individual customers a priori because their number changes over time.

The third, and perhaps the most fundamental difference, is that in the above mentioned papers the firm was still maximizing individual’s CLV (with the complications introduced by the per-person budget or per-person salesforce limitations). A capacity constraint requires a fundamentally different framework: one in which the firm must realize that “if demand exceeds capacity at any time, then an opportunity [here, to sell the product/service] is lost”, Rust and Chung (2006). This creates a non-trivial cross-customer and inter-temporal endogeneity, to address which one should not just maximize CLV of every individual separately. Customers are no longer the limited resource – the capacity to serve them is, and thus the firm must operate such as to extract maximum value from its capacity, not necessarily its customers.

Building a model that accounts for the above endogeneities responds well to Rust and Chung’s suggestion that “we need a model with the following characteristics:” (1) personalized marketing interventions, (2) multiple interventions, (3) maximizes CLV4, (4) addresses endogeneity. They also point out that “a model that puts all of these elements together” is “missing.” Sun (2006) further develops Rust and Chung’s argument by suggesting that “[i]f a firm with the goal of maximizing long-term profit, CRM should be formulated as a stochastic dynamic control problem under demand uncertainty with the firm as the decision maker which makes dynamic marketing intervention decisions.”

Our paper provides a start at addressing all of the above mentioned concerns. The firm in our dynamic programming model maximizes its total net present value of the profits from sales to heterogenous customers over an infinite time horizon (following the typical dynamic programming terminology we refer to this value as the profit-to-go). In each period the firm observes the number of customers of different types, and decides on the personalized retention spending directed toward each customer during that period. Such spending reduces the current period’s profit, but increases the chance that the customer will be retained next period. In addition to this customer-specific retention spending the firm also decides on acquisition spending for the period; such spending (e.g., on advertising in broadcast media) generates a mixture of new (candidate, potential, etc.) customers arriving in random order next period. After accommodating the renewing customers, the firm accepts new customers on a first-come first-served basis. In our initial analysis the firm faces a hard capacity constraint: If the total number of customers exceeds some exogenously given \( \bar{N} \) then the firm bumps or denies service to the remaining newly acquired (candidate) customers5.

We then extend the analysis to the case with soft constraint: i.e., when the firm has an option to increase its capacity by \( E \) units at the total cost \( c_E \) with lag \( \Delta \geq 1 \).

To isolate the effect of the capacity limit, we consider a situation where the firm’s customer relationships would otherwise be independent were it not for the capacity constraint; that is, the behavior of a customer is only affected by the firm’s marketing spending dedicated towards that customer and not by any other factor (such as, for example, the number of other customers served or spending directed towards them). We then compare two versions of the same firm: one with

---

4 Note that one has to be careful and not interpret Rust and Chung’s “maximizes CLV” to mean maximizing separately each individual’s CLV; in a capacity constraint environment, maximizing one customer’s CLV could mean taking a unit of capacity away from another customer, i.e., decreasing that other customer’s CLV. The goal therefore is to maximize the total value of the firm; equivalently the sum of the CLVs of all customers served.

5 One could also consider a case where instead of bumping a customer the firm would incur a penalty if the number of customers exceeds capacity. Although we do not do so in this paper, it is of interest for future research. The two cases are equivalent if the penalty is high enough, e.g., exceeds the present value of contributions from a customer.
unlimited capacity and another with limited\textsuperscript{6} capacity. Our model is purposefully stylized so that for the firm with unlimited capacity the optimal spending is the one that separately maximizes each customer’s CLV; as mentioned earlier we refer to such spending as \textit{CLV-maximizing}. We consider a customer retention situation (where a customer that is not retained is considered lost for good) and in the case with unlimited capacity the resulting expressions for CLV are effectively identical to those used by other authors, e.g., Blattberg and Deighton (1996) and Gupta and Zeithaml (2006).

To determine the optimal spending for the case with limited capacity in Section 3 we formulate the problem as a dynamic program. In Section 4 we study the dynamics of spending as well as the limiting case of the firm at capacity and contrast them with the CLV-maximizing spending. In Section 5 we provide numerical illustrations. In Section 6 we discuss a customer selection problem and present additional numerical examples. Section 7 discusses capacity extension in the case with a soft constraint. Our conclusions are presented in Section 8.

1.1. Contributions of the paper
First and foremost our contribution is in proposing and solving the dynamic marketing interventions model for a CRM-focused firm with a significant endogeneity caused by the supply-side dependency. We focus on a specific form of supply side dependency – a constraint on the number of customers the firm can serve, but the dynamic framework we develop is quite general because it captures multiple personalized marketing interventions, maximizes the firm’s long-term value and addresses the endogeneity issues. As argued above, such models are of general interest, and in particular, the models with supply side dependencies have received limited attention so far. We obtain five main results:

\textbf{Dynamics of Spending.} We show that optimal acquisition and retention spending for the firm with limited capacity is increasing in the firm’s capacity and is decreasing in the number of customers. This stands in contrast to optimal spending for the firm with unlimited capacity in an identical situation. The firm with unlimited capacity spends the same CLV-maximizing amount for a given individual regardless of the number of customers or its spending on others. In both cases, the firm spends more to retain higher-value customers than lower-value ones, but in the case with limited capacity the amount spent changes over time depending on the current customer mix and on how close the firm is to the capacity limit. Note that the spending changes even though the customers’ behavior does not; the spending therefore changes \textit{because} of the capacity constraint.

\textbf{Magnitude of Spending Compared to the CLV-maximizing Spending.} Very surprisingly, the optimal retention spending of the firm with limited capacity could exceed the CLV-maximizing spending. This happens because for the firm with limited capacity balances its acquisition and retention spending across heterogenous customers in a fundamentally different way. The firm could find it profitable to overspend on high-value customers and decrease the lifetime value of each relationship in order to improve its customer mix by getting more high-value customers to occupy its scarce capacity and simultaneously save on the acquisition. This effect is observed when the time value of money is high (equivalently, when the discount factor is small); in those cases the degradation of CLV from overspending is relatively small (because of the high time value of money) and hence the improvement in customer mix and immediate acquisition savings play a bigger role\textsuperscript{7}. In the opposite case when the discount factor is high and the firm is well below capacity limit,

\textsuperscript{6}To avoid trivial cases, an implicit assumption in our paper is that the capacity limit is below the number of customers that the firm would grow to if it managed its acquisition and retention without a constraint on size. In that sense “limited capacity” means any capacity level below such unconstrained number of customers, and “unlimited” means the reverse.

\textsuperscript{7}Such situations can be found, for example, in the emerging markets, where the high cost of capital, political instability and vague property rights discourage firms from aggressively acquiring customers if it will take long time to reap the benefits from those customers.
numerically we see that the optimal spending is close to the CLV-maximizing amounts. That is, if the firm with a high discount factor and lots of unutilized capacity ignores, by mistake or on purpose, its capacity constraint and selects spending so as to maximize each customer’s individual CLV, then it’s spending is not far from optimal. That, however, is not true if the firm is close to capacity limit – in that case the optimal spending is much smaller than the CLV-maximizing amounts.

**Role of CLV in Optimizing Customer Mix.** At the same time, even when the firm is at capacity, it still spends on acquisition and retention. This finding can also be counterintuitive for managers. As one expert (who wished to remain anonymous) pointed to us in a written exchange: “clearly, if a firm is almost at capacity there is no reason for it to spend money on marketing in general, and acquiring or retaining customers in particular.” We note that such a point of view is rather naïve; we show the firm should balance it spending across customer types and time which is most cases implies non-zero spending. In a similar way, it is suboptimal to aim at having only the high-value customers. As we show, at optimality the spending across customer types must be balanced and although CLV is not the only factor that determines spending, it is useful in optimizing the customer mix.

**Role of CLV in Customer Selection.** We also consider customer selection and show that the firm with limited capacity may be better off by selecting a customer with a lower CLV, or even one with lower per period contribution. This happens because when a firm with limited capacity allocates a unit of capacity to a customer, it acquires his or her CLV, but incurs the opportunity cost from not being able to allocate the same unit to another customer (now or at any point within the duration of relationship with that customer). Equivalently, when losing a customer the firm loses his or her CLV, but gains the opportunity value of an available unit of capacity (oddly, losing a customer could be beneficial for the firm from the long-term value perspective). As we show, such interplay between the CLV of the customer and the opportunity cost could lead to the cases when the firm with limited capacity may be better off choosing a customer with a lower CLV. This finding is surprising and provocative, especially because it calls for looking at the customer selection problem in a new way.

**Capacity Expansion.** Lastly, we consider possible capacity expansion. We show that expansion decision is of the threshold type: intuitively, the firm should expand when it becomes large enough. Interestingly, however, the increase in the profit-to-go because of expansion (the value of expansion) is largest when the firm still has significant unutilized capacity. This happens because the larger the firm is, the more suboptimally it treats each individual customers (recall, that for a given capacity constraint the optimal spending decreases in the number of customers) and thus the more value from each customer relationship is “left on the table”. If the firm waits too long then this value is lost, thus it is optimal to expand earlier, in anticipation of future limitations. The timing of expansion is related to cost such that the higher the cost is, the later the firm expands. This is driven by the time value of money: expanding early implies that some capacity will not be utilized for longer time, so when the cost is high the firm balances the opportunity lost from suboptimal treatment of its current customers with the cost paid for the capacity that temporarily stands idle. Finally, we observe an interesting behavior of spending during the time window between the decision to expand and the period when the additional capacity becomes operational: the firm drastically reduces spending on the low-value customers, while increasing the spending on the high-value ones. This is driven by rebalancing customer mix: the firm with increased capacity has a different optimal customer mix (it is optimal to have more high-value customers when capacity is large) and hence the firm wants to keep more of its high-value customers and lose some of the low-value ones in preparation for capacity increase.
2. Literature Review

The marketing literature that estimates and uses CLV for various business decisions is very rich. The most common use of CLV is to guide marketing spending, e.g., Blattberg and Deighton (1996), but CLV has been applied to many other problems as well, e.g., Berger et al. (2006), Gupta and Lehman (2006), Wiesel et al. (2008). CLV is also touted as a concept to be used to shape the marketing strategy of some firms, e.g., Rust et al. (2000).

