A Lagrangian Approach to Customer Relationship Management: Variable-Rate Pricing Strategy

9 August 2006

Joseph George Caldwell, PhD
Lagrangian Solutions
503 Chastine Drive
Spartanburg, SC 29301
Tel. (001)(864)439-2772
E-mail: jcaldwell9@yahoo.com
Website: http://www.foundationwebsite.org

Contents

1. Introduction

This paper describes an optimization-based approach to customer relationship management (CRM), and illustrates the methodology with an application to the problem of determining a variable-rate pricing model for banking loan products. The methodology employed to solve the problem is Everett's Generalized Lagrange Multiplier (GLM) method. This optimization methodology is ideally suited to determining the solution to CRM problems for several reasons: it is the most appropriate optimization methodology for strategy determination or resource allocation problems involving a very large number of independent resource-allocation opportunities, or "targets"; it can easily handle difficult analytical conditions, including objective ("payoff") functions that are nonconvex, nonlinear, and discontinuous; it is ideally suited to solution on a digital computer – solutions are available very quickly.

This paper consists of five additional sections. Section 2, Description of the Problem, describes the general problem to be solved – customer relationship management – with specific reference to the problem of determining an optimal pricing policy for bank loans. Section 3, Mathematical Formulation of the Problem, formulates the problem in mathematical terms. Section 4, Solution of the Problem, describes the GLM methodology for solving the mathematical problem. Section 5 describes the design of a computer-program model for implementing the methodology in the variable-rate pricing example.

2. Description of the Problem

"Customer relationship management" refers to the collection of policies and procedures that an organization adopts and applies in dealing with its customers. The term is used mainly by organizations having very large numbers of customers, such as a bank, credit-card company, insurance company or financial advisory service. The number of customers or potential customers in these applications is very large, often in the millions. Any organization dealing with such large numbers of customers cannot deal individually with each customer or potential customer, and needs automated procedures for conducting operations such as mass mailings to solicit business, extend offers to existing customers, or modify the terms of accounts. Because of the very large numbers involved, even small improvements in the policies or procedures used to deal with customers can result in significant increases in profit. For this reason, large banks and other organizations having large customer bases invest heavily in scientific research, such as in statistical models of customer behavior, statistical decision functions and optimization models.

The past couple of decades has seen the application of a variety of technical methodologies such as statistical models and artificial-intelligence models to describe customer behavior (or potential-customer behavior) and optimization models to determine good strategies for dealing with customers. These methodologies the use of logistic-regression models to estimate the probability that a customer or non-customer will respond favorably to a mass mailing offer, and the use of neural-networks to estimates the likelihood of loan default on the basis of observed customer characteristics such as credit score, income, age, or marital status.

At present, a massive amount of effort is invested by customer-relationship-management organizations in the conduct of statistical analysis of customer data to describe customer

behavior as a function of customer and product / service characteristics. This includes not only efforts to develop sophisticated analytical models based on statistical experimental-design methodology, but also the use of "exploratory data analysis" or "data mining" to detect apparent relationships and thereby suggest hypotheses that may be investigated by methodologies oriented toward identification of causal relationships. In contrast, a much smaller amount of effort is invested in the development of optimization or control models to achieve specified objectives, such as maximization of income or profitability subject to constraints such as funds availability or return on investment; or to support risk management. This is partly due to the fact that the formulation and solution of optimization problems is often more difficult than the formulation and solution of statistical estimation or statistical decision problems, and partly due to the fact that the "statistical analysis industry" (represented by SAS, SPSS, CART, etc.) is much larger than the "optimization" industry, and the availability of appropriately skilled personnel to apply statistical-analysis software, which is highly automated, is much greater than the availability of appropriately skilled personnel to develop optimization models.

The problem to be addressed in this article is the problem of determining an appropriate action to be taken with respect to each customer (or non-customer) so as to maximize (or minimize) a specified quantity, subject to specified constraints. For example, it may be desired to conduct a mass mailing to non-customers to invite them to accept a bank credit card, subject to a fixed budget (number of letters to be mailed). A limited amount of information is known about each non-customer, such as credit score, marital status, location (e.g., urban/rural status), and number of credit cards already using. A "traditional" way of dealing with this problem is to use historical data (from earlier mailings) to estimate the probability (likelihood) that a potential mail recipient will accept the offer, using a logistic regression model. Then, all of the potential mail recipients are ranked in decreasing order of the estimated probability (i.e., from the most likely to the least likely), and as many addressees as desired are selected from the top of the list. The terms of the offer (interest rate, fees, credit limit) are typically indicated in the offer letter, and they are usually the same for all recipients (in a particular mailing).

This approach maximizes the number of favorable responses to the mailing. Since it maximizes a quantity, it is an optimization problem, but it is a rather trivial one. The bank's fundamental objective is not to maximize the number of people responding (favorably) to the offer, but to maximize its income, or return on investment, or stockholder equity. It would be considerably more relevant to the bank's ultimate objective to select mail recipients in such a way as to maximize one of these quantities, rather than to maximize the favorable-response rate. Using the traditional approach of sending to those who are most likely to accept the offer, it could happen that the least creditworthy of the target population responds, so that the bank obtains lots of new customers, but these customers will not generate much income for the bank.

