Chapter 24
Churn Management

Abstract While database marketing activities such as cross-selling, up-selling, frequency reward, and customer tier programs focus on developing the customer, there is always the fear that in the midst of these efforts, the customer will decide to leave the company, i.e., "churn." We discuss the approaches that can be used to control customer churn, focusing on proactive churn management, where the customer is contacted ahead of when he or she is predicted to churn, and provided a service or incentive designed to prevent the customer from churning. We review predictive modeling of customer churn, and present a framework for developing a proactive churn management program.

24.1 The Problem

Using a simple retention model, the lifetime value of a customer is:

$$LTV = \sum_{t=1}^{\infty} \frac{m_t r^{t-1}}{(1+\delta)^{t-1}} \tag{24.1}$$

Customer churn management focuses on the retention component, r. At the customer level, churn refers to the probability the customer leaves the firm in a given time period. At the firm level, churn is the percentage of the firm's customer base that leaves in a given time period. Churn is therefore one minus the retention rate:

$$\text{Churn} = c = 1 - \text{Retention Rate} = 1 - r \tag{24.2}$$

High customer churn is a concern for any industry where a simple retention lifetime value model is applicable, i.e., where customers can leave and not naturally return without a significant re-acquisition effort. This includes many services such as magazine and newsletter publishing, investment services, insurance, electric utilities, health care providers, credit card providers, banking, Internet service providers, telephone service providers, online services, and cable service providers.

There are two major types of customer churn: "voluntary" and "involuntary." Involuntary churn refers to the *company* deciding to terminate the relationship with the company, typically because of poor payment history. Voluntary churn refers to the *customer* deciding to terminate the relationship with the company. Hadden et al. (2006) further distinguish between "deliberate" voluntary churn, where the customer is dissatisfied or has received a better competitive offer, and "incidental" voluntary churn, where the customer cancels service because he or she no longer needs the product or has moved to a location where the company does not offer service.

Table 24.1 displays reported total churn rates for various industries and companies. The table shows it is not uncommon for between 20% and 50% of the customers who begin the year with a company to leave by the end of the year.[1] Table 24.1 suggests two observations: (1) Churn seems to be decreasing in the US wireless telecom industry.[2] This could be due to industry consolidation as well as better service. (2) Nascent industries such as digital services have high churn rates. This could be due to competition, poor service, or using steep promotions to acquire customers.

A few calculations illustrate the importance of churn rates of the magnitude reported in Table 24.1. First, consider the expected lifetime of the customer, that is, how long the customer stays with the company. This can be shown to equal the reciprocal of the churn rate (Appendix 24.1), i.e.,

$$\text{Expected Lifetime of Customer} = \frac{1}{c} \tag{24.3}$$

Figure 24.1 graphs expected lifetime as a function of the churn rate. The figure shows that if the annual churn rate is 50%, the average customer is with the company for 2 years. At a 40% churn rate, this becomes 2.5 years; at 30% it's 3.3 years; and at 20% churn, it's 5 years. So within the range of the data reported in Table 24.1, a company with a relatively low churn rate keeps its customers for roughly 2 1/2 additional years compared to a company with a relatively high churn rate. What's more, the function is convex: The 20% churn company starts with a customer lifetime of 5 years, but can move to 6.6 years by decreasing its churn rate by just 5 percentage points, to 15%. If it can decrease churn all the way to 10%, average customer lifetime doubles, to 10 years.

Second, we can look at the lifetime value of the customer. Restating Equation 24.1 in terms of churn rather than retention, and assuming constant profit

---

[1] Note that one can extrapolate the monthly churn rates to annual churn rates by calculating $1 - (1 - c)^{12}$. This calculation assumes that the customer base on which the monthly churn rate is calculated retains that same churn rate for 12 consecutive months.

[2] The latest statistics are for "post-paid" customers, who typically have lower churn levels than "pre-paid" customers. But post-paid customers are the dominant customer type. In addition, Prudential Equity Group (2006) reports that the *post-paid* churn rate among the top four wireless carriers has on average steadily declined since 2002.

Table 24.1 Churn rates reported in the trade press

Industry	Year	Company/industry	Annual churn (%)	Reference
Internet service	2001	America Online	21	Kolko (2002)
Internet service	2001	Earthlink	34	Kolko (2002)
Internet service	2001	AT&T WorldNet	36	Kolko (2002)
Internet service	2001	NetZero/Juno	46	Kolko (2002)
Internet service	2001	MSN	57	Kolko (2002)
Internet service	2001	Industry range	38.7–63.2[a]	Pierce (2001)
Internet service	2000	Earthlink	47	Pierce (2001)
Internet service	2002	Industry range	31–39[a]	Yang (2002)
Wireless telephone	2000	Industry average	26.2[a]	Davidor (2000)
Wireless telephone	4Q1999	Industry med.	23.4[a]	Young (2000)
Wireless telephone	2001	Industry range	23.4–28.7[a]	Pierce (2001)
Wireless telepone	3Q2001	Verizon	31[a]	Fitchard (2002)
Wireless telephone	3Q2001	Cingular	34[a]	Fitchard (2002)
Wireless telephone	3Q2001	AT&T	37[a]	Fitchard (2002)
Wireless telphone	3Q2001	Sprint PCS	31[a]	Fitchard (2002)
Wireless telphone	3Q2001	VoiceStream	46[a]	Fitchard (2002)
Wireless telephone	3Q2001	Nextel	28[a]	Fitchard (2002)
Wireless telephone	1Q06	Verizon	10[a,b]	Prudential Equity Group, LLC (2006)
Wireless telephone	1Q06	Cingular/ATT	18[a,b]	Prudential Equity Group, LLC (2006)
Wireless telephone	1Q06	Sprint/Nextel	22[a,b]	Prudential Equity Group, LLC (2006)
Wireless telephone	1Q06	T-Mobile	22[a,b]	Prudential Equity Group, LLC (2006)
Satellite TV	1999	Pegasus Com.	17	Henderson (1999)
Satellite radio	2002	XM Satellite	20[a]	Wachovia Capital Markets, LLC (2006)
	2003	XM Satellite	19[a]	Wachovia Capital Markets, LLC (2006)
	2004	XM Satellite	29[a]	Wachovia Capital Markets, LLC (2006)
	2005	XM Satellite	28[a]	Wachovia Capital Markets, LLC (2006)
Financial services	1996	UK industry average	6	Supply Management (1998)
Financial services	1997	UK industry average	9	Supply Management (1998)
Financial services	2001	UK industry average	20–30	Fisher (2001)
Digital services	2005	Audible, Inc.	46[a]	Kaufman Bros (2005)
Digital services	2005	Netflix	39[a]	Needham & Company, LLC (2006)
Digital services	2005	HouseValues, Inc.	52[a]	Thomas Weisel Partners (2005)

[a] Numbers were provided on a monthly basis. Annual churn calculated as $1-(1-c)^{12}$. (See Footnote 1).

[b] Churn among "post-paid" as opposed to "pre-paid" customers.

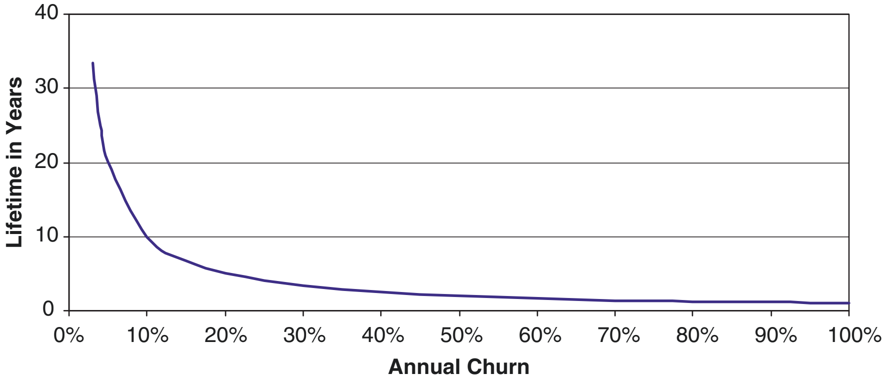

Fig. 24.1 Expected lifetime as function of churn rate (Equation 24.3).

contribution, yields:

$$LTV = \sum_{t=0}^{\infty} \frac{m(1-c)^{t-1}}{(1+\delta)^{t-1}} = \frac{m(1+\delta)}{(\delta+c)} \tag{24.4}$$

where:

m = Annual profit contribution per customer.
c = Annual churn rate.
δ = Annual discount rate.

