# Loan Pricing Modeling

Gonzalo Lezma

May 18, 2021

# Contents


**1 Risk Based Pricing** **5**

1.1 Preliminary definitions . . . 5

1.2 Payment schedule in a risk free world . . . 6

1.2.1 The Constant Installment Function . . . 6

1.2.2 The Balance Function: . . . 7

1.2.3 Amortization factor: . . . 8

1.3 Introducing Risk Events in Payment schedule . . . 8

1.3.1 Default and Prepayment probabilities and the survival function: . . . 9

1.4 Alternative setups for Incremental Profit computation . . . 10

1.4.1 Standard model . . . 10

1.4.2 Prepayment dependent on initial balance . . . 11

1.5 Terms included in the incremental profit (CLV) . . . 13

1.5.1 Interest on loans: . . . 13

1.5.2 Cost of Funds: . . . 14

1.5.3 Equity Benefit (Capital Rebate) and Equity Capital Charge: . . . 14

1.5.4 Loss from Default: . . . 15

1.5.5 Fees and Servicing Costs: . . . 15

1.5.6 Recovery costs . . . 15

1.6 Incremental Profit Definition (CLV): . . . 15

1.6.1 Incremental Profit Function: . . . 16

**2 Willingness to Pay Modeling (WTP):** **17**

2.1 The Price-Response Function . . . 17

2.2 Segmented vs Join Estimation . . . 17

2.3 On the flexibility of the logistic function . . . 19

2.4 Expected effect of features on client's WTP . . . 20

2.5 Estimation WTP with Black Box Models - Neural Networks . . . 21

**3 Price optimization (CLV+WTP):** **23**

3.1 Price optimization without constraints . . . 23

3.2 Price optimization with competing objectives: The efficient Frontier . . . 23

**4 Survival models** **25**

4.0.1 Standard survival setup . . . 25

4.0.2 Survival setup in presence of competing risks . . . 25

**5 Explainable ML** **27**
5.1 Introduction . . . . . 27

**6 The PricingPy Python Library** **29**
6.1 Implementation in vectorial form . . . . . 29
6.2 Class Diagram . . . . . 30

# Chapter 1

# Risk Based Pricing

## 1.1 Preliminary definitions

We will start by characterizing the payouts for a constant installment loan under the assumption that there is no risk. In this world, there is only the contractual cash-flow and no risk of any kind (e.g. prepayment risk, default risk,etc). In section (1.3) we will introduce those risk factors from first principles. [1]

The connection between the risk free world and the real world will be established through the theorem of telescopic amortizations in section (1.3). To the best of my knowledge, this theorem has not been established anywhere else but it turns out to be extremely useful for two reasons: First, using this theorem we can easily establish a closed form mathematical representation of the Economic Profit of a loan from first principles even allowing for different modeling choices for the sequence of computations.

Second, the use of the theorem allow for such a simple representation that its programmatic implementation does not need any for-loops of any kind, freeing the space for using the vectorization approach that is widely used in data intensive applications in statistics and deep learning.

[1] The reader familiarized with modern mathematical asset pricing will recognize the similarity of this approach with the typical Q and P measure approach, - i.e. studying first the asset price under a risk neutral world (measure Q) and then connecting with the real world which is non-risk-neutral (P measure)

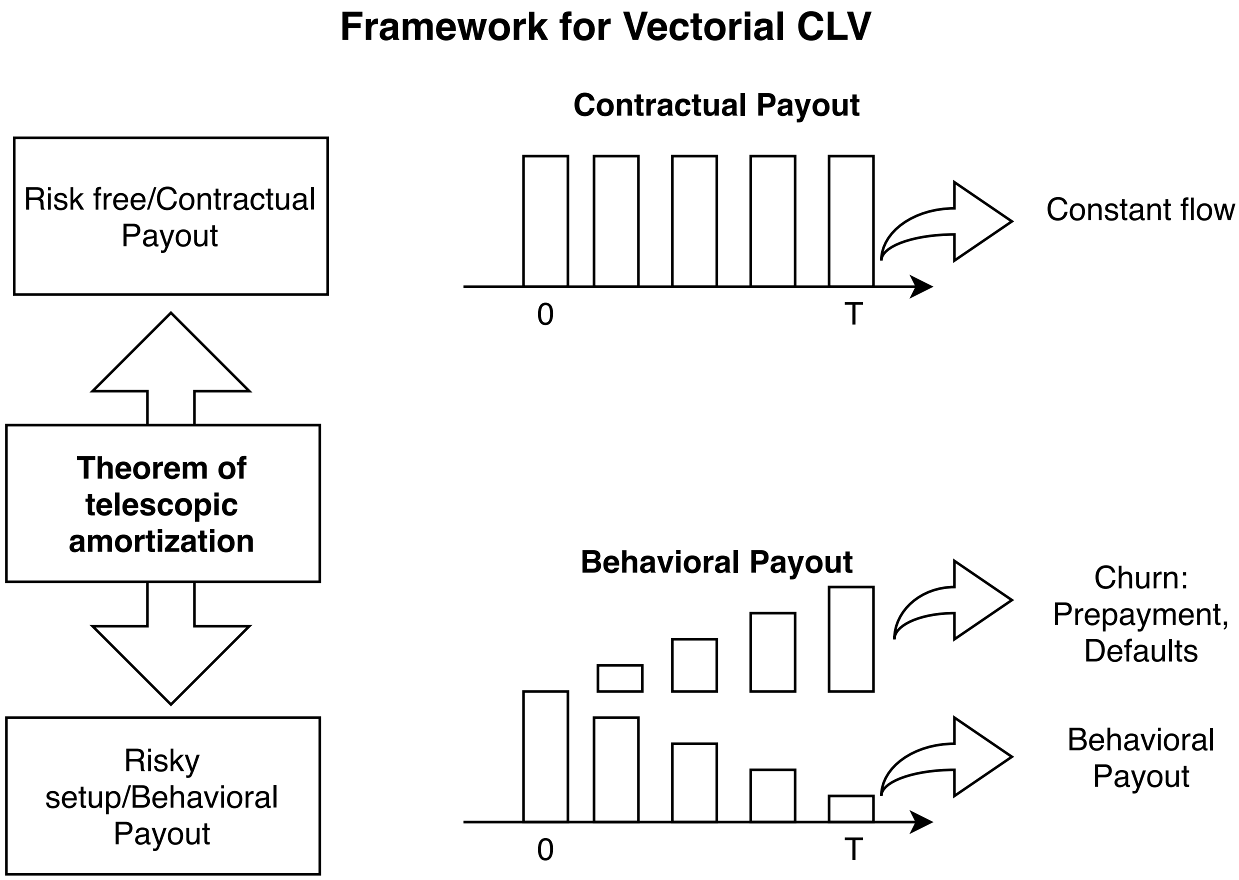



Figure 1.1: Vectorial setup

Finally lets state some basic notation for rates, discount factors and maturity.

