



HAL
open science

Provisioning against borrowers default risk

Geoffrey Nichil, Pierre Vallois

► **To cite this version:**

Geoffrey Nichil, Pierre Vallois. Provisioning against borrowers default risk. Insurance: Mathematics and Economics, 2016, 66, pp.29-43. 10.1016/j.insmatheco.2015.10.004 . hal-01224520

HAL Id: hal-01224520

<https://hal.science/hal-01224520>

Submitted on 4 Nov 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Provisioning against borrowers default risk

Tuesday 10th March, 2015

Geoffrey Nichil and Pierre Vallois.

Abstract

This paper focuses on the risk of loan default from the point of view of an insurer required to indemnify a bank for losses resulting from a borrower defaulting. The main objective of this paper is to model the provision (or claim reserve) against the risk of borrowers defaulting. Unlike traditionally used models, our model depends on specific information concerning the borrowers (amount borrowed and term of loan). Our approach will also take into account three kinds of dependence: the dependence between each claim amount (by taking into account the real estate price), the dependence between the date of default and the claim amount, and the dependence between the number of defaults and the claim amount. Both theoretical and applied, our model allows the calculation of the mean, the variance, and the law of the provision. The amount of data available allows us to estimate all the parameters and to calculate the mean and the variance plus the quantiles of the provision.

Key words: *Borrower default risk; Individual stochastic provisioning; Poisson point process; Geometric Brownian motion; Time of default; Quantile.*

1 Introduction and summary

1) Overview:

In this paper, it is assumed that an insurance company is required to indemnify a bank against losses incurred by a borrower defaulting on payments on real estate loans. The insurance company under consideration works in the French market. A borrower is considered to be in a state of “claim” at a date T when he or she can no longer pay their monthly loan installments.

The main issue for the insurance company is to evaluate, at time t_0 , potential future claim amounts from borrowers whose loans are drawn down prior to t_0 and who have not defaulted prior to t_0 . This amount is the “provision”. It is important to note that in the event of default, the bank can get the money directly from the borrower or try to repossess the property and sell it. Obviously the provision is unknown at time t_0 as it depends on possible future events. The main aim of this paper is to propose a stochastic model which is not only simple but also focuses on the key factors which affect it.

The claim amount is the key to our model. For one borrower and for a fixed default date, this amount is equal to $\max(S_T^1 - S_T^2; 0)$, where S_T^1 is the outstanding amount owed by the borrower at date T and S_T^2 is the property sale amount (see Section 2.2). The amount S_T^1 is proportionate to the borrowed amount; the proportionality coefficient depends on the term of the loan and the date of default (i.e. the maturity). The quantity S_T^2 is proportionate to the borrowed amount; the proportionality coefficient represents the market price fluctuation, the borrower’s deposit, and the selling costs of the property. This is modeled by a geometric Brownian motion. We assume that the process representing the market price fluctuation is the same for each borrower, so there is a dependence between each claim amount. We will model the law of the couples (Maturity of the loan, Term of the loan) by a Poisson point process (see Section 2.3), which will allow us to define the provision as the sum of individual claims (see Definition 2.5). For more details about individual models see [1], [7], [11], [12], [14], and [15]. In Section 2.4 we also examine the links between our model and those of Cramer - Lundberg and Mack. Unlike those, ours considers three different dependencies, which is one of the main points of focus of this paper:

1. The dependence between individual claim amounts (because the market price fluctuation is common for each borrower),
2. The dependence between claim amounts and default dates,
3. The dependence between the term of loan and the date of default.

In Section 3, by making additional assumptions, we will calculate the mean (see Theorem 3.2) and the variance (see Theorem 3.4) of the provision. In addition, our model allows us to go even further than the calculation of these two parameters and provides an algorithm which permits us to simulate the provision (see Theorem 3.6). The amount of data available on each borrower allows us to give a numerical value to the parameters of our model (see Section 4.2). When the parameters are estimated, we will obtain a numerical value for the mean, the variance, the quantile of the provision, and we will also give a sensitive analysis of the provision mean (see Section 4.3).

All proofs will be given in Section 5. Additional information can be found in Section 6, with an index of notations at the end of the paper.

2) Existing methods:

Models normally used by insurance companies to calculate such provisions are collective models (Chain Ladder, Mack; see [8], [9], and [16]) because they aggregate data and do not allow for the calculation of a single claim amount. The disadvantage of these models is two-fold: firstly the aggregation of data implies an over-parameterization and secondly there is no reason to assume that the accident years (or drawdown years in our case) are independent. There is a second type of model; seldom used by insurance companies which provides this type of guarantee; these are models derived from the ruin theory, the best known of which is the Cramer - Lundberg model (see [13]). In this model the provision is a random sum of individual claims concerning a given borrower. The drawback of such models is that they do not take into account the dependence between each claim amount, the dependence between a claim amount and the date of the claim, nor the dependence between the number of claims and the amount of claim.

3) Conclusion:

This paper describes an alternative method for provisioning adapted to the specific insurance contracts considered. It defines an individual stochastic model which takes into account the borrowers' characteristics together with a number of dependencies. Both theoretical and applied, this model gives explicit formula for the mean and the variance, and we also can simulate the provision. The latter enables us to calculate different risk indicators including confidence interval and quantile.

However, we have made a couple of assumptions which in everyday circumstances, would not hold true. Firstly, that the interest rate charged for each borrower is identical and secondly, that any default is immediate and irreversible.

Acknowledgements

The authors are grateful to the referee for helpful comments and suggestions that led to substantial improvements of the paper, and also to Stuart Armour who assisted with the translation into English.

2 Loss provision modeling

In Section 2.1 we will introduce notations and assumptions. Then, in Section 2.2, we will define a claim amount. In Section 2.3 we will model the couples (Maturity of the loan, Term of the loan) by a Poisson point process (see assumption (A_6)) and we will define the fluctuation set of these couples (see Definition 2.3). This allows us to model the provision by totaling the random sum of individual claim amounts reported on random dates (see Definition 2.5). This is the key concept and will allow us to make a quantitative analysis (mean, variance, and simulations). In Section 2.4, we will examine the links between our model and those of Cramer - Lundberg and Mack. Finally in Section 2.5 we will discuss the main assumptions that we made in all previous sections (assumptions (A_5) , (A_7) , and (A_8)).

2.1 Notations and assumptions

In this paper we will assume a continuous time. The time t is defined as a fraction of year and t^* is one month. We denote:

- B ("B" for "borrower") the borrower set. The borrower by j : $j \in B$,
- I^j ("I" for "inception") the drawdown date of the borrower j : $I^j \in [0, t_0]$, where t_0 is the last date under consideration. I^j is the date on which the funds are released.
- D^j (resp. M^j) the positive r.v.¹ representing the loan term (resp. loan amount) for borrower j .

¹r.v.: random variable ; r.v.'s: random variables.

The couple (D^j, M^j) is associated to each borrower j . Borrower j draws down his loan in month i if $i = \lfloor \frac{I^j}{t^*} \rfloor$, where $\lfloor x \rfloor$ represents the floor of x . In the following we will assume that a loan drawn down in month i is drawn down at the beginning of this month, that is to say:

$$(A_1) : \text{If borrower } j \text{ draws down his loan at date } I^j \text{ which corresponds to month } i, \text{ we will assume} \quad (2.1)$$

that: $I^j = t^*i$.

We assume that the duration of loans is capped by an unknown quantity $\bar{D} > 0$ such as:

$$(A_2) : \begin{cases} D^j \leq \bar{D}, j \in B, & (2.2) \\ t_0 < I^j + \bar{D}, j \in B. & (2.3) \end{cases}$$

In the rest of the paper we will consider real estate loans with fixed monthly payments and fixed interest rates for the term of the loan and for every borrower. We will use r to denote the exponential rate for all borrowers. Finally we will assume that early payments are not possible. Therefore we will assume that:

$$(A_3) : \begin{cases} \text{Monthly repayments are fixed,} & (2.4) \\ r \text{ is the exponential and deterministic rate,} & (2.5) \\ \text{Early partial payments are prohibited.} & (2.6) \end{cases}$$

When a borrower defaults (not definitive a priori), the bank has two possibilities: make a claim or wait for the borrower to pay the missing monthly installment. In the second situation, the number of missing monthly installments can increase. However, we will assume that the bank, in the event of default, submits the claim immediately. Note that a default could represent only a portion of the monthly installment but for simplicity we will assume a default concerns the entire monthly installment. Therefore we will assume that:

$$(A_4) : \begin{cases} \text{Default is submitted immediately as a claim,} & (2.7) \\ \text{Default concerns the entire monthly installment.} & (2.8) \end{cases}$$

The bank can recover the borrower's debt in the event of default. This amount is called the recoveries amount. There are two types of recovery:

1. The bank may resell the property financed by the loan. Resale will be automatic if the bank took a mortgage guarantee but less certain if it didn't.
2. The bank may also be repaid by the borrower: inheritance, return to credit, etc..

Recoveries from borrowers are not so frequent. A lack of funds is the most common cause of default and recoveries usually result from the sale of the property. Once again for simplicity, we will assume that borrower recoveries are always zero and that any property sale is made immediately at the time of default. In reality this is seldom, if ever, the case. We assume that the property's value at time T is proportionate to the borrowed amount, the market price variation, the borrower's deposit, and the resale costs (including notary fees, legal procedures fees, etc.). We will assume that market price variation depends only on the time interval between drawdown date I^j and default date T . We can summarize all the assumptions put forward as follows:

$$(A_5) : \begin{cases} \text{Borrower recoveries are zero,} \\ \text{Property recoveries received at date } T \text{ are:} \\ \quad \gamma M^j R_{T-I^j}, j \in B, I^j \leq T \leq I^j + D^j, \end{cases} \quad (2.9)$$

where $\gamma \in \mathbb{R}_+$ is a coefficient independent of borrowers (see (2.33)), and $(R_t)_{t \geq 0}$ represents the market price fluctuation with $R_0 = 1$. For numerical applications we will assume that $(R_t)_{t \geq 0}$ is the geometric Brownian motion with drift μ and volatility σ so that $R_0 = 1$, that is to say:

$$R_t = \exp \left\{ \sigma B_t + \mu t \right\}, \quad (2.10)$$

where $(B_t)_{t \geq 0}$ is a standard Brownian motion. Details about assumption (A_5) are given in Section 2.5.