Figure 24.2 shows lifetime value as a function of churn rate assuming $m = \$500$ per year and $\delta = 14\%$. The firm with a 50% churn rate has a lifetime value of \$891 per customer. The 40% churn firm has LTV = \$1,056; 30% translates to LTV = \$1,295; and 20% churn implies LTV = \$1,676. Assume acquisition costs in the wireless telephone or ISP industry of roughly \$300–400

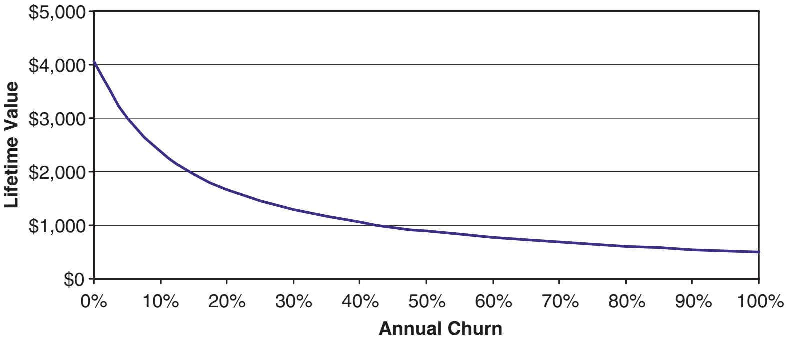

Fig. 24.2 Lifetime value of customer as function of churn rate (Equation 24.4)*.
* Assumes $\delta = 0.14$ and $m = \$500$.

per customer (Fitchard 2002; Pierce 2001). Companies then would be making money, but the 20% churn company has an ROI of ($1,676–$400)/$400 = 319%, while the 50% churn company has an ROI of ($891–$400)/$400 = 123%, which is quite different.

Figure 24.2 shows a convex relationship between churn rate and lifetime value – firms enjoy increasing returns with respect to decreasing churn. A telecom company with 20% annual churn starts with an LTV of $1,676, but can increase it to $1,966 by reducing churn to 15%. On a base of 5,000,000 subscribers, that's $1.45 billion in additional profits!

Of course it is difficult to decrease churn and it costs money, but one sees the opportunity for increased profits by managing churn effectively, and conversely, the threat to company viability if churn escalates.

24.2 Factors that Cause Churn

The first step toward predicting and remedying churn is to understand the factors that cause or alleviate it. Table 24.2 provides a categorization of these factors: customer satisfaction, switching costs, customer characteristics, marketing efforts, and competition.

More satisfied customers should be less likely to churn, i.e., have higher lifetime durations (Anderson and Sullivan 1993; Hallowell 1996; Bolton 1998). In fact, Anderson and Sullivan (1993) found that companies with higher satisfaction levels enjoy lower elasticities of retention with respect to quality, i.e., these firms are insulated from short-term negative deviations in quality. There is some evidence that gross measures of overall satisfaction may not be perfectly predictive of churn (Kon 2004). Therefore, satisfaction should be measured with respect to particular components of the product, e.g., perceived service quality, whether the product is meeting expectations, how well the product fits customer needs, and price. Service quality and expectations are of course fundamental issues in services marketing (Anderson and Sullivan 1993).

"Fit-to-needs" is another important issue in customer satisfaction. For example, when companies use strong promotional incentives for acquisition, it succeeds in acquiring the customers, but acquires the wrong customers, i.e., customers for whom this is not really their best choice. Another point regarding fit-to-needs is the belief that the one-size-fits-all approach to services marketing often leads to lack of fit and results in lower satisfaction and churn. Companies in online retailing, Internet service provision, satellite TV, telecommunications, news and information, banking, and shipping are trying to customize their products to the customer to prevent this lack of fit problem and create customer satisfaction (Verton 2001; Kreitter 2000; Kleinman 2000; Bitner et al. 2000).

Table 24.2 Factors hypothesized to cause or alleviate churn

- Customer satisfaction
  - Service quality
  - Fit-to-needs
  - Meeting expectations
  - Price
- Switching costs
  - Physical
  - Psychological
- Customer characteristics
  - Risk aversion
  - Variety seeking
  - Deal proneness
  - Mavenism
- Marketing
  - Reward programs
  - Promotions
  - Price
  - Customized products
- Competition
  - Within category
  - Between category

Keaveney and Parthasarathy (2001) find intriguing evidence that the information source customers use to choose a particular service influences churn. This is consistent with evidence that acquisition channel is related to customer retention (Verhoef and Donkers 2005). Keaveney and Parthasarathy attribute their finding to the influence that information source has on customer expectations and perceived fit-to-needs. They find that customers are less likely to churn if they use external or experiential information in making their choices. This information leads to more accurate expectations, more knowledgeable choice, and a potential bias to confirm the positive "priors" set up by the information. However, information from peers was associated with more churn, since the information provided there is vicarious and less convincing.

Unfortunately, rarely does the database marketer have direct measures of customer satisfaction. For this reason, prior purchase and usage behavior variables, which are usually measurable, often serve as surrogates for satisfaction. Reinartz and Kumar (2003) find that sales volume and cross-buying are negatively associated with churn,[3] while returns and focused buying are positively associated. Purchase frequency has a U-shaped relationship with

[3] Note cross-buying might also create switching costs, making it less likely the customer will churn.

churn, with very light and very heavy users having higher churn likelihood. The light user might be clearly unhappy with the service, while the heavy users might be so involved in the product that they are more demanding, hence less satisfied and always on the look-out for a better alternative. Li (1995) also finds that cross-buying is negatively related to churn but finds sales volume is positively related.

As additional indicators of satisfaction, companies sometimes measure previous customer contacts with the call center, or complaints and objective measures of customer quality. Hadden et al. (2006) include these variables in a predictive churn model, and we will discuss the results in Sect. 24.3.1. Coussement and Van den Poel (2007b) find that the "emotionality" displayed in customer e-mails to the company can be measured and improve churn prediction.

Switching costs, or the lack of them, are another reason for churn. Switching costs can take two forms – psychological and physical – but both have the same theme: If it does not cost the customer much to switch to a competitor, the customer is more likely to churn. Psychological switching costs include psychological barriers to switching such as inertia, brand pull, familiarity, and perceptions of a relationship with the current company. Physical switching costs include real inconveniences due to switching. Consider for example Internet portals. The customer might have invested much effort in customizing the home page of their chosen portal (e.g., AOL). Switching to a new portal would require those efforts to be repeated Kolko 2002).

An interesting natural experiment regarding physical switching costs occurred in the wireless telephone industry, regarding "number portability." Before the fall of 2003 customers who switched carriers would have to change their telephone numbers as well. As of fall of 2003, customers were allowed to retain their old numbers. The wireless industry braced for a huge increase in churn (La Monica 2003). This was supported by a Forrester study (Golvin 2002), which found that 22% of cell-phone users were interested in switching carriers if they could retain their current phone number, while only 9% were interested in churning if they could not keep this number. Interestingly, they found that high volume callers would be the most likely to switch under number portability. This makes sense in that these customers are spending much money and looking for a better deal, but without portability, the cost of switching is enormous because they would have many people to notify of their new number.

Figure 24.3 displays customer churn for the six major carriers during the period when number portability was instituted (November 2003). The results reveal, quite surprisingly, that churn did not generally increase. The possible exception is ATT (perhaps not unexpectedly, ATT was later to merge with Cingular). But besides ATT, the major carriers experienced virtually no increase in churn.