These papers, however, – and this is crucial – do not consider capacity constraints. A capacity constraint creates a dependency (endogeneity) among customer relationships, which naturally leads to the dynamics of spending. Indeed the goal of the firm that is well under capacity is to grow as quickly and profitably as possible, but the goal of the firm that is close to its capacity limit is to optimize the customer mix and stay at capacity as cheaply as possible.

The paper by Lewis (2005) is one of the few works that studies the dynamics of spending in the CLV context (in his paper the firm uses price discounts to manage retention). Lewis observes empirically that retention probabilities increase with customers’ tenure with the firm. He then numerically optimizes the discounts offered to retain customers and shows that the optimal discounts should be decreasing over time (i.e., the longer the customers stays with the firm, the smaller discount should he or she be offered). Although based on drastically different approach, we also show that retention spending decreases over time as the firm grows towards its capacity limit. The strength of our result is that it is analytical as opposed to Lewis’s numerical solution. Overall, however, both results offer strong support for the use of decreasing spending/discounts.

Related to that discussion is the work of Fader and Hardie (2007). In contrast with Lewis, who assumed/observed that customers became more loyal over time, they considered the case where individual consumers have different yet time-invariant retention functions, but that the firm only knows the distribution of consumer’s retention functions/probabilities. Then Fader and Hardie showed that the retention probabilities of the firm’s existing customers increase over time (intuitively, customers that were retained are more likely to have higher retention probabilities), even though individuals themselves are not becoming more loyal. We recognize the plausibility of Fader and Hardie’s assumption and use their framework in Section 6.2 to show that our results are robust with respect to the uncertainty in retention probabilities.

Lastly, there are several works (also mentioned in the introduction) that examine marketing spending for firms with limited resources. In Berger and Bechwati (2001) the constraint is on the total marketing spending per customer per period, whereas we consider firms limited in the number of customers served. Liu et al. (2007) consider the effects of limited sales force per customer in a general CLV framework. Roemer (2006) considers a seller with the option to switch from one customer to another should it become in the seller’s economic interest to do so. Implicit in Roemer’s model is the assumption that the seller is at capacity. Roemer, however, includes the value of the option to switch customers as a component of the CLV of the first customer. We agree with Roemer that decisions about acquiring the first customer should account for the option of reusing the unit of capacity with the second customer. But rather than including the reuse in CLV, we model it explicitly in a dynamic programming model presented next.

3. Model(s)

In this section we formulate two models. We first present a simplified model, which we use for discussion, to build intuition, and in the numerical analysis. Our theoretical results, however, hold for a much more general model, which we present second.

In our models the firm is observing how many customers of different types it has and decides on the personalized retention spending and total (non-personalized) acquisition spending. We acknowledge that in practice the set of tools is wider and includes customer growing activities, pricing, referrals, and the like. In attempt to keeping the model parsimonious and directly comparable to
the well-known CLV models we do not consider those activities in this paper, but doing so is of interest for future research.

We emphasize, however, that our model applies to retention situations where the customers that leave are considered lost for good; for example in the Opera Company C that we mentioned in the introduction, the likelihood that the last year’s season ticket holder will renew the subscription this year reaches 90%, while the likelihood that someone who did not renew this year will renew next year is only 4%. Our model does not directly apply to the capacity constraint industries where non-purchase is not a signal of the end of the relationship, e.g., airlines. Such firms focus on transactions and for them pricing is the main tool. For example, airlines would routinely vary prices by a factor of up to ten, something that a performing arts company would never do, because that would destroy the long-term relationship with the customer. We assume that price is given exogenously, which is consistent with the latter. We also assume that the firm is a monopoly. Thus our model is better suited for the cases when the firm offers a highly differentiated product and has no direct competitors; Opera Company is also a good example for that.

3.1. Simplified model

We assume that there are two types of customers: “high” (H) and “low” (L). In period $t = 1, 2, \ldots$ the firm serves $N^H_t$ H-type and $N^L_t$ L-type customers for a total of $N_t = N^H_t + N^L_t$ customers. Let $\rho_t$ be the fraction of H-types. Note that there is a one-to-one correspondence between the pair $(N^H_t, N^L_t)$ and the pair $(N_t, \rho_t)$. We will use either pair to denote the state in the dynamic program described below.

We assume that H- and L-type customers contribute\(^8\) $M^H$ and $M^L \leq M^H$ respectively, on average per period, and that in period $t$ spending $R^H_t$ on retention per H-type customer results in a retention rate of $r^H(R^H_t)$, and spending $R^L_t$ on an L-type customer results in retention rate $r^L(R^L_t)$. We assume that $r^H(\cdot)$ and $r^L(\cdot)$ are increasing concave functions and that $r^L(x) \leq r^H(x)$ and $\partial r^H/\partial x \geq \partial r^L/\partial x$ for all $x > 0$. These assumptions are consistent with the idea that higher spending customers are more loyal, and were also supported by our discussions with the management of company C mentioned in the introduction. Similarly, spending $A_t$ results in $n(A_t)$ new (candidate, potential) customers\(^9\) who arrive in period $t + 1$, a fraction $\xi$ of which are of H-type. We assume $n(\cdot)$ is an increasing concave function. Then the profit in period $t$ is:

$$g_t(N_t, \rho_t, R^H_t, R^L_t, A_t) = N_t \rho_t [M^H - R^H_t] + N_t (1 - \rho_t) [M^L - R^L_t] - A_t \quad (1)$$

$$\equiv N^H_t [M^H - R^H_t] + N^L_t [M^L - R^L_t] - A_t \quad (2)$$

Period $t+1$ demand is:

$$D_{t+1} = N_t \rho_t r^H(R^H_t) + N_t (1 - \rho_t) r^L(R^L_t) + n(A_t) \equiv N^H_t r^H(R^H_t) + N^L_t r^L(R^L_t) + n(A_t), \quad (3)$$

so that the number of customers served in period $t + 1$ is $N_{t+1} = \min[\tilde{N}, D_{t+1}]$, where $\tilde{N}$ is firm’s capacity. The fraction of H-types is therefore:

$$\rho_{t+1} = \frac{N_t \rho_t r^H(R^H_t) + \xi \min[n(A), \tilde{N} - (N_t \rho_t r^H(R^H_t) + N_t (1 - \rho_t) r^L(R^L_t))] \min[N, D_{t+1}]}{N_{t+1}} \equiv \frac{N^H_{t+1}}{N_{t+1}} \quad (4)$$

\(^8\) Contribution amount $M$ is the net of the price charged, cost of goods/services sold, cost of funds and all other direct expenses, excluding marketing costs.

\(^9\) Such an assumption reflects a practice of acquiring customers using broadcast media; it is also supported by the discussions with the management of C. C approaches its current and recent customers individually with personalized offers to renew and simultaneously advertises in mass media to attract new customers. Note also that we assume that a new cohort of prospects arrives each period. This is also consistent with the experience of C: because of the population migration each year there is a number of potential opera lovers who move into the city.
reflecting proportional bumping of new customers in the case of excess/overflow demand; i.e., that
the firm first accommodates its renewing customers and then allocates the remaining capacity on a
first-come first-served basis to newly acquired (candidate) customers who arrive in random order.

To simplify the discussion we presented deterministic time-invariant acquisition and retention
processes. All our structural results continue to hold when acquisition and retention processes
depend on time (i.e., when $M^H$, $M^L$, $r^H$, $r^L$ and $n$ are indexed with subscript $t$ to reflect, for
example, price changes, population growth, etc.) and/or are stochastically increasing and concave
random variables. For example, in our numerical analysis, Section 5, we consider binomial retention
with $\tilde{N}$, as before.

Let $\pi_t(N_t, \rho_t, R^H_t, R^L_t, A_t) = g_t(N_t, \rho_t, R^H_t, R^L_t, A_t) + \beta \pi_{t+1}^*(N_{t+1}, \rho_{t+1})$.

$\pi_{t+1}^*(N_{t+1}, \rho_{t+1}) = \max_{R^H_{t+1}, R^L_{t+1}, A_{t+1}} \pi_{t+1}(N_{t+1}, \rho_{t+1}, R^H_{t+1}, R^L_{t+1}, A_{t+1})$ is the optimal profit-to-go from state $(N_{t+1}, \rho_{t+1})$ onwards, $\beta \in [0, 1]$ is the firm’s discount factor, $N_{t+1} = \min[\tilde{N}, D_{t+1}]$ with $D_{t+1}$ given by (3) and $\rho_{t+1}$ is given by (4).

Note that the dynamic program (5) could be equivalently specified by defining states as $(N^H_t, N^L_t)$. That is, in fact, the construct we use next in the general model.

3.2. General model

Let $\tilde{N}_t$ be the vector of the number of customers of different types, which we refer to as the state.
Let $\tilde{S}_t$ be the vector of spending(s), which we refer to as the decision. In the simplified model, for
example, we have $\tilde{N}_t = \{N^H_t, N^L_t\}$ and $\tilde{S}_t = \{R^H_t, R^L_t, A_t\}$. But in general there can be any number
of types of customers; in the limit, each customer can be of his or her own type.