2.2 Claim amount

In this section we will define the claim amount concerning a borrower who has a default at time T . This is the key notion of our model.

Definition 2.1. Let us consider borrower j who has defaulted at date T . We denote:

1. $S_T^{1,j}$ the sum, at date T , of unpaid capital until initial maturity, unpaid interest until the date of default, and additional interest for late payment,
2. $S_T^{2,j}$ the recoveries amount.

The claim amount is denoted by $S(T - I^j, D^j, M^j, R_{T-I^j})$ and defined by:

$$S(T - I^j, D^j, M^j, R_{T-I^j}) := (S_{T-I^j}^{1,j} - S_{T-I^j}^{2,j})_+, \quad I^j \leq T \leq I^j + D^j. \quad (2.11)$$

It is clear that (2.11) implies that the claim amount cannot be negative and if the sum of unpaid capital, unpaid interest, and additional interest is lower than the recoveries then the insurance company will not pay anything to the bank. The claim amount does not represent the actual amount of loss but only the amount that the insurance company will repay to the bank in the event of default.

Under assumptions (A_1) to (A_5) , the quantities $S_t^{1,j}$ and $S_t^{2,j}$ can be expressed in terms of individual loan criteria for borrower j (D^j and M^j), γ (see (2.33)), and the market price fluctuation $(R_t)_{t \geq 0}$.

Proposition 2.2. Let assumptions (A_1) to (A_5) hold.

1. We have:

$$S_t^{1,j} = M^j \phi_t(D^j), \quad (2.12)$$

$$S_t^{2,j} = \gamma M^j R_t, \quad (2.13)$$

where γ and $(R_t)_{t \geq 0}$ are introduced in assumption (A_5) and:

$$\phi_t(d) := \begin{cases} [e^{rt^*}(1+t^*p) - t^*p] \frac{e^{rd} - e^{r(t-t^*)}}{e^{rd} - 1}, & \text{if } 0 < t \leq d, \\ 1, & \text{if } t = 0, \end{cases} \quad (2.14)$$

with $p \in \mathbb{R}_+$ represents coefficient due to late payment interest (see (5.7) for more details).

2. So by combining (2.11), (2.12), and (2.13) we have:

$$S(t, d, m, u) := m(\phi_t(d) - \gamma u)_+. \quad (2.15)$$

Proof. See Section 5.1. □

2.3 Characteristic of claims, a model of provision

Let $T_{i,j}$ be the r.v. representing the maturity of the loan drawn down during month i by borrower j . We deduce from notations introduced in Section 2.1:

$$t^*i \leq T_{i,j} \leq t^*i + D^j. \quad (2.16)$$

It is important to distinguish two different situations in the previous inequality:

- $T_{i,j} = t^*i + D^j$, if the borrower repays his loan normally until the initial maturity,
- $T_{i,j} < t^*i + D^j$, if the borrower defaults before the initial maturity.

The fact that a borrower may or may not default leads to a partition of the “fluctuation set” of the couples (Maturity of the loan, Term of the loan).

Definition 2.3. We denote $A_{t_0+}^i$ (respectively $A_{t_0-}^i$) as the “fluctuation set” of the couples (Maturity of the loan, Term of the loan) concerning loans drawn down during month i and which have defaulted in the period $]t_0, t^*i + D[$ (respectively $[t^*i, t_0]$):

$$A_{t_0+}^i := \left\{ (d, t) \in (\mathbb{R}_+)^2 : t_0 < t < t^*i + d; d \leq \bar{D} \right\}, \quad (2.17)$$

$$A_{t_0-}^i := \left\{ (d, t) \in (\mathbb{R}_+)^2 : t^*i \leq t \leq t_0; t < t^*i + d; d \leq \bar{D} \right\}, \quad (2.18)$$

where \bar{D} is introduced in assumption (A_2) . An illustration of these sets is given in Appendix 6.1.

Remark 2.4.

1. Defaults at the end of a loan are largely unknown and none have been recorded by the bank we are considering. In addition, any defaults occurring at this time would result in insignificant claims so we do not take into account claims submitted at date $T_{i,j} = t^*i + D^j$.
2. It is clear that relation (2.16) implies that when a claim occurs, there is a dependence between the loan duration and the date of default.

Now we would like to consider loans drawn down during month i . We assume:

$$(A_6) : \text{The cloud of points } (T_{i,j}, D^j)_{j \in B} \text{ is a Poisson point process (PPP for short, see [2] and [4]) with intensity } \Lambda_i. \quad (2.19)$$

Consequently we assume that the number of loans in a portfolio is random.

If we assume that the drawdown rules do not change from one month to another then a default can only be explained by a change in the borrower's personal situation, the month of drawdown does not affect the default. So we will assume that:

$$(A_7) : \text{The PPP } \left((T_{i,j}, D^j)_{j \in B} \right)_{i \geq 0} \text{ are independent,} \quad (2.20)$$

and:

$$(A_8) : \begin{cases} \text{The r.v.'s } (T_{i,j}, D^j)_{j \in B}, (M^j)_{j \in B}, \text{ and the process } (R_t)_{t \geq 0} \text{ are independent,} \\ \text{The r.v.'s } (M^j)_{j \in B} \text{ are independent and identically distributed (i.i.d.).} \end{cases} \quad (2.21)$$

$$(2.22)$$

In Section 2.5 we will discuss in details assumptions (A₇) and (A₈).

Now we will define the provision over $]t_0, t^*i + \bar{D}[$ in a stochastic way as a sum of individual claims amounts.

Definition 2.5. *Provision over $]t_0, t^*i + \bar{D}[$.*

1. The claims amount for all loans drawn down during month i , $t^*i \leq t_0$, and submitted as claims in $]t_0, t^*i + \bar{D}[$ is called "provision for month i ", denoted by $P_{t_0}^i$ ("P" for "provision") and defined as the sum of each individual claim amount:

$$P_{t_0}^i := \sum_{j \geq 1} \mathbb{1}_{\{(T_{i,j}, D^j) \in A_{t_0+}^i\}} S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j} - t^*i}), \quad (2.23)$$

where $A_{t_0+}^i$ is defined by (2.17) and $S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j} - t^*i})$ is given by (2.15) and represents the claim amount at time $T_{i,j}$ concerning borrower j who drew down his loan during month i .

2. The claims amount for all loan drawn down prior to t_0 and occurring as claims in $]t_0, t_0 + \bar{D}[$ is called the "total provision", denoted by P_{t_0} , and defined as the sum of monthly provisions:

$$P_{t_0} := \sum_{i=0}^{\lfloor \frac{t_0}{T^*} \rfloor} P_{t_0}^i. \quad (2.24)$$

Remark 2.6.

1. Determining the total provision as the sum of monthly provisions is a constraint imposed by the insurance company under consideration.
2. We have not included discounting in the modeling of the claim amount because we use the date of default instead.

2.4 Relation with other models

In Cramer - Lundberg's model (see [13]) the provision over $]t_0, t_1]$, $0 \leq t_0 < t_1 < +\infty$, is modeled by:

$$\sum_{j \geq 1} Y^j \mathbb{1}_{\{t_0 < T_j \leq t_1\}},$$

where T_j is the default time of j -th claim and $(Y^j)_{j \geq 1}$ is a sequence of i.i.d. r.v.'s representing the claim amounts. The r.v.'s $(Y^j)_{j \geq 1}$ are assumed to be independent of $(T_j)_{j \geq 1}$. So this model does not allow us

to consider the dependence between the date of default and the claim amount, the dependence between the borrower's situation and the claim amount or the dependence between all the claim amounts.

In the Mack model (see [9]), $C_{i,k+1}$ is the claim amount, for the accident year i (corresponding to the drawdown date in our case) and the development year k , submitted as claims in $]i; i+k+1]$. Therefore, according to Definition 2.5 this quantity can be expressed as:

$$C_{i,k+1} = C_{i,k} + \sum_{j \geq 1} \mathbb{1}_{\{(T_{i,j}, D^j) \in A_{(i+k+1)+}^i \setminus A_{(i+k)+}^i\}} S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j} - t^*i}). \quad (2.25)$$

Let $\mathcal{F}_{i,k}$ be the information available at date $i+k$, i.e.:

$$\mathcal{F}_{i,k} := \sigma \left[(T_{i,j}, T_{i,j} \leq it^* + k)_{j \in B}; (R_t)_{t \leq it^* + k}; (D^j)_{j \in B: I^j = t^*i}; (M^j)_{j \in B: I^j = t^*i} \right]. \quad (2.26)$$

Assume that (2.10) holds, since $(T_{i,j}, D^j)_{j \in B}$ is a PPP and conditionally to $\mathcal{F}_{i,k}$:

$$(R_{s+it^*+k}; 0 \leq s \leq 1) \stackrel{(d)}{=} (R_{it^*+k} R'_s; 0 \leq s \leq 1), \quad (2.27)$$

where R' is an independent copy of R , it follows that:

$$\mathbb{E} [C_{i,k+1} | \mathcal{F}_{i,k}] = C_{i,k} + \beta_{i,k}(R_k), \quad (2.28)$$

where:

$$\beta_{i,k}(x) := \mathbb{E} \left[\sum_{j \geq 1} \mathbb{1}_{\{(T_{i,j}, D^j) \in A_{(i+k+1)+}^i \setminus A_{(i+k)+}^i\}} M^j f(T_{i,j} - t^*i, D^j, x R_{T_{i,j} - t^*i - k}) \right]. \quad (2.29)$$

So our model is drastically different from the Mack model, where the main assumption is:

$$\mathbb{E} [C_{i,k+1} | \mathcal{F}_{i,k}] = f_k C_{i,k}, \quad (2.30)$$

where f_k is a deterministic coefficient independent of i . Our model, from a collective view point, is an additive model where β is a r.v. which takes into account the month of drawdown i and the development date k .

2.5 Discussion concerning assumptions (A_5) , (A_7) , and (A_8)

For more details concerning assumptions (A_1) to (A_4) and (A_6) see [10].

