On customer lifetime value

Martin J. Crowder, David J. Hand
Imperial College London

and

Wojtek J. Krzanowski
University of Exeter

September 2005

Reasons for analysing customer data:

- to understand customer behaviour
- to predict customer behaviour
- and perhaps to intervene to change customer behaviour

- that is, to *manage* customer behaviour

Key point: intervention changes behaviour

So we want to:

- predict behaviour under different interventions/actions
- choose that intervention which optimises some outcome

This work is part of a major programme aimed at guiding the decisions and actions to be made in managing customers

The past

- emphasis on customer recruitment

The (ideal) present

- equal emphasis on development of long-term relationships

But is this what we see?

- still see 0% offers aimed at recruiting CC customers

Those more likely to move to you are more likely to churn

- still see gradually degrading rates offered to existing customers compared with new ones, presumably in the hope that inertia will lead them not to churn

Still see organisations in which the customer acquisition arm is not linked to the risk assessment arm

Conflicting incentives:

(i) the number of new customers through the door

vs

(ii) the lifetime profit on the customers

Pointless to create attractive offer and loyalty at the cost of profit !

And, of course, still see many organisations where acquisition decision is based on estimated default probability rather than expected overall profitability

Not so unreasonable?

- default probability more accurately predicted?
- lifetime profitability more sensitive to external factors?

Customer lifetime value (CLV) is the likely return on a customer over the course of their relationship with the company

Kotler and Armstrong (1996): A customer is profitable if their '*revenues over time exceed, by an acceptable amount, the company costs of attracting, selling, and servicing that customer*'

Initial estimate of expected CLV to guide selection decisions

Estimate of CLV under different circumstances

Standard accept/reject models focus on accepts: they don't tell you the loss arising from incorrectly rejecting customers

CLV models give a handle on this

Key ideas:

- long-term (loyalty)
- relationships
- understanding (not a series of separate one-off decisions)
- life stages
- intervention/action changes behaviour

Answer the question of how a customer would have behaved if you had acted differently

- counterfactual
- selectivity bias

And then choose the action that is ‘best’

Traditional CLV models have forms which are variants/extensions/generalisations of

$$CLV = C\sum_{i=0}^{n}\frac{(1-r)^{i}}{(1+d)^{i}} - M\sum_{i=0}^{n}\frac{(1-r)^{i-1}}{(1+d)^{i-0.5}}$$

C is mean annual gross income per customer
M is mean annual cost per customer
n the period over which CLV is to be assessed
r is the churn rate
d is the annual discount rate

Assumes everything is known!

Fails to take account of interventions changing behaviour!

Fundamental points:

- All actions have consequences

- Predictions of what will happen under A do not tell you what will happen under B

- Models for the effect of treating type 1 customers with action A, and type 2 customers with action B, tells nothing about how type 1s behave under action B and how type 2s behave under action A

- One ‘action’ may be ‘do nothing’

Types of behaviour:

(1) ***Single decisions***

e.g. application scoring: will they / won't they default ?
e.g. give a loan vs do not give a loan

e.g. good track record on product A
offer better terms for product B

e.g. likely to churn - choose action A
not likely to churn - choose action B

Difficulties arising from biased data samples:
Iding Wu's work on how to choose optimal action

(2) ***Dynamic decision sequences***

e.g. behavioural scoring

Monitor behaviour over time
choose *sequence* of actions which is most effective

This talk: special case of single decision
to maximise single overall measure of profit: the *Customer Lifetime Value*

Intervening to maximise CLV

Basic idea:

Combine revenue rate of accrual with lifetime distribution

Intervention decisions based on

- measures of current behaviour
- models of future behaviour under current actions
- models of future behaviour under alternative actions

Simple case discussed here:

Single terminate/continue decision based on customer’s value record to date

Retain good customers, drop bad ones

Induce good customers to stay

Discourage bad customers from staying

All customers are ‘good’ at the right price?
But the ‘right price’ is an intervention

Simple basic model:

$v(t)$ is rate of accrual of value

$F(t)$ is distribution of lifetime

Expected lifetime value is

$$E\left\{\int_0^T v(s)\,ds\right\} = \int_0^\infty v(s)\left[1 - F(s)\right]ds$$

Model *F* using survival analysis

- predict lifetime from initial covariates
- and from further information as it becomes available
- explore effect of interventions
- find optimal time for intervention

Generalisation (1):

$v(t)$ subject to random changes, shocks, etc

$\Rightarrow$ $v(t)$ is stochastic

Generalisation (2):