Let $g_t(\tilde{N}_t, \tilde{S}_t)$ be the expected single period profit from making decision $\tilde{S}_t$ in state $\tilde{N}_t$. Let
$F_{(\tilde{S}_t, \tilde{S}_t, \tilde{N})}(w)$ be the distribution function for state $w$ in period $t+1$ given state $\tilde{N}_t$ and decision $\tilde{S}_t
in period t and capacity $\tilde{N}$. Let $\beta$ be, as before, the discount factor. We have the following model:

$$\pi_t(\tilde{N}_t, \tilde{S}_t, \tilde{N}) = g_t(\tilde{N}_t, \tilde{S}_t) + \beta \int \pi_{t+1}^*(w, \tilde{N}) dF_{(\tilde{N}_t, \tilde{S}_t, \tilde{N})}(w)$$

where as before $\pi_{t+1}^*(w, \tilde{N}) = \max_{\tilde{S}_{t+1}} \pi_{t+1}(w, \tilde{S}_{t+1}, \tilde{N})$.

We assume that single-period profit $g_t$ is an increasing function in $\tilde{N}$ and $\tilde{N}_t$. Both these assumptions
are plausible: the more customers the firm has, and the more capacity it has to serve them, the
higher should its immediate profit be.

We assume that the distribution $F$ is stochastically increasing in $\tilde{N}$, $\tilde{N}_t$ and $\tilde{S}_t$. These
assumptions are intuitive: with more capacity, the more customers the firm serves today, and the better
it treats them, the more customers it should have tomorrow. We also assume that $F_{(\tilde{N}_t, \tilde{S}_t, \tilde{N})}(w)$ is
stochastically supermodular in $(\tilde{N}, \tilde{S}_t)$. This assumption is also plausible and reflects the following
logic. Since the capacity limit $\tilde{N}$ restricts the number of customers that can be served, when firm
size $|\tilde{N}_t|$ is close to capacity, additional spending will be effectively “wasted.” An increase in capacity
allows the firm to better accommodate newly arriving customers—hence supermodularity in capacity and spending. Next we use both models to characterize optimal spending.

4. Optimal Spending: Analytical Results

We first discuss the optimal spending for the firm with unlimited capacity. The results we provide
are well-known, but they are nevertheless useful as a benchmark for the optimal spending for the
firm with limited capacity. These results are also helpful in establishing that in the limiting case
with $\tilde{N} = \infty$ our setup is in fact identical to conventional constant customer retention CLV models.
4.1. Firm with Unlimited Capacity: CLV-maximizing Spending

For the firm with unlimited capacity (and otherwise independent customer relationships as we assumed in Section 3) maximizing its total value is equivalent to maximizing separately the CLVs of each individual customer. Following our notation, and suppressing (for the moment) type H- and L-indices, CLV is the expected net present value from the cash flow of $M - R_t$, which the firm enjoys until the customer leaves. Since we assume memoryless time-invariant retention process and because capacity is not limited and customer relationships are independent, the firm will maximize this net present value by spending the same amount in every period, i.e., the time index and because capacity is not limited and customer relationships are independent, the firm will maximize this net present value by spending the same amount in every period, i.e., the time index can be omitted as well. The number of periods the customer stays with the firm has a Geometric distribution with parameter $1 - r(R)$. Therefore:

$$
CLV(R) = \sum_{d=0}^{\infty} r(R)^d (1 - r(R)) \sum_{i=0}^{d} \beta^i (M - R) = (M - R) \left( 1 - r(R) + r(R)(1 - r(R))(1 + \beta) + r(R)^2(1 - r(R))(1 + \beta + \beta^2) + \ldots \right) = (M - R) \left( 1 + \beta r(R) + (\beta r(R))^2 + \ldots \right) = \frac{M - R}{1 - \beta r(R)},
$$

(7)

implying that $CLV^H(R^H) = \frac{M^H - R^H}{1 - \beta r(R)}$ and $CLV^L(R^L) = \frac{M^L - R^L}{1 - \beta r(R)}$. It is easy to verify that these expressions are identical to the expressions for CLV in Blattberg and Deighton (1996) and Gupta and Zeithaml (2006).

Subject to our very reasonable assumption that $r(R)$ is increasing and concave, $CLV(R)$ is quasi-concave and so the optimal CLV-maximizing retention spending $R^*_{CLV}$ satisfies:

$$
\frac{\partial CLV}{\partial R}|_{R = R^*_{CLV}} = \beta CLV(R^*_{CLV}),
$$

(8)

where indices H and L are omitted for notational convenience. This expression has a succinct managerial interpretation that CLV is “the limit on retention spending.” (Note that from the properties of the inverse functions, the inverse of the derivative in the left-hand side can be interpreted as the marginal cost of ensuring the desired retention rate).

Further, since by spending $A$ the firm acquires on average $\xi n(A)$ type-H customers with $CLV^H(R^*_CLV)$ and $(1 - \xi)n(A)$ type-L customers with $CLV^L(R^*_CLV)$, the firm will therefore select its acquisition spending so that to maximize $n(a)\left[\xi CLV^H(R^*_CLV) + (1 - \xi)CLV^L(R^*_CLV)\right] - A$. Subject to our very reasonable assumption that $n(A)$ is increasing and concave, the optimal CLV-maximizing acquisition spending $A^*_CLV$ therefore satisfies:

$$
\frac{\partial n}{\partial A}|_{A = A^*_CLV} = \xi CLV^H(R^*_CLV) + (1 - \xi)CLV^L(R^*_CLV),
$$

(9)

which, same as above, can be interpreted as “the limit on acquisition spending.”

The results in equations (8) and (9) and their interpretations are well-known. They provide simple managerially relevant guidance for selecting optimal spending, and are one of the reasons why the concept of CLV became popular with both marketing practitioners and academics.

A direct application of CLV-maximizing spending [equations (8) and (9)] to a firm with limited capacity, however, is obviously problematic, because the above and other existing models do not consider that at some point the firm will not be able to serve all its customers. We discuss the solution to the limited capacity case next.
4.2. Firm with Limited Capacity: Dynamics of Spending

Our main result in this Section is that optimal spending is increasing in the firm’s capacity and decreasing in the number of customers. We have the following Proposition (a proof of which is in the Appendix).

PROPOSITION 1. The optimal spending, \( \hat{S}_t^* \), for the general model discussed in Section 3.2 is increasing in the firm’s capacity, \( \bar{N} \).

Note that the result of the above proposition holds for every (given) vector of the number of customers, \( \tilde{N}_t \). Since an increase in capacity for a given number of customers is analogous to a decrease in the number of customers for a given capacity level, we therefore have the following Corollary.

COROLLARY 1. The optimal spending, \( \hat{S}_t^* \), is decreasing in the firm’s number of customers, \( \tilde{N}_t \).

The above results have an appealing managerial interpretation: The firm should decrease its spending as it approaches capacity and should increase its spending if more capacity becomes available. The direction of these results is quite intuitive, but the magnitude of the decrease is not. The advantage of our approach is that we presented a model that can determine the optimal spending. Numerically, in Section 5, we find that the decrease in spending can be very large, but at the same time, even when at capacity, the firm still spends actively on both retention and acquisition.

Driving the results in Proposition 1 and Corollary 1 is the supermodularity of the profit-to-go function in \((\bar{N}, \hat{S}_t)\). From a managerial perspective, supermodularity is very plausible because a firm with a potentially binding capacity constraint will not be able to fully reap the benefits from additional marketing spending. Aggressive spending in the face of potentially binding capacity constraint is likely to result in more customers showing up than the firm can serve. Since these excess/overflow customers are bumped, the profitability of spending is therefore less that it would otherwise be with a larger capacity limit. The more capacity the firm has, the fewer customers it will bump, and the more beneficial is any increase in spending; hence supermodularity.

We note that the firm faces the abovementioned effects of capacity even when customers (both current and new/candidate/potential prospects) are oblivious to the capacity constraint as we assume in this paper to isolate the effect of capacity. In practice, the supermodularity described above could be magnified by a decrease in the current customers’ responsiveness to the firm’s marketing spending as they become aware of how “crowded” the firm is and the associated decrease in quality of the firm’s service. Prospects might also anticipate an increased likelihood of being bumped and thus be less responsive to the firm’s acquisition efforts. Rephrasing the well-known Yogism, customer reaction to crowding could lead to a situation where “Nobody reads their ads anymore, it’s too crowded!” and supermodularity would be magnified.

We lastly return to the simplified model with two customer types, H and L. Specifically, note that in (1) - (5) the model is defined using the state space \((N_t, \rho_t)\). We have the following result:

PROPOSITION 2. Optimal spending in the simplified model (1) - (5), \( A_t^*, R_t^H \) and \( R_t^L \) increases in \( \bar{N} \) and decreases in \((N_t, \rho_t)\). Further, at each state \( R_t^H \geq R_t^L \).

That is, the effects of capacity and firm size in the simplified model with two customer types are the same as in the general model: the optimal spending in the simplified model is increasing in \( \bar{N} \) for any \((N_t, \rho_t)\), and is decreasing in \( N_t \) (correspondingly \( \rho_t \)) for any \( \bar{N} \) and \( \rho_t \) (correspondingly \( N_t \)). This flexibility is convenient in the discussion of the firm at capacity below and numerical illustrations in Section 5.
4.3. Firm (Optimally) at Capacity: Optimizing Customer Mix

We say that the firm is at capacity if it is in the state with \( N_t = \bar{N} \). Note, however, that depending on the customer mix, \( \rho_t \), the firm at capacity can still exhibit some dynamics of spending until it eventually reaches the state with the optimal customer mix \( \rho^* \) at which the firm’s profit-to-go is maximized globally over all states. Note that bumping is clearly not optimal at such globally optimal state \((\bar{N}, \rho^*)\) and further, that state maps onto itself, time indices can also be omitted. Therefore the firm’s problem can be rewritten as follows:

\[
\pi^*(\bar{N}, \rho^*) = \max_{(A, R^H, R^L)} \frac{1}{1 - \beta} \left[ \bar{N} \rho^* (M^H - R^H) + \bar{N} (1 - \rho^*) (M^L - R^L) - A \right] \tag{10}
\]

subject to

\[
\begin{align*}
\bar{N} &= \bar{N} \rho^* r^H(R^H) + \bar{N} (1 - \rho^*) r^L(R^L) + n(A) \\
\rho^* &= \frac{\bar{N} \rho^* r^H(R^H) + \bar{N} (1 - \rho^*) r^L(R^L) + \xi n(A)}{\bar{N}} \tag{11}
\end{align*}
\]

Observe that while discount factor \( \beta \) clearly influences the magnitude of the firm’s profit-to-go, it does not influence the optimal policy. The firm that is optimally at capacity chooses its spending so as to remain optimally at capacity, and thus its problem is equivalent to that of maximizing per period profit while staying (indeed indefinitely) optimally at capacity.