(A_5) Recoveries from borrowers are not so frequent. A lack of funds is the most common cause of default and recoveries usually result from the sale of the property. This is why we assume that borrower recoveries are nil. On the drawdown date, we assume that the property's value is equal to the amount borrowed plus the borrower's deposit. We can reasonably assume that the borrower's deposit is proportionate to the borrowed amount since this is usually a condition of the loan. We assume that the proportionality coefficient is fixed and that it is the same for all borrowers. Then we denote:

$$a \in \mathbb{R}_+ \text{ the coefficient representing the borrower's deposit.} \quad (2.31)$$

We can further assume that the property's value at time t depends on market price; improvements or damage to the house are not taken into account. We assume in particular that the property's value at time t is equal to the property's value at the drawdown date multiplied by a r.v. R_{T, I^j} ("R" for "Real estate"). This represents the market price variation between times I^j and T . This variable is assumed to be identical for all borrowers. Again to simplify, we assume that coefficient R_{T, I^j} depends only on the time interval between drawdown date I^j and default date T , that is to say:

$$R_{T, I^j} = R_{T - I^j},$$

with $R_0 = 1$. Property sale in the event of default incurs costs, including notary fees, legal procedure fees, etc.. These costs are assumed to be proportional to the property's value at default time. The proportionality coefficient is assumed to be identical for all borrowers. Then we denote:

$$c \in [0; 1] \text{ the coefficient representing selling costs of property.} \quad (2.32)$$

So the amount realized from the potential sale of the property in the event of default at date T is:

$$(1+a)M^j R_{T - I^j} - c(1+a)M^j R_{T - I^j} = \gamma M^j R_{T - I^j},$$

where:

$$\gamma := (1 - c)(1 + a). \quad (2.33)$$

(A₇) A priori one borrower has no link between another borrower. In addition, the drawdown rules do not change from one month to another then a default can only be explained by a change in the borrower's personal situation, the month of drawdown has no affect on the default.

(A₈) Assumption (2.22) is confirmed because loans are relatively homogeneous. The insurance company under consideration has drawn up a profile covering loans and borrowers and if the loan and the borrower do not match this profile then cover is cancelled. This profile concentrates the amount borrowed by applying an upper and lower lending limit (see Section 4.2 for more informations about descriptive statistics). This leads us to assume they will have the same law. Finally the independence assumption between each loan amount is justified because a link between two separate borrowers is extremely rare.

Now we discuss assumption (2.21) in detail. $(R_t)_{t \geq 0}$ represents the fluctuation of the property value due to a price fluctuation in the real estate market. In some cases it is possible that real estate prices will fluctuate in relation to the relative value of the property in question. For example, there will be a bigger fluctuation for cheaper houses than there will be for expensive houses and this is significant where the lending book covers a wide range of property values. However, for the insurance company under consideration, as we have mentioned above, the loan amounts are relatively concentrated and this is why we assume that:

$$(M^j)_{j \in B} \text{ and } (R_t)_{t \geq 0} \text{ are independent.} \quad (2.34)$$

In addition the loan amount does not impact on a submission of a claim: a claim is due to a change in the personal situation of the borrower not in the characteristics of the loan, this is why we assume:

$$(T_{i,j}, D^j)_{j \in B} \text{ and } (M^j)_{j \in B} \text{ are independent.} \quad (2.35)$$

See also Appendix 6.4 for an illustration of the scatter plot of the loans duration and loans amount.

Economic trigger events (e.g. unemployment, inflation, etc.) have an impact on real estate price and default rate. In this paper we only consider a retail loans portfolio for financing principal residential properties (we have excluded rental loans and corporate loans) and in this context, real estate price fluctuations have, in the European market, no impact on the default rate. This is the framework of our study.

In addition for the insurance company considered, the number of claims to date is around 10.000 and can be broken down as follows:

1. 84.70% - changes in the life of the borrower (divorce, death of one borrower, etc.),
2. 8.65% - fraud,
3. 6.84% - other reasons.

With these remarks, we can reasonably assume that:

$$(T_{i,j}, D^j)_{j \in B} \text{ and } (R_t)_{t \geq 0} \text{ are independent.} \quad (2.36)$$

Even so, this assumption could not be applied in the US market (see strategic default,).

3 Results

We will start by calculating the mean and the variance of the provision, then we will deal with the cumulative distribution function (c.d.f.) of this r.v..

3.1 Mean and variance of the provision

Notation 3.1. *It is convenient to adopt the following notation:*

$$F_1(t, x, \mu, \sigma) := x \Phi \left[\frac{\ln(x) - \mu t}{\sigma \sqrt{t}} \right] - e^{(\mu + \frac{\sigma^2}{2})t} \Phi \left[\frac{\ln(x) - (\mu + \sigma^2)t}{\sigma \sqrt{t}} \right], \quad (3.1)$$

where $K \in \mathbb{R}_+$ and Φ is the c.d.f. of the standard normal distribution.

Theorem 3.2. *Let assumptions (A₁) to (A₈) hold. We also assume that $(R_t)_{t \geq 0}$ is the geometric Brownian motion with drift μ and volatility σ defined by (2.10). So we have:*

$$\mathbb{E}[P_{t_0}^i] = \gamma \mathbb{E}[M^1] \int_{A_{t_0+}^i} F_1\left(t - t^*i, \frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma\right) d\Lambda_i(t, d), \quad (3.2)$$

$$\mathbb{E}[P_{t_0}^i] = \gamma \mathbb{E}[M^1] \sum_{i=0}^{\lfloor \frac{t_0}{i^*} \rfloor} \int_{A_{t_0+}^i} F_1\left(t - t^*i, \frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma\right) d\Lambda_i(t, d), \quad (3.3)$$

where functions F_1 and $\phi(\cdot)$ are defined by (3.1) and (2.14) respectively.

The proof is given in Section 5.2. Formulas (3.2) and (3.3) are not only theoretical because when the measure Λ_i is the Lebesgue measure and when the parameters are given, it is possible to calculate a numerical value of the provision mean, see Section 4.3.

Now we will focus on the calculation of the provision variance.

Notation 3.3. *Let $(\alpha, \beta) \in \mathbb{R}^2$. We denote:*

$$\tilde{d}(x, y, z, t) := \frac{\ln(x) - yt}{z\sqrt{t}}, \quad (3.4)$$

$$F_2(t, d) := 2 \frac{\phi_t(d)}{\gamma} F_1\left(t, \frac{\phi_t(d)}{\gamma}, \mu, \sigma\right) - F_1\left(t, \frac{[\phi_t(d)]^2}{\gamma^2}, 2\mu, 2\sigma\right), \quad (3.5)$$

$$F_3(t, d, t', d', x) := \frac{e^{\mu t + \sigma \sqrt{t}x - \frac{\sigma^2}{2}t}}{\sqrt{2\pi}} \left(\frac{\phi_t(d)}{\gamma} - e^{\mu t + \sigma \sqrt{t}x} \right) F_1\left(t' - t, \frac{\phi_{t'}(d')}{\gamma e^{\mu t + \sigma \sqrt{t}x}}, \mu, \sigma\right), \quad (3.6)$$

where functions F_1 and $\phi(\cdot)$ are defined by (3.1) and (2.14) respectively.

Theorem 3.4. *Let assumptions (A₁) to (A₈) hold. We also assume that $(R_t)_{t \geq 0}$ is the geometric Brownian motion with drift μ and volatility σ defined by (2.10). The provision variance of month i is:*

$$\begin{aligned} \mathbb{V}[P_{t_0}^i] &= \gamma^2 \mathbb{E}[(M^1)^2] \int_{A_{t_0+}^i} F_2(t - t^*i, d) d\Lambda_i(t, d) + \gamma^2 [\mathbb{E}(M^1)]^2 \\ &\times \left[\int_{(A_{t_0+}^i)^2} \left\{ \int_{-\infty}^{\tilde{d}\left(\frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma, t-t^*i\right)} F_3(t - t^*i, d, t' - t^*i, d', x) dx \right\} \right. \\ &\left. d\Lambda_i(t, d) d\Lambda_i(t', d') - \left\{ \int_{A_{t_0+}^i} F_1\left(t - t^*i, \frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma\right) d\Lambda_i(t, d) \right\}^2 \right]. \end{aligned} \quad (3.7)$$

The proof of this result is given in Section 5.3. The variance of the total provision can also be determined, see Appendix 6.2.

Formula (3.7) is not easy to prove. This is due to the fact that the claim amounts are not independent due to the process $(R_t)_{t \geq 0}$. Therefore, we have to take into account the correlation between two claim amounts.

Remark 3.5. *By not assuming that $(R_t)_{t \geq 0}$ is the geometric Brownian motion defined by (2.10) we get:*

$$\begin{aligned} \mathbb{V}[P_{t_0}^i] &= \mathbb{V}[N_i(A_{t_0+}^i)] \mathbb{E}\left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1}-t^*i})\right)^2\right] \\ &+ \left[\mathbb{E}[N_i(A_{t_0+}^i)]\right]^2 \left\{ C^i - \left[\mathbb{E}[S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1}-t^*i})]\right]^2 \right\}, \end{aligned} \quad (3.8)$$

where the couple $(T_{i,1}, D^1)$ represents a generic point of the PPP $(T_{i,j}, D^j)_{j \in B}$ and:

$$N_i(A) := \sum_{j \geq 1} \mathbb{1}_{\{(T_{i,j}, D^j) \in A\}}, \quad A \subseteq A_{t_0-}^i \cup A_{t_0+}^i, \quad (3.9)$$

$$C^i := \left[\mathbb{E}[M^1]\right]^2 \mathbb{E}\left[h\left(T_{i,1} - t^*i, D^1, T_{i,2} - t^*i, D^2\right)\right], \quad (3.10)$$

$$h(t, d, t', d') := \mathbb{E}\left[\left(\phi_t(d) - \gamma R_t\right)_+ \left(\phi_{t'}(d') - \gamma R_{t'}\right)_+\right], \quad (3.11)$$

where the couple $(T_{i,2}, D^2)$ represents another generic point of the PPP $(T_{i,j}, D^j)_{j \in B}$ and independent of $(T_{i,1}, D^1)$. Result (3.8) (proved in Section 5.3) allows us to identify sources of risk. The first term of this expression represents the variability of the number of claims around the mean of the number of claims. The second term of (3.8) is, in the classical case of the compound Poisson process (independence of claim amount), equal to the variance of the claim amount. So, the second term represents variability of the claim amount around its mean.