In order to target a mass mailing in such a fashion as to maximize bank income or profitability, it is necessary to develop a mathematical model that expresses these quantities as functions of customer characteristics and offer characteristics, and to tailor the mailing and offer to customers in a way to maximize the quantity of interest. In mathematical terminology, the quantity to be maximized, such as bank income or return on investment, is called the *objective function* or *payoff function*.

In general, the problem of customer relationship management involves allocating resources to a target population (customers or potential customers) in such a way as to maximize a specified objective function, subject to various constraints.

This article will illustrate the optimization-based approach to customer relationship management by means of an example in which it is desired to determine a pricing strategy for bank loan products (e.g., credit cards, installment loans, mortgages) that maximizes stockholder value added, subject to a constraint on the return on the capital allocated as a provision against loss. The "pricing" of the loan is specified by two quantities: (1) the decision to make the loan; and (2) the interest rate to be charged. The pricing strategy is determined by specifying the loan decision (i.e., to extend the loan or not extend the loan) and the interest rate for each customer. Since the interest rate may vary for each customer (or class of customers), this pricing policy is referred to as a "variable-rate pricing" strategy. (This example does not address the problem of determining a strategy for setting fees, which may be substantial for some loan products, such as credit cards – it deals only with determining an optimal strategy for maximizing interest income, not fee income.)

The motivation for adopting a variable-rate pricing (VRP) strategy is that bank profits may be increased by making credit decisions and interest rates sensitive to customer characteristics (i.e., to the risk). Credit may be extended to riskier customers at higher rates, which offset the increased losses associated with those customers, or it may be extended to premium customers at lower rates, to retain them or attract more of them. With this approach, the customer base may be increased, and higher-risk customers will pay higher rates, with the end result that total bank income may be increased. In order to produce higher income, however, it is necessary to determine rules for the credit decision and the interest rate in a satisfactory way. In order to maximize bank income, it is necessary to determine an optimal rule for making the credit decision and for setting the interest rate.

Under the so-called Basel II Accord (the *International Convergence of Capital Measurement and Capital Standards - A Revised Framework*) of the Bank for International Settlements, international banks are to relate the size of their reserve capital (provision for loan defaults) to risk. Large international banks take risk into account in pricing their products, and the increased risk sensitivity is reflected in the willingness of these banks to lend to higher-risk borrowers at higher prices. An advantage of variable-rate pricing to higher-risk customers is that some that were previously excluded from consideration have a chance to establish a good credit history. (See Wikipedia entry, *Basel II* at http://en.wikipedia.org/wiki/Basel_II ; also, *International Convergence of Capital Measurement and Capital Standards, A Revised Framework*, Updated November 2005 at http://www.bis.org/publ/bcbs118.pdf .

3. Mathematical Formulation of the Problem

We will illustrate the problem of optimization-based customer-relationship management by means of a specific example, viz., the determination of an optimal variable-rate pricing strategy for loans. In practice, a different pricing stragegy would be developed for each different type of credit product, such as credit cards, installment loans, and secured loans (e.g., home mortgages).

Let us denote by the symbol x a pricing stragegy for a single customer, and by X the pricing strategy for all customers. The pricing strategy consists of a credit decision (i.e., whether to extend the loan to the customer) and an interest rate for the loan if it is extended. If x_i denotes the pricing stragegy for customer i, then $X = (x_1, x_2,…,x_n)$, where n denotes the total number of customers. Let $h_i(x_i)$ denote the payoff resulting from using strategy x_i with the i-th customer.

Let H denote the total payoff from all customers, and H(X) denote the total payoff (e.g., net income, or net income after taxes) if strategy X is used. Then

$$H(X) = \Sigma_i(h_i(x_i)),$$

where Σ_i denotes summartion over the index i.

In the absence of any constraints, the objective of the bank would be to determine the stratgegy X so as to maximize the total payoff, H(X). But, in this resource-constrained, risk-sensitive world, there are always constraints. First, the amount of funds available to any particular business application are always constrained in some fashion. Second, a bank must pay in some way for the funds that it loans, either by means of interest payments to depositors, or a return to stockholders, or as a "discount rate" to to the central (reserve) bank from which it borrows funds. What it pays for this capital (i.e., what rate) is the "cost of capital." For the bank to continue to make loans of a certain type, it is necessary that the return on the loans that it makes be greater than the cost of the funds that it uses for the loans. To avoid the risk of bank default, banking regulations require that a bank must set aside a "capital allocation," or "capital requirement," for each type of loan (credit) product, such that the risk of losing the entire capital allocation because of loan defaults is very small. In recognition of these conditions, we shall address the problem of determining X so as to maximize H(X) subject to a constraint on the return on investment of the capital allocation (the total capital set aside as provision for bad loans).

The amount of the capital allocation varies by credit product type, depending on the level of risk associated with that type. The size of the capital allocation is determined by a risk analysis, such as a risk-adjusted performance measurement analysis.