| Element | Notation |
|---|---|
| Client Interest Rate | $r$ |
| Cost of Funds Rate,Fund Transfer Pricing FTP, TT | $r_c$ |
| Discount Rate | $r_d$ |
| Contractual Maturity | $T$ |

## 1.2 Payment schedule in a risk free world

### 1.2.1 The Constant Installment Function

We can define a $c_f(.)$ function to compute constant installments by defining the following equality involving $c$ (the constant installments), $\bar{B}(1)$ (the lended principal), $r$ (the loan interest rate) and $T$ the contractual maturity.

$$\bar{B}(1) = \frac{c}{(1+r)} + \frac{c}{(1+r)^2} + \frac{c}{(1+r)^3} + ... + \frac{c}{(1+r)^T} \tag{1.1}$$

Setting $\delta = \frac{1}{1+r}$

$$\begin{aligned}\bar{B}(1) &= c\delta[1 + \delta + \delta^2 + \delta^3 + \delta^4 + ... + \delta^{T-1}] \\ &= c\delta\left[\frac{1}{1-\delta} - \delta^T\frac{1}{1-\delta}\right] = c\delta\left(\frac{1-\delta^T}{1-\delta}\right)\end{aligned} \tag{1.2}$$

Solving for $c$ we establish the definition for $c_f(r,T)$

$$c = \bar{B}(1)\left(\frac{1-\delta}{\delta}\right)\frac{1}{1-\delta^T} = B(1)\left[r\frac{(1+r)^T}{(1+r)^T-1}\right] \tag{1.3}$$

$$\boxed{c_f(r,T) := \frac{r(1+r)^T}{(1+r)^T-1}} \implies c = \bar{B}(1)c_f(r,T) \tag{1.4}$$

### 1.2.2 The Balance Function:

The Current Balance function $\bar{B}(t)$ is the remaining balance left at time $t-1$ for a loan with principal $B = \bar{B}(1)$ and in absence of any prepayment or default risk [2]. For a constant installments loan, the remaining balance at time $t$ for a loan with principal (Balance at t=0) B is given by

$$\bar{B}(t) = B\frac{(1+r)^T - (1+r)^{t-1}}{(1+r)^T-1} \tag{1.5}$$

To establish equation (1.5) we start from first principles noticing that the remaining balance is the capitalized previous balance minus the installment

$$\bar{B}(t) = \bar{B}(t-1)(1+r) - c \tag{1.6}$$

Equation (1.6) is a difference equation of first order that can be solved recursively with initial value given by the principal $B$ as follows:

$$\begin{aligned}&\bar{B}(1) = B \\ &\bar{B}(2) = \bar{B}(1)(1+r) - c = \bar{B}(1+r) - c \\ &\bar{B}(3) = \bar{B}(2)(1+r) - c = \bar{B}(1+r)^2 - c(1+r) - c \\ &\bar{B}(4) = \bar{B}(3)(1+r) - c = \bar{B}(1+r)^3 - c(1+r)^2 - c(1+r) - c \\ &\bar{B}(5) = \bar{B}(4)(1+r) - c = \bar{B}(1+r)^4 - c(1+r)^3 - c(1+r)^2 - c(1+r) - c\end{aligned}$$

$$\bar{B}(t) = \bar{B}(1+r)^{t-1} - c\sum_{s=0}^{t-2}(1+r)^s \tag{1.7}$$

[2] The acute reader will notice that the definition of $\bar{B}(t)$, in terms of the remaining balance left at time $t-1$ , was given so that we can state such a simple equation for the earned contractual interest $\bar{I}(t) = r\bar{B}(t)$ and eliminate dependence on past indexes $t$.

Using the fact that: If $S_n = 1 + x + x^2 + x^3 + ... + x^n \implies S_n = \frac{1-x^{n+1}}{1-x}$

$$\bar{B}(t) = \bar{B}(1+r)^{t-1} - c\left[\frac{(1+r)^{t-1}-1}{r}\right] \tag{1.8}$$

Replacing (1.4) in (1.8)

$$\bar{B}(t) = \bar{B}(1)\left[\frac{(1+r)^T-(1+r)^{t-1}}{(1+r)^T-1}\right] \tag{1.9}$$

$$\boxed{B_f(t,r,T) = \left[\frac{(1+r)^T-(1+r)^{t-1}}{(1+r)^T-1}\right]} \implies \bar{B}(t) = \bar{B}(1)B_f(t,r,T) \tag{1.10}$$

### 1.2.3 Amortization factor:

Finally, we define the amortization factor noticing the constant installment should be equal to the interest payment plus the amortization.

$$A_f(t,r,T) = c_f(r,T) - rB_f(t,r,T) \tag{1.11}$$

$$\boxed{A_f(t,r,T) = \frac{r(1+r)^{t-1}}{(1+r)^T-1}} \tag{1.12}$$

**Corollary 1.** Using (1.6) and (1.11) we can state that:

$$1 - \sum_{s=1}^{t-1} A_f(s,r,T) = B_f(t)$$

## 1.3 Introducing Risk Events in Payment schedule

Last section we defined financial math operators to define a loan payment schedule in absence of any risk events. In this section we introduce risk events including defaults and prepayments. Lets start with the Theorem of Telescopic Amortizations which will be key to connect the risk free schedule with the risky schedule.