3.2 The law of $P_{t_0}^i$

It is useful to order the loans maturities as follows:

$$T_{i,(1)} < T_{i,(2)} < \dots < T_{i,(n)} < \dots \quad (3.12)$$

We denote the loan duration associated to the date $T_{i,(j)}$ by $D^{(j)}$.

Theorem 3.6. *Let assumptions (A_1) to (A_8) hold. We also assume that $(R_t)_{t \geq 0}$ is the geometric Brownian motion with drift μ and volatility σ defined by (2.10).*

1. Recall that:

$$\mathcal{L}\left[N_i(A_{t_0+}^i)\right] = \mathcal{P}\left[\Lambda_i(A_{t_0+}^i)\right].$$

2. The conditional law of $P_{t_0}^i$ given $N_i(A_{t_0+}^i) = n$ is:

$$\mathcal{L}\left[P_{t_0}^i | N_i(A_{t_0+}^i) = n\right] = \mathcal{L}\left[f_n\left((T_{i,(j)})_{1 \leq j \leq n}, (D^{(j)})_{1 \leq j \leq n}, (M^j)_{1 \leq j \leq n}, (G^j)_{1 \leq j \leq n}\right)\right], \quad (3.13)$$

where:

$$f_n\left((t_{i,(j)})_{1 \leq j \leq n}, (d^{(j)})_{1 \leq j \leq n}, (m^j)_{1 \leq j \leq n}, (g^j)_{1 \leq j \leq n}\right) := \sum_{j=1}^n m^j \left\{ \phi_{t_{i,(j)} - t^*i}(d^{(j)}) - \gamma \prod_{k=1}^j r^k \right\}_+, \quad (3.14)$$

$$r^k = \exp\left[\sigma \sqrt{t_{i,(k)} - t_{i,(k-1)}} g^k + \mu(t_{i,(k)} - t_{i,(k-1)})\right], \text{ with this agreement: } t_{i,(0)} = 0, \quad (3.15)$$

$$(G^j)_{1 \leq j \leq n} \text{ is a collection of i.i.d. r.v.'s with standard normal distribution and independent of } \left((T_{i,(j)}, D^{(j)})_{1 \leq j \leq n}, (M^j)_{1 \leq j \leq n}\right). \quad (3.16)$$

Proof. See Section 5.4. □

Clearly the law of the provision is not explicit. However, we formulated the law of the provision as a complicated function but dependent on a sequence of r.v.'s which can be easily simulated. Therefore using the Monte Carlo method we can obtain an approximation of the quantile (see Section 4.3 and Appendix 6.3 for more details).

4 Numerical application

The theoretical model that we have presented above has been created for application and this stochastic modeling takes into account the borrowers' characteristics. The amount of data available on each borrower (amount borrowed, term, and default date) allows us to estimate the parameters of our model (see Section 4.2). The theoretical analysis presented in Section 3, with Theorems 3.2, 3.4, and 3.6, allows us to calculate numerically the mean, the variance, the quantile of the provision, and also to give some sensitivity analysis (see Section 4.3). To achieve this goal, we need to specified the choice of the measure Λ_i .

4.1 Choice of the intensity Λ_i

Partition $A_{t_0-}^i$ and $A_{t_0+}^i$ (see (2.18) and (2.17)), of the fluctuation set of the couples (Maturity of the loan, Term of the loan), concerns loans which have defaulted.

Definition 4.1. *We denote $C_{t_0+}^i$ (resp. $C_{t_0-}^i$) as the "fluctuation set" of the couples (Maturity of the loan, Term of the loan) concerning loans drawn down during month i and which have matured in the period $]t_0, t^*i + \bar{D}]$ (resp. $[t^*i, t_0]$):*

$$C_{t_0+}^i := \left\{ (d, t) \in (\mathbb{R}_+)^2 : t = t^*i + d; t > t_0; d \leq \bar{D} \right\}, \quad (4.1)$$

$$C_{t_0-}^i := \left\{ (d, t) \in (\mathbb{R}_+)^2 : t = t^*i + d; t \leq t_0; d \leq \bar{D} \right\}. \quad (4.2)$$

An illustration of this set is given in Appendix 6.1.

Having disposed of this preliminary step, we can now choose the intensity Λ_i . We will assume that:

$$\Lambda_i := \lambda_1 Leb_{1|C_{t_0-}^i \cup C_{t_0+}^i} + \lambda_2 Leb_{2|A_{t_0-}^i \cup A_{t_0+}^i}, \quad (4.3)$$

where:

- λ_1 and λ_2 are two real and strictly positive parameters,
- $Leb_{1|C_{t_0-}^i \cup C_{t_0+}^i}$ (resp. $Leb_{2|A_{t_0-}^i \cup A_{t_0+}^i}$) is the Lebesgue measure on \mathbb{R} (resp. \mathbb{R}^2) restricted on $C_{t_0-}^i \cup C_{t_0+}^i$ (resp. $A_{t_0-}^i \cup A_{t_0+}^i$).

Note that parameter λ_1 represents loans which have matured without default. This parameter will only help us to estimate parameter λ_2 (and also h) but not to determine the mean (or the variance) of the provision.

4.2 Estimation of parameters

We will start by listing the parameters:

- The mean of borrowed amount: $\mathbb{E}(M^1)$,
- The monthly rate r defined by (2.5),
- The coefficient γ associated to the borrower's deposit and the selling costs of property, see (2.33),
- The penalty coefficient p , in relation to late payment, see (5.7),
- Coefficients μ and σ of the geometric Brownian motion defined by (2.10),
- Parameters λ_1 and λ_2 linked to the intensity Λ_i of the PPP, see (4.3),
- Parameter \bar{D} representing the maximal duration, see (2.2).

At date t_0 , we dispose of three different sources of data: data shared by all borrowers, data linked to claims, and data exogenous to the insurance company. Note that the first two sources are private because they are supplied by the insurance company while third is public. For each borrower j having drawn down his loan during month i so that $it^* \leq t_0$, we have the term d^j , the amount m^j and the maturity $t_{i,j}$ of the loan. We also dispose of the number b_i of loans drawn down during month i , the number n_i of defaults submitted until date t_0 and the number v_i of loans which have matured prior to t_0 . So the number of loans "in force" at date t_0 , that is to say loans which have not defaulted or matured, is given by $b_i - n_i - v_i$. In the following, the total number of loans drawn down prior to t_0 is denoted by:

$$b := \sum_{i=0}^{\lfloor \frac{t_0}{t^*} \rfloor} b_i. \quad (4.4)$$

Finally, we have exogenous data provided by INSEE, by the broker "Empruntis", and by the French authority ACPR. INSEE gives an index of house prices in France which is denoted by u_t at date t . The broker "Empruntis" and ACPR provide annual statistics about real estate loans, in particular the mean of the borrower's deposit and the mean of actuarial loan rate.

So the mean of borrowed amount will be estimated by empirical mean. The rate of each loan will be estimated with the actuarial rate given by "Empruntis" (actuarial rate estimate to 5%). Coefficient γ (see Section 2.5) is determined by the statistics from "Empruntis" and "ACPR" (coefficient a , see (2.31), representing the borrower's deposit is determined at 20%) and by "expert judgement" (coefficient c , see (2.32), representing selling costs of property is fixed at 10%). Coefficient p (see (5.7)) will be estimated by "expert judgement" at 2. Coefficients μ and σ will be estimated with INSEE's data or by expert judgment in the case of stress scenarios.

We turn now to the estimation of λ_1 , λ_2 , and the maximum term \bar{D} . The maximum-likelihood estimators are denoted by $\hat{\lambda}_1$, $\hat{\lambda}_2$, and $\hat{\bar{D}}$.

Proposition 4.2. *Let assumptions (A₁) to (A₈) hold. Then $x = \hat{\lambda}_1$, $y = \hat{\lambda}_2$, and $z = \hat{\bar{D}}$ are solutions of the following system:*

$$(S) : \begin{cases} \sum_{i=0}^{i_0} \left[\frac{v_i}{x} + \frac{(b_i - n_i - v_i)\sqrt{2}(z - t_0 + t^*i)}{x\sqrt{2}(z - t_0 + t^*i) + \frac{y(z - t_0 + t^*i)^2}{2}} \right] = \sqrt{2}(i_0 + 1)z, \\ \sum_{i=0}^{i_0} \left[\frac{n_i}{y} + \frac{(b_i - n_i - v_i)(z - t_0 + t^*i)^2}{2x\sqrt{2}(z - t_0 + t^*i) + y(z - t_0 + t^*i)^2} \right] = \frac{(i_0 + 1)z^2}{2}, \end{cases} \quad (4.5)$$

$$\sum_{i=0}^{i_0} \left[\frac{n_i}{\left(z - \frac{1}{2}(t_0 - t^*i)\right)} + \frac{(b_i - n_i - v_i) \left(\sqrt{2}x + y(z - t_0 + t^*i)\right)}{x\sqrt{2}(z - t_0 + t^*i) + \frac{y(z-t_0+t^*i)^2}{2}} \right] = (x\sqrt{2} + yz)(i_0 + 1),$$

where:

$$i_0 := \lfloor \frac{t_0}{t^*} \rfloor \in \mathbb{N}^*. \quad (4.6)$$

The proof of Proposition 4.2 is given in Section 5.5. Note that the above system can be resolved numerically using the software *R*. Here we summarize the estimation:

Parameter	Notation	Estimation	Source
$\mathbb{E}(M^1)$	\bar{m}	$\frac{1}{b} \sum_{i=1}^b m^i$	Insurance company
r	\hat{r}	$\ln \left(1 + \frac{5}{100} \frac{1}{12} \right)$	Empruntis
γ	$\hat{\gamma}$	1.20×0.9	Expert judgement and Empruntis
p	\hat{p}	2	Expert judgement
μ	$\hat{\mu}$	$\frac{1}{i_0} \sum_{i=1}^{i_0} [\ln(u_i) - \ln(u_{i-1})]$	INSEE or expert judgement
σ	$\hat{\sigma}$	$\sqrt{\frac{1}{i_0} \sum_{i=1}^{i_0} [\ln(u_i) - \ln(u_{i-1}) - \hat{\mu}]^2}$	INSEE or expert judgement
$\lambda_1, \lambda_2, \bar{D}$	$\hat{\lambda}_1, \hat{\lambda}_2, \hat{\bar{D}}$	Maximum likelihood	Insurance company

(4.7)

Now we want to estimate numerically all the parameters. To do this we will apply the estimation method introduced previously on data collected by a French insurance company from January 2000 to December 2013 for one particular bank: 23,067 contracts and 330 claims will be analyzed. Firstly, we give some descriptive statistics concerning the contracts involved:

	Mean	Median	Standard deviation	Minimum	Maximum
Borrowed amount	100,946	98,691	53,096	10,100	249,928
Loan duration	207	204	58	36	328

Illustration 1: Descriptive statistics.