Let us denote the capital allocation as CA, and the cost of capital as CC. Let NI denote the net income for the loans, after deduction of all expenses except the charge for the capital allocation. In this article, we shall ignore taxes, and focus attention on pretax income. (Whether attention focuses on pretax income or posttax income has no significant effect on methodology for determining the pricing strategy.) We define two other terms: the return on capital, or ROC, which is defined as

$$ROC = NI/CA$$

and the shareholder value added, SVA, which is defined as

$$SVA = NI - CC\ CA .$$

The return on assets, ROA, is defined as

$$ROA = NI/A$$

where A, assets, is the sum of the loan account balances. The ratio A/CA is called the financial leverage, FA, so that the relationship of ROC to ROA is ROC = FL ROA.

The four quantitites, net income (NI), return on capital (ROC), return on assets (ROA) and stockholder value-added (SVA) will be referred to as performance measures. Once the optimal pricing strategy is determined, the value of these performance measures will be determined for

the optimal pricing strategy, the bank's current pricing strategy, and any other user-specified pricing strategy of interest.

Net income is determined by a variety of income and expense components. In developing an optimization model to maximize net income, it is necessary to know the relationship of net income to customer / account characteristics and the pricing policy. That is, it is necessary to estimate net income for a customer as a function of the customer's characteristics (e.g., credit score) and the loan terms (interest rate, fee provisions). To do this it is necessary to estimate the relationship of demand for loans and risk of default to customer characteristics (such as credit score) and loan product characteristics (such as interest rate). In this example, we shall be concerned only with determining a policy for determining the credit decision and the interest rate, and not with other loan terms, such as fee schedules. (If the methodology were to be extended to include determination of optimal fee schedules, then demand and risk models would have to be available for fee income.) The following formula expresses net income, NI, as a function of various components of income and expense:

NI = Net Interest Income (interest income – cost of funds – provision for credit loss)

   + Fee Income – Direct Operating Expenses (cost of origination, cost of servicing)

   - Indirect Operating Expenses (indirect operating expenses and overhead) .

In determining an optimal policy for loan pricing (credit decision and interest rate), we are most concerned with the components of NI that are affected by loan pricing. These include net interest income, cost of servicing and fee income. The other components of NI (indirect costs) are necessary to calculate the value of NI correctly, but they are not affected by the pricing. (They do affect the optimal pricing strategy, since they affect the value of ROC, and pricing policy is determined by a constraint on ROC.)

The objective of extending loans is to maximize net income, subject to whatever constraints are imposed on the operation. The three principal constraints to loan operations are (1) the total amount of capital available is limited to a certain amount; (2) the (total) return on capital must exceed a particular level; and (3) the risk of loss of the capital allocation must be very small. In this illustrative example, we shall assume that the size of the capital allocation has already been determined in a fashion so that the risk of loss of the capital is very low, and so no constraint will be included in this imposed to reflect that concern (i.e., the probability of loss exceeding the capital allocation). Also, it will be assumed that, as long as the return on capital meets or exceeds the prescribed level, there is not constraint on the amount of capital available, i.e., capital is an "infinite resource." It is also assumed that there is no requirement to utilize all of the capital allocated to the operation, i.e., there is no penalty for failure to make loans up to the maximum level permitted by the capital allocation.

The problem of variable-rate pricing may be focused on determining the pricing strategy for a single credit product or for a group of credit products. The pricing strategy for a particular loan product will differ in these two cases, since in the first case the constraint on ROC must be met by that particular product, while in the second case the constraint on ROC must be met by the entire group of products (in which case the ROC may be less on some products and higher on others, such that the ROC for the total group of products meets the constraint).

In view of these remarks, the problem to be addressed here is to maximize net income subject to the single constraint that the return on allocated capital exceed a specified level. If NI(X)

denotes the net income generated by pricing strategy X, and if ROC(X) denotes the return on capital under pricing strategy X, then, in mathematical terminology, the problem is to determine pricing strategy X so as to maximize NI(X), subject to the constraint that ROC(X) >= CC (where CC denotes the cost of capital). Or, more compactly, determine the pricing strategy, X*, such that

$$NI(X^*) = \max_X NI(X)$$

subject to

$$ROC(X^*) >= CC ,$$

where $\max_X NI(X)$ denotes the maximum value of NI(X) with respect to X.

The constraint ROC(X*) >= CC may be written as

$$ROC(X^*) = NI(X^*) / CA(X^*) >= CC ,$$

or as

$$SVA(X^*) = NI(X^*) - CC\ CA(X^*) >= 0 .$$

4. Solution of the Problem

This section will present a solution to the optimization problem defined in the preceding section, for the situation in which the loan pricing may be done independently for each customer account. This situation applies if the income ("payoff") for a particular account depends only on the pricing of that account and not on the pricing of other accounts. In this situation, a very powerful optimization methodology is available to determine the solution to the problem (i.e., to determine the optimal pricing strategy). That methodology is the Generalized Lagrange Multiplier (GLM) method developed by Hugh Everett III. The GLM methodology is a very powerful technique for solving "cell-separable" constrained-optimization problems in which the quantity (the *objective function*, or *payoff function*) to be maximized (or minimized) may be represented as a sum of a large number of independent terms (such as income from customer accounts). It is much more flexible than most constrained-optimization methodologies, since it can handle objective (payoff) functions that are nonlinear, nonconvex, and discontinuous.