Using the Lagrangian method, and after some algebraic simplifications, it is possible to establish that the optimal retention spending, \( R^H \) and \( R^L \), and acquisition spending, \( A^\ast \), in addition to the constraints (11) satisfy the following conditions:

\[
(M^H - R^H) - (M^L - R^L) = \frac{1 - r^H(R^H)}{\partial r^H / \partial R^H} - \frac{1 - r^L(R^L)}{\partial r^L / \partial R^L} \tag{12}
\]

and

\[
\frac{\partial n}{\partial A^\ast} = \frac{1 - r^H(R^H)}{\xi ((M^H - R^H) - (M^L - R^L)) + \frac{1 + (1 - \xi) r^H(R^H) + \xi r^L(R^L)}{\partial r^L / \partial R^L}_{|R^L}} \tag{13}
\]

While equation (13) is hard to interpret (the optimal acquisition spending seems to have a complex relationship with other model parameters), equation (12) has an appealing managerial interpretation. It implies that for a firm that is optimally at capacity, retention spending on marginal H- and L-type customers should be balanced such that the additional profit the firm receives from an H-type customer beyond what it receives from an L-type equals the additional cost of retaining that customer. Intuitively, equation (12) holds because if for a given customer mix, \( \rho \), the value the firm gets from a marginal H-type customer is, for example, larger than the cost, then such customer mix cannot be optimal: the firm could obtain additional profit by spending more on Hs and less on Ls, as a result increasing \( \rho \).

Equation (12) has another, arguably less intuitive, interpretation through its connection to CLV. By rearranging the terms, (12) is equivalent to:

\[
[1 - r^H(R^H)] \left( \frac{M^H - R^H}{1 - r^H(R^H)} - \frac{1}{\partial r^H / \partial R^H} \right) = [1 - r^L(R^L)] \left( \frac{M^L - R^L}{1 - r^L(R^L)} - \frac{1}{\partial r^L / \partial R^L} \right) \tag{14}
\]

Recall from (7) that \( \frac{M - R}{1 - r^H} \) equals the customer’s CLV but with \( \beta = 1 \). We refer to it as the non-discounted CLV and denote it by \( CLV|_{\beta=1} \). Consider the differences in the round braces on each side of the equation, and observe that they are in the form of \( CLV|_{\beta=1} - \frac{1}{\partial r^H / \partial R^H} \). Suppose the firm indeed had \( \beta = 1 \). Then from (8) if it had unlimited capacity, its CLV-maximizing retention spending would be such that \( CLV|_{\beta=1} - \frac{1}{\partial r^H / \partial R^H} = 0 \), i.e., to set these differences equal zero. Equation (14) would then obviously hold. From Proposition 2, however, the optimal spending for the firm
with limited capacity is smaller than CLV-maximizing (here with $\beta = 1$); in fact, because functions $r(\cdot)$ are increasing and concave the differences are positive. These positive differences represent the degree of deviation from the optimality in CLVs due to limited capacity – the opportunity loss from not being able to retain the firm’s customers so as to maximize each customer’s CLV. Condition (14) therefore implies that:

**Proposition 3.** In order to optimize the customer mix the firm at capacity should select its retention spending so as to balance the deviations from optimal non-discounted CLVs across customers of different types, weighted by the corresponding churn probabilities.

More importantly, however, $CLV|_{\beta=1}$ is above the customer’s actual CLV with the firm’s actual $\beta < 1$, which we denote as $CLV|_{\beta<1}$, and as a result the CLV-maximizing spending with $\beta = 1$ is also larger than that with $\beta < 1$. Therefore if $\beta$ is small enough then even though $CLV|_{\beta=1} - \frac{1}{\partial r/\partial R|_{R=AtCapacity}} > 0$ it could be that $CLV|_{\beta<1} - \frac{1}{\partial r/\partial R|_{R=AtCapacity}} \leq 0$, in which case by Proposition 2 and because functions $r(\cdot)$ are increasing and concave:

**Proposition 4.** The optimal retention spending for the firm with limited capacity could exceed the CLV-maximizing spending.

The reason for this rather provocative result is in the fundamentally different relationship between the acquisition and retention spending for firms with limited and unlimited capacity. A firm with unlimited capacity optimizes its spending using backwards induction: assuming that a customer is acquired it optimizes retention so as to maximize the customer’s CLV and then uses this optimal CLV in order to optimize acquisition spending. As a result retention spending influences acquisition but not the other way around. For a firm with limited capacity acquisition and retention spending are related in both directions. Specifically, the firm with limited capacity has the ability to decrease acquisition spending and use a fraction of those savings to increase retention spending on high-value customers. By doing so the firm would lose some value over the lifetime of each high-value customer, but it will improve the customer mix and hence gain short-term value from having a higher-value customer occupying a unit of its scarce capacity effectively next period. When $\beta$ is small (equivalently, the time value of money is large), the benefit from having more high-value customers sooner plus the instantaneous acquisition savings outweigh the long-term losses on each high-value customer (because the present value of losses is small when $\beta$ is small). Therefore the firm could find such overspending profitable.

This discussion also suggests that the high-value customers could add value to the firm with limited capacity beyond what is captured in their individual CLVs even thought the behavior of customers is unchanged. Spending over their CLV-maximizing amount allows the firm to improve overall profitability by simultaneously reducing spending on low-value customers and cutting acquisition spending, thus improving the customer mix and allocating a larger portion of its scarce capacity to high-value customers. Next we illustrate this and other results with numerical examples.

### 5. Optimal Spending: Numerical Analysis

In this section we provide illustrations of the phenomena discussed earlier. We use the following stylized example, say of a new performing arts company. For ease of comparisons with existing CLV models we patterned our example after the model presented in Blattberg and Deighton (1996). Suppose that currently the firm spends $2,500 a year (here year$ symbolizes period) to acquire 15 new (candidate, potential) customers. Each customer contributes $1,000 with the firm and occupies one unit of capacity per year. With probability 50% a new customer is of type H, and otherwise she is of type L. The firm currently spends $150 on retention per an H-type customer, which results in an 80% retention probability for an H-type. Similarly, spending of $150 on an L-type customer results in a 60% retention probability. The firm uses a $\beta = 90\%$ discount factor.
new customers (50% of which are L-types). These observations illustrate that the firm with a high point in the graph with less than 3 periods). Second, even when completely full with only H-type customers (the frontmost than 50 customers per period, thus even with unbounded spending it cannot reach its capacity in the same the firm with unlimited capacity (note that by construction the firm cannot acquire more

important. First, when the firm has no customers at all, the firm with $N \overline{=}$ 100 and $N = 0$ use a discrete state space

methods. The results are given in the two rightmost columns of Table 1.

Given the model assumptions, the firm should increase acquisition spending about ten times, to $20,925$. The resulting 47 acquired customers a year is the optimal number because at this point

worth on average 0.6784. With those parameters, it is easy to verify that if the firm continues what it is currently doing, it will converge to having 38 H- and 18 L-type customers (for the total $N = 56$, and $\rho \approx 67\%$) which from (7) have CLVs $3,035$ and $1,847$ respectively.

If the firm wants to optimize its value and has no capacity limitations, then from (8) its optimal retention spending should be $252$ and $177$ for H- and L-types, respectively, leading to the optimal CLVs of $3,215$ and $1,854$ – a minor improvement. However, a major change comes with respect to the acquisition spending. In particular, spending $2,500/15 = $167 to get something that is worth on average $0.5 \times 3,036 + 0.5 \times 1,848 = $2,442 is clearly what the firm wants to do more of. Given the model assumptions, the firm should increase acquisition spending about ten times, to $20,925$. The resulting 47 acquired customers a year is the optimal number because at this point the incremental acquisition cost equals the expected CLV of the acquired customer. By doing so, however, the firm will eventually grow to 222 customers; see Table 1.

To illustrate the impact of limited capacity we consider two “versions” of the original firm: one with $\bar{N} = 100$ and another with $\bar{N} = 150$.

For the cases at capacity we solve the non-linear program as per Section 4.3 using gradient methods. The results are given in the two rightmost columns of Table 1.

For the dynamics of spending we assume binomial acquisition and retention processes and solve the resulting stochastic dynamic program as per Section 3.1 using value iteration. To do so we use a discrete state space $N^H_t, N^L_t \in \{0, 1, 2, \ldots, \bar{N}\}$ and controls $R^H \in \{0, 25, 50, \ldots, 350\}$, $R^L \in \{0, 25, 50, \ldots, 225\}$, and $A \in \{0, 2500, 5000, \ldots, 25000\}$.