There are several possible reasons for the lack of churn experienced after number portability. First, there still may have been switching costs because the customer would have to purchase a new phone and the new phone would

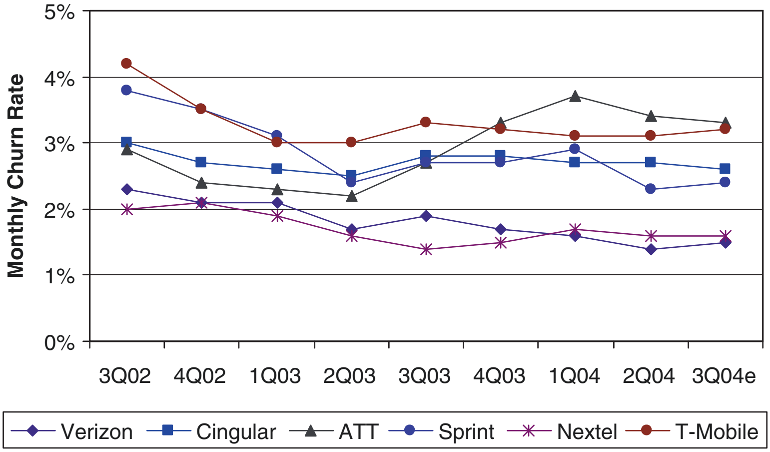

Fig. 24.3 Customer churn in the wireless telephone industry before and after the implementation of number portability (November 2003) (From Jefferies Equity Research 2004).

need the old number programmed into it. Second, when all is said and done, customers by November 2003 were not really so dissatisfied with their service that they wanted to churn (intentions do not always translate into behavior (Morwitz 1997)). Third, companies may have increased marketing efforts during this period (e.g., see The Economist (2007)).

Customer characteristics also play a role in churn. Risk takers, variety seekers, innovators, shopping mavens, and deal prone consumers might be more likely to churn. Keaveney and Parthasarathy (2001) found that risk taking was positively related to churn, and churn was also higher for lower income and lower levels of education. Reinartz and Kumar (2003) found that rural and higher income customers were less likely to churn. Li (1995) found that married and higher income customers were less likely to churn, and car owners and frequent movers were more likely to churn. Hallowell (1996) also found that higher income customers were less likely to churn. The higher loyalty from higher income customers may be due to lower price sensitivity.

Company marketing efforts ideally should have an impact on churn. This includes special services, loyalty programs, and price. Reinartz and Kumar (2003) found marketing efforts and loyalty programs associated with lower churn, although Reinartz and Kumar (2000) found that catalog mailing costs relative to total sales were not different between long and short lifetime customers. Li (1995) found that discount programs were associated with less churn. One common assumption is that matching the customer to the right pricing plan (e.g., for telecom) is associated with less churn. However, Lambrecht and Skiera (2006) find that customers have a systematic bias toward selecting a flat rate plan over a pay-per-use plan, even if they would save money on the pay-for-use plan. What's more, they find that using the flat fee plan (the "wrong" plan in terms of dollars spent) is not associated with higher churn. The authors attribute this to an "insurance effect" desired by customers. These results caution companies to be careful in encouraging customers to adopt new pricing plans to head off churn.

Competition is the fifth major category of causes for churn. Competition can come from both within and outside the product category. For example, dial-up ISP providers need to worry other dial-up providers as well as broadband (Kolko 2002). Online bankers have to worry about regular banking (Ensor 2002). The conventional wisdom is that competitive offers and opportunities are a major cause of churn (Elstrom 2002; Whiting 2001), but there has been little empirical verification on this, perhaps due to the fact that a given company often has little direct information on competitive offers.

In any given situation, all five factors above may be at work. Consider for example the Internet Service market. Many companies in this business relied on ubiquitous promotions and deals to sign up subscribers. Customer satisfaction and fit-to-needs were not always high, but high switching costs prevented churn. Many of the first ISP's used dial-up service, which did not satisfy needs for speed and convenience. So when a competitive category (broadband) offered better service, these ISP's developed a significant churn problem (Yang 2002). Forrester Research hypothesized that the "savviest" users (the market mavens) were most likely to churn (Kolko 2001). This churn problem continued despite ample marketing efforts to arrest it.

24.3 Predicting Customer Churn

A key step in proactive churn management is to predict which customers are most likely to churn. While predictive models can be developed for this purpose, their accuracy is hampered by lack of access to direct measures of the causes for churn. Companies therefore use behavioral measures such as recency, frequency, and monetary value as surrogates. They may also have records on customer complaints or previous offers extended to customers, as well as current prices paid, previous retention efforts, or acquisition source. They might have good measures of physical switching costs such as number of products used by the customer. They typically would not have good data on customer psychographics such as risk-taking, but they often have demographic measures that can link well to psychographics (e.g., Ailawadi et al. 2001). Finally, they typically have no data on competitive activity.

There are two types of churn prediction models, "Single Future Period" and "Time Series." In Single Future Period models, predictors are available for one or several time periods before the "churn period," where we observe whether or not the customer churned. For example, predictors might be measured for February and March, and those data are used to predict whether the customer churns in April. Often a month is skipped between the predictor and churn periods because it will take a month for the analyst to predict would-be churners and then contact the would-be churners with an offer to prevent the customer from churning in the subsequent month.

Time Series models are based on data where we observe predictors and churn simultaneously as they occur period to period. So, each customer would be observed for several periods. Each period potential predictors would be collected, as well as whether or not the customer churned. Obviously, one still must use data from period $t-1, t-2$, etc., to predict for period t or later, but the major difference is that churn is observed over time rather than in a single future period.

24.3.1 Single Future Period Models

An excellent example of a single future period churn model is from the credit card industry, provided by Advanced Software Applications as a demo for their ModelMax software.[4] This software uses an exploratory/stepwise procedure to select predictors, and then estimates a neural net to predict the dependent variable. Predictors were collected for 6 months prior to the target month when churn was observed. Predictors included demographics (age, occupation, credit rating), current product ownership (various other services offered by the bank), and product usage variables (RFM measures such as number of purchases, credit card balance, purchase amounts, etc.). The usage variables were calculated for each of the 6 months prior to the churn month. For example, "Credit card balance in month 6" refers to the customer's credit card balance in the month before the churn month. In all, there were 105 potential predictors. The model selected six predictors:

- *Credit card balance in month 6*: Users with low balances are more likely to churn. It may be difficult for those with high balances to transfer those balances to a new credit card (a switching cost).
- *Interest charged in month 3*: Customers with low interest charges, but not zero, are more likely to churn. This may be because such customers view interest as nuisance charges.
- *Purchase level in month 6*: Customers with low purchase levels are more likely to churn. This makes sense as these customers are probably less satisfied with the card, or have in fact adopted a new card and are phasing out the old one.
- *Household age*: Younger customers are more likely to churn. This may be due to the venturesome nature of younger people in credit card market, since it is relatively new to them. They are not set in their ways. They are also more likely to be courted by competitors.
- *Payment year-to-date month 2*: This is an overall measure of monetary value. The finding is that customers in the mid-range are more likely to churn. It might be that high-monetary value customers are satisfied with

[4] The authors are grateful to Advanced Software Applications for allowing us to use this example.

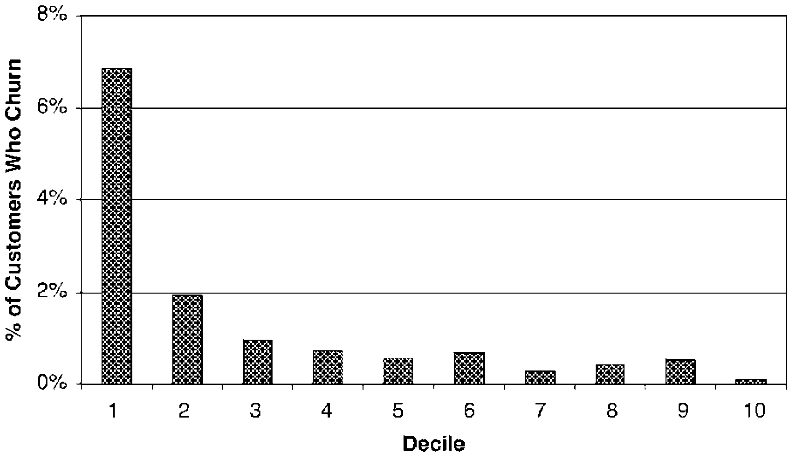

Fig. 24.4 Calibration data lift chart for neural net churn model – credit card company*. * This example provided courtesy of Advanced Software Applications. Model estimated using ModelMax software.

the card, which is why they use it so much, while it is not worth it for low-monetary value customers to switch because there is little at stake (e.g., a low interest rate would not save them much money because their spending level is low).