Under the GLM approach, the solution to the constrained optimization problem posed above, viz., determine strategy X* such that

$$NI(X^*) = \max_X(NI(X))$$

subject to

$$NI(X^*) - CC\ CA(X^*) >= 0 ,$$

is the same as the solution to the following *unconstrained* optimization problem:

Determine strategy X* such that

$$\text{NI}(X^*) - \lambda\,[\text{CC CA}(X^*) - \text{NI}(X^*)] = \max_X [\text{NI}(X) - \lambda\,\{\text{CC CA}(X) - \text{NI}(X)\}]$$

where λ denotes a lagrange multiplier (nonzero number), the value of which is determined so that the constraint $\text{NI}(X^*) - \text{CC CA}(X^*) >= 0$ is satisfied. The tremendous power of the GLM method is that it is much easier to solve the unconstrained extremization problem than the original constrained extremization problem.

It is customary to simplify the preceding expression by combining terms. The right-hand-side of the expression may be written as

$$\max_X [(1 - \lambda)\,\text{NI}(X) - \lambda\,\text{CC CA}(X)] \ .$$

For a specified value of λ, the value of X that maximizes this expression is the same as the value that maximizes

$$\text{NI}(X) - \lambda'\,\text{CA}(X)$$

where $\lambda' = \text{CC}\,\lambda\,/\,(1 - \lambda)$ is chosen so that the constraint $\text{NI}(X) >= \text{CC CA}(X)$ is satisfied. For simplicity, the constant λ' is relabelled as λ. With this reformulation, the problem is to determine strategy X^* such that

$$\text{NI}(X^*) - \lambda\,\text{CA}(X^*) = \max_X [\text{NI}(X) - \lambda\,\text{CA}(X)]$$

$$= \max_X [\Sigma_i(\text{NI}_i(x_i)) - \lambda\,\Sigma_i(\text{CA}_i(x_i))\,] \ ,$$

where λ is chosen so that the constraint $\text{NI}(X) >= \text{CC CA}(X)$ is satisfied.

In order to solve this problem, the essential step is to take into account the assumption that the income for the i-th account is a function of the pricing for that account only, and not of the pricing of other accounts. Under this assumption, the problem has a "cell-separable" payoff function, and my be solved by the GLM method. This assumption is reasonable for accounts owned by different customers, but it may not be reasonable for accounts associated with the same customer. For example, if a customer is denied a home mortgage loan, or such a high price is charged for this mortgage that he declines, he may decide to cancel all of his other accounts with the bank (e.g., his credit card accounts, or perhaps other mortgages). (In this case, the expected net income for the mortgage loan would not be zero, but negative, if the decision were made not to extend the loan.)

Under the assumption that the net income (the "payoff") for the i-th account is independent of the net income for other accounts, then the price strategy x_i for each account may be determined independently for every account. In this case, the expression $\max_X [\Sigma_i(\text{NI}_i(x_i)) - \lambda\,\Sigma_i(\text{CA}_i(x_i))]$ may be maximized by maximizing each term $\text{NI}_i(x_i) - \lambda\,\text{CA}_i(x_i)$ independently, i.e.,

$$\max_X [\Sigma_i(\text{NI}_i(x_i)) - \lambda\,\Sigma_i(\text{CA}_i(x_i))] = \Sigma_i(\max_{x_i} [\text{NI}_i(x_i) - \lambda\,\text{CA}_i(x_i)]) \ .$$

That is,

$$\text{NI}(X^*) - \lambda\,\text{CA}(X^*) = \Sigma_i(\max_{x_i} [\text{NI}_i(x_i) - \lambda\,\text{CA}_i(x_i)]) \ .$$

The optimization problem is hence solved by determining, *independently for each account*, the price strategy x_i that maximizes the expression $NI_i(x_i) – \lambda\, CA_i(x_i)$, which is called a *lagrangian function*. To do this, all that is required is to determine λ so that the constraint $NI(X) \geq CC$ $CA(X)$ is satisfied. This is done by using an interative numerical method (such as Newton's method) to find the minimum value of λ for which the constraint is satisfied. (There may be multiple values of λ for which the constraint is satisfied. The reason why we seek the minimum value of λ is that it is desired to maximize net income, not just find a solution that satisfies the constraint.)

The original problem of determining a pricing strategy that maximizes total net income subject to the constraint on return on capital allocation has now been reduced to the simpler problem of determining, independently for each account, a pricing strategy that maximizes the account lagrangian function. To solve this problem, it is necessary to know the relationship of net income (NI_i) and capital allocation (CA_i) to the account pricing and account characteristics. The value of CA_i is known – it is either zero if no loan is extended or a fixed percentage of the loan value (the percentage factor being the same for all accounts of a given loan type). It remains to specify the relationship of NI_i to account pricing and account characteristics.