Figure 1 presents a 3D plot of the optimal acquisition spending for the $\bar{N} = 100$ case. The general decreasing pattern of spending is evident – the optimal spending is decreasing in both $N^H$ and $N^L$ from $20,000$ down to $10,000$ in steps dictated by our discretization. Two observations are important. First, when the firm has no customers at all, the firm with $\bar{N} = 100$ spends almost the same the firm with unlimited capacity (note that by construction the firm cannot acquire more than 50 customers per period, thus even with unbounded spending it cannot reach its capacity in less than 3 periods). Second, even when completely full with only H-type customers (the frontmost point in the graph with $N^H = 100$ and $N^L = 0$) the firm still spends very significantly on acquiring new customers (50% of which are L-types). These observations illustrate that the firm with a high

<table>
<thead>
<tr>
<th>$\bar{N} = 100$</th>
<th>$\bar{N} = 150$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$A$</td>
<td>$\bar{N} = 100$</td>
</tr>
<tr>
<td>$R^H$</td>
<td>$20,925$</td>
</tr>
<tr>
<td>$R^L$</td>
<td>$70.06$</td>
</tr>
<tr>
<td>$n(A)$</td>
<td>$15.00$</td>
</tr>
<tr>
<td>$r^H()$</td>
<td>$0.800$</td>
</tr>
<tr>
<td>$r^L()$</td>
<td>$0.600$</td>
</tr>
<tr>
<td>$CLV_H$</td>
<td>$3,036$</td>
</tr>
<tr>
<td>$CLV_L$</td>
<td>$1,848$</td>
</tr>
<tr>
<td>$\rho$</td>
<td>$0.6667$</td>
</tr>
<tr>
<td>Number of customers</td>
<td>56.25</td>
</tr>
</tbody>
</table>

Table 1 Comparison between the firm at capacity and uncapacitated firm.

The firm’s marketing research indicates that the retention probability for H-types can be boosted to at most 90%, but will not fall below 60% even with no retention spending. The corresponding estimates for L-types are 70% and 40%. With respect to acquisition, research shows that at least 1, but no more than 50 new customers can be acquired per year. Finally, research shows that in all three cases the relationship between spending and the outcome can be described by an exponential function. This leads to the following estimates: $n(A) = 1 + 49 \times [1 - \exp(-0.00013459A)]$, $r^H(R^H) = 0.6 + 0.3[1 - \exp(-0.007324R^H)]$ and $r^L(R^L) = 0.4 + 0.3[1 - \exp(-0.007324R^L)]$.

With those parameters, it is easy to verify that if the firm continues what it is currently doing, it will converge to having 38 H- and 18 L-type customers (for the total $N = 56$, and $\rho \approx 67\%$) which from (7) have CLVs $3,035$ and $1,847$ respectively.
discount factor that initially ignores the capacity limit is spending nearly optimal. But as the firm grows, it should decrease its spending very significantly. What is important, though, is that even at capacity the firm still actively spends. This contrasts with the logic of some marketers, as per our discussion in the introduction, who may intuitively guess to not spend at all.

Similar 3D plots could be presented for the optimal retention spending and for the $\bar{N} = 150$ case, but they are visually similar. To gain additional insight, we present several bisections of those plots. Figures 2 (a) and (b) present bisections over the states with $N^H = N^L$ as a function of $N = N^H + N^L$ for the cases with $\bar{N} = 100$ and $\bar{N} = 150$, respectively (solid lines), and the corresponding CLVs for H- and L-type customers for the cases with unlimited capacity (upper dashed lines) and at capacity (lower dashed lines) from Table 1. The scales on those plots are maintained, thus a horizontal comparison between (a) and (b) is valid. Same as with the acquisition spending, the decreasing pattern is evident, and so are the limiting cases. In addition, we observe that, as suggested by our theoretical results, the firm with more capacity spends more on retention, particularly when it has few customers. Note also, that the firm with more capacity spends more even for the comparable amount of empty capacity (e.g., the $\bar{N} = 150$ firm’s spending at $N_t = 50$ is larger that the $\bar{N} = 100$ firm’s spending at $N_t = 0$). At the same time, note that when the firm is well under capacity it spends almost at the CLV-maximizing levels.

Figure 2 (c) considers the case with $\bar{N} = 100$ and $\beta = 0.5$ in order to illustrate that the optimal retention spending for the firm with limited capacity may exceed CLV-maximizing spending when discount factor is small. In that case, it is easy to verify that the CLV-maximizing retention spending is 63.92 and 44.40 for H- and L-types respectively. We therefore immediately note that from Table 1 optimal retention spending for H-types for the case when firm is optimally at capacity is 70.06 which is above CLV-maximizing. But more so, from Proposition 2 the optimal spending is decreasing in firm size, therefore spending is even higher when the firm is below capacity; which is illustrated on Figure (c). Note that the firm in (c) is not optimally at capacity, thus the spending drops below 70.06; we discuss optimality next.

Finally, note that even when the firm reaches capacity in the cases depicted on Figure 2 (a) – (c), the firm has $\rho$ of 50%, while from Table 1 we know that for $\bar{N} = 100$ and $\bar{N} = 150$ the optimal $\rho$ is 67% and 70%, respectively for (a/c) and (b). Thus even with $N_t = \bar{N}$ the firm is still transitioning to the optimal state. To do so it must increase the number of H-type customers. But since $\xi = 0.5$, half of the newly acquired customers are L-types. The only way to increase the H-types count is to spend more so as to encourage more of them stay with the firm. That however does not mean abandoning spending on L-types. Even thought the firm is attempting to increase the proportion of H-types in the mix, it still spends on L-types. This happens because the first few dollars of
retention spending are very effective at increasing retention – so effective that it is in the firm’s best interest to take advantage. As a result the firm spends more on H-types that also bring more value, and spends less on L-types that are less valuable so as to balance the marginal profitabilities as suggested earlier.

To further illustrate this balance, Figure 2 (d) presents a bisection over states with $N^H + N^L = \bar{N}$ as a function of $\rho = N^H / \bar{N}$ for the $\bar{N} = 150$ case. We see that if the firm finds itself with too many H-type customers, then it spends much less, “milking” the customers in order to improve short-term profitability and lose some H-type customers and restore the balance (note that the optimal
retention spending on H-types is less than $147 (the “at capacity” optimum) when $\rho > 0.67$, and for L-types the spending is lower than the “at capacity” optimum of $64$ for $\rho > 0.76$). That, again, contradicts the typical managerial intuition, which suggests that the firm should strive to have only “good” customers. Rather, as we show, the firm should strive to have the balance of customers of different types. That is driven by both the need to balance retention spending across customers, and by the constant proportion of H- and L-type customers among new acquisitions. The latter, arguably, can be improved if the firm can select among potential candidates.

6. Customer Selection for Firms at Capacity

Suppose that the firm has a single unit of capacity available for the coming period (e.g., C’s last seat in a certain seating area), and must decide whether to accept or reject a candidate customer. We call the candidate customer $i$, and assume he or she is characterized by $M_i$, $R_i$ and $r_i$ which combine according to (7) to CLV$_i$. For ease of exposition we assume $R_i$ is known in advance and will be spent throughout the life of customer $i$. We initially assume that the retention probability equals $r_i$, and in Section 6.2 allow the retention probability to be Beta-distributed with mean $r_i$ as per Fader and Hardie (2007). If the firm rejects customer $i$, then it faces business as usual (i.e., the mix of H- and L-type customers discussed earlier). Our main question is: Should the firm accept $i$ if his or her CLV is larger than that of the firm’s usual customer?

If the firm selects $i$ it receives an immediate value $M_i - R_i$. With probability $r_i$ it finds itself in exactly the same spot next period, with customer $i$ occupying one of its $N$ units of capacity. With probability $1 - r_i$ customer $i$ does not return which frees up a unit of capacity for the firm. Let $\tau$ be the value of the available unit of capacity$^{10}$. Then the value gained by accepting $i$ is:

$$\pi_i = (M_i - R_i) + \beta r_i \pi_i + \beta (1 - r_i) \tau = CLV_i + \frac{\beta (1 - r_i)}{1 - \beta r_i} \tau.$$  \hspace{1cm} (15)

This equation has the following managerial interpretation. By accepting customer $i$, the firm gets the CLV for customer $i$ together with an open unit of capacity at some point in the future. The value of that open unit is $\tau$. The multiplier of $\tau$ accounts for the delay in the freeing up of the unit of capacity and the resulting decrease in its present value; it captures the opportunity cost of giving the unit of capacity to customer $i$. The higher the retention rate, the smaller is the multiplier because the unit gets freed up later, and so the opportunity cost is higher.

Note that CLV$_i$ depends on $M_i$, while the delay multiplier and $\tau$ are independent of $M_i$. Therefore neither focusing on the immediate profitability, nor on the CLV is optimal. What also matters is the opportunity cost of how long the customer can be expected to occupy a unit of capacity (this cost is measured by the delay multiplier). Therefore comparing two customers $i$ and $j$ it is possible to have situations where CLV$_i > CLV_j$ but $\pi_i < \pi_j$ if the multiplier for $j$ is sufficiently larger than for $i$. In other words,

**Proposition 5.** The firm at capacity could prefer a customer with a lower CLV. Further, it could prefer a customer that contributes less per period.

It is interesting to note that for customers with (nearly) equal CLVs, the firm prefers the one with the lowest retention rate, because such a customer leaves the firm earlier and hence incurs a smaller opportunity cost (has a higher multiplier for $\tau$). In addition, for equal CLVs, a lower retention rate implies higher spending, and so the firm naturally prefers a higher-spending fickle customer to low-spending loyal customer. For non-equal CLVs, though, a customer with higher immediate profitability is not necessarily the preferred one. An approach that focuses only on per-period

$^{10}$Note that $\tau$ is a constant, which can be easily computed, for example, as a difference between the optimal value of the firm with capacities $N$ and $N - 1$. 

profitability ignores the opportunity cost and the associated with it cost of future retention and acquisitions. For example, if a customer with lower $M_i$ is very loyal, then the firm needs to spend less on acquisition in the long run, and in expectation these savings over time could compensate for a lower per period profitability. We illustrate both these points in the example below.