Note that to solve the system (4.5) we used an annual time unit in order to get a sufficient number of claims by “grouping year”. So this data leads to the following estimates:

Parameter	\bar{m}	\hat{r}	$\hat{\gamma}$	\hat{p}	$\hat{\mu}$	$\hat{\sigma}$	$\hat{\lambda}_1$	$\hat{\lambda}_2$	$\hat{\bar{D}}$
Value	100,946	0.05	1.08	2	-0.08	0.05	86.43	0.27	286

Illustration 2: Synthesis estimation parameters.

Note that parameters related to the geometric Brownian motion were fixed by expert judgement. These parameters represent the evolution of real estate over 30 years: it seems more appropriate to set these parameters rather than using data to estimate them. The choice made corresponds to a 30% drop in prices of real estate over a 5 years period. This scenario partially matches the benchmark used by the European Central Bank for stress tests in the banking sector ; This exercise is called “Asset Quality Review” or AQR.

4.3 Mean of the provision and sensitivity analysis

In this section we consider month $i = 0$ and we will only focus on the provision mean ; a similar analysis to the provision variance (or quantile) could be obtained in the same manner.

Firstly we will compare the theoretical mean, obtained with Theorem 3.2, to the empirical mean, obtained with Theorem 3.6. We will conclude this section with a sensitivity analysis of the provision mean according to a fluctuation of some parameters.

We obtain the numerical value of the provision mean (see Theorem 3.2 and formula (4.3)) by replacing the value of parameters by their estimators, that is to say:

$$\hat{\mathbb{E}}[P_{t_0}^0] = \lambda_2 \bar{m} \int_0^{\hat{D}} \left[\int_{t_0}^y \left\{ \phi_t(y) \Phi \left[\tilde{d} \left(\frac{\phi_t(y)}{\hat{\gamma}}, \hat{\mu}, \hat{\sigma}, t \right) \right] - \hat{\gamma} e^{(\hat{\mu} + \frac{\hat{\sigma}^2}{2})t} \Phi \left[\tilde{d} \left(\frac{\phi_t(y)}{\hat{\gamma}}, \hat{\mu} + \hat{\sigma}^2, \hat{\sigma}, t \right) \right] \right\} dt \right] dy, \quad (4.8)$$

where $\phi(\cdot)$ and \tilde{d} are defined by (2.14) and (3.4) respectively.

In addition Theorem 3.6 allows Monte Carlo simulations of the provision. However, we need to know the law of the borrowed amount M^j . For loans drawn down during month i we have b_i observations of the borrowed amount. b_i is actually around 300, so it is reasonable to replace the c.d.f. of M^1 by the empirical c.d.f.. This enables us to get an algorithm to simulate the provision (see Appendix 6.3): we denoted the l -sample obtained with this algorithm by (p_1^0, \dots, p_l^0) . Consequently we get the empirical mean which is denoted by $\mathbb{E}^{emp}[P_{t_0}^0]$ and defined by:

$$\mathbb{E}^{emp}[P_{t_0}^0] := \frac{1}{l} \sum_{k=1}^l p_k^0. \quad (4.9)$$

In the graph below we compare the empirical mean (blue curve) and the theoretical mean (green curve) obtained with the set of parameters presented in Illustration 2 (we could chose another set of parameter of course). The analysis date t_0 is on the x axis and the provision means are on the y axis.

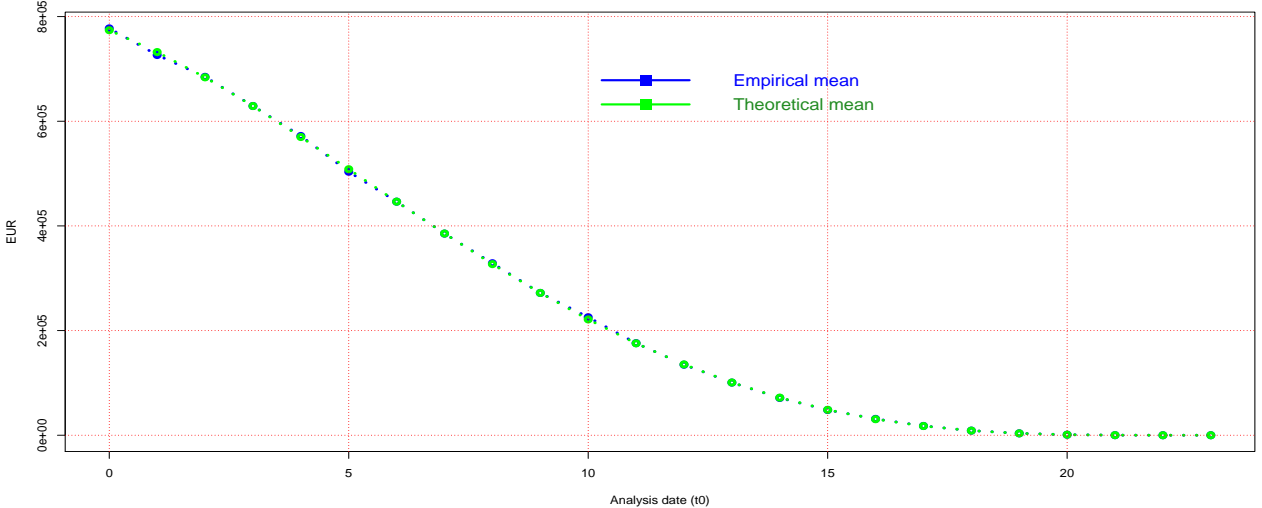


Illustration 3: Empirical mean and theoretical mean.

A few observations:

1. The theoretical mean and the mean obtained by simulation are extremely close whatever analysis date.
2. As expected, the provision mean decreases with the analysis date but this decrease is not linear.
3. Finally, from 2016 the mean of the provision is almost zero. This allows us to determine numerically the time t_0^* (in this example $t_0^* = 16$ years) from which the mean of the provision becomes insignificant:

$$t_0^* := \inf \{ t_0 \geq 0 : \hat{\mathbb{E}}[P_{t_0}^0] \leq \beta \}, \quad (4.10)$$

where β is set very small. This provides the insurance company the maximum provisioning duration for each drawdown month, i.e. the period during which the company should record a provision.

Now we give a sensitivity analysis of the provision mean. Because we get an explicit formula for the provision mean (see formula (3.2)) it is possible to calculate explicitly the provision mean derivative. However we only show in this paper the provision mean fluctuation according to the fluctuation of our parameters (see [10] for the calculation of the derivative). On the graph below you can see the fluctuation of the provision mean according to the fluctuation of one parameter (others parameters are fixed):

1. λ_2 (see (4.3)) top left: $\lambda_2 \in [0.1; 1.1]$,

2. γ (see (2.33)) top right: $\gamma \in [0.1, 1.4]$,
3. μ (see (2.10)) bottom left: $\mu \in [-0.1; -0.01]$,
4. r (see (2.5)) bottom right: $r \in]0; 0.15]$.

The y axis is the provision mean and the x axis is the value of the parameters. We plot each fluctuation for $t_0 = 0$ (blue curve), $t_0 = 5$ years (green curve), and $t_0 = 10$ years (red curve). The black vertical curves represent the provision mean value for the set of parameters presented in Illustration 2.

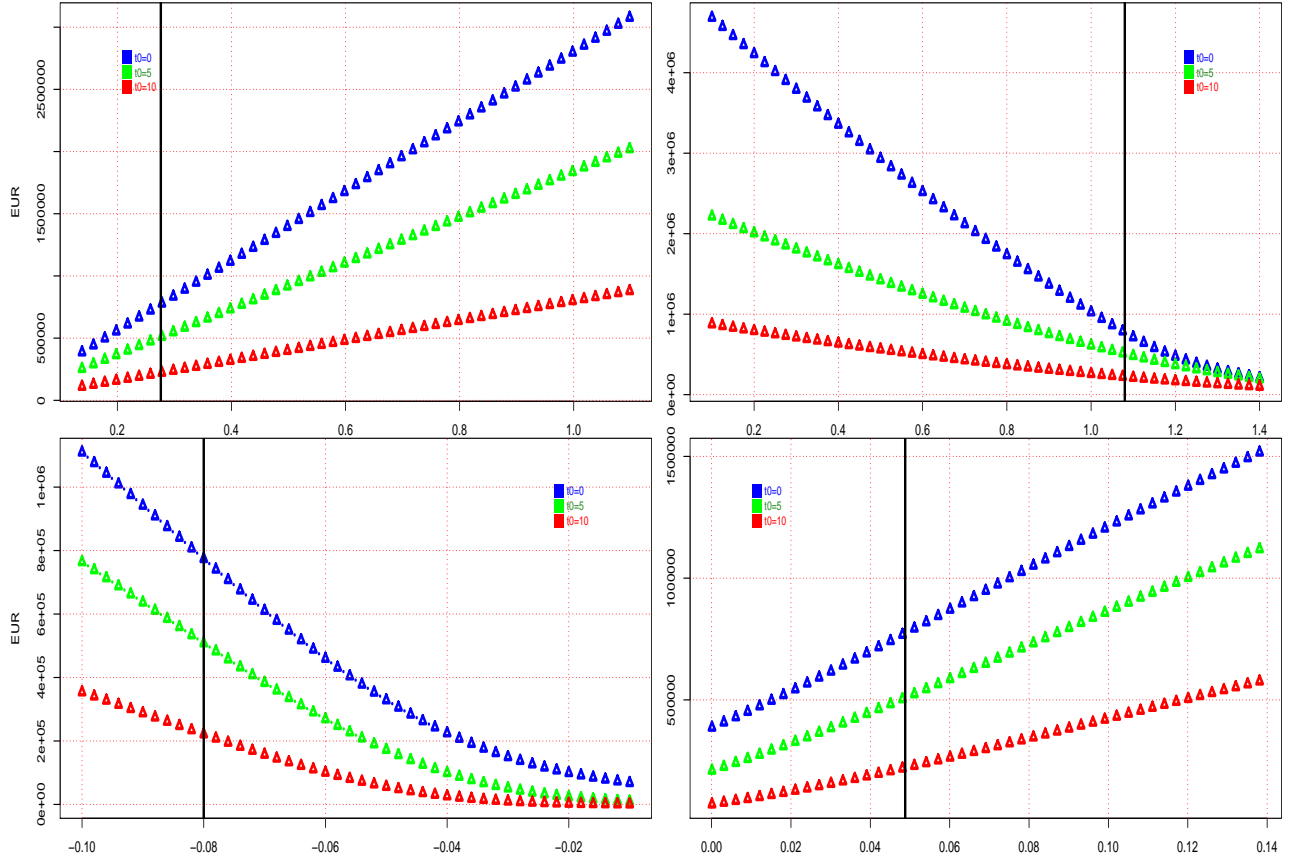


Illustration 4: Sensitivity analysis of the provision mean.