Customers with similar descriptions behave differently

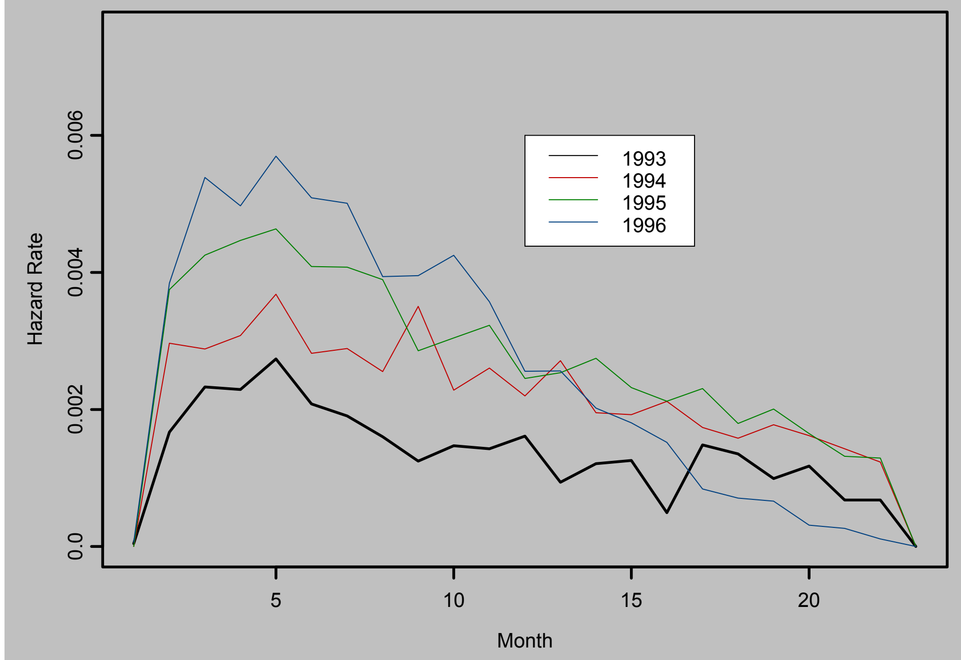
So customer parameters drawn from a distribution

Parenthetical example: hazard functions for 24 Month Loans

144119 accounts   5.2% bads

1993 (22795 obs 3.0% bads); 1994 (32343 obs 4.8% bads)

1995 (41077 obs 5.9% bads); 1996 (47904 obs 5.8% bads)



The simple assumption of exponential lifetime distributions, with constant hazard, is unreasonable

But an assumption of constant hazard ***for each customer*** seems reasonable, if we assume ***some people are riskier than others***

So let these constants differ randomly between customers

Let the constant hazard for people with characteristic vector *x* be

$\lambda(x) = \lambda\psi(x)$ with $\psi(x) = \exp\left(x^T\beta\right)$

$$P_F\left(r > T \mid \lambda, x\right) = e^{-\lambda\psi(x)r}$$

and suppose that $\lambda = \Gamma\left(\nu, \lambda_0\right)$ i.e. $\lambda \sim \lambda_0^{\nu}\lambda^{\nu-1}e^{-\lambda_0\lambda} / \Gamma(\nu)$

Then

$$P_F\left(r > T \mid x\right) = \int e^{-\lambda\psi(x)r}\ \lambda_0^{\nu}\lambda^{\nu-1}e^{-\lambda_0\lambda} \Big/ \Gamma\left(\nu\right) d\lambda$$

$$= \left(1 + \exp\left(x'\beta^*\right)T\right)^{-\nu}$$

This is the cdf of a Pareto distribution with hazard function

$$h_F\left(T \mid x\right) = \frac{\nu \exp\left(x'\beta^*\right)}{1 + \left(x'\beta^*\right)T}$$

which is very like the loan hazard curves

Account opened at time 0

Accumulated income by time *t* is $V_1(t)$ $\qquad V_1(0)=0$

(or discount: $V_1(t)=\int_0^t e^{-rs}dV_1'(t)$)

$V_1(t)$ usually m.i., with jumps (e.g. compound Poisson)

Accumulated cost by time *t* is $C_1(t)$ $\qquad C_1(0)=0$

Assume major costs arise from staffing and overheads, not directly attributed to individual customers. *C* includes such costs

Customer can terminate at time T_1, with distribution $F_1(T_1)$

Strategy

Review customer at time r_1
Specify threshold v_1
If $V_1(r_1) \le v_1$ then terminate agreement
If $V_1(r_1) > v_1$ then renew agreement on more favourable terms

Second phase then has
- new income process $V_2(t)$
- new cost process $C_2(t)$
- new account lifetime T_2

(Possibility of customer declining new offer if $P(T_2 = 0) > 0$)