**Theorem 1.1 (Telescopic Amortizations).** If A is the function defined in (1.12) then:

$$\prod_{s=1}^{t-1}(1 - A_f(1,r,T-s+1)) = 1 - \sum_{s=1}^{t-1} A_f(s,r,T) = B_f(t,r,T)$$

*Proof.* Lets define

$$E_1 = \prod_{s=1}^{t-1}(1 - A_f(1,r,T-s+1)), E_2 = 1 - \sum_{s=1}^{t-1} A_f(s,r,T)$$

and

$$\delta = 1/(1+r)$$

Working on $E_1$ and setting $\xi = 1+r$:

$$1 - A_f(1, r, T-s+1) = \frac{(1+r)^{T-s+1} - 1 - r}{(1+r)^{T-s+1} - 1} = \frac{\xi^{T-s+1} - \xi}{\xi^{T-s+1} - 1}$$

$$\begin{aligned}
\implies E_1 =& \frac{(\xi^T - \xi)}{(\xi^T - 1)} \times \frac{(\xi^{T-1} - \xi)}{(\xi^{T-1} - 1)} \times \frac{(\xi^{T-2} - \xi)}{(\xi^{T-2} - 1)} \times ... \times \frac{(\xi^{T-t+1} - \xi)}{(\xi^{T-t+1} - 1)} \\
=& \xi^t \frac{(\xi^{T-1} - 1)}{(\xi^T - 1)} \times \frac{(\xi^{T-2} - 1)}{(\xi^{T-1} - 1)} \times \frac{(\xi^{T-3} - 1)}{(\xi^{T-2} - 1)} \times ... \times \frac{(\xi^{T-t} - 1)}{(\xi^{T-t+1} - 1)} \\
=& \frac{\xi^T - \xi^t}{\xi^T - 1} \qquad (1.13)
\end{aligned}$$

Working on $E_2$:

$$A_f(s, r, T) = \frac{r(1+r)^{s-1}}{(1+r)^T - 1} = (\xi - 1)\frac{\xi^{s-1}}{\xi^T - 1}$$

$$\begin{aligned}
\implies E_2 =& 1 - \frac{(\xi - 1)(1 + \xi + \xi^2 + \xi^3 + ... + \xi^{t-1})}{\xi^T - 1} \\
=& \frac{\xi^T - \xi^t}{\xi^T - 1} \qquad (1.14)
\end{aligned}$$

$\therefore E1 = E2$

Finally, the second equality follows directly from Corollary 1. □

### 1.3.1 Default and Prepayment probabilities and the survival function:

We define de conditional probabilities $p_p(t)$ and $p_d(t)$ $\forall t \in \{1, 2, ...T\}$ as follows:

$p_p(t)$ : Probability that the loan will prepay at time t given that it has survived to that point
$p_d(t)$ : Probability that the loan will default at time t given that it has survived to that point

Given these definitions we can establish $S(t)$, the probability that a loan survives until period $t$ using the pigeon hole principle.

$$\begin{aligned}
S(t) =& (1 - p_d(1) - p_p(1)) \times (1 - p_d(2) - p_p(2)) \times ... \times (1 - p_d(t) - p_p(t)) \\
=& \Pi_{s=1}^t (1 - p_d(s) - p_p(s)) \qquad (1.15)
\end{aligned}$$

# 1.4 Alternative setups for Incremental Profit computation

## 1.4.1 Standard model

In this model prepayments/full prepayment, default probability are expressed as conditional probabilities. This probabilities are conditioned on the running active balance i.e. the balance that has not been prepaid or defaulted upon.

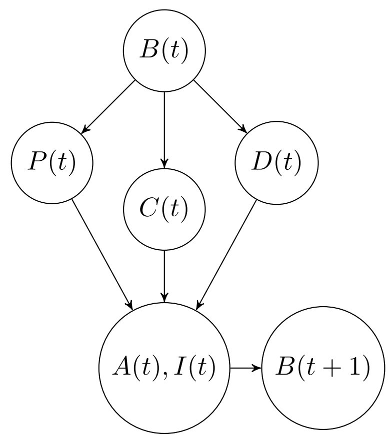



Figure 1.2: Computation Graph

| Variable | Notation | Calculation |
|---|---|---|
| Balance in presence of risk | $B(t)$ | $B(t)$ |
| Default | $D(t)$ | $p_d(t)B(t)$ |
| Full Prepayment | $C(t)$ | $p_c(t)B(t)$ |
| Prepayment | $P(t)$ | $p_p(t)B(t)$ |
| Amortization | $A(t)$ | $(B(t)-D(t)-C(t)-P(t))A_f(1,r,T-t+1)$ |
| Interest | $I(t)$ | $(B(t)-D(t)-C(t)-P(t))r$ |
| Principal | $B(1)$ | $B$ |
| Risky balance next period | $B(t+1)$ | $B(t)-D(t)-C(t)-P(t)-A(t)$ |

Table 1.1: Computation for Model 1: Standard Model

Notice that we defined $B(t)$ as the Loan Balance subject to risk (Behavior affected Loan Balance), as opposed to $\bar{B}(t)$, which is the Risk Free Loan Balance (Contractual Balance). Given this definitions we can compute the recursive form for the balance function $B(t)$

$$\begin{aligned} B(t+1) =& B(t)[1-p_d(t)-p_c(t)-p_p(t)-(1-p_d(t)-p_c(t)-p_p(t))A_f(1,r,T-t+1)] \\ =& B(t)(1-p_d(t)-p_c(t)-p_p(t))(1-A_f(1,r,T-t+1)) \end{aligned} \tag{1.16}$$

Notice that (1.16) is a first order equation in difference which can be easily solved as.

$$B(t) = \prod_{s=1}^{t-1}(1 - A_f(1, r, T - s + 1))(1 - p_d(s) - p_c(s) - p_p(s))B \tag{1.17}$$

Using theorem (1.3) we can state the behavioral balance $B(t)$ as a function of the contractual balance.

$$\begin{aligned} B(t) &= \prod_{s=1}^{t-1}(1 - p_d(s) - p_c(s) - p_p(s))\bar{B}(t) \\ &= S(t-1)\bar{B}(t) \end{aligned}$$

$$\boxed{B(t) = S(t-1)\bar{B}(t)} \tag{1.18}$$

Equation 1.18 states a very simple relation between the theoretical/contractual balance $\bar{B}(t)$ and the behavioral balance $B(t)$. We have derived this equation from a first principles approach through the Theorem of Telescopic Amortization (1.3). This equation represent a very powerful shortcut not only for using intuition, since the behavioral balance can be thought of as the contractual balance adjusted by the survival probability, but also for implementing the model programatically using vectorization instead of recursive loops over the different points in the payment schedule, the last alternative can be very hard to maintain and compute, not to mention its proneness to error.