As expected only μ and γ allow the provision mean to reduce (the provision mean decreases when the parameter increases), otherwise the provision mean increases when r and λ_2 increase.

5 Proof

5.1 Proof of Proposition 2.2

Let us consider borrower j having drawn down his loan at date I^j for an amount M^j and a term D^j . The constant rate of return is denoted by \tilde{r} . So:

$$(1 + \tilde{r})^{t^*} = e^{rt^*}. \quad (5.1)$$

From [3], pages 172 to 178, we obtain:

1. Capital repaid by borrower j at date t is denoted by C_t^j and defined by:

$$C_t^j := \begin{cases} (e^{rt^*} - 1)M^j \left(\frac{e^{r(t-I^j-t^*)}}{e^{rD^j} - 1} \right), & \text{if } I^j < t \leq I^j + D^j, \\ 0, & \text{if } t = I^j. \end{cases} \quad (5.2)$$

2. Interest repaid by borrower j at date t is denoted by U_t^j and defined by:

$$U_t^j := \begin{cases} (e^{rt^*} - 1)M^j \left(\frac{e^{rD^j} - e^{r(t-I^j-t^*)}}{e^{rD^j} - 1} \right), & \text{if } I^j < t \leq I^j + D^j, \\ 0, & \text{if } t = I^j. \end{cases} \quad (5.3)$$

3. Monthly installment paid by borrower j at date t is denoted by ME_t^j and defined by:

$$ME_t^j := \begin{cases} (e^{rt^*} - 1)M^j \left(\frac{e^{rD^j}}{e^{rD^j} - 1} \right), & \text{if } I^j < t \leq I^j + D^j, \\ 0, & \text{if } t = I^j. \end{cases} \quad (5.4)$$

Now we assume that this borrower defaults at the deterministic date T . The assumption (A_4) implies that the bank reports the claim immediately. Definition 2.1, formulae (5.1), (5.2), (5.3), and (5.4) allow us to determine the different factors making up the claim amount:

1. The unpaid capital from date T until the loan maturity, $I^j + D^j$, is:

$$\sum_{t=T}^{I^j+D^j} C_t^j = \begin{cases} M^j \left(\frac{e^{rD^j} - e^{r(T-I^j-t^*)}}{e^{rD^j} - 1} \right), & \text{if } I^j < T \leq I^j + D^j, \\ M^j, & \text{if } T = I^j. \end{cases} \quad (5.5)$$

2. The unpaid interest for the monthly installment at date T is:

$$U_T^j = \begin{cases} (e^{rt^*} - 1)M^j \left(\frac{e^{rD^j} - e^{r(T-I^j-t^*)}}{e^{rD^j} - 1} \right), & \text{if } I^j < T \leq I^j + D^j, \\ 0, & \text{if } T = I^j. \end{cases} \quad (5.6)$$

3. Any missed monthly installments will incur penalty interest charges proportional to the delay. This penalty interest is calculated from the date of the last payment until the claim is submitted. The penalty interest is equal to the monthly interest multiplied by a coefficient. The multiplicative coefficient due to late payment interest is fixed for every borrower and this coefficient will not change. It will be denoted by p (“ p ” for “penalty”). The additional interest on the unpaid monthly installments only is:

$$t^*pU_T^j = \begin{cases} t^*p(e^{rt^*} - 1)M^j \left(\frac{e^{rD^j} - e^{r(T-I^j-t^*)}}{e^{rD^j} - 1} \right), & \text{if } I^j < T \leq I^j + D^j, \\ 0, & \text{if } T = I^j. \end{cases} \quad (5.7)$$

The additional interest is multiplied by the number of months that the monthly payment remains unpaid: in this case we multiply by t^* because the duration of unpaid interest until the reporting date is one month (from date $T - t^*$ to date T) according to the assumption (A_5) .

4. The recoveries received at date T is:

$$\gamma M^j R_{T-I^j}, \quad (5.8)$$

where γ is defined by (2.33). See assumption (A_4) for more details.

By combining (5.5), (5.6), and (5.7) we prove (2.12). Formula (2.13) is given by (5.8).

5.2 Proof of Theorem 3.2

The proof of Theorem 3.2 is based on the following lemma:

Lemma 5.1. *Let $K \in \mathbb{R}_+$. We consider $(\alpha, \beta) \in \mathbb{R}^2$ and $(B_t)_{t \geq 0}$ a standard Brownian motion. Then:*

$$\mathbb{E} \left[\left(K - \exp[\alpha B_t + \beta t] \right)_+ \right] = F_1(t, K, \beta, \alpha), \quad (5.9)$$

where the function F_1 is defined by (3.1).

Proof of Lemma 5.1. Analysis similar to that in [6] page 303 shows that (5.9) is proved. \square

Having disposed of this preliminary step, we can now prove the Theorem 3.2. Let assumptions (A_1) to (A_8) hold. We assume that $(R_t)_{t \geq 0}$ is the geometric Brownian motion defined by (2.10). $(T_{i,j}, D^j)_{j \in B}$ is a PPP,

so given $N_i(A_{t_0+}^i) = k$ the couples $(T_{i,j}, D^j)_{1 \leq j \leq k}$ are i.i.d., with law $\frac{1}{\Lambda_i(A_{t_0+}^i)} \Lambda_i|_{A_{t_0+}^i}$, where N_i is defined by (3.9). So:

$$\begin{aligned} \mathbb{E}[P_{t_0}^i] &= \sum_{k \geq 0} \mathbb{P}[N_i(A_{t_0+}^i) = k] \mathbb{E}\left[\sum_{j=1}^k S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j}-t^*i})\right] \\ &= \mathbb{E}[N_i(A_{t_0+}^i)] \mathbb{E}\left[S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1}-t^*i})\right] \end{aligned} \quad (5.10)$$

$$= \mathbb{E}[N_i(A_{t_0+}^i)] \mathbb{E}[M^1] \int_{A_{t_0+}^i} \mathbb{E}\left[\left\{\phi_{t-t^*i}(d) - \gamma R_{t-t^*i}\right\}_+\right] \frac{d\Lambda_i(t, d)}{\Lambda_i(A_{t_0+}^i)}, \quad (5.11)$$

where $\phi(\cdot)$ is defined by (2.14). By combining (5.11) and (5.9) we have:

$$\mathbb{E}[P_{t_0}^i] = \gamma \mathbb{E}[M^1] \int_{A_{t_0+}^i} F_1\left(t - t^*i, \frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma\right) d\Lambda_i(t, d).$$

This establishes (3.2). It follows easily that formula (3.3) is proved.

5.3 Proof of Theorem 3.4

Let assumptions (A_1) to (A_8) hold. Our proof starts with a technical lemma:

Lemma 5.2. *Let assumptions (A_1) to (A_8) hold.*

1. For all $j \neq l$ we have:

$$\mathbb{E}\left[S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j}-t^*i}) S(T_{i,l} - t^*i, D^l, M^l, R_{T_{i,l}-t^*i})\right] = C^{i,i}, \quad (5.12)$$

where:

$$C^{i,i'} := \left[\mathbb{E}[M^1]\right]^2 \mathbb{E}\left[h\left(T_{i,1} - t^*i, D^1, T_{i',2} - t^*i', D^2\right)\right]. \quad (5.13)$$

For all $i \neq i'$ we have:

$$\mathbb{E}\left[S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j}-t^*i}) S(T_{i',l} - t^*i', D^l, M^l, R_{T_{i',l}-t^*i'})\right] = C^{i,i'}. \quad (5.14)$$

2. Let $K \in \mathbb{R}_+$. Consider $(\alpha, \beta) \in \mathbb{R}^2$ and $(B_t)_{t \geq 0}$ a standard Brownian motion. Then:

$$\mathbb{E}\left[\left\{\left(K - \exp[\alpha B_t + \beta t]\right)_+\right\}^2\right] = 2KF_1(t, K, \beta, \alpha) - F_1(t, K^2, 2\beta, 2\alpha), \quad (5.15)$$

where F_1 is defined by (3.1).

3. We have:

$$\begin{aligned} \mathbb{V}[P_{t_0}^i] &= \mathbb{V}\left[N_i(A_{t_0+}^i)\right] \mathbb{E}\left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1}-t^*i})\right)^2\right] \\ &\quad + \left[\mathbb{E}[N_i(A_{t_0+}^i)]\right]^2 \left\{C^{i,i} - \left[\mathbb{E}[S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1}-t^*i})]\right]^2\right\}, \end{aligned} \quad (5.16)$$

where $C^{i,i}$ is defined by (5.13).