- *Cash advances month 4*: Customers who do not use this feature are more likely to churn. Customers who use the feature may be more satisfied with their card, or would have a higher switching cost to switch cards since they'd have to learn how to use the cash advance feature with their new card.

Figure 24.4 shows the lift chart for this model. It is very strong, with 5.3 to 1 top-decile lift for the 40,000-customer calibration data. There was no appreciable degradation in prediction on the 10,000-customer validation data. This performance is much better for example than for the wireless telephone churn models reported by Neslin et al. (2006a) and discussed later in this section. There can be a variety of reasons. One could be the relative maturity of the credit card category, so behaviors are stable and easier to predict. Obviously, the reason why churn model accuracy differs by category, and indeed why predictive model accuracy differs by category, is an important topic for future research.

Lemmens and Croux (2006) investigate the use of boosting and bagging (Chapter 19) in predicting churn. They use the same data as those used in the Teradata Churn Prediction Tournament described subsequently in this section (Neslin et al. 2006a). The authors show that both boosting and bagging of a decision tree model improve performance significantly over logistic regression. Since the true churn rate was approximately 2% for these data (i.e., churn is a "rare event", see Chapter 10; Donkers et al. 2003), the authors use choice-based sampling to create a balanced calibration dataset. The authors investigate ways to adjust for this when the calibration data model is applied to holdout data, which reflects the true proportion of churners (2%). They find a simple intercept correction performs best. The authors find that how long the customer has owned his or her handset is the most important predictor of churn. This makes sense. Given the rapid changes in handset technology and features, customers who have owned the same handset for a long time are ripe for a competitive offer featuring a new handset if they switch.

Wei and Chiu (2002) also address the issue of rare events in modeling customer churn. In their data, the average monthly churn rate is 1.5–2%. The authors develop a "multi-classifier class-combiner" approach to address this issue. Say 1:X is the ratio of churners in the sample, i.e., with a 2% churn rate, 1:X = 1:50. The researcher desires a ratio of 1:Y, where Y = 1 means a 50–50 split between churners and non-churners. The researcher creates X/Y databases as follows. Each database contains all available churners and N/(X/Y) non-churners. The same churners are used in each sub-database, while the non-churners are evenly divided among the sub-databases. A model is estimated for each sub-database. Assume that n_{c_1} models predict the customer will churn, while n_{c_2} models predict the customer will not churn. We predict the customer is a churner if $wn_{c_1} > (1 - w)n_{c_2}$, where w is a weight such that $w \in [0, 1]$. If $w = 0.5$, the customer is classified as a churner according to majority rule. If $w > 0.5$, churn predictions are given heavier weight, and we are more likely to predict the customer is a churner.

The authors test various sub-database ratios (1:2, 1:4, 1:8, and 1:16) and compare the false alarm and miss rates for these ratios to a single sample, for all w's between 0 and 1 in 0.01 increments. The authors find of course that as w increases, the multi-classifier class-combiner approach has fewer false alarms (Prob(Not churn | Predict Churn) decreases) but the miss rate is much higher (Prob(Churn | Predict Not Churn) increases). This is because a high w "biases" the method to predict more non-churners. Most importantly, the multi-classifier class-combiner approach is superior to the single-sample approach for all sub-database ratios tested. The authors prefer the 1:2 ratio because it provides a greater spread in miss rates as w varies, giving the analyst more leeway as to where to establish a cut-off. The comparison to the single-sample approach is shown in Table 24.3. The results show that for a given miss rate, the multi-classifier class-combiner yields a lower false alarm

Table 24.3 Performance of multi-classifier class combiner in predicting customer churn (From Wei and Chieu 2002, adapted from Figure 1, p. 108)

Method	False alarm rate (%)	Miss rate (%)
Single sample	21	48
Multi-classifier class combiner (Miss rate = miss Rate for single sample)	15	48
Multi-classifier class combiner (false alarm rate = false alarm rate for single sample)	21	40

rate, or for a given false alarm rate, it yields a lower miss rate. Depending on the value of w, the multi-classifier class-combiner method can be superior in terms of false alarms and miss rates (see the authors' graph on p. 108).

Hadden et al. (2006) demonstrate the use of predictors that, while not direct measures of satisfaction, are very good surrogates. In particular, the authors had 24 variables representing "Complaints" (type, number, etc.), "Repairs" (type of problem, how long it took make the repair, etc.), and "Provisions" (appointments broken before making the repair, days after the promised date that the repair was resolved, etc.). The authors do not identify the industry but state that the company was "one of the largest in the world in its domain," (p. 105), which explains why the company kept such careful data.

The authors compare neural nets with Bayesian architecture (Li and Wen 2005; Yu and Dayan 2005), neural nets with feed-forward back propagation, decision trees, and linear regression. The neural nets used all 24 variables ("inputs" in neural net parlance) along with two hidden layers (see Chapter 18). The decision tree method used was CART ("classification and regression tree," see Chapter 17) and while all 24 variables were made available, the final tree utilized only seven variables. To select variables for the regression model, the authors calculated the standard error rate (total number of errors divided by total number of classifications) for each of the 24 variables, and decided to retain 14.

The authors find that the models differ in accuracy. In their validation data, 30% of customers were churners; 70% were non-churners. They use a cut-off for each model's prediction (0.70) to classify customers and the results are in Table 24.4.

The NN–Bayesian model achieved a "lift" of $70/30 = 2.33$ (the customers it predicted to churn were more than twice as likely to churn as average). However, the NN–Bayesian method did little better than average in predicting a customer was not going to churn, i.e., the lift would be $75/70 = 1.07$. Regression and decision trees appear to balance the two errors the best.

While, clearly, Complaints, Repairs, and Provisions data are quite useful in predicting churn, similar to what we observed earlier, there was little overlap across models in the variables determined to be most important. For example, neural networks found that "resolution time" was among the seven most important variables; none of the other methods found this to be in the top seven.

Table 24.4 Performance of alternative methods for predicting churn (From Hadden et al. 2006)

Method	P(Churn | predict churn) (%)	P(No churn | predict no Churn) (%)	Overall accuracy (%)
NN–Bayesian	70	75	74
NN–standard	55	79	72
Decision tree	66	88	82
Regression	51	94	81

This may be because the predictors are highly correlated and in that context, we might not expect the top seven of 24 variables to have much overlap.

Hadden et al.'s study is important because it shows that data very "close" to customer satisfaction – complaints, repairs, and repair provisions – have predictive power. Consistent with this observation, Coussement and Van den Poel (2007b) find that measures of customer emotionality in e-mails they send to the firm provide additional predictive power to predicting churn for newspaper subscriptions. More work is needed to show the incremental contribution of these data over usage history variables typically also used as surrogates of customer satisfaction.

Coussement and Van den Poel (2008) compare the accuracy of support vector machines (SVM), random forests (Breiman 2001; a form of bagging, see Chapter 19), and logistic regression in predicting churn for a Belgian newspaper. Their predictor variables include measures of client–company interactions such as complaint behavior; renewal variables such as whether the last subscription was renewed before it expired; customer characteristics, and previous purchase and usage behavior variables such as length of the current subscription. The authors use two approaches for estimating the SVM, each differing in the criterion used for a grid search – one is based on percentage correctly classified (PCC) using a threshold; the other is based on the area under the ROC (receiver operating characteristic curve; see Chapter 11), which they call AUC. The authors find that random forests perform the best. The differences do not appear markedly different (top-decile lifts were 4.48 for logistic, 4.75 for random forests, 4.21 for $\mathrm{SVM_{PCC}}$, and 4.49 for $\mathrm{SVM_{AUC}}$). However, as we will see in Sect. 24.4.2, differences in lift on the order of a tenth of a point can have an important impact on the profitability of churn management programs.