**Corollary : Interest in model 1.** Given the interest rate definition in Table (1.1) and equation (1.18) we can state that:

$$\boxed{I(t) = S(t)\bar{B}(t)r} \tag{1.19}$$

### 1.4.2 Prepayment dependent on initial balance

This model is a variation of the previous one in which the prepayment probability is a proportion of the initial balance/principal $B(1)$. In this setup, the prepayment amount will be define in function of the marginal probability of default $p_p^m(t)$ and not the conditional probability $p_p(t)$

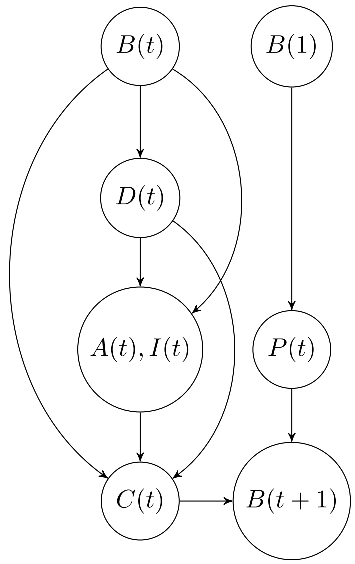



Figure 1.3: Computation Graphs: Model 2

| Variable | Notation | Calculation |
|---|---|---|
| Balance in presence of risk | $B(t)$ | $B(t)$ |
| Default | $D(t)$ | $p_d(t)B(t)$ |
| Amortization | $A(t)$ | $(B(t) - D(t))A_f(1, r, T - t + 1)$ |
| Interest | $I(t)$ | $(B(t) - D(t))r$ |
| Full Prepayment | $C(t)$ | $p_c(t)[B(t) - D(t) - A(t)]$ |
| Principal | $B(1)$ | $B$ |
| Prepayment | $P(t)$ | $p_p^m(t)B$ |
| Risky balance next period | $B(t + 1)$ | $B(t) - D(t) - C(t) - P(t) - A(t)$ |

Table 1.2: Computation for Model 2: Prepayment Dependent on Initial Balance

Given this definitions we can compute the recursive form for the balance function $B(t)$

$$
\begin{aligned}
B(t+1) &= B(t)[1-p_d(t)-p_c(t)(1-p_d(t)-(1-p_d(t))A_f(1,r,T-t+1))-(1-p_d(t))A_f(1,r,T-t+1)]-p_p^m(t)B(1) \\
&= B(t)[1-p_d(t)-p_c(t)+p_d(t)p_c(t)+p_c(t)(1-p_d(t)A_f(1,r,T-t+1)-(1-p_d(t))A_f(1,r,T-t+1)]-p_p^m(t)B(1) \\
&= B(t)[(1-p_d(t))(1-p_c(t))-(1-p_d(t))(1-p_c(t))A_f(1,r,T-t+1)]-p_p^m(t)B(1) \\
&= B(t)(1-p_d(t))(1-p_c(t))(1-A_f(1,r,T-t+1))-p_p^m(t)B(1) \qquad (1.20)
\end{aligned}
$$

Notice that (1.20) is a first order equation in difference of type $x(t + 1) = \alpha(t)x(t) + \beta(t)$ where $\alpha(t) := (1 - p_d(t))(1 - p_c(t))(1 - A_f(1, r, T - t + 1))$ , $\beta(t) := -p_p^m(t)\bar{B}(1)$, $x(t) := \bar{B}(t)$ and the initial condition given by the lended principal $x(1) = \bar{B}(1) = B$. This equation can be easily solved as:

$$x(t) = x_1 \prod_{s=1}^{t-1} \alpha(s) + \sum_{k=1}^{t-2} \left[ \beta(k) \prod_{s=k+1}^{t-1} \alpha(s) \right] + \beta(t-1) \tag{1.21}$$

Plugging back the definitions of $\alpha(t)$, $\beta(t)$ and ,$x(t)$ we get:

$$\begin{aligned} B(t) =& \prod_{s=1}^{t-1} (1 - p_d(s))(1 - p_p(s))(1 - A_f(1, r, T - s + 1))B \\ &- \sum_{k=1}^{t-2} \left[ p_p^m(k)B \prod_{s=k+1}^{t-1} (1 - p_d(s))(1 - p_c(s))(1 - A_f(1, r, T - s + 1)) \right] - p_p^m(t-1)B \end{aligned} \tag{1.22}$$

Using the theorem of Telescopic Amortizations (1.3) and defining $\tilde{S}(t) := \Pi_{s=0}^{t}(1-p_d(s))(1-p_c(s))$ we can state the behavioral balance $B(t)$ as a function of the contractual balance $\bar{B}(t)$.

$$\begin{aligned} B(t) &= \tilde{S}(t-1)\bar{B}(t) - \sum_{k=1}^{t-2} \left[ p_p^m(k)B \frac{\tilde{S}(t-2)}{\tilde{S}(k)} \frac{\bar{B}(t-1)}{\bar{B}(k+1)} \right] - p_p^m(t-1)B \\ &= \tilde{S}(t-1)\bar{B}(t) - \sum_{k=1}^{t-1} \left[ p_p^m(k)B \frac{\tilde{S}(t-1)}{\tilde{S}(k)} \frac{\bar{B}(t)}{\bar{B}(k+1)} \right] \\ &= \tilde{S}(t-1) \left( 1 - \sum_{k=1}^{t-1} \frac{p_p^m(k)B}{\tilde{S}(k)\bar{B}(k+1)} \right) \bar{B}(t) \end{aligned}$$

We define conveniently $S(t)$ as:

$$S(t-1) := \tilde{S}(t-1) \left( 1 - \sum_{k=1}^{t-1} \frac{p_p^m(k)B}{\tilde{S}(k)\bar{B}(k+1)} \right)^+ \tag{1.23}$$

So that we can state that:

$$\boxed{B(t) = S(t-1)\bar{B}(t)} \tag{1.24}$$

Equation (1.23) is analogous to equation (1.18) with an additional term that represents a weighted sum of the last prepayment amounts where the last prepayment has a weight of one.