Customer Lifetime Value (CLV) is

$$CLV = \begin{cases} V_1(T_1) - C_1(T_1) & \textit{if } T_1 \le r_1 \\ V_1(r_1) - C_1(r_1) & \textit{if } T_1 > r_1 \textit{ and } V_1(r_1) \le v_1 \\ V_1(r_1) - C_1(r_1) + V_2(T_2) - C_2(T_2) & \textit{if } T_1 > r_1 \textit{ and } V_1(r_1) > v_1 \end{cases}$$

Put $V_1 = ZU_1$ and $V_2 = ZU_2$

with U_i the common shape of the cumulative income curve in phase *i*

and *Z* customer-specific multiplier: we take $Z \sim N\left(1, \sigma_Z^2\right)$

(So a ‘good’ customer has a large positive value of *Z*, yielding a high rate of income accrual)

Hence

$$CLV = \left\{ZU_1\left(T_1\right) - C_1\left(T_1\right)\right\} I\left(T_1 \le r_1\right) + \left\{ZU_1\left(r_1\right) - C_1\left(r_1\right)\right\} I\left(T_1 > r_1\right)$$
$$+ \left\{ZU_2\left(T_2\right) - C_2\left(T_2\right)\right\} I\left(T_1 > r_1\right) I\left\{ZU_1\left(r_1\right) > v_1\right\}$$

Assume U_1, U_2, T_1, T_2, and *Z* independent

From which

$$E\left(CLV\right) = \int_0^{r_1} E\left\{U_1\left(t\right) - C_1\left(t\right)\right\} dF_1\left(t\right) + E\left\{U_1\left(r_1\right) - C_1\left(r_1\right)\right\}.\left[1 - F_1\left(r_1\right)\right]$$
$$+E\left[\int_{v_1/U_1\left(r_1\right)}^{\infty} \left\{zU_2\left(T_2\right) - C_2\left(T_2\right)\right\} dF_Z\left(z\right)\right].\left[1 - F_1\left(r_1\right)\right]$$

Note: suppose customer does not terminate. Then $P\left(T_1 = \infty\right) = 1$ and

$$E\left(CLV\right) = E\left\{U_1\left(r_1\right) - C_1\left(r_1\right)\right\} + E\left[\int_{v_1/U_1\left(r_1\right)}^{\infty} \left\{zU_2\left(T_2\right) - C_2\left(T_2\right)\right\} dF_Z\left(z\right)\right]$$

Summary so far:

- At time r_1 compare difference of customer's accumulated value and cost with threshold v_1;
- If $V_1\left(T_1\right) > v_1$ enter phase 2;
- If $V_1\left(T_1\right) \leq v_1$ terminate

Now need to find values for r_1 and v_1
So as to maximise $E\left(CLV\right)$

Estimation:

In phase 1 we can estimate:

- *(i) the distribution of **T1**, using survival analysis*
- *(ii) **U1**, which is a property of the process in phase 1 and not of the individuals*
- *(iii) **Z**, which is an individual property*

In phase 2, we can estimate:

- *(i) the value of each individual, **V2** = Z*U2*
- *(ii) then using the known value of Z for that person we can find an unbiased estimate of **U2**.*

However, *we cannot obtain an unbiased estimate of* ***T2***

Unless

(i) we make distributional assumptions

or

(ii) assign some people to phase 2 who have estimated $V_1\left(r_1\right) \leq v_1$

The assignment to phase 2 must use only information available in phase 1 (the '*no sequential confounders*' condition)

For example, selecting a random subset of those with $V_1\left(r_1\right) \leq v_1$ *will do*

Example

Assume $E\left\{U_1(t)\right\} = \beta_1 t$ and $E\left\{U_2(t)\right\} = \beta_2 t$

Note 1: the ***average*** rate of income accrual is constant over time: the actual rate may be highly variable

Note 2: the ***individual*** rate of income accrual is $ZU_1(t)$ (with its customer specific multiplier)

Assume $0 < \beta_2 < \beta_1 \Rightarrow$ a better deal is offered to valued customers, to retain them

Assume $E\left\{C_1(t)\right\} = \gamma_1 t$ and $E\left\{C_2(t)\right\} = \gamma_2 t$ (often $\gamma_1 = \gamma_2$)

Then

$$E(CLV) = (\beta_1 - \gamma_1)\left\{\int_0^{r_1} tF_1(t) + r_1\left[1 - F(r_1)\right]\right\}$$

*

$$+\left\{\beta_2 \sigma_Z E_1 + (\beta_2 - \gamma_2)(1 - E_2)\right\}\mu_2\left[1 - F(r_1)\right]$$

where

$E_1 = E\left[\phi\left\{\left(\frac{v_1}{U_1(r_1)} - 1\right)\Big/\sigma_Z\right\}\right]$ and $E_2 = E\left[\Phi\left\{\left(\frac{v_1}{U_1(r_1)} - 1\right)\Big/\sigma_Z\right\}\right]$

What distributions might be reasonable?