Neslin et al. (2006a) conducted a "churn modeling tournament" to investigate the best ways to predict churn. They made available data from a wireless service provider to anyone interested in participating in the tournament. The data consisted of a 100,000-customer calibration dataset including 171 predictors including customer characteristics, previous cell phone usage, and previous contact variables (calls to the customer care group, etc.). The data were of the typical single future period format – the predictors were collected over a 3-month period, then 1 month was skipped, and then churn was observed in the fifth period. There were two holdout datasets – one compiled at the same time as the calibration dataset ("current" validation data); the other compiled roughly 3 months later ("future" validation data).

The tournament attracted 44 entries from 33 participants. Roughly half were academics; half were practitioners. The top-decile lift results for the two validation databases were as follows:

Validation data	Mean	Std. Dev	Minimum	Maximum
Current	2.14	0.53	1.07	2.90
Future	2.13	0.53	1.19	3.01

There are two very important conclusions from the above results: (1) There is wide variation in achieved lift (from roughly 1–3). This means that method matters – not every entry achieved the same lift. We will show how to calculate the importance of lift in financial terms in Sect. 24.4.2. (2) There is very little fall-off between the current and future validation databases. This means that the shelf life of a churn prediction model is at least 3 months.

Perhaps most important is the authors' analysis of the methods used to make the predictions. As discussed in Chapter 10, they use factor analysis to uncover five general approaches to modeling churn:

1. *Logit*: These entrants used logistic regression as the statistical model, exploratory data analysis (EDA) and stepwise regression for variable selection, and allocated relatively less time to preparing the final prediction files. Practitioners tended to be more highly associated with this approach.
2. *Decision Tree*: These entrants heavily relied on decision trees as the statistical model. They were far less likely to use EDA and stepwise procedures, but allocated a lot of their time to estimation. This is quite consistent with the decision tree method, which requires careful pruning, etc., of the decision tree. Users of this approach spent more time in general on the task and subdivided the calibration data into estimation and holdout samples.
3. *Practical*: These entrants did not favor any particular statistical model. They relied heavily on common sense in selecting variables and allocated more time than average on downloading data but less time on average on the entire task. They did not subdivide the data as the Decision Tree entrants did. Users of this approach tended to be practitioners.
4. *Discriminant*: These entrants used discriminant analysis as their statistical technique. They allowed less time to data cleaning and more time to estimation, and tended to use many variables in the final model.
5. *Explain:* These entrants did not favor any particular statistical model and tended to explore several statistical techniques. They reported that they relied on theory, factor analysis, and cluster analysis to selecting variables, and tended to use fewer variables in their final models. Although the following is subjective, by their reliance on theory, factor analysis, and cluster analysis, it is as if these entrants were trying to come up with a parsimonious explanation of churn as well as trying to predict it.

The authors found that the Logit and Decision Tree approaches predicted most accurately, followed by Practical and Explain, with Discriminant lagging behind. But their major message is that predicting churn, as well as predictive modeling in general, is about much more than the statistical technique. It entails allocation of time, methods for selecting variables, and details such as subdividing the calibration data into estimation and holdout samples. For researchers, this is important because it suggests the importance of topics such as variable selection and holdout samples, but also the need to study predictive modeling holistically rather than as a statistical model. For practitioners, the implications are that logistic and decision tree-based approaches can both work well, but require different management. Decision tree approaches require much time for estimation; hence managers need to make sure their analysts have sufficient, in fact, more than sufficient, computing power. Logistic tree approaches will probably entail stepwise, so managers need to question analysts as to what variables were omitted and how this might affect the interpretation of the results.

This tournament is exciting in its use of a contest to find the best ways to predict churn, and its drawing on both academics and practitioners. However, the study has important limitations that suggest future research. First, there was not ample representation of machine learning techniques. It turns out that the winning entry was a machine learning version of decision trees, but there were not many entries like this and no entries of support vector machines and genetic algorithms, etc. Second, the data were so highly correlated that the authors could not tease out the individual contribution of statistical model, time allocation, etc. These variables tended to clump together to form "approaches," and this is instructive. Still, it would be nice to know the individual contributions of the components of the approaches. Third, other details such as treatment of missing data could not be explored in depth.

In summary, the message of Neslin et al. (2006a) is that churn modeling requires an *approach*, not just a statistical technique, and that while different approaches can be successful, not all approaches are successful.

24.3.2 Time Series Models

Hazard models provide a particularly attractive method of predicting churn using time series data (see Chapter 15). Lu (2001) employs a hazard model estimated using the LIFEREG routine in SAS.[5] The model does not allow for time-varying covariates but the user can assume a parametric baseline hazard function. This permits a closed form expression for survival probabilities, which in this case are the probability a customer does not churn up to a certain number of months from the origin of the data. In this application, Lu assumed a log-normal distribution for the baseline hazard function.

Lu collected data on 41,374 customers who were active as of August 2000 and observed over time whether each one churned. Predictors included demographics, internal customer data (RFM type data on call behavior, plus data on plan type, customer segmentation code, ownership of other company products, billing disputes, late fee charges, etc.), and customer contact data (inbound calls to the company's call center and outbound mail contacts with the customer). These 212 variables were narrowed to 29 by first omitting variables that did not significantly correlate with churn, and then using a stepwise procedure.

[5] See also Van den Poel and Larivière (2004) for an application of hazard models to predicting churn for a financial services company.

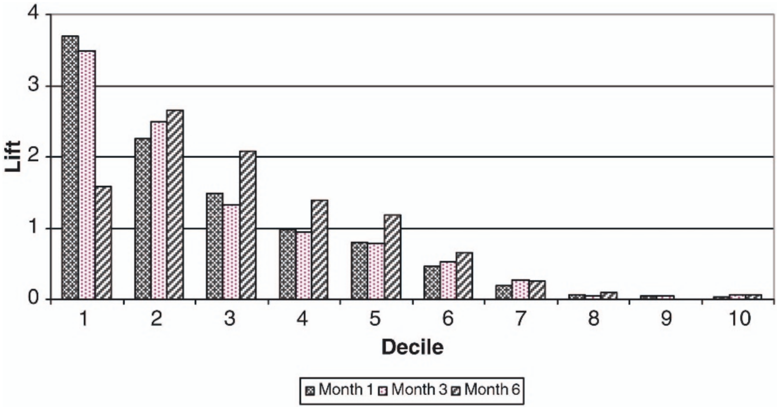

Fig. 24.5 Lift charts for hazard churn model depending on time horizon for prediction. Base churn probabilities were not available, but decile-by-decile lift could be calculated from cumulative gains charts that were provided. (From Lu 2001)

Survival probabilities were calculated to create lift charts, shown in Fig. 24.5. The figure shows the lift for predicting the probability that a customer will churn in a given future month – month 1, 3, or 6. There are two interesting findings in Lu's lift charts. First, the top decile lift is the highest for month 1, next highest for month 3, and noticeably lower for month 6. This may be because the author did not use time varying covariates (the covariates were defined as of time zero and did not change), and it is inherently more difficult to predict churn 6 months out than 1 or 3 months out.[6] Second, the top decile lift in month 3 is 3.5 to 1, which is better than most of the entries in the Teradata Churn Prediction Tournament described in the previous section (Neslin et al. 2006a), which predicted for month 2. This better performance may of course be due to the fact that these are different data. Or it could be due to the hazard model, which makes use of the time series nature of the data.[7]

Yan et al. (2001) propose an interesting approach to time series data that (1) combines several consecutive windows of data, and (2) can be used even if the dependent variable is not available for some observations. We discuss the first feature in more detail since it is more relevant to most churn modeling situations.