## 1.5 Terms included in the incremental profit (CLV)

### 1.5.1 Interest on loans:

Independently of the model, we saw that $B(t) = S(t-1)\bar{B}(t)$ so given any of the computations graphs we can state.

$$LI(t) = S(t)\bar{B}(t)r \tag{1.25}$$

### 1.5.2 Cost of Funds:

This term represents the amount of balance the lender has to finance through the cost of funds $r_c$. We follow the setup suggested by Phillips (2018):

$$COF(t) = S_c(t)\bar{B}_c(t)r_c \tag{1.26}$$

where:

$$S_c(t) = \Pi_{s=0}^{t}[1 - p_p(s) - (1 - LGD(s))p_d(s)] \tag{1.27}$$

This setup assumes that, if the client prepays, the bank also prepays the remaining capital that is owed. If the client defaults, the bank uses the recovered capital (i.e.$(1 - LGD(t))$) to make a prepayment to the remaining owed capital and still has to keep making interest payments on the amount defaulted.

Think about this scenario in the following way: At each time, there are three possible states of nature that will determine the remaining balance next period: 1) The client does not default nor prepays with probability $(1 - p_p(t) - p_c(t))$, in this case the remaining balance is still B(t), 2) The client prepays with probability $p_p(t)$, thus the bank prepays an the remaining balance and the banks due balance becomes zero, 3) The client defaults with probability $p_d(t)$, thus the bank pays back only $(1 - LGD(t) \times B(t)$ (which means the bank will not need to make interest payments for that capital anymore) but it will still have to make interest payments for the capital amount equal to $LGD(t) \times B(t)$. It follows that the expected value of the remaining balance will be $(1 - p_p(t) - (1 - LGD(t))p_d(t))B(t)$. [3]

### 1.5.3 Equity Benefit (Capital Rebate) and Equity Capital Charge:

Since any loan has to be financed by both debt and capital, let's assume the fraction of capital the lender has to maintain for each lended dollar is $\alpha_t$. To be consistent with the left and right side of the balance sheet, we need to take into account both the additional revenue of investing the amount of required capital (Equity Benefit) and also the equity charge that the shareholder demands.

$$EB(t) = \alpha_t S(t)\bar{B}(t)r_c \tag{1.29}$$

$$EC(t) = \alpha_t S(t)\bar{B}(t)r_e \tag{1.30}$$

[3] Observe that if we state

$$COF(t) = S(t)\bar{B}_c(t)r_c \tag{1.28}$$

the ALM unit will have to fund the behavioral expected balance, i.e the remaining balance after deducting prepayments (either full or partial) and defaults. However in this case we would be missing the interest payments that the ALM has to make for the part that could not be recovered at default (i.e. $LGD(t) \times B(t)$) since it has to payback the borrowed funding

### 1.5.4 Loss from Default:

So far, we have included the effect of the client's behavior on the interest income and outcome. In this term, we will also incorporate the loss of capital in which the lending entity incur when a client defaults.

$$EL(t) = p_d(t)LGD(t)S(t)\bar{B}(t) \tag{1.31}$$

### 1.5.5 Fees and Servicing Costs:

The lending entity can get additional source of revenue or incur in costs that do not depend on the lended amount can be written as:

$$F(t) = fS(t) \tag{1.32}$$

$$SC(t) = \sigma S(t) \tag{1.33}$$

### 1.5.6 Recovery costs

In case of default the bank recovers an amount of capital equal to $(1 - LGD(t)) \times B(t)$ which the bank uses to prepay the funding used to back the loan as we saw in section ()

$$C(t) = c \times p_d(t)S(t) \tag{1.34}$$

## 1.6 Incremental Profit Definition (CLV):

The net present value is given by:

$$NPV(x(t), r_d, T) = \sum_{t=1}^{T} \frac{x(t)}{(1 + r_d)^t} \tag{1.35}$$

| Element | Notation | Calculation |
|---|---|---|
| Lending Interest | $LI$ | $PV(LI(t), r_d, T)$ |
| Cost of Funds | $COF$ | $PV(COF(t), r_d, T)$ |
| Equity benefit | $EB$ | $PV(EB(t), r_d, T)$ |
| Fees | $LI$ | $PV(F(t), r_d, T)$ |
| Ancillary profit | $A$ | – |
| Origination cost | $OC$ | – |
| Commision | $COM$ | – |
| Servicing Costs | $SC$ | $PV(SC(t), r_d, T)$ |
| Expected Loss | $EL$ | $PV(EL(t), r_d, T)$ |
| Collection costs | $C$ | $PV(C(t), r_d, T)$ |
| Equity charge | $EC$ | $PV(EC(t), r_d, T)$ |

| Element | Notation | Calculation |
|---|---|---|
| Net Interest Income | $NII$ | $LI - COF + EB$ |
| Total Income | $TI$ | $NII + A + F$ |
| Net Income before tax | $NIBT$ | $TI - OC - COM - SC - LD - C$ |
| Net Income after tax | $NIAT$ | $(1 - \tau) \times NIBT$ |
| Incremental profit | $IP$ | $NIAT - EC$ |

### 1.6.1 Incremental Profit Function:

Define the incremental profit function as:

$$\pi(p) = IP(p, r_d, B, T, \theta) \tag{1.36}$$

Given definition (1.36) we can perform several types of computations. For example, to compute the IRR of a given rate $p$ we set $IP(p, irr, B, T, \theta) = 0$. To compute the minimun rate that covers all costs (and risks) and yields a profitability of $r_d$ we set $IP(r^{min}, r_d, B, T, \theta) = 0$. Notice that given this setup, charging the minimum rate does not mean the lending entity is not making money, it just means we are charging enough to reach a target profitability rate.

# Chapter 2

# Willingness to Pay Modeling (WTP):

## 2.1 The Price-Response Function

Each price-response function specifies the demand that the lender would experience at each price, which will depend on:

- $D$ Number of interested applicants, accepted by the bank's creditworthyness criteria, who would achieve a positive surplus from taking the loan from the lender at the offered price.