The net income from a customer account depends on a number of factors, including the decision to extend the loan, the interest rate, the event of a default if the loan is extended, and the amount of the loss in the event of default. The decision to extend the loan and the interest rate are under the control of the bank, and will be made according to the strategy determined by the solution to the optimization problem described above (which strategy will take into account customer characteristics). The other two variables (the event of a default and the loss in the event of a default) are random variables. Hence the relationship of net income to pricing strategy and customer characteristics will be a stochastic (probabilistic, statistical) one, not a deterministic one. This relationship will typically be expressed as a table or formula that specifies the expected (mean, average) net income as a function of pricing strategy (loan decision, interest rate) and customer characteristics (e.g., credit score).

The customer (or potential customer) characteristics of interest in this application include all available data (known facts) about the customer that may affect loan income or expense. For a simple model, or for a potential customer, this information might be a single variable – the individual's credit score. Whatever customer (or potential customer) data are available, it is necessary to know the relationship of net income to those variables. These relationships would typically be expressed as statistical tables or statistical regression functions that specify the expected value of net income as a function of the available explanatory variables. These relationships would be derived from a statistical analysis of available data, or from expert judgment for situations in which no historical data were available (e.g., estimation of net income for an interest rate / credit score combination that had not been used before).

The pricing strategy, x_i, consists of two components – the credit decision and the interest rate. We shall represent the credit decision by a numerical variable, d_i, whose value is 0 if the decision is made not to extend the loan and 1 if the decision is made to extend the loan. We denote the interest rate as r_i. In terms of d_i and r_i, we shall denote the pricing strategy as $x_i = (d_i, r_i)$. Let us denote the collection of customer (or potential customer) variables as y_i (a vector, $y_i = (y_{i1}, y_{i2},\ldots,y_{in})$). For simplicity in this illustrative example, we shall assume that there is a single customer characteristic that is to be taken into account in determining pricing strategy, namely, the customer's credit score, which we shall denote as cs_i. That is, y_i in this example will have but a single component, $y_{i1} = cs_i$.

8

In the formulation of the optimization problem presented above, net income was represented as a deterministic variable, and the variables over which the account lagrangian was to be maximized were not explicitly shown. Representing net income as an expected value, the problem is to determine the pricing strategy $x_i = (d_i, r_i)$ so as to maximize the lagrangian function

$$E(NI_i(x_i,y_i)) - \lambda\, E(CA_i(x_i,y_i)) = E(NI_i(d_i,r_i,cs_i)) - \lambda\, E(CA_i(d_i,r_i,cs_i)) ,$$

where the symbol $E(.)$ represents the expectation (expected value) operator. (Note: The formula for determining net income and capital allocation might differ from customer to customer, and for this reason the index i is appended to each symbol, viz., NI_i and CA_i. In general, however, the functional form will usually be the same for all accounts of a particular loan type. It is conceivable, however, that different functions might be used if different variables are known for some customers and not others (e.g., the only variable known for potential customers might be credit score, but many additional variables, such as income and account age, would be known for customers, and so a different formula could used to calculate net income in these two cases.)

In this example, we shall assume that the capital allocation is the same rate for all customers of a given loan type, and not dependent on the individual customer's credit score or the interest rate to be charged. In this case, $CA_i(d_i,r_i,cs_i)$ is a deterministic quantity, and so $E(CA_i(d_i,r_i,cs_i)) = CA_i(d_i,r_i,cs_i)$. If $d_i = 0$ (i.e., the decision is made to not make the loan) then $CA_i = 0$. If $d_i = 1$, then $CA_i = CC$ times (amount of loan to the i-th customer).

We shall now present a formula for determining the value of the expected net income, $E(NI_i(x_i,y_i))$.

Note that if a decision is made to not make a loan, then it does not matter what the interest rate is – the value of the interest rate matters only if the loan is made. Hence, the two components of the loan pricing – the credit decision and the interest rate – may be considered to be independent. If a decision is made to not make the loan, i.e., $d_i = 0$, then the net income is zero (no matter what the interest rate). The lagrangian function is hence maximized, as a function of both d_i and r_i, by maximizing it as a function of r_i in the case $d_i = 1$ (the loan is made), comparing the resulting value to zero (corresponding to $d_i = 0$, the loan is not made), and then choosing the combination of (d_i, r_i) corresponding to the larger of these two quantities. For this reason, we shall present in the discussion that follows only the expression for expected net income conditional on making the loan (i.e., $d_i = 1$).

In the case of some loans (e.g., installment loans, mortgages), the amount of the loan is known – initially it is the full amount of the loan, and subsequently it is the balance owed. For credit cards, the amount of the loan is not known when the offer is made to the customer. All that is known is the maximum amount (credit limit) that will be extended, and the balance, which may fluctuate from day to day. For loans that are not yet made, the analysis will focus on the expected size of the loan to be made. For loans that are already made, the analysis will focus on the loan balance (current market value of the loan). This quantity will be denoted by VAL_i for the i-th customer.

The net income for an account equals the sum of all of the revenues associated with the account minus the sum of all expenses associated with the account. The account revenues include net interest income, fee income, insurance revenue, and capital earnings. The account expenses include provision for credit loss, capital cost, loan acquisition costs, insurance, and overhead. In order to be able to calculate ROC, *all* components of revenue and expense must

9

be included in the calculation, not just the direct ones, or the ones that are affected by the loan pricing. For simplicity of presentation, we shall combine all of the income other than net interest income (NII) into a single quantity, "other income," denoted by OI_i for the i-th customer, and all of the expenses that are not affected by the loan pricing into a single quantity, "other expenses," denoted by OE_i for the i-th customer. Each of these two quantities must be disaggregated into two components – one that is dependent on pricing and account characteristics, and the other that is independent of these variables.