4. If $(R_t)_{t \geq 0}$ is the geometric Brownian motion defined by (2.10) then:

$$\mathbb{E}\left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1}-t^*i})\right)^2\right] = \gamma^2 \mathbb{E}\left[(M^1)^2\right] \int_{A_{t_0+}^i} F_2(t - t^*i, d) \frac{d\Lambda_i(t, d)}{\Lambda_i(A_{t_0+}^i)}, \quad (5.17)$$

$$\begin{aligned} C^{i,i'} &= \gamma^2 \left[\mathbb{E}[M^1]\right]^2 \int_{A_{t_0+}^i \times A_{t_0+}^{i'}} \left\{ \int_{-\infty}^{\tilde{d}\left(\frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma, t-t^*i\right)} F_3\left(t - t^*i, d, t' - t^*i', d', x\right) dx \right\} \\ &\quad \frac{d\Lambda_i(t, d) d\Lambda_{i'}(t', d')}{\Lambda_i(A_{t_0+}^i) \Lambda_{i'}(A_{t_0+}^{i'})}, \end{aligned} \quad (5.18)$$

where functions \tilde{d} , F_2 , and F_3 are defined by (3.4), (3.5), and (3.6).

The proof of this lemma will be made at the end of this section. Having disposed of this preliminary step, we can now prove the Theorem 3.4. We want to calculate the variance of the monthly provision. By combining (5.16), (3.2), (5.17), and (5.18) we obtain:

$$\begin{aligned} \mathbb{V}\left[P_{t_0}^i\right] &= \gamma^2 \mathbb{E}\left[(M^1)^2\right] \int_{A_{t_0+}^i} F_2\left(t-t^*i, d\right) d\Lambda_i(t, d) + \gamma^2 \left[\mathbb{E}\left[M^1\right]\right]^2 \\ &\quad \times \left[\int_{(A_{t_0+}^i)^2} \left\{ \int_{-\infty}^{\bar{d}\left(\frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma, t-t^*i\right)} F_3\left(t-t^*i, d, t'-t^*i, d', x\right) dx \right\} \right. \\ &\quad \left. d\Lambda_i(t, d) d\Lambda_i(t', d') - \left\{ \int_{A_{t_0+}^i} F_1\left(t-t^*i, \frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma\right) d\Lambda_i(t, d) \right\}^2 \right]. \end{aligned}$$

This establishes (3.7). Let us come back to the proof of Lemma 5.2.

Proof of Lemma 5.2.

1. By combining (2.15) and (3.11) we obtain that $\mathbb{E}\left[S(T_{i,j}-t^*i, D^j, M^j, R_{T_{i,j}-t^*i})S(T_{i,l}-t^*i, D^l, M^l, R_{T_{i,l}-t^*i})\right]$ is equal to:

$$\begin{aligned} &\mathbb{E}\left[M^j M^l \left\{ \phi_{T_{i,j}-t^*i}(D^j) - \gamma R_{T_{i,j}-t^*i} \right\}_+ \left\{ \phi_{T_{i,l}-t^*i}(D^l) - \gamma R_{T_{i,l}-t^*i} \right\}_+\right] \\ &= \left[\mathbb{E}\left[M^1\right]\right]^2 \mathbb{E}\left[\left\{ \phi_{T_{i,j}-t^*i}(D^j) - \gamma R_{T_{i,j}-t^*i} \right\}_+ \left\{ \phi_{T_{i,l}-t^*i}(D^l) - \gamma R_{T_{i,l}-t^*i} \right\}_+\right] \\ &= \left[\mathbb{E}\left[M^1\right]\right]^2 \mathbb{E}\left[h\left(T_{i,1}-t^*i, D^1, T_{i,2}-t^*i, D^2\right)\right] = C^{i,i}, \end{aligned}$$

as $(T_{i,j}, D^j)_{j \in B}$ is a PPP then the r.v.'s $(T_{i,j}, D^j)_{j \geq 1}$ are i.i.d.. This establishes (5.12).

In the same manner $\mathbb{E}\left[S(T_{i,j}-t^*i, D^j, M^j, R_{T_{i,j}-t^*i})S(T_{i',l}-t^*i', D^l, M^l, R_{T_{i',l}-t^*i'})\right]$ is equal to:

$$\begin{aligned} &\mathbb{E}\left[M^j M^l \left\{ \phi_{T_{i,j}-t^*i}(D^j) - \gamma R_{T_{i,j}-t^*i} \right\}_+ \left\{ \phi_{T_{i',l}-t^*i'}(D^l) - \gamma R_{T_{i',l}-t^*i'} \right\}_+\right] \\ &= \left[\mathbb{E}\left[M^1\right]\right]^2 \mathbb{E}\left[\left\{ \phi_{T_{i,j}-t^*i}(D^j) - \gamma R_{T_{i,j}-t^*i} \right\}_+ \left\{ \phi_{T_{i',l}-t^*i'}(D^l) - \gamma R_{T_{i',l}-t^*i'} \right\}_+\right] \\ &= \left[\mathbb{E}\left[M^1\right]\right]^2 \mathbb{E}\left[h\left(T_{i,1}-t^*i, D^1, T_{i',2}-t^*i', D^2\right)\right] = C^{i,i'}. \end{aligned}$$

This establishes the formula (5.14).

2. If $x \geq 0$ then:

$$\left((K-x)_+\right)^2 = 2K(K-x)_+ - (K^2-x^2)_+.$$

Applying Lemma 5.1 and the last observation we can conclude:

$$\begin{aligned} \mathbb{E}\left[\left\{\left(K - \exp[\alpha B_t + \beta t]\right)_+\right\}^2\right] &= 2K \mathbb{E}\left[\left(K - \exp[\alpha B_t + \beta t]\right)_+\right] - \mathbb{E}\left[\left(K^2 - \exp[2\alpha B_t + 2\beta t]\right)_+\right], \\ &= 2K F_1\left(t, K, \beta, \alpha\right) - F_1\left(t, K^2, 2\beta, 2\alpha\right). \end{aligned}$$

This establishes the formula (5.15).

3. First we calculate the second moment of the monthly provision. Then we can deduce the variance of the monthly provision. We have:

$$\begin{aligned} \mathbb{E}\left[\left(P_{t_0}^i\right)^2\right] &= \sum_{k \geq 1} \mathbb{P}\left[N_i(A_{t_0+}^i) = k\right] \mathbb{E}\left[\left\{\sum_{j=1}^k S(T_{i,j}-t^*i, D^j, M^j, R_{T_{i,j}-t^*i})\right\}^2\right] \\ &= A + B, \end{aligned} \tag{5.19}$$

where:

$$A := \sum_{k \geq 1} \mathbb{P}\left[N_i(A_{t_0+}^i) = k\right] \mathbb{E}\left[\sum_{j=1}^k \left(S(T_{i,j}-t^*i, D^j, M^j, R_{T_{i,j}-t^*i})\right)^2\right],$$

and:

$$B := \sum_{k \geq 1} \mathbb{P} \left[N_i(A_{t_0+}^i) = k \right] \mathbb{E} \left[2 \sum_{1 \leq j < l \leq k} S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j} - t^*i}) S(T_{i,l} - t^*i, D^l, M^l, R_{T_{i,l} - t^*i}) \right].$$

We begin by calculating A . Analysis similar to that in the proof of Theorem 3.2, shows that:

$$\begin{aligned} A &= \sum_{k \geq 1} \mathbb{P} \left[N_i(A_{t_0+}^i) = k \right] \sum_{j=1}^k \mathbb{E} \left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right)^2 \right] \\ &= \sum_{k \geq 1} \mathbb{P} \left[N_i(A_{t_0+}^i) = k \right] k \mathbb{E} \left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right)^2 \right] \\ &= \mathbb{E} \left[N_i(A_{t_0+}^i) \right] \mathbb{E} \left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right)^2 \right]. \end{aligned} \quad (5.20)$$

Now we calculate B . Expression (5.12) implies:

$$\begin{aligned} B &= \sum_{k \geq 1} \mathbb{P} \left[N_i(A_{t_0+}^i) = k \right] \times 2 \sum_{1 \leq j < l \leq k} C^{i,i} \\ &= \sum_{k \geq 1} \mathbb{P} \left[N_i(A_{t_0+}^i) = k \right] k(k-1) C^{i,i} \\ &= \mathbb{E} \left[(N_i(A_{t_0+}^i))^2 \right] C^{i,i} - \mathbb{E} \left[N_i(A_{t_0+}^i) \right] C^{i,i} \\ &= \left[\mathbb{E} \left[N_i(A_{t_0+}^i) \right] \right]^2 C^{i,i}, \end{aligned} \quad (5.21)$$

where $C^{i,i}$ is defined by (5.13). By combining (5.19), (5.20), and (5.21) we obtain:

$$\mathbb{E} \left[(P_{t_0}^i)^2 \right] = \mathbb{E} \left[N_i(A_{t_0+}^i) \right] \mathbb{E} \left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right)^2 \right] + \left[\mathbb{E} \left[N_i(A_{t_0+}^i) \right] \right]^2 C^{i,i}. \quad (5.22)$$

Let us come back to the calculation of the monthly provision variance. Combining (5.10) and (5.22) we conclude that:

$$\begin{aligned} \mathbb{V} \left[P_{t_0}^i \right] &= \mathbb{E} \left[(P_{t_0}^i)^2 \right] - \left[\mathbb{E} \left[P_{t_0}^i \right] \right]^2 \\ &= \mathbb{V} \left[N_i(A_{t_0+}^i) \right] \mathbb{E} \left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right)^2 \right] \\ &\quad + \left[\mathbb{E} \left[N_i(A_{t_0+}^i) \right] \right]^2 \left\{ C^{i,i} - \left(\mathbb{E} \left[S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right] \right)^2 \right\}. \end{aligned}$$

This establishes the formula (5.16).

4. Assumption (A₈) and (2.15) imply:

$$\mathbb{E} \left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right)^2 \right] = \mathbb{E} \left[(M^1)^2 \right] \int_{A_{t_0+}^i} E \left[\left\{ (\phi_{t-t^*i}(d) - \gamma R_{t-t^*i})_+ \right\}^2 \right] \frac{d\Lambda_i(t, d)}{\Lambda_i(A_{t_0+}^i)}.$$

Applying (5.15) and (2.10) we obtain:

$$\mathbb{E} \left[\left(S(T_{i,1} - t^*i, D^1, M^1, R_{T_{i,1} - t^*i}) \right)^2 \right] = \gamma^2 \mathbb{E} \left[(M^1)^2 \right] \int_{A_{t_0+}^i} F_2(t - t^*i, d) \frac{d\Lambda_i(t, d)}{\Lambda_i(A_{t_0+}^i)},$$

where F_2 is defined by (3.5). So formula (5.17) is proved.