- $F(p)$ Number of accepted applicants who **take-up** the offered loan at a given price $p$.

In most lending markets, the final price is not known to the client at the time she applies for the loan so we assume that the number of clients who apply for a loan is not influenced by the price.

$$d(p) = D\bar{F}(p) \tag{2.1}$$

$d(p)$ is the number of the loans offered by a lender that would be taken up at the price p. $D$ is the number of successful applicants for the loan, and $\overline{F}(p)$ is the take-up rate, which is defined as the fraction of successful applicants who will take up the loan at price $p$

$$\bar{F}(p) = \int_p^\infty f(w)dw \tag{2.2}$$

## 2.2 Segmented vs Join Estimation

One of the most popular techniques to estimate a WTP model is the logistic regression approach. For $n$ segments, the segmented estimation assumes each segments has its own demand function so we need to estimate $2n$ parameters.

$$\bar{F}_i(p_i) = \frac{e^{a_i+b_ip_i}}{1+e^{a_i+b_ip_i}} = \frac{1}{1+e^{-a_i-b_ip_i}} = \sigma(a_i + b_ip_i) \tag{2.3}$$

Where $\sigma(x) = \frac{1}{1+e^{-x}}$ For $n$ segments, the join estimation assume we can estimate one single price response function that includes all explanatory variables within it (including price)

$$\bar{F}(p_i; x_i) = \frac{1}{1 + e^{a+bp_i+\theta^T x_i}} \tag{2.4}$$

Notice that in the two previous equations there is a subtle difference: In the segmented model (2.3), the price coefficient varies with each segment however in (2.4) the price coefficient is the same for all segments. This difference is not restrictive. To ilustrate that, lets consider we want to segment according to 2 features. The client's income (high or low) and its channel preference (digital or physical). These two features define 4 segments[1]. The segmentation aproach will perform four diferent logistic regressions, one for each segment as shown in Figure (2.1)

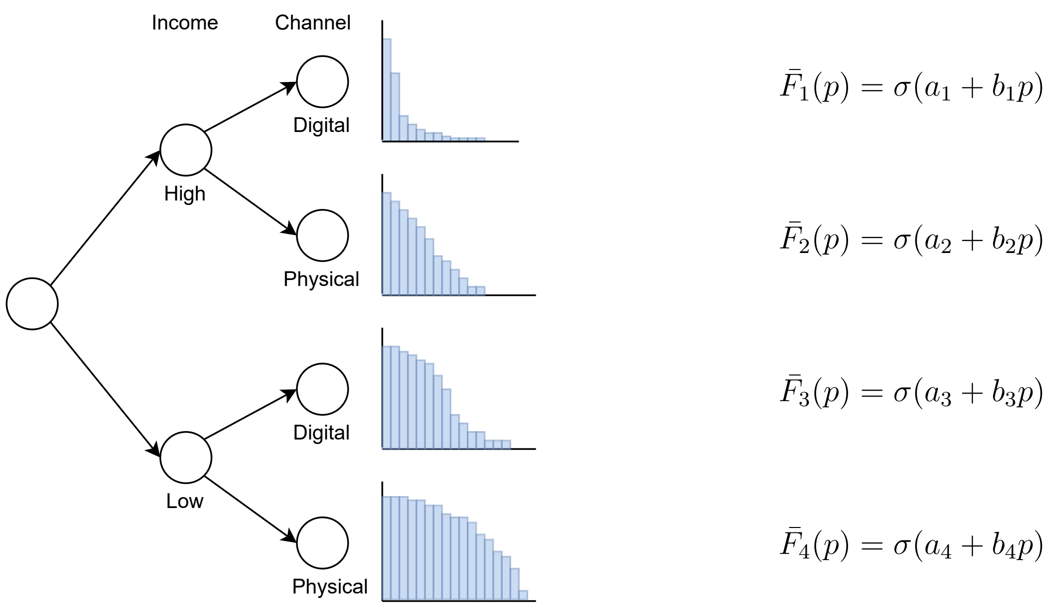



Figure 2.1: Segmented estimation

A more convenient way to achieve the same models would be to include the segments as another feature in a "bigger" but unique equation.

$$\begin{aligned}
\bar{F}(p) = \quad & \sigma(a_1 I(x_1 = \mathrm{H} \wedge x_2 = \mathrm{D})) + b_1 I(x_1 = \mathrm{H} \wedge x_2 = \mathrm{D}))p \\
& a_2 I(x_1 = \mathrm{H} \wedge x_2 = \mathrm{P})) + b_2 I(x_1 = \mathrm{H} \wedge x_2 = \mathrm{P}))p \\
& a_3 I(x_1 = \mathrm{L} \wedge x_2 = \mathrm{D})) + b_3 I(x_1 = \mathrm{L} \wedge x_2 = \mathrm{D}))p \\
& a_4 I(x_1 = \mathrm{L} \wedge x_2 = \mathrm{P})) + b_4 I(x_1 = \mathrm{L} \wedge x_2 = \mathrm{P}))p)
\end{aligned} \tag{2.5}$$

The decision about wheter to chose one aproach over the other is completely up to the model building team, however if you consider that it is better to work on a systematic framework

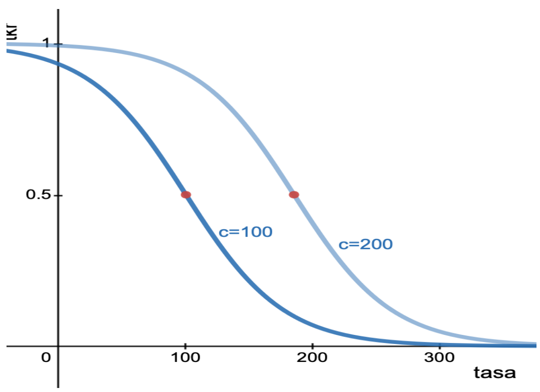
[1] Common methods to find relevant segments include Classification Trees and non supervised clustering methods such as K-Means or DBScan

instead of an ad hoc approach with automation bias you should probably chose the join estimation over the other. Another advantage of the join model is the posibility to learn parameters not only from each individual regression but from the entire data (e.g. lets say we have a digital system which has not much data of volumes -probably because it was only configure to give say $1000 ticket loans- by generating the joint model we will be able to transfer the learning about the relationship between amount and prices from the other channels and use it to our new channel.