Phase 1 lifetime: T_1 ~ Weibull

Phase 1 income accrual: compound Poisson, with Gamma increment distribution

This yields $U_1\left(r_1\right)$ ~ Gamma

Or Phase 1 income accrual: Wiener process

A real application

- 359 customers
- Discrete time, spanning 1-12 months
- Mean time to termination $\hat{\mu}_1 = 3.6$ months
- Assume $\mu_2 = 3\mu_1$
- Assume $E\left\{C_1\left(j\right)\right\} = \gamma_1 j, \quad \gamma_1 = 5.0$

$\Delta V_{1j} = V_1\left(j\right) - V_1\left(j-1\right)$, for $j = 1,\ldots,12$

ΔV_{1j} is often 0 in these data

So model

$$\Delta U_{1j} = \begin{cases} 0 & \text{with probability } \left(1 - p_U\right) \\ gamma\left(\xi, \eta\right) & \text{with probability } p_U \end{cases}$$

$$Z \sim N\left(1, \sigma_Z^2\right)$$

so that $E\left\{V_1\left(j\right)\right\} = p_U \xi \eta j = \beta_1 j$

Skipping details of parameter estimation:

$$\hat{\beta}_1 = 8.21$$

For phase 2 take

$$E\left\{V_2\left(j\right)\right\} = \beta_2 j, \quad \beta_2 = 7.5$$

$$E\left\{C_2\left(j\right)\right\} = \gamma_2 j, \quad \gamma_2 = 5.0$$

Discrete time version of * yields

$$r_1 = 9 \;\; v_1 = 20 \quad P\left(T_1 < 8\right) = 0.89$$

and

$$E\left(CLV\right) = 11.72$$

Note:

1) assessment at 9 months is late ⇒ high drop out (89%)
This is because of the high variability in V_1

2) A flat peak ⇒ so precise r_1 is not crucial

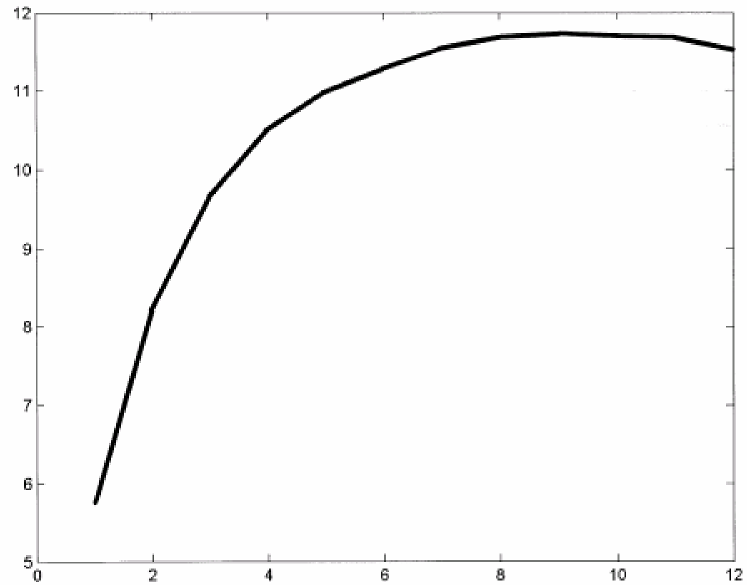


Fig. 3: E(CLV) vs r1

Extension to multiple decision points

The dynamic programming perspective

$$P_n(S) = \sup_a \left[K(S, a, T-n) + E\left(P_{n-1}(Y) \mid S, a\right) \right]$$

- $P_n(S)$ is the maximum CLV after *n* decisions given that the customer begins in condition *S*

- $K(S, a, T-n)$ is the profit (or, negative, loss) resulting from the immediate choice of action a

But, of course, we have to *estimate* the distributions

$\Rightarrow$ the reinforcement learning perspective

Dynamic programming: each step is a trade-off between

(i) immediate gain

(ii) positioning to enable future gain

Reinforcement learning: each step is a trade-off between

(i) immediate gain

(ii) positioning to enable future gain

(iii) information about shape of future distributions

Conclusions

- Two parts to customer lifetime modelling
  - the lifetime distribution
  - the rate of value accrual
- rate of value accrual is stochastic
- both lifetime and value accrual depend on active interventions
- intervention changes likely outcome
- we’ve described an optimal intervention for a simple case
- extensions to more elaborate situations are possible

END

d.j.hand@imperial.ac.uk

http://stats.ma.ic.ac.uk/djhand/public_html/