6.1. Example of Selecting Customers

Continuing with the example from Section 5, suppose that the firm with $\bar{N} = 100$ finds itself with one empty unit of capacity and three candidates walk in the door vying to become customers. For reasons that will become apparent, we refer to the three candidates as the hare, the tortoise, and the usual (here, H-type). The important characteristics of the three candidates are displayed in Table 2. For simplicity, in this section we assume that the firm spends the same amount on retention ($70.06 as is optimal for the H-type) for all three candidates – the question is which customer to select.

Table 2 presents CLVs and $\pi$ values for the three candidates. Here $\tau$ is computed as per Section 4.3 as a difference between the optimal value of the firm with capacities $\bar{N} = 100$ and $\bar{N} = 99$.

The most important observation is that although the CLV of the hare is lowest, her $\pi$ is the highest. Thus the hare is the preferred candidate. The hare has the highest $\pi$ because her multiplier, see (15), more than makes up for her lower CLV: the firm does not get as much value directly from the hare as from the other two candidates, but what works to the hare’s advantage is the fact that the hare will use up a valuable unit of capacity for fewer periods, and hence by selecting the hare the firm incurs a smaller opportunity cost. “Live fast and die young” is the kind of customer the firm at capacity should prefer, even if that customer has a low CLV.

The above discussion, however, does not imply that retention no longer matters and that the firm should select customers based on immediate profitability. For example, if the firm must choose between the tortoise and the usual, then the firm should choose tortoise: the tortoise has a higher $\pi$ despite the fact that the firm makes only $780 = $850-$70 per period from the tortoise compared with $930 from the usual customer. This is because the firm spends less on acquisition in the long run if it chooses the tortoise. Choosing the usual, the firm needs to acquire a new customer (for its $N^{th}$ unit of capacity) approximately every five periods, but choosing a tortoise, only every hundred. This gets accounted for in $\pi$ but is ignored if one focuses only on immediate profitability.

6.2. Uncertainty in Customer Selection

The above analysis assumed that the probability of retaining a given customer is known. In practice, this probability could be uncertain. In this section, similar to Fader and Hardie (2007), we assume that individual (uncertain) retention probabilities follow a Beta distribution. We study whether the qualitative conclusions we obtained earlier continue to hold – in particular, whether the firm at capacity may still prefer a customer with a lower CLV.

To be consistent with our customer selection example we consider a fixed $R$. We assume that the retention probability, $r_i$, for customer $i$ follows a Beta($a_i, b_i$) distribution. We further assume the same level of uncertainty across customers, i.e., that $a_i + b_i = c$ for some $c$ and all $i$. The value of $c$ could be interpreted as the number of observations $a_i$ of which resulted in a successful retention, i.e., the larger is $c$ the smaller is the uncertainty in $r_i$. Note $E[r_i] = a_i/c$.

From Fader and Hardie (2007), assuming that $r_i \sim Beta(a_i, b_i)$ implies that the expected value of the retention probability of the customer $i$ acquired $d$ periods ago and retained for $d-1$ periods
is \( r_i^d = \frac{a_i + d - 1}{a_i + b_i + d - 1} \). Note that Fader and Hardie (2007) assumed that the churn rate (one minus the retention rate) is Beta distributed; compared with their expression for \( r_i^d \) we therefore switch the parameters in the equation. Using this equation, it is easy to obtain the expected CLV and value of unit \( \pi \) as a discounted sum. We implement this computation in Excel.

Table 3 presents the expected CLVs and \( \pi \)s for the random hare and usual customers as a function of the uncertainty in the retention probability. Several observations can be made from examining the table. As the variance goes to zero (i.e., \( c \to \infty \)), CLV and \( \pi \) estimates converge to those in Table 2. As the variance grows (i.e., \( c \) decreases), the value of both CLV and \( \pi \) could be significantly different from those with no variance. This is because uncertainty has an asymmetric effect: the average of CLVs at \( r = 1 \) and \( r = 0 \) is much larger than the CLV with the “average” \( r = 0.5 \).

The main observation, however, is that the value of selecting the hare is always larger than the value of selecting the usual, i.e., \( \pi_{\text{hare}} > \pi_{\text{usual}} \), while the CLV of the hare is always smaller. Therefore, the qualitative conclusion that the firm at capacity may be better off by selecting a customer with a lower CLV proves to be robust, at least in this example, with respect to (non-memoryless) uncertainty and heterogeneity in retention.

And even beyond this example, such robustness should not be that surprising. It is based on realizing that there is a value in the unit of capacity after the customer leaves. The value of the unit today (\( \pi \)) is the value we get from the customer to whom we give this unit to (CLV) plus the unit’s future value (\( \tau \)), appropriately discounted for the number of periods the customer will occupy it. Since the CLV and length of customer relationship depend differently on retention spending and customer’s per-period contribution with the firm, it is therefore reasonable that the same values of CLV can produce different values of \( \pi \), effectively regardless of how exactly CLV is computed. The above example illustrates this phenomenon – the values of CLV are different due to uncertainty in retention probability, but the message remains the same: the firm at capacity may prefer a customer with a lower CLV.

### Table 3 Effects of uncertainty in retention rate on the CLVs and \( \pi \)s for a random hare and usual (H-type) customers.

<table>
<thead>
<tr>
<th>c=a+b</th>
<th>a</th>
<th>b</th>
<th>( \sigma^2 )</th>
<th>CLV</th>
<th>( \pi )</th>
<th>a</th>
<th>b</th>
<th>( \sigma^2 )</th>
<th>CLV</th>
<th>( \pi )</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.25</td>
<td>0.75</td>
<td>0.0937</td>
<td>3,423.02</td>
<td>8,744.94</td>
<td>0.720</td>
<td>0.279</td>
<td>0.100</td>
<td>4,206.50</td>
<td>7,279.29</td>
</tr>
<tr>
<td>2</td>
<td>0.5</td>
<td>1.5</td>
<td>0.0625</td>
<td>2,932.50</td>
<td>8,421.93</td>
<td>1.440</td>
<td>0.559</td>
<td>0.067</td>
<td>4,034.28</td>
<td>7,636.74</td>
</tr>
<tr>
<td>4</td>
<td>1</td>
<td>3</td>
<td>0.0375</td>
<td>2,684.64</td>
<td>8,257.22</td>
<td>2.881</td>
<td>1.118</td>
<td>0.040</td>
<td>3,496.76</td>
<td>7,534.20</td>
</tr>
<tr>
<td>16</td>
<td>4</td>
<td>12</td>
<td>0.0110</td>
<td>2,531.91</td>
<td>8,155.71</td>
<td>11.526</td>
<td>4.473</td>
<td>0.011</td>
<td>2,868.64</td>
<td>7,344.87</td>
</tr>
<tr>
<td>256</td>
<td>64</td>
<td>192</td>
<td>0.00073</td>
<td>2,492.71</td>
<td>8,129.65</td>
<td>184.422</td>
<td>71.577</td>
<td>0.00078</td>
<td>2,658.24</td>
<td>7,280.93</td>
</tr>
<tr>
<td>65,536</td>
<td>16,384</td>
<td>49,152</td>
<td>2.86E-06</td>
<td>2,490.25</td>
<td>8,128.02</td>
<td>47.212</td>
<td>18.323</td>
<td>3.07E-06</td>
<td>2,644.63</td>
<td>7,276.79</td>
</tr>
<tr>
<td>256</td>
<td>64</td>
<td>192</td>
<td>0.00073</td>
<td>2,492.71</td>
<td>8,129.65</td>
<td>184.422</td>
<td>71.577</td>
<td>0.00078</td>
<td>2,658.24</td>
<td>7,280.93</td>
</tr>
<tr>
<td>65,536</td>
<td>16,384</td>
<td>49,152</td>
<td>2.86E-06</td>
<td>2,490.25</td>
<td>8,128.02</td>
<td>47.212</td>
<td>18.323</td>
<td>3.07E-06</td>
<td>2,644.63</td>
<td>7,276.79</td>
</tr>
</tbody>
</table>

The main observation, however, is that the value of selecting the hare is always larger than the value of selecting the usual, i.e., \( \pi_{\text{hare}} > \pi_{\text{usual}} \), while the CLV of the hare is always smaller. Therefore, the qualitative conclusion that the firm at capacity may be better off by selecting a customer with a lower CLV proves to be robust, at least in this example, with respect to (non-memoryless) uncertainty and heterogeneity in retention.

And even beyond this example, such robustness should not be that surprising. It is based on realizing that there is a value in the unit of capacity after the customer leaves. The value of the unit today (\( \pi \)) is the value we get from the customer to whom we give this unit to (CLV) plus the unit’s future value (\( \tau \)), appropriately discounted for the number of periods the customer will occupy it. Since the CLV and length of customer relationship depend differently on retention spending and customer’s per-period contribution with the firm, it is therefore reasonable that the same values of CLV can produce different values of \( \pi \), effectively regardless of how exactly CLV is computed. The above example illustrates this phenomenon – the values of CLV are different due to uncertainty in retention probability, but the message remains the same: the firm at capacity may prefer a customer with a lower CLV.

### 7. Capacity Expansion

Up until this point we have considered the case when the firm operates with a “hard” capacity constraint, \( \bar{N} \), that limits the number of customers it can serve now or at any point in the future. In some situations, however, firms may have the ability to expand capacity (now or at some point in the future) should it become profitable to do so. In such cases we say that the firm operates with a “soft” capacity constraint. When should the firm expand its capacity? What is the value of capacity expansion, and how does this value depend on the time and cost of expansion? And lastly, if expansion involves a time lag, how should the firm’s marketing spending change during that time? These are the questions we address in this Section.