## 2.3 On the flexibility of the logistic function

To illustrate the shapes the logistic function can take it is useful to express the logistic formula in the form proposed by Richards [1] .

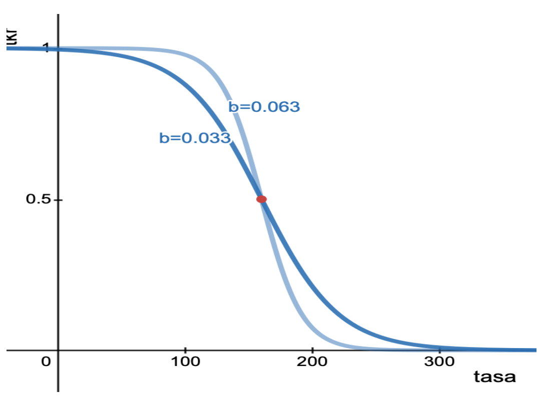
$$\bar{F}(x) = \sigma(-b(x - c))$$




$$\bar{F}_1(p) = \sigma(-b_1(p - 100))$$

$$\bar{F}_2(p) = \sigma(-b_2(p - 200))$$

Figure 2.2: Center/Bias effect




$$\bar{F}_1(p) = \sigma(-0.033(p - 150))$$

$$\bar{F}_2(p) = \sigma(-0.063(p - 150))$$

Figure 2.3: Pseudo slope effect

From the previous analysis we can conclude that if we want to be flexible enough to capture not only differences in the center effect but also the difference in the pseudo slope we should ensure different segments affect not only the intercept coefficient but also the price coefficient. To accomplish that effect a useful specification is the following.

$$\bar{F}(p) = \sigma(c + x_f^t \beta_f + [p, x_p^t]\beta_p) \tag{2.6}$$

Notice that (2.5) is a special version of the more general form (2.6)

## 2.4 Expected effect of features on client's WTP

Although the set of variables available for pricing discrimination varies from country to country, we list a set of variables and their expected effects on the WTP for each customer since.

Price: The effect of an increase in interest rate is a decrease in the take-up probability. The degree to which this effect takes place is given by the price elasticity. [2]

Amount: Clients expect to receive a better loan offer- a lower rate- when they ask for a bigger ticket. Whenever the bank offers lower rates for bigger ticket there is an incentive for clients to borrow a bigger amount to what is needed and then prepay the excess. This effect can be mitigated for some clients since they will pass first to an underwriting decision process, some clients can qualify for a \$10,000 dollar loan but not for a \$20,000 dollar loan. Another way to mitigate this risk is by capturing the prepayment behavior for these clients into the average prepayment for all portfolio. In this case the rate will be greater for many other clients that will pay for the bad prepayment behavior of a few clients.

Channels: A useful framework to assess if clients will have less willingness to pay for a given channel is considering the shopping time available for each client. If the client is known to be operating intensely in digital channels then it is possible that they will have less WTP because they will be more efficient looking at different rates and loan offers, on the other hand, for client who are not so prone to use digital channels they will have less time to spend shopping for rates.

Age: The shopping time framework is usefull to shed light about the expected effecto of age, Phillips(2011) states that young people and old people tend to have less WTP since they can spend more time shopping for raates whereas middle-aged people have less time to shop so they are more prone to have a higher WTP

Income, Geolocation: Any

[2]Common counter examples include Giffen goods: The scottish economist Robert Giffen, based on his observations of European people in the eighteen century pointed out that when the price of bread rose it drained so much the purchase power of people that they could not afford other goods such as meat and consequently opted out to buy more bread

## 2.5 Estimation WTP with Black Box Models - Neural Networks

# Chapter 3

# Price optimization (CLV+WTP):

Consider the incremental profit function given by $\pi(p)$ in equation (1.36) and the take-up rate function given by $d(p) = D\bar{F}(p)$ in equation (2.1). Moreover, following section (2.2), lets consider we are interested in optimizing across $N$ client segments.

## 3.1 Price optimization without constraints

$$p^* = \arg\max_p \sum_{i=1}^{N} D_i \overline{F}_i(p_i)\pi(p_i) \tag{3.1}$$

## 3.2 Price optimization with competing objectives: The efficient Frontier

$$p^* = \arg\max_p \sum_{i=1}^{N} D_i \bar{F}_i(p_i)\pi(p_i) \tag{3.2}$$

$$s.t.$$

$$\sum_{i=1}^{N} D_i \bar{F}_i(p_i) = q \tag{3.3}$$

# Chapter 4

# Survival models

### 4.0.1 Standard survival setup

Let T be a positive random variable in $1, 2, 3, ...$

$$S(t) = P(T > t) \tag{4.1}$$
$$F(t) = 1 - S(t) = 1 - P(T > t) = P(T \leq t) \tag{4.2}$$

### 4.0.2 Survival setup in presence of competing risks

We define the cumulative incidence function as:

$$CIF_k(t) = P(T \leq t, D = k) \tag{4.3}$$
$$= \sum_k P(T \leq t, D = k) = P(T \leq t) \tag{4.4}$$

In order to see what is the relationship between the CIF function and the usual conditional probability of default (death) we state the definition of conditional probability and use the fact that the event $T = t + 1 \wedge T > t$ is equal to $T = t + 1$ standalone.

$$p_k(t+1) = p(T = t+1, D = k/T > t) = \frac{P(T = t+1, D = k)}{P(T > t)} \tag{4.5}$$
$$= \frac{P(T \leq t+1, D = k) - P(T \leq t, D = k)}{1 - \sum_k P(T \leq t, D = k)} \tag{4.6}$$

As an example lets consider that $D = 1$ represents default and $D = 2$ represents prepayment then the conditional probabilities of default and prepayment are given by:

$$p_d(t+1) = \frac{CIF_d(t+1) - CIF_d(t)}{1 - CIF_d(t) - CIF_p(t)} \tag{4.7}$$

# Chapter 5

# Explainable ML

## 5.1 Introduction

# Chapter 6

# The PricingPy Python Library

## 6.1 Implementation in vectorial form

In the previous chapter we showed the full model and expressed in such a simple final equation that it can be stated directly in vectorial form in Python or any programmatic language with support for matrix algebra.