Assume the company has $T+1$ "windows" available, e.g., January–March for predictors; April to observe churn would be one window, February–April for predictors; May to observe churn would be the next, etc. This is essentially a time series of observations divided into $T+1$ databases. A predictive model

[6] Note the 2nd, 3rd, and other deciles look better for the 6-month predictions, but that's because fewer churners were grouped into the first decile, so they were more likely to be in other deciles.

[7] Even though in this case the author did not use time-varying predictors, his hazard model is inherently dynamic due to the log-normal hazard function.

could be estimated for each window, as in the Single Future Period approach. The prediction works as follows:

$$P[c|x_i] = \sum_{t=0}^{-T} w(t)P_t[c|x_i] \tag{24.5}$$

where:

x_i = Vector of attributes for customer i applicable at the prediction period.
$P[c|x_i]$ = Predicted probability customer i churns during the prediction period, given x_i.
$P_t[c|x_i]$ = Predicted probability customer i churns using the model estimated for the tth window, $t = 0, \ldots, -T$, where $t = 0$ signifies the most recent window, and $t = -T$ is the oldest model.
$w(t)$ = Set of weights.

and:

$$w(t) = \frac{e^{-\sigma_{t-1}}}{Z} \tag{24.6}$$

where:

σ_{t-1} = Mean square error using the model $t-1$ to predict for model 0's dataset.
Z = Normalization constant so that the w's sum to 1.

The prediction of whether the customer will churn in the current period is a weighted average of predictive models estimated from previous data windows, where the weights are proportional to a model's ability to predict the most recent data window. One would expect the weights to decline as we move back in time, but the weights are determined empirically by σ. Note this is very similar to the multi-classifier class-combiner approach of Wei and Chiu (2002) discussed above. The windows are the sub-samples, and Equations 24.5 and 24.6 describe how the different predictions are combined.

The authors compare three models: (1) the proposed "Combined" model based on three windows, (2) a single model using all the data from four periods before the prediction period, and (3) a single model based on one period before the prediction period. Model 2 is similar to a time series model since customers may churn in each period. Model 3 is similar to the Single Future Period approach since there is only one period when the customer might churn. The models are based on between 67,278 and 72,431 customers per window, and 71 potential predictors (customer credit classifications, location, use of various services, monthly service charges, monthly usage rate of various services, monthly number of dropped calls, and monthly number of customer service calls, etc.) were derived to predict churn for a 2-month period during which churn was roughly 6%. All the estimated models were neural nets with 71 inputs, one hidden layer, and 10 hidden nodes.

The results were that the Combined model and Model 2 predicted equally well, and together, both performed better than Model 3. The authors advocate their Combined model over Model 2 because the Combined model has smaller data requirements per model and is easier to estimate and utilize. This of course depends on the model used for Model 2. It isn't clear that a Combined model using neural nets for each model would be easier, or superior, to a hazard model. In any case, Yan et al.'s approach is very interesting and shows the potential of combining models into a single prediction, a theme that runs throughout the machine learning literature (Chapter 19).

Bonfrer et al. (2007) propose a Brownian motion model for predicting customer churn and apply it to a telecom company. The model assumes that weekly usage rate for customer $i(\mathrm{x_{it}})$ follows a first-order Markov process. Future usage can be predicted from (1) usage in the previous period, (2) a drift parameter μ_i related to the mean of the process, and (3) a volatility parameter σ_i related to the standard deviation of the process. Customers are assumed to churn if they hit zero consumption. If weekly usage follows a normal distribution, and $\mathrm{x_{it}} > 0$, then the time until the customer churns follows an inverse-Gaussian distribution, and the cumulative distribution, or the probability the customer churns within time τ of the current time t can be written as:

$$F(\tau|x_{it}, \mu_i, \sigma_i) = 1 - \Phi\left(\frac{\mu_i t + x_{it}}{\sigma_i \sqrt{t}}\right) + \exp\left(\frac{-2x_{it}\mu_i}{\sigma_i^2}\right)\Phi\left(\frac{\mu_i t - x_{it}}{\sigma_i \sqrt{t}}\right) \tag{24.7}$$

where Φ is the cumulative standard normal distribution. The probability of churning is increasing in the drift parameter (μ) and decreasing in recent consumption (x). A customer with negative drift and low recent consumption is a churn candidate. Higher volatility generally signals trouble for a customer with positive drift, but can actually lessen the chances of churn if drift is negative. These results make sense.

Since the model only has two parameters – drift and volatility – it can be estimated at the individual customer level. This makes the model easy to implement and avoids heterogeneity assumptions. The authors test their model and find it achieves roughly 2 to 1 top-decile lift in holdout data, similar to the predictive accuracy found in the churn modeling tournament discussed in Sect. 24.3.1. Overall, the simplicity and interpretability of the Brownian motion model makes it a promising method for future application and development.

24.4 Managerial Approaches to Reducing Churn

24.4.1 Overview

Approaches to reducing churn can be first categorized as targeted or untargeted. Untargeted approaches try to increase customer satisfaction or increase customer-switching costs by improving the product, advertising, or using loyalty programs. Targeted approaches, in contrast, identify customers most likely to churn and attempt to "rescue that customer." The key difference between targeted and untargeted approaches is that the targeted approaches identify potential churners and take action, whereas the untargeted approaches do not single out potential churners.

There are two types of targeted approaches: reactive and proactive. Reactive approaches wait for the customer to identify him/herself as a likely churner, usually when the customer calls to cancel service. A proactive approach identifies, in advance, customers most likely to churn, diagnoses the reason for the potential churn, and targets an appropriate action or incentive to induce the customer to stay.

Reactive churn management programs have the advantage of perfect prediction, or at least near-perfect. The customer has called and is about to cancel service, and then corrective action is taken. Because of the perfect prediction, the company can afford a significant incentive to keep the customer. However, at this point a strong incentive is *necessary*, and the company is basically bribing the customer to stay. This costs a lot in the short term and also may "train" the customer to call whenever he or she has an attractive competitive offer, expecting the company to match or exceed that offer.

Proactive churn management programs use predictive models to identify would-be churners, and hence have imperfect predictive accuracy – the accuracy depends on the quality of the churn model. Even with a 5 to 1 lift ratio, if the overall churn rate is 2%, only 10% of customers in the high churn segment are real churners. As a result, the company can't spend as much money per incentive as in a reactive program, since some of it may be wasted. On the other hand, the company may not need to spend as much, because things haven't gotten so bad yet that the customer is walking out the door.

Another potential concern for proactive churn management programs is that they might stimulate non-would-be churners to contemplate churning. Perhaps the customer is identified to be at risk for churning based on a predictive model, but really wasn't going to churn. However, the proactive contact and accompanying offer might stimulate "need recognition" (Engel et al. 1995, Chapter 5) and set in motion the customer's decision process for deciding whether to churn. Another way of looking at it is that the customer had a "latent need" to churn that was identified by the predictive model, but it took the proactive contact to enable the customer to recognize that need. This idea is advanced by Berson et al. (2000, pp. 282–295), who describe a UK telecom company case where indeed, the apparent result of a proactive program was to reduce churn among customers who responded to an offer designed to reduce churn, but increase churn among those who did not respond.[8]

---

[8] The authors are indebted to Professor Charlotte Mason of University of North Carolina for bringing this example to their attention.