As mentioned, net income is a random variable that depends on a number of stochastic variables, including the probability that the customer accepts the loan, the probability of loan default, and the expected size of the loss in the event of default. We shall denote the probability that the customer accepts the loan as $pacc_i$. We denote the probability of loan default as $pdef_i$. We shall represent the expected size of the loss as the expected proportion of the loan that is lost, $plos_i$. These quantities are estimated from historical data or from expert opinion (e.g., a panel of loan officers may be able to provide insight on demand that may not be available from historical data). They may be expressed as tables or as equations, and they may be as simple or as complex as desired (depending on the availability of historical data and analytical resources). In a simple application of this methodology, it may be, for example, that the probability that the customer accepts the loan is a function of the interest rate, that the probability of default is a function of the customer's credit score, and that the expected proportion of loss in event of default is also a function of the credit score.

What characteristics are taken into account in determining the models for demand (probability of acceptance) and default depends on what data are available about the customer or potential customer. The data elements that are available for potential customers or new customers may differ from those that are available for existing customers. For example, all that may be known for potential customers in a mass mailing may be the customer's credit score, whereas for existing customers the available customer / account characteristics may include many additional variables, such as annual income, age, marital status and occupation.

Using the preceding notation, a general expression for the expected net interest income from the i-th account, if the loan is offered at interest rate r_i and is accepted, is:

$$E(NII_i(d_i=1, r_i)) = VAL_i \ pacc(r_i) \ [(1 - pdef_i) \ r_i - pdef_i \ plos_i \ ]$$

where, to summarize the notation,

$E(.)$ denotes the expectation (expected value) operator

index i refers to the i-th customer

d_i = loan decision (1 corresponding to extending the loan, 0 to denying the loan)

r_i = interest rate

$NII_i(d_i, r_i)$ = net interest income (a function of d_i and r_i, and whatever other customer characteristics are included on the right-hand-side of the equation, such as credit score, cs_i)

VAL_i = loan balance

pacc(r_i) = probability that the customer accepts the loan (a function of r_i, and possibly of customer characteristics as well, such as credit score)

pdef$_i$ = probability of loan default (may be a function of customer characteristics, such as credit score)

plos$_i$ = expected proportion of loan lost in event of default (may be a function of customer characteristics, such as credit score) .

The expression for expected net income, E(NI$_i$), is expected net interest income plus other income minus other expenses:

$$E(NI_i(d_i=1, r_i)) = VAL_i \ pacc(r_i) [(1 - pdef_i) r_i - pdef_i \ plos_i] + OIi - OEi .$$

If the decision is made not to extend the loan (i.e., d_i=0), the expected net income is zero.

Once the functions pacc, pdef and plos have been determined (by analysis of historical data), it is possible to calculate the expected net income for each account. For a specified value of the lagrange multiplier, λ, the credit decision (value of d_i) and the interest rate, r_i, may be determined so as to maximize the value of the lagrangian function

$$E(NI_i(x_i,y_i)) - \lambda \ E(CA_i(x_i,y_i)) .$$

The solution to the constrained optimization problem is obtained by adjusting the value of λ so that the constraint on ROC is satisfied. For problems involving a constraint on a single resource (e.g., capital allocation, as in this example), algorithms for adjusting λ so as to satisfy the constraint are easy to develop, and they converge very quickly to a solution.

Note that under the preceding formulation of the problem of determining a variable-rate pricing strategy, a different credit decision and interest rate are determined for each customer, as a function of the customer's exact credit score. It may be that this strategy is much more "detailed" than what is actually desired, particularly when conducting "exploratory" analysis to investigate a variety of pricing strategies. If it is desired to determine a pricing strategy that determines a credit decision and interest rate as a function of *categories* of credit scores ("credit bands"), then all that is required to be done is to solve the problem using models of pacc, pdef and plos that depend on the customer's credit category, and not on the customer's exact credit score. In this case, it is not necessary to iterate over the entire customer account set in determining the optimal solution, but simply over a frequency table of loan-size by credit-score categories (in which case the value of total net income and total return on capital are determined by weighting the net income and captial allocation values for each loan-size by credit-score category by the number of accounts in each category).

Such a simplification greatly increases the speed of the numerical algorithm for determining the optimal solution, since the optimization is performed over a small number of loan-size by credit-score categories, rather than over the set of all customers or potential customers. (For a mass credit-card mailing to non-customers, the same initial loan amount may be used for all customers, such as $5,000, so that there is but a single loan amount. In this case, the optimization is done simply over the total number of credit bands, e.g., 20.) Also, it may be desired to restrict the interest rate to a small number of discrete values (e.g., 3 or 5), rather than allow a continuum of values. Such a restriction is readily accommodated by the GLM methodology, simply by determining the maximum of the account lagrangian function over the

restricted set of values. With such simplifications, the result of the optimization could be a simple table that shows the optimal credit decision and interest rate as a function of credit score band.