First, we (re)formulate our model to account for the possibility of capacity expansion. As we argued in the introduction, a distinct characteristic of capacity expansion as opposed to other
forms of limited resources (such as a budget constraint), is that expansion is “lumpy” (discrete and discontinuous): it involves some upfront one-time fixed (installation, acquisition, etc.) cost, $c_E$, and adds some rather significant extra capacity $E$ possibly with some time lag $\Delta \geq 1$. Such expansions are also typically “one-off” opportunities; e.g., it is hard to imagine that the doctor’s office will move every few years to expand capacity by, say 10 units; a far more realistic scenario is that it will move once and expand by, say 50, even if some of this newly added capacity will not be immediately utilized. Thus we consider only a single possible expansion, which requires the following modification to our model.

Observe that after the capacity is expanded, the firm’s profit-to-go is given by the function
\[ \pi_t(\tilde{N}_t, \tilde{S}_t, \tilde{N} + E) \] as defined in (6), i.e., the function is not changed, just the capacity parameter changes. Before capacity is expanded, however, the firm’s profit-to-go function will be different because it must incorporate the option of future expansion.

Let $\pi^E_t(\tilde{N}_t, \tilde{S}_t, \tilde{N})$ denote the firm’s profit-to-go in the case with an option to expand, and let $t^E$ denote the time period when the firm expands (a decision variable, in addition to the decision vector for spending, $\tilde{S}_t$). Then by the above logic:

\[
\pi^E_t(\tilde{N}_t, \tilde{S}_t, \tilde{N}) = \pi_t(\tilde{N}_t, \tilde{S}_t) - c_E \times 1_{t=t^E} + \beta \left\{ \begin{array}{ll}
\int \pi^E_{t+1}(w, \tilde{N} + E) dF_{(\tilde{N}_t, \tilde{S}_t, \tilde{N} + E)}(w), & \text{if } t \geq t^E + \Delta - 1; \\
\int \pi^E_{t+1}(w, \tilde{N}) dF_{(\tilde{N}_t, \tilde{S}_t, \tilde{N})}(w), & \text{if } t < t^E + \Delta - 1.
\end{array} \right.
\]

(16)

where as before $\pi^E_{t+1}(w, \tilde{N}) = \max_{\tilde{S}_{t+1}} \pi^E_{t+1}(w, \tilde{S}_{t+1}, \tilde{N})$ is the optimal profit-to-go.

By the same argument as in the proof of Proposition 1, the difference between the optimal profit-to-go after capacity is expanded and before is supermodular in $(\tilde{N}, \tilde{S})$, while for each of the two cases the optimal spending is decreasing in $\tilde{N}_t$. Therefore the difference, to which we refer as the value of expanding capacity, is increasing in $\tilde{N}_t$ and as a result:

**PROPOSITION 6.** The optimal capacity expansion policy is of the threshold form; that is, there exists a set of state vectors $\{\tilde{N}^E\}$ such that $t^E$ is the first period in which $\tilde{N}_t \geq \tilde{N} \in \{\tilde{N}^E\}$.

And an immediate corollary to the increasing difference is that:

**COROLLARY 2.** The value of expanding capacity increases in the state $\tilde{N}_t$, and hence also in the firm size, $|\tilde{N}_t|$.

Since the above results and formulation are stated for the general model, they also hold for the simplified models as well. Specifically, the threshold policy takes the following form: for any $N^H_t$ there exists a threshold function, $N^{LE}_t(N^H_t)$, such that the firm should expand its capacity as soon as it finds itself in the state $(N^H_t, N^L_t)$ with $N^L_t \geq N^{LE}_t(N^H_t)$. Such a threshold function, however, may be managerially difficult to work with.

A heuristic approach could be to decide on the expansion based on the total number of customers, $N_t = N^H_t + N^L_t$, i.e., on the norm of the state vector rather than the vector itself. This is certainly more managerially relevant: the firm expands as soon as its number of customers hits some $N^E$. In our simulations we saw that the optimal (expected) profit-to-go using this heuristic was within 1.5% of the optimum, starting from states in which the initial numbers of H- and L-type customers were unbalanced by a factor of up to four. This is because the firm is continuously and actively optimizing its customer mix and hence seeing grossly unbalanced customer mix is unlikely; that is, the retention and acquisition spending is selected such that the ratio of H- and L-types is as close to optimum as possible, thus deciding based on $N_t$ is nearly optimal.

Next we use this heuristic to illustrate our results numerically based on the same parameters as in the previous sections with baseline capacity of 100 that can be expanded by 50 units. Figure 3 (a) shows how the profit-to-go of the firm that follows the above heuristic depends upon the value of $N^E$ for three costs, $c_E$. 
We make the following observations: When the cost of expanding is very small, it is optimal to expand immediately, when the cost is very large it may never be optimal to expand, and for the intermediate costs, the higher is the cost, the later should the firm expand. However, if ever expanding, it is not optimal to wait until the capacity constraint becomes binding – the firm should expand when it still has some unutilized capacity.

The latter observation is driven by the following logic: since the firm’s spending (for a given $N$) is decreasing in the number of customers, as the firm approaches the capacity limit, it is treating its customers more and more suboptimally, leaving more and more value it can extract from each existing customer relationship “on the table”. Expanding capacity pushes the firm further away from the capacity limit, making it optimal to treat customers better and hence the firm recoups this previously uncaptured value. Since this value (captured if the firm expands but lost if it does not) increases as the firm approaches $N$ it is natural that the optimal $N^E$ is below $N$. Note, however, that there is an upward “tick” in the profit-to-go immediately preceding $N$. That happens because the above retention effect is partly offset by bumping. When the firm is very close to $N$, it loses value not only because of “suboptimal” retention but also because of occasional bumping.

We also note that the value of the unit of additional capacity is much larger than the CLV of the expected customers who will occupy it. For example, from Table 1, the CLV of the average customer is $2,534$, while the firm in our example expands for $c_E < 165,000$, i.e., with the per-unit cost of increased capacity up to $165,000/50 = 3,300$. Note that this is even larger than the CLV of the H-type customer, $3,036$. By this point in the paper that should not be very surprising – by the same logic as in Section 6, CLV measures only the value from the first customer to occupy the unit of added capacity, while the firm can generate additional value from the unit even after the first customer leaves.

Our second numerical illustration deals with the spending during the time-lag for capacity expansion (construction period, etc.) Figure 3 (b) presents the optimal spending for the case when the decision to expand from 100 to 150 was made in period 1 but the capacity becomes available in period 5. The main observation is that during that time window the firm drastically changes its spending on the L-type customers. That happens because increased capacity implies the change in the optimal customer mix: from Table 1 for $N = 100$ the optimal $\rho = 0.6784$ while for $N = 150$ it is $0.7074$. Since new acquisitions have $\rho = 0.5$, to improve customer mix the firm must spend more on H-types (to retain more) and less on L-types (to lose some) – which is exactly what we see.

8. Conclusions
This paper considers a situation where a CRM-focused firm operates in the limited capacity environment; by which we mean that the number of customers the firm can serve is limited by some
The firm dynamically decides on customer-specific retention spending and on the per-period acquisition spending in order to maximize its total net present value from sales to heterogeneous customers over an infinite time horizon. To isolate the effects of the capacity constraint, we assume that customer relationships are otherwise independent. That is, were it not for the capacity constraint, the firm’s optimal spending would be selected so that to separately maximize each customer’s lifetime value. We seek to understand how optimal spending given the capacity constraint compares to the CLV maximizing amounts.

We find that the optimal spending for the firm with limited capacity increases in capacity and decreases in firm size, which contrasts to the constant (across time, number of customers, customer mix, etc.) spending that is optimal for the firm with unlimited capacity. Counterintuitively, however, the optimal retention spending for the firm with limited capacity can exceed the CLV-maximizing spending. As we discuss, that happens because of the complex relationship between acquisition and retention spending, which makes it beneficial for the firm with limited capacity to overspend on some customers and decrease the lifetime value of each relationship in order to improve the customer mix and save on the acquisition spending. We show that this effect is particularly strong in the cases with small discount factor; in those cases the present value of the lifetime value losses is small, and hence the immediate savings play a bigger role.

At the same time, when the discount factor is high and the firm has only a few customers, its optimal spending is effectively equal to the CLV-maximizing amounts. That is, if the firm with high discount factor and ample unutilized capacity ignores, by mistake or on purpose, its capacity limit and spends so as to separately maximize the CLVs of its individual customers, then it operates nearly optimal. As the firm grows, however, it should decrease its spending. Numerically we see that the spending decreases significantly, but certainly not all the way to zero, as some managers seem to think.

Once the firm reaches its capacity limit, it can further increase its value by optimizing the customer mix, and as a result converge to the case of the firm optimally at capacity. We consider this case and provide implicit functional equations for the optimal spending. In particular, we show that a non-discounted CLV (i.e., computed without accounting for the time value of money) is helpful in determining the optimal spending for firm at capacity: the firm should select spending such that to balance the deviations from the non-discounted CLV-maximizing spending of different customers.

We also discuss customer selection. We show that selecting customers based on their CLVs is misleading. In addition to customer CLV the firm should also consider the opportunity cost of tying-up a unit of capacity. That cost is not proportional to CLV, and as a result, the firm may be better off by selecting a customer with a lower CLV. Likewise, the firm may be better off by selecting a customer that contributes less to the firm each period. We also show that these findings are robust with respect to the uncertainty in retention probability.

Lastly, we extended our model to the case when the firm can expand its limited capacity. A distinct characteristic of limited capacity, as opposed to other forms of limited resources (such as for example a budget constraint), is that expansion is “lumpy” and discontinuous: it involves a significant upfront fixed cost, which adds a significant number of units of capacity, possibly with a delay. We show that the optimal expansion policy is of the threshold form. At optimality the decision involves a vector of customer counts, but a simple managerial heuristic is to expand when the total number of customers becomes large enough. We show that it is not optimal to wait with expansion until capacity becomes binding: the firm should expand even in anticipation of capacity limit because it allows for extracting more value out of each customer relationship. Decision for when to expand is naturally related to cost: the larger is the cost, the later should the firm expand; in any case, the value of added unit of capacity is much larger than the CLV of the customer who will occupy it. Finally, if expansion involves the time lag, then the firm’s spending changes
drastically during this time lag: it spends more on high-value customers and less on low-value, which happens because expansion changes the optimal customer mix and hence the firm needs to rebalance its customer mix.