For illustration purposes, consider the computation given in equation (1.16). To implement that we will define the following vectors:

$$T = \begin{bmatrix} T_1 & \dots & T_m \end{bmatrix} \tag{6.1}$$

$$r = \begin{bmatrix} r_1 & \dots & r_m \end{bmatrix} \tag{6.2}$$

$$disb = \begin{bmatrix} d_1 & \dots & d_m \end{bmatrix} \tag{6.3}$$

$$T^{mat} = \begin{bmatrix} 1 \\ 2 \\ \vdots \\ T^{max} \end{bmatrix} \tag{6.4}$$

With the definitions above, $\bar{B}$ can be diretly computed using equation (1.16) and array broadcasting operators in numpy:

$$\bar{B} = disb * \frac{(1+r)^{T^{max}} - (1+r)^{T^{mat}-1}}{(1+r)^{T^{max}} - 1} \tag{6.5}$$

With this approach $\bar{B}$ results of order $T^{max} \times m$ as expected.

$$\bar{B} = \begin{bmatrix} b_{1,1} & \dots & b_{1,m} \\ \vdots & \ddots & \\ b_{T^{max},1} & & b_{T^{max},m} \end{bmatrix} \tag{6.6}$$

## 6.2 Class Diagram

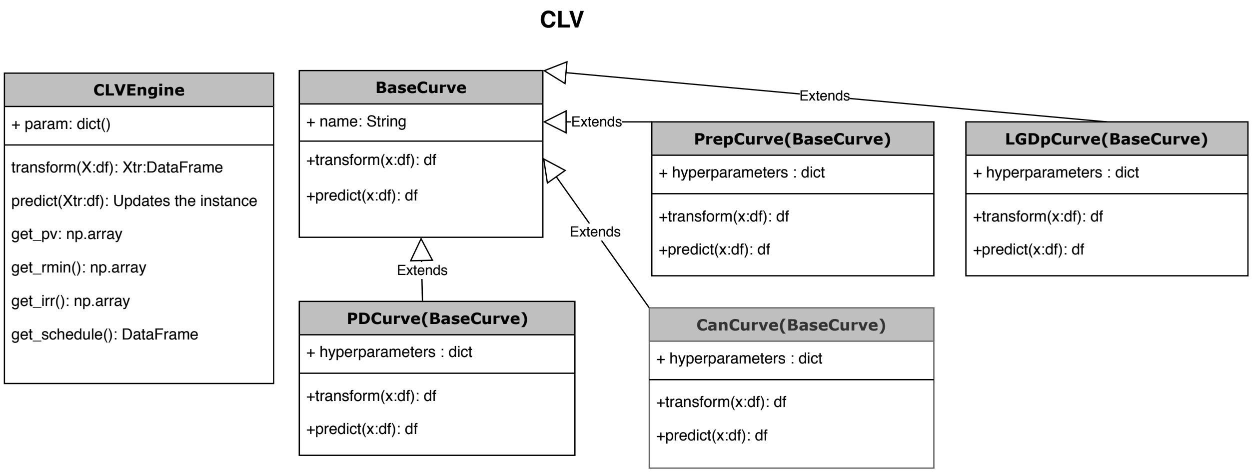



Figure 6.1: CLV Class Diagram

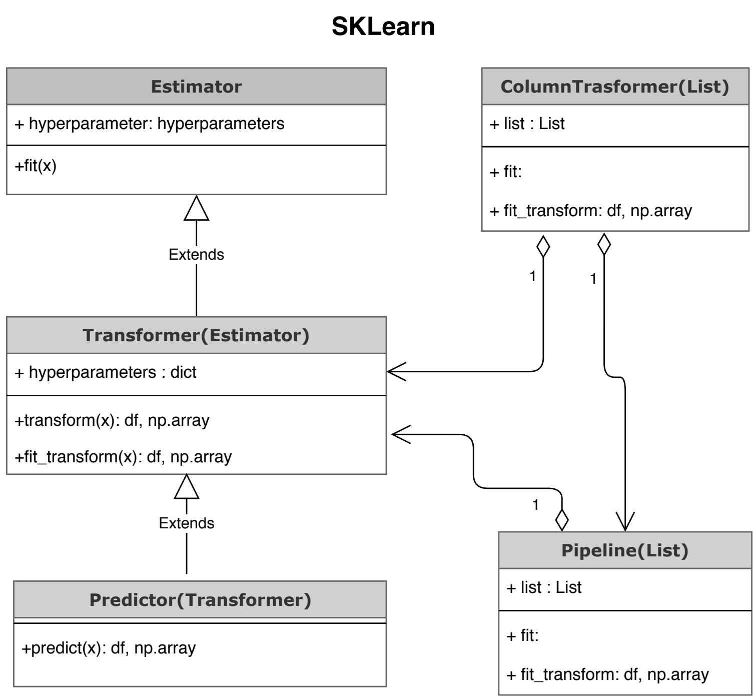



Figure 6.2: SKLearn Class Diagram

```python
import pandas as pd
import numpy as np
import source.engine.CLV_Engine as pricing
import source.engine.utils as ut

# Loading Data
filename = 'inputs.xlsx'
full_filename = Path('.').resolve() / 'data'/ filename
X = pd.read_excel(full_filename, sheet_name = 'inputs')

# Creating CLV engine to compute CLV
eng = ppr.CLVEngine()
Xtr = eng.transform(X)
eng = eng.predict(Xtr) # computes behavioral curves and update all
    behavioral curves

# Computations
PV = eng.get_pv()
RMIN = eng.get_rmin()
IRR = eng.get_irr()
```

Listing 6.1: Python API

# Bibliography

[1] Albert Einstein. "Zur Elektrodynamik bewegter Körper. (German) [On the electrodynamics of moving bodies]". In: *Annalen der Physik* 322.10 (1905), pp. 891–921. DOI: http://dx.doi.org/10.1002/andp.19053221004.