If the optimization is performed over a frequency table of customers rather than over a sample of customer records in order to determine results quickly, the optimization may be speeded up even more by restricting the interest rates to a small number such as three or five. If the optimization is performed over a sample of records in order to determine a pricing strategy that is "fine-tuned" to each customer's characteristics, it would likely be desired to allow for a finer selection of interest rates, such as ten.

It may be desired to investigate the relationship of the optimal pricing strategy (or any other pricing strategy) with respect to customer characteristics, such as demographic characteristics (referred to as "segmentation variables" in the banking industry). When working with a customer sample, it is easy to determine the relationship of features of the strategy (such as the credit decision) with respect to any known customer characteristic, such as age, race, sex, marital status, income, or location. If the optimal pricing strategy is determined by optimizing over a frequency table of customer characteristics (rather than over a sample of customer records), then determination of strategy characteristics with respect to customer characteristics requires two steps – the first step is to determine the optimal strategy, and the second is to scan through a sample of customer records to talley customer characteristics with respect to the strategy characteristic of interest (e.g., credit decision).

Under the preceding GLM solution of the variable-rate pricing problem, the credit decision is not necessarily monotonic with respect to credit score, i.e., it is possible that credit could be extended for a particular score, but not for every higher score. This is unlikely to happen, but is possible if the distribution of customers by credit score is "unusual" in that historical data indicate a significant proportion of customers having a higher default rate than customers with lower credit scores. To avoid this anomaly, it is desirable to modify the solution so that it produces a credit-decision "cutoff" point, such that credit is extended to all customers having a credit score above the cutoff point, and no credit is extended to customers having credit scores lower than the cutoff point.

5. Computer Program Model

The GLM methodology is implemented very easily through the use of numerical algorithms programmed on a digital computer, using a programming language or application development system such as Microsoft Visual Basic or a database development program such as Microsoft Access. The program can be set up to accept user input parameters, such as the value of the constraint on the total capital allocation.

At a minimum, the program would be set up to determine the optimal pricing strategy for a single credit product type and specified customer (or potential customer) data set. A more comprehensive model might include all of the bank's credit products.

The model input data include the customer characteristics and specifications of the demand and default functions (pacc, pdef, plos). As explained previously, the customer data may be specified in the form of a file containing a record for each customer (or potential customer), or in the form of a table that presents the frequency distribution of customers by known

characteristics, such as credit score. The demand and default functions may, therefore, be specified as tables or as continuous functions. These functions enable the calculation of expected net interest income for each customer (or "cell" in the frequency distribution table). In the simple model described above, pacc would be specified as a function (table) of credit score and interest rate, and pdef and plos would be specified as a function of credit score. The probability of default would be specified as a function of time (e.g. probability of default in 60 or 90 days, as a function of credit score). In addition, the model input data include specification of all income and expense components other than interest, including the portion that depends on pricing and account characteristics (e.g., account service cost) and the portion that does not (e.g., overhead).

In summary, the model input data include:

1. Whether the optimization is to be done using a sample of records for individual customers or using an account probability distribution over credit scorebands.

2. The demand and default functions, pacc, pdef, and plos, as functions of credit score (or scoreband).

3. Specification of all income and expense components other than interest.

As output data, the model should display, in addition to all input parameters, the optimal pricing strategy (credit decision, interest rate), net income, return on capital allocation, return on assets and stockholder value added. The output should include a presentation of the frequency distribution of assets by customer characteristics (such as credit score). For an application in which the only customer characteristic on which these functions depend is credit score band, the optimal pricing strategy (credit decision, interest rate) depend only on credit score band, and the optimal pricing strategy may be displayed very easily, on a single screen.

The following are examples of model output that may be produced, under the assumption that the various model functions depend on a single customer characteristic, viz., credit score (this output would be presented for each loan type). For each output described below, the user should have the option of specifying, (1) for the optimal pricing strategy, the minimum acceptable ROC, and whether the model should determine the optimal credit decision independently for each credit scoreband or as a credit-score cutoff; and (2) for a user-specified pricing strategy: (a) the interest rate (either a specified constant interest rate or the optimal constant interest rate); and (b) the credit decision as a cutoff score, either as a specified cutoff or as an optimally determined cutoff.

In summary, the computer model should produce all of the following output, in the case in which the optimization is done over a frequency table of accounts by credit score:

1. A table displaying all model input parameters that depend on credit score, by credit score (e.g., the table row headings are credit score band (interval) and the table column headings are parameter name, and the table entry is the parameter value).

2. A table displaying all model input parameters that do not depend on credit score.

3. A graph displaying each of the following model parameters as a function of credit score: (1) 60-day default probability; (2) 90-day default probability; (3) proportion of loss in event of default; (4) probability of acceptance (of the loan offer by the customer); (5)

expenses (loan service / collection costs) that are dependent on customer attributes and pricing strategy parameters

4. Asset distribution (frequency or proportion of assets) by credit score, for the following pricing strategies: (1) current pricing strategy; (2) current pricing strategy, but no credit-decision cutoff; (3) current pricing strategy, but no credit-decision cutoff and all customers accept the loan offer; (4) optimal pricing strategy; (5) user-specified pricing strategy (other than current pricing strategy). (All graphics specified in items 1-4 may be displayed on the same output page, by providing the user options by option buttons ("radio" buttons).