In terms of the future research and potential extensions, this paper offers several possibilities. First, price discounts could be used as a retention device; our current model needs to be adjusted to accommodate discounts as they are only “spent” on those customers that renew; we do so in the follow-up paper. Second, a possible dependency between acquisition and retention effectiveness in the sense of Thomas (2001) could be of interest for future research; e.g., acquisition could be more effective in the periods when significant retention discounts are provided. Similar effect could be driven by competition as well: e.g., bumping customers in the presence of a competitor may have adverse effects on acquisition; likewise competition for acquiring high-value customers may have adverse effects on retention (e.g., Klemperer 1995). Third, one could explicitly consider the case where instead of bumping a customer and losing a sale, the firm would incur a penalty (either in the same period, or, more interestingly, in the future because of the decrease in customer satisfaction). In either of the above cases, it could be valuable to consider a more elaborate state space within the dynamic programming formulation; in that case, however, it may not be immediately clear what a comparable CLV model would be and how the results of dynamic optimization of spending would compare with the CLV-maximizing spending suggested by that model. Finally, our models apply only to the customer retention situations; it could be of interest to consider “customer migration” situations in which a non-response (e.g., to a promotion) does not imply the end of the relationship. While we expect that the general findings of this paper will apply in many of the above mentioned extensions, future research is needed to provide more detailed results.

References


Appendix. Proofs

Proof of Proposition 1. Recall that if a distribution $J_x(y)$ is stochastically increasing (concave, supermodular, etc.) in parameter $x$, then the expectation of any increasing function $j(y), \int j(y) dJ_x(y)$, is also increasing (concave, supermodular, etc.) in $x$ (e.g., see Topkis 1998, proposition 3.9.1).

Further, because $\bar{N}$, $\bar{S}$, and $\bar{F}$ are vectors, with respect to the above result, “increasing” means that at least one component of the vector in increasing, while other components do not decrease (see Topkis 1998, Section 2.2). Consider arbitrary components $N_i$ and $S_j$ that we will increase, and assume other components do not change (this can be done without loss of generality, because an increase in multiple components can be represented as a series of single-component-only increases). From the above result, because $g$ is increasing in $N_i$ and $N$ and $F$ is stochastically increasing in $N_i$ and $N$, by induction $\pi_t^*+1$ is increasing in $N_i$ and $\bar{N}$. Therefore the optimal $\bar{N}_t$ is also concave. Therefore the optimal $\pi^*_t$ and $\bar{S}_t$ satisfy Lagrangian FOC:

$$
\frac{\partial^2 \pi}{\partial N \partial S_t} = \beta \int \frac{\partial \pi^*_t}{\partial N} \frac{\partial}{\partial S_t} dF + \beta \int \pi^*_t \frac{\partial^2}{\partial N \partial S_t} dF \geq 0
$$

because, in addition to the increasing properties established above, in the first term $F$ is stochastically increasing in $S_t$, and in the second term $F$ is stochastically supermodular in $(\bar{N}, S_t)$. Since components $N_i$ and $S_j$ were selected arbitrary, $\pi^*_t(\bar{N}, S_t)$ is supermodular in $(\bar{N}, S_t)$ and therefore the optimal spending $\bar{S}_t^*(N)$ is increasing in $N$. Q.E.D.

Proof of Proposition 2. Let $D(N, \rho, S)$ be the maximum demand that the firm with $N$ customers with aggregate quality $\rho$ today can face next period if it spends $S$ in total:

$$
D(N, \rho, S) = \max_{A, R^H, R^L, S} \text{such that } \max_{N, \rho R^H + N(1-\rho)R^L + A = S}[N\rho r^H(R^H) + N(1-\rho)r^L(R^L) + n(A)]. \tag{17}
$$

Observe that if $D(\cdot)$ is increasing, then $\min[N, D(\cdot)]$ is supermodular in $(\bar{N}, \cdot)$. Lemma 1 establishes the increasing and some useful properties of $D(\cdot)$ which we then use to prove the proposition.

**Lemma 1.** Let $R^H(N, \rho, S)$, $R^L(N, \rho, S)$ and $A^*(N, \rho, S)$ be the solutions to the demand function optimization problem (17). Then:

(a) $R^H(N, \rho, S)$, $R^L(N, \rho, S)$ and $A^*(N, \rho, S)$ exist and are unique;

(b) $R^H(N, \rho, S)$, $R^L(N, \rho, S)$ and $A^*(N, \rho, S)$ increase in $S$ and decrease in $N$ and $\rho$;

(c) $R^H(N, \rho, S) \geq R^L(N, \rho, S)$;

(d) $D(N, \rho, S)$ is increasing in $N$, $\rho$ and $S$.

Proof of Lemma 1. Part (a). Because $n(\cdot)$, $r^L(\cdot)$ and $r^H(\cdot)$ are concave, $N\rho r^H(R^H) + N(1-\rho)r^L(R^L) + n(A)$ is also concave. Therefore the optimal $A^*$, $R^H$ and $R^L$ satisfy Lagrangian FOC:

$$
\begin{aligned}
&\begin{cases}
N \rho R^H + N(1-\rho)R^L + A = S \\
\frac{\partial n}{\partial A} = \frac{\partial r^L}{\partial A} = \frac{\partial r^H}{\partial A} = -\lambda.
\end{cases}
\end{aligned}
$$

Since $n(\cdot)$, $r^L(\cdot)$ and $r^H(\cdot)$ are increasing and concave, the optimal $A^*(\lambda)$, $R^L(\lambda)$ and $R^H(\lambda)$ exist and are unique. Further, it is easy to see that all three optimal decisions are increasing in $\lambda$ (note that $\lambda \leq 0$). Therefore, the optimal $A^*(N, \rho, S)$ also exists and is unique, and so $A^*(N, \rho, S)$, $R^L(N, \rho, S)$ and $R^H(N, \rho, S)$ exist and are unique.

Part (b). The result follows by implicit differentiation of the constraint for $D$ (first equation in the FOC above) at the optimal $A^*(N, \rho, S)$, $R^L(N, \rho, S)$ and $R^H(N, \rho, S)$. The cases with $N$ and $S$ are straightforward. For the case of $\rho$, $\partial A^*/\partial \rho$ satisfies:

$$
N(R^H - R^L) + \frac{\partial A^*}{\partial \rho} \left( \frac{\partial A^*}{\partial \lambda} + N \left( \rho \frac{\partial R^H}{\partial \lambda} + (1-\rho) \frac{\partial R^L}{\partial \lambda} \right) \right) = 0
$$

and so

$$
\frac{\partial A^*}{\partial \rho} = -\frac{N(R^H - R^L)}{\frac{\partial A^*}{\partial \lambda} + N \left( \rho \frac{\partial R^H}{\partial \lambda} + (1-\rho) \frac{\partial R^L}{\partial \lambda} \right)} \leq 0
$$

(18)

if $R^H \geq R^L$, which we show next.

Part (c). By assumption $r^L(\cdot)$ and $r^H(\cdot)$ are increasing concave and at each point the slope of $r^L$ is higher than the slope of $r^H$. Since at the optimal $R^H$ and $R^L$, the slopes of $r^L$ and $r^H$ are equal (to $-\lambda$), it therefore follows that $R^H \geq R^L$. 


Part (d). We show only the proof for the property in \( \rho \); properties in \( N \) and \( S \) can be established similarly. By differentiating \( D \) over \( \rho \) at \( A^*(\lambda^*(N, \rho, S)) \), \( R^L(\lambda^*(N, \rho, S)) \) and \( R^H(\lambda^*(N, \rho, S)) \) we obtain:

\[
\frac{\partial D}{\partial \rho} = N(r^H - r^L) + N \left( \rho \frac{\partial R^H}{\partial \lambda} \frac{\partial \lambda}{\partial \rho} + (1 - \rho) \frac{\partial R^L}{\partial \lambda} \right)
\]

where the first equation holds because at the optimum \( \frac{\partial n}{\partial A} = \frac{\partial r^L}{\partial R^L} = \frac{\partial r^H}{\partial R^H} = -\lambda \), the second follows from (18), and the inequality follows from (c) because \( r^L \) and \( r^H \) are concave, \( r^H \) has a steeper slope and because at optimum slopes are equal to \(-\lambda\). Q.E.D.

Proof of Proposition 2 continued. Since by the assumption \( \frac{\partial r^H}{\partial x} \geq \frac{\partial r^L}{\partial x} \), and \( \rho, \xi \leq 1 \), it is not difficult to verify that \( N_{t+1} \) and \( \rho_{t+1} \) are increasing in \( (N_t, \rho_t, S_t) \). Further, for all \( R^H, R^L, A \) such that \( N_t \rho_t R^H + N_t(1 - \rho_t) R^L + A = S_t \) for any given \( S_t \), \( g_t \) is increasing in \( N_t \) and \( \rho_t \). Combining these results with (d) of the Lemma, the simplified model satisfies the assumptions of the general model discussed in Section 3.2 with decision \( S_t \) and state space \( N_t = N^H + N^L \). That is, the total optimal marketing spending \( S_t^* = N_t \rho_t R^H_t + N_t(1 - \rho_t) R^L_t + A_t^* \) is increasing in the firm’s capacity, \( N \) and is decreasing in firm’s size \( N_t \). From (b) both acquisition and retention components of spending also increase with \( N \) and decrease with \( N_t \). Further, from (c) at all times the firm spends more on retaining high-value customers. Q.E.D.