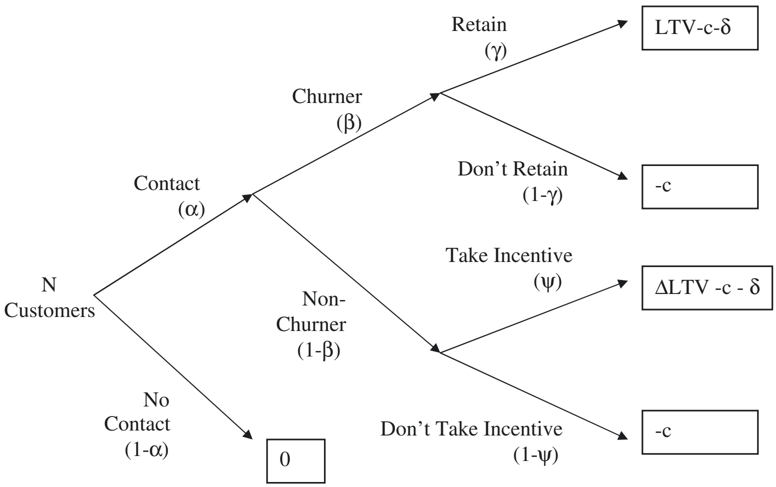

Fig. 24.6 Profitability framework for proactive targeted churn management program.

24.4.2 A Framework for Proactive Churn Management

The above discussion identifies some of the trade-offs inherent in a proactive churn management program, e.g., the predictive accuracy of the model, the effectiveness of the offer, and the potential wastage of offers made to customers who weren't going to churn. These issues can be quantified using the model portrayed in Fig. 24.6, which uses the following quantities:

N = Total number of customers.
α = The probability a customer is contacted as part of the churn management program. For example, if the program is to contact all customers in the top decile identified by a churn predictive model, then $\alpha = 0.10$.
β = The probability the customer is a churner, given the customer is contacted. For example, if all customers in the top decile are contacted and 15% of them are churners, then $\beta = 0.15$.
γ = The probability the customer is rescued, given he or she is a churner. This is the effectiveness of the incentive. For example, if a would-be churner has a 50% chance of being rescued by the offer, $\gamma = 0.50$.
ψ = The probability a non-churner takes the incentive. That is, a customer who isn't going to churn may take the incentive, e.g., if it is a price reduction or a free gift, ψ could be quite large, possibly 100%.
Δ = Percentage increase in lifetime value among non-churners who take the incentive. Although a non-churner has not been prevented from churning, having received an incentive could increase their retention probability or their purchase level of the company's services. This could be interpreted as a "delight the customer" effect (Rust and Oliver 2000; Hatch 2002).

The free offer is an unexpected surprise, one that enhances the customer's satisfaction with the company.

c = Contact cost, i.e., the cost of contacting a customer with a churn management offer. This might be \$0.50 if contact is made by mail.

δ = Cost of the incentive. If for example the incentive is a free cell phone, δ would equal the out-of-pocket cost of the cell phone.

LTV = Lifetime value of the customer. This differs of course by customer.

Given these definitions, one can calculate the total profit for a proactive churn management program by summing the appropriate probabilities times payoffs in Fig. 24.6. The result is:

$$\begin{aligned}\Pi &= N\{\alpha\beta\gamma(LVC - c - \delta) + \alpha\beta(1-\gamma)(-c) + \alpha(1-\beta)\psi(\Delta LVC - c - \delta) \\ &\quad + \alpha(1-\beta)(1-\psi)(-c)\} \\ &= N\alpha\{(\beta\gamma + (1-\beta)\psi\Delta)LVC - \delta(\beta\gamma + (1-\beta)\psi) - c\} \qquad (24.8)\end{aligned}$$

Equation 24.8 has a simple interpretation. The $\beta\gamma$ and $(1-\beta)\psi\Delta$ terms within the {} brackets represent the incremental profits of the program. The $\beta\gamma$ term is due to the rescued churners; the $(1-\beta)\psi\Delta$ term is due to delighted non-churners. The $\delta(\beta\gamma + (1-\beta)\psi)$ and c terms represent the incremental costs of the program. The $\delta(\beta\gamma + (1-\beta)\psi)$ term represents incentive costs, first among churners, then among non-churners. The c term represents contact costs, which are incurred among all α customers who are contacted.[9]

Equation 24.8 shows how key factors come together to determine the profitability of a churn management campaign. Note for example that if a customer's lifetime value is small, Equation 24.8 can easily turn out to be negative. Similarly, if the rescue probability (γ) is small and we can't assume any delight effect (Δ), the campaign may not be worth it.

The total investment in the program is $\alpha(\delta(\beta\gamma + (1-\beta)\psi)) - c)$, the third and fourth terms in Equation 24.8. Equation 24.8 divided by investment therefore is ROI for the program. One can then set an ROI requirement, say ROI $> \rho$ and solve for the incentive value δ. This yields the maximum the firm could spend on an incentive, namely:

$$\delta < \frac{(\beta\gamma + (1-\beta)\psi\Delta)LTV - c(1+\rho)}{(1+\rho)(\beta\gamma + \psi(1-\beta))} \qquad (24.9)$$

Figure 24.7 graphs maximum allowable incentive (δ) as a function of predictive accuracy (β) for two different values of incentive effectiveness (γ), assuming $c = \$0.50, LTV = \$2,000$, required ROI $= \rho = 15\%$, non-churner

[9] Note we are assuming that the only cost among non-churners who do not accept the offer is the contact cost. If as in the UK telecom case described by Berson et al. (2000), churn rates actually increase among this group, the cost will be higher than the contact cost. Importantly, Fig. 24.6 provides a framework that with some extension could be used to analyze the impact of this possibility.

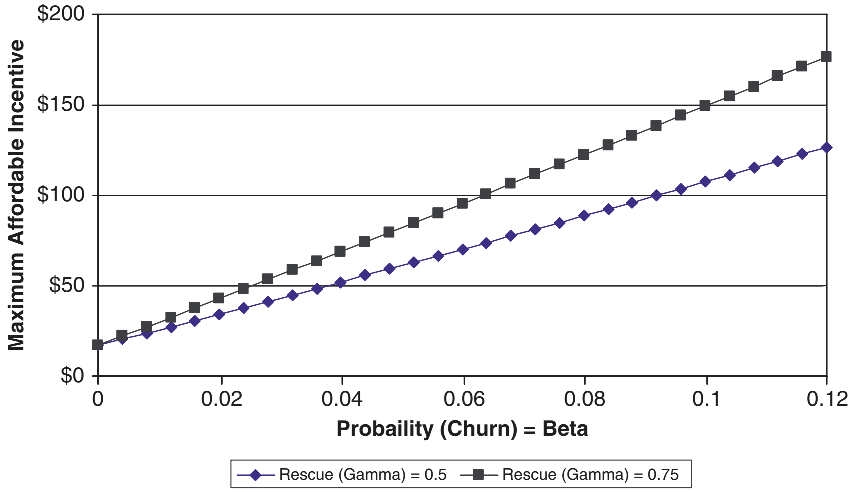

Fig. 24.7 Maximum possible incentive as function of churn model accuracy (β) and retention (Rescue) effectiveness (γ).
Calculations based on Equation 24.9 with $c = \$0.50$, $LTV = \$2,000$, $\rho = 0.15$, $\Delta = 0.01$, and $\psi = 100\%$.

acceptance rate $\psi = 100\%$, and delight effect $\Delta = 1\%$. The figure shows that given these parameters, one can spend more money on incentives to the extent that the churn model is more accurate and that would-be churners respond to the incentive.

For example, if, 5% of the customers in the targeted group were churners ($\beta = 0.05$), and the rescue rate from the program is 50% ($\gamma = 0.50$), the company could afford to spend \$61.03 per customer in a churn management program. Even though the lifetime value of the customer is \$2,000, the company can only afford to spend much less than that per customer, because of low predictive accuracy and because only half the churners will be swayed by the incentive. Of course, the rescue parameter γ is really a function of δ, and one could incorporate this explicitly as well.

Interestingly, the positive relationship between predictive accuracy and profit in Fig. 24.7 does not hold in general. For example, taking the derivative of Equation 24.8 with respect to predictive accuracy β yields $(\gamma - \psi\Delta) > 0$ as a necessary condition for increased accuracy to improve profits. If many non-churners take the incentive, and doing so delights them, we are better off targeting non-churners with the incentive, i.e., having poor predictive accuracy. This highlights the importance of the delight phenomenon and the need to understand its magnitude in a churn management context.