5. On a single page, the following: At the top of the page, a three-dimensional graph showing the distribution of expected net income, by credit score, for a range of interest rates (i.e., the same interest rate for all credit bands). At the bottom of the page, a graph showing the interest rate by credit score, for any selection of (1) optimal pricing strategy; (2) current pricing strategy; (3) user-specified strategy.

6. On a single page, the following: At the top of the page, a graph showing the expected return on capital (NI/CA) if credit is extended. At the bottom of the page, a graph showing the optimal credit decision (acceptance rule), for any selection of (1) optimal pricing strategy; (2) current pricing strategy; (3) user-specified strategy.

7. On a single page, the following: The distribution, by credit score, of any selected performance measure: (1) net income (NI); (2) return on capital (ROC); (3) return on assets (ROA); or (4) shareholder value added (SVA).

8. On a single page, the following: For any selected performance measure (NI, ROC, ROA, SVA), the net income by pricing strategy (optimal, current, or user-specified) and product type. (Product on one axis, performance measure by strategy on the other.)

9. On a single page, the following: For any selected performance measure (NI, ROC, ROA, SVA), the net income by pricing strategy (optimal, current, or user-specified) (summed over all product types).

10. A single table summarizing all numerical outputs by product type. A different table may be presented for each selection of pricing strategy (optimal, current, or user-specified) and output measurement scale (dollars or basis points).

If the optimization is done over a sample of customer records, then the model output should include, in addition to the above, the following:

1. The credit decision (extend the loan or not) and the interest rate, under the optimal pricing strategy, the bank's current pricing strategy, and a user-specified pricing strategy.

2. Graphs of various dependent variables, such as mean optimal interest rate, mean optimal credit decision, and frequency of occurrence as a function of various dependent variables (account characteristics) such as credit score, race, and account age. For example, it may be desired to know the proportion of Asians (or Whites, or Blacks, or Hispanics) who are denied credit under the bank's current pricing strategy vs. the optimal pricing strategy, and the mean interest rate charged to them under the bank's current pricing strategy vs. the optimal pricing strategy.

Joseph George Caldwell, PhD
Consultant in Statistics, Operations Research and Information Technology

Professional Profile:

Career in management consulting, research, and teaching. Directed projects in strategic planning, policy analysis, program evaluation, economics, public finance, statistics, operations research / systems analysis, optimization (Everett's Generalized Lagrange Multiplier method) and information technology for US, state and foreign governments, and US and foreign organizations. Areas of expertise include health, education, vocational rehabilitation, welfare, public finance (tax policy analysis, cost-benefit analysis, Medicaid and AFDC financing), agriculture, civil rights, economic development, energy, environment, population, and defense (US Army, Navy, Air Force, Department of Defense). Considerable overseas experience.

Experience includes monitoring and evaluation of national health, education and welfare programs in the US and developing countries (design of national sample surveys and statistical monitoring systems); development of computer management information systems (e.g., the Personnel Management Information System for the Malawi civil service and the Education Management Information System for Zambia); the development of optimal strategies and optimal resource-allocation systems in defense applications (e.g., missile defense, air defense, naval general-purpose forces); and operations-research analysis in industrial applications. Banking applications include the development of statistical models in support of customer relationship management for credit-card operations; development of an optimization model for determining locations for automated banking machines (ATMs); development of a variable-rate pricing model; direction of the "Year 2000" project for a central bank; direction of a disaster-avoidance project for a central bank; introduction of standards-based quality management in the information technology (IT) operations of a central bank; and management of all of information-technology functions of a central bank (hardware, software, personnel, facility).

2001-	Management Consultant (US Agency for International Development, Zambia; United Nations Development Program, Timor-Leste)
1999-2001	Director of Management Systems, Bank of Botswana (Botswana's central bank)
1991-1998	Management Consultant (Wachovia Bank, Charlotte, NC; US Agency for International Development, Egypt, Malawi, Ghana; Asian Development Bank, Bangladesh; TD Canada Trust Bank, Toronto, Canada)
1989-1991	President, Vista Research Corporation, Tucson, Arizona
1982-1991	Director of Research and Development and Principal Scientist, US Army Electronic Proving Ground's Electromagnetic Environmental Test Facility / Bell Technical Operations Corporation and Combustion Engineering; Adjunct Professor of Statistics, University of Arizona; Principal Engineer, Singer Systems and Software Engineering; Arizona
1964-1982	Consultant or employee to firms in South Carolina, North Carolina, Virginia, Maryland, District of Columbia, Haiti, Philippines

Education:

PhD, Statistics, University of North Carolina at Chapel Hill, Chapel Hill, NC, 1966
BS, Mathematics, Carnegie Mellon University, Pittsburgh, PA, 1962
Graduate of Spartanburg High School, Spartanburg, SC, 1958

Contact Information: 503 Chastine Drive, Spartanburg, SC 29301-5977 USA, e-mail
jcaldwell9@yahoo.com , website http://www.foundationwebsite.org .