Neslin et al. (2006a) show how the framework also allows one to calculate the financial value of predictive accuracy. For simplicity and to be conservative, they assume $\psi = 1$ (i.e., all contacted customers accept the incentive) and $\Delta = 0$ (there is no delight effect). Then, they express predictive accuracy β as a function of lift (Chapter 10) as follows: Let β_0 be the base churn rate in the data, and λ be the lift among the customers contacted by the churn management program. Then $\beta = \lambda\beta_0$. Substituting this into Equation 24.8 and re-arranging terms, they obtain:

$$\Pi = N\alpha\{[\gamma LTV + \delta(1-\gamma)]\beta_0\lambda - \delta - c\} \tag{24.10}$$

The terms in front of the lift parameter (λ) are the slope of profits with respect to predictive accuracy. That is,

$$GAIN = N\alpha\{[\gamma LTV + \delta(1-\gamma)]\beta_0\} \tag{24.11}$$

represents the impact on profits of increasing lift by one unit. For example, consider a campaign with the following parameters:

$N = 5{,}000{,}000$ customers
$\alpha = 0.10$ (10% of the customer base is to be contacted)
$LTV = \$500$ (lifetime value of rescued customer)
$\gamma = 0.10$ (10% of contacted would-be churners stay with the company)
$\delta = \$50$ (incentive cost)
$\beta_0 = 0.018$ (base monthly churn rate)

Then, the gain in the profitability of a single churn management campaign, per *half* unit increase in lift (e.g., from 2.5 to 3.0; note 0.5 is the standard deviation in performance among the entries in Neslin et al.'s churn modeling tournament), is $5{,}000{,}000 \times 0.10 \times \{[0.10 \times \$500 + \$50 \times (1 - 0.10)] \times 0.018 = \$427{,}500$. As we see, the framework can illustrate the profit impact of better churn prediction, and in this example, we see that the impact can be substantial.

The framework integrates key concepts of churn predictive accuracy, incentive effectiveness, delight effects, lifetime value, and costs into an expression for the profitability of a proactive churn management program. The model provides diagnostics regarding the maximum a company could spend on an incentive, which increases as:

- Predictive accuracy (β) increases (as long as $(\gamma - \psi\Delta) > 0$; see above).
- Incentive effectiveness (γ) increases.
- Lifetime value of customers (LTV) increases.
- The impact of the incentive on churners (the delight factor Δ) increases.
- Contact costs (c) decrease.
- Required ROI from the program decreases.

In addition, the framework can be used to calculate the benefits of improved churn prediction. This is important for a firm that needs to decide whether to invest in expanding its churn analytics department or purchasing additional data.

An important limitation of the framework is that it is a single-campaign model. It does not explicitly answer questions such as how soon before the customer is predicted to churn should the customer be contacted, or how many times to contact a customer.

24.4.3 Implementing a Proactive Churn Management Program

Implementing a proactive churn management program consists of four steps: (1) identifying potential churners, (2) understanding why they might churn, (3) designing an appropriate contact/offer strategy for the churners, and (4) monitoring and evaluating results.

The first two steps can be accomplished by a predictive model. The predictive model identifies the customers most likely to churn, and provides diagnostics for why they might churn. The third step requires creativity to consider the various incentives that one might implement. Inevitably this will entail costs, and the framework in Fig. 24.6 can help identify the upper bound of how much one can spend on incentives. As Equation 24.9 shows, one will spend more for higher lifetime value customers. Another consideration here is that the incentive should be such that it does not stimulate "need recognition" among those who do not accept the offer (Berson et al. 2000).

Another important challenge is to decide how many customers to contact, i.e., what should be the value of α? This could be decided in two ways. First would be to use a budget. If \$X are allocated to the campaign and we anticipate the values for rescue (γ), predictive ability (β), contact cost (c), and non-churner acceptance (ψ), total expenditures are $N\alpha\{(\beta\delta+(1-\beta)\psi)+c\}$. One can adjust α so that the total expenditure is within budget. Second would be to quantify the relationship between α and β: β is the percentage of contacted customers who are would-be churners, and will naturally decrease as a function of the number of customers contacted (since we are using a predictive model to identify churners and ordering customers in terms of their likelihood of churning). One could then calculate profits of the campaign for different combinations of α and β and choose the combination that maximizes profits.

Evaluating a churn management program is best accomplished using a field test. For example, estimating the rescue rate γ is challenging because rescued churners cannot be identified with certainty. Following is how a field test based on the framework in Fig. 24.6 could be used for this purpose. Let

A = Group that receives the churn offer (treatment group).
B = Group that does not receive the offer (control group).
N = Number of customers assigned to each group (assumed equal).

R_g = Churn rate in group g during the churn management campaign (either A or B).
O = Number of offers accepted in the treatment group.
LTV_g = Average lifetime value per customer in group g calculated a reasonable amount of time after the campaign is finished.

The impact of the program would be estimated by $N(LTV_A - LTV_B)$. To estimate the rescue rate, we express the churn rates in each group as:

$$R_A = \beta(1-\gamma) \tag{24.12a}$$

$$\mathrm{R_B} = \beta \tag{24.12b}$$

Using β from the control group, we can solve Equation 24.12a for the rescue rate γ. Also, the total number of offers accepted in the treatment group equals the number of would-be churners who accept plus the number of non-would-be churners who accept. This yields:

$$O = N\{\beta\gamma + (1-\beta)\psi\} \tag{24.12c}$$

Since we know N, β and γ, we can solve for ψ.

This illustrates how a field test can be used to provide an overall evaluation of a churn management campaign. In addition, the test provides estimates for key parameters such as the rescue rate to be used for future planning. Providing an overall evaluation is crucial; one cannot take for granted that the churn management program will work. As discussed earlier regarding the UK telecom case (Berson et al. 2000), it is even possible for $N(LTV_A - LTV_B)$ to be negative!

Another issue in implementing a proactive churn program is how early to intervene.[10] This may require a trade-off between churn model accuracy (β) and rescue effectiveness (γ). Earlier intervention may rescue more would-be churners (higher γ), but predictive accuracy will be lower (see discussion of Lu 2001 in Sect. 24.3.2). Intervening later might benefit from higher model accuracy, but a rescue offer may be less effective or have to be more costly to achieve a desired rescue level (γ). The point is, the question is not only *what* the churn management offer should be, but *when* it should be delivered.

A final issue is the single campaign versus multiple contact approach. For example, consider the customer who receives an offer in March and does not churn, but appears in the top decile risk group in May. Should that customer receive a second incentive in May? There may be a chronically high churn risk group for the company, and under this policy, they will be receiving offers every few months. This could erode their profitability, or at least their loyalty, as we train them that they can always expect new offers and deals from the company. One way investigate this would be to include previous churn rescue offers in the predictive churn model. This would

[10] The authors benefited from discussions with Wagner Kamakura and Carl Mela (both at Duke University) on this issue.

tell managers whether previous offers enhance or diminish long-term loyalty. Another approach would be to formulate and implement an optimal contact model (Chapter 28). In this model, the current period decision to offer an incentive to a would-be churner would take into account the future impact of that offer.

24.5 Future Research

We have discussed the churn problem, its causes, how predictive models can identify churners and diagnose the reasons they churn, and how companies can manage churn. There are many exciting findings and methods developed to address this topic, but there are several avenues for future research on churn, as follows:

- *Predicting churn*: What are the key variables? Which models work best? How high a lift is feasible and how does it vary by product category?
- *Churn programs*: We need field tests to show that churn management programs can work.
- *Profitability framework*: We need parameter estimates of the framework in Fig. 24.6 so that we can understand the factors that feed into churn program profitability, and how they vary by program and product category.
- *Dynamic optimization*: We need a optimal contact model to decide which customers should be contacted when, on an ongoing basis, to minimize churn. This would take into account the impact of current incentives on future responsiveness, lifetime value, and churn likelihood. This would be a substantial contribution to the literature and to the practice of effective churn management.