Now we want to calculate $C^{i,i'}$ which is defined by (5.13). We start by evaluating $h(t, d, t', d')$, for $t \neq t'$, where h is defined by (3.11). Assume for example $t' > t$. Let $0 < t_1 < \dots < t_n$. Definition of R . (see (2.10)) shows that given $R_{t_1} = x_1, \dots, R_{t_{n-1}} = x_{n-1}$, R_{t_n} has the same law as $x_{n-1} R_{t_n - t_{n-1}}$, i.e.:

$$\mathcal{L} \left[R_{t_n} | R_{t_1} = x_1, \dots, R_{t_{n-1}} = x_{n-1} \right] = \mathcal{L} \left[x_{n-1} R_{t_n - t_{n-1}} \right]. \quad (5.23)$$

In particular,

$$\mathcal{L} \left(R_t, R_{t'} \right) = \mathcal{L} \left(R_t, R'_{t'-t} \right), \quad (5.24)$$

where:

$$R'_{t'-t} := \exp [\sigma(B_{t'} - B_t) + \mu(t' - t)]. \quad (5.25)$$

So for $t' \geq t$:

$$\begin{aligned} h(t, d, t', d') &= \gamma^2 \mathbb{E} \left[\left\{ \frac{\phi_t(d)}{\gamma} - R_t \right\}_+ \left\{ \frac{\phi_{t'}(d')}{\gamma} - R_{t'} \right\}_+ \right] \\ &= \gamma^2 \mathbb{E} \left[\left\{ \frac{\phi_t(d)}{\gamma} - R_t \right\}_+ \left\{ \frac{\phi_{t'}(d')}{\gamma} - R_t R'_{t'-t} \right\}_+ \right] \\ &= \gamma^2 \mathbb{E} \left[\left\{ \frac{\phi_t(d)}{\gamma} - R_t \right\}_+ \tilde{h}(R_t) \right], \end{aligned} \quad (5.26)$$

where:

$$\tilde{h}(z) := z \mathbb{E} \left[\left\{ \frac{\phi_{t'}(d')}{z\gamma} - R_{t'-t} \right\}_+ \right].$$

Formula (5.9) implies:

$$\tilde{h}(z) = z F_1 \left(t' - t, \frac{\phi_{t'}(d')}{z\gamma}, \mu, \sigma \right). \quad (5.27)$$

Let us come back to the calculation of h . Using (3.4), (5.26), and (5.27) we obtain:

$$\begin{aligned} h(t, d, t', d') &= \gamma^2 \mathbb{E} \left[\left\{ \frac{\phi_t(d)}{\gamma} - R_t \right\}_+ R_t F_1 \left(t' - t, \frac{\phi_{t'}(d')}{R_t \gamma}, \mu, \sigma \right) \right] \\ &= \gamma^2 \int_{-\infty}^{\tilde{d}(\frac{\phi_t(d)}{\gamma}, \mu, \sigma, t)} \frac{e^{\mu t + \sigma \sqrt{t} x - \frac{x^2}{2}}}{\sqrt{2\pi}} \left(\frac{\phi_t(d)}{\gamma} - e^{\mu t + \sigma \sqrt{t} x} \right) F_1 \left(t' - t, \frac{\phi_{t'}(d')}{\gamma e^{\mu t + \sigma \sqrt{t} x}}, \mu, \sigma \right) dx \\ &= \int_{-\infty}^{\tilde{d}(\frac{\phi_t(d)}{\gamma}, \mu, \sigma, t)} \gamma^2 F_3 \left(t, d, t', d', x \right) dx, \end{aligned} \quad (5.28)$$

where the function F_3 is defined by (3.6). Formulas (3.10) and (5.28) imply:

$$C^{i, i'} = \gamma^2 \left[\mathbb{E}[M^1] \right]^2 \int_{A_{t_0+}^i \times A_{t_0+}^{i'}} \left\{ \int_{-\infty}^{\tilde{d}(\frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma, t-t^*i)} F_3 \left(t-t^*i, d, t-t^*i', d', x \right) dx \right\} \frac{d\Lambda_i(t, d) d\Lambda_{i'}(t', d')}{\Lambda_i(A_{t_0+}^i) \Lambda_{j'}(A_{t_0+}^{i'})}.$$

This establishes the formula (5.18) and achieves the proof of Lemma 5.2. \square

5.4 Proof of Theorem 3.6

To simplify the proof we will denote:

$$T_{i,j} = T_j.$$

First we order the dates of default $(T_j)_{1 \leq j \leq n}$ as follows:

$$T_{(1)} = T_{j_1} < \dots < T_{(n)} = T_{j_n}, \quad (5.29)$$

where $(j_1, \dots, j_n) \in B^n : j_k \neq j_l \forall k \neq l \in \{1, \dots, n\}$ and with agreement $T_{(0)} = 0$. The term and the amount associated to the loan which defaulted at date $T_{(j)}$ are denoted by $D^{(j)}$ and $M^{(j)}$ respectively. Then, for example, $D^{(1)} = D^{j_1}$.

With Definition 2.5 and Proposition 2.2 we can rewrite the provision for month i :

$$P_{t_0}^i = \sum_{j \geq 1} \mathbb{1}_{\{(T_{(j)}, D^{(j)}) \in A_{t_0+}^i\}} M^{(j)} \left\{ \phi_{T_{(j)} - t^*i}(D^{(j)}) - \gamma R_{T_{(j)} - t^*i} \right\}_+.$$

The identity in law (5.23) implies:

$$\mathcal{L} \left[\left(R_{T_{(1)}}, \dots, R_{T_{(n)}} \right) \right] = \mathcal{L} \left[\left(\exp \left(\sigma \sqrt{T_{(1)}} G^1 + \mu T_{(1)} \right), \dots, \prod_{k=1}^n \exp \left(\sigma \sqrt{T_{(k)} - T_{(k-1)}} G^k + \mu (T_{(k)} - T_{(k-1)}) \right) \right) \right], \quad (5.30)$$

where $(G^j)_{1 \leq j \leq n}$ is a collection of r.v.'s i.i.d. with standard normal distribution and independent of $\left((T_j, D^j)_{1 \leq j \leq n}, (M^j)_{1 \leq j \leq n} \right)$.

So the law of $P_{t_0}^i$ given $N_i(A_{t_0+}^i) = n$ is:

$$\mathcal{L}\left[P_{t_0}^i | N_i(A_{t_0+}^i) = n\right] = \mathcal{L}\left[f_n\left((T_{(j)})_{1 \leq j \leq n}, (D^{(j)})_{1 \leq j \leq n}, (M^{(j)})_{1 \leq j \leq n}, (G^j)_{1 \leq j \leq n}\right)\right],$$

where:

$$f_n\left((t_{(j)})_{1 \leq j \leq n}, (d^{(j)})_{1 \leq j \leq n}, (m^{(j)})_{1 \leq j \leq n}, (g^j)_{1 \leq j \leq n}\right) := \sum_{i=1}^n m^{(j)} \left\{ \phi_{t_{(j)} - t^* i}(d^{(j)}) - \gamma \prod_{k=1}^j r^k \right\}_+,$$

$$r^k = \exp\left[\sigma \sqrt{t_{(k)} - t_{(k-1)}} g^k + \mu(t_{(k)} - t_{(k-1)})\right], \text{ with this agreement: } t_{(0)} = 0.$$

In other way, random subscripts j_1, \dots, j_n defined by (5.29) do not depend on T_1, \dots, T_n . Yet $(M^j)_{1 \leq j \leq n}$ are independent of $(T_j)_{1 \leq j \leq n}$, so:

$$\mathcal{L}\left[(M^{j_1}, \dots, M^{j_n})\right] = \mathcal{L}\left[(M^1, \dots, M^n)\right]. \quad (5.31)$$

Therefore:

$$\mathcal{L}\left[P_{t_0}^i | N_i(A_{t_0+}^i) = n\right] = \mathcal{L}\left[f_n\left((T_{(j)})_{1 \leq j \leq n}, (D^{(j)})_{1 \leq j \leq n}, (M^j)_{1 \leq j \leq n}, (G^j)_{1 \leq j \leq n}\right)\right].$$

5.5 Proof of Proposition 4.2

Our proof starts with a technical lemma:

Lemma 5.3.

1. Consider the collection $(Z_i, Z'_i, Z''_i)_{0 \leq i \leq i_0}$ of r.v.'s satisfying:

- (a) The r.v.'s $\{Z_i, Z'_i, Z''_i, 0 \leq i \leq i_0\}$ are independent,
- (b) The marginal distributions are:

$$\mathcal{L}(Z_i) = \mathcal{P}(f_i(\theta)), \quad \mathcal{L}(Z'_i) = \mathcal{P}(g_i(\theta)), \quad \mathcal{L}(Z''_i) = \mathcal{P}(h_i(\theta)), \quad (5.32)$$

where $f_i, g_i,$ and h_i are $C^1(\mathbb{R}^3)$ functions and $\theta = (\theta_1, \theta_2, \theta_3) \in \mathbb{R}^3$.

Let $(z_i, z'_i, z''_i)_{0 \leq i \leq i_0}$ be a realisation of the collection $(Z_i, Z'_i, Z''_i)_{0 \leq i \leq i_0}$. The three likelihood equations are:

$$\sum_{i=0}^{i_0} \left[z_i \frac{\partial f_i(\theta)}{\partial \theta_k} + z'_i \frac{\partial g_i(\theta)}{\partial \theta_k} + z''_i \frac{\partial h_i(\theta)}{\partial \theta_k} \right] - \sum_{i=0}^{i_0} \frac{\partial}{\partial \theta_k} \left[f_i(\theta) + g_i(\theta) + h_i(\theta) \right] = 0, \quad k = 1, 2, 3. \quad (5.33)$$

2. Let Leb_1 (respectively Leb_2) the Lebesgue measure on \mathbb{R} (respectively on \mathbb{R}^2). So:

$$Leb_1(C_{t_0-}^i) = \sqrt{2}(t_0 - t^*i), \quad (5.34)$$

$$Leb_2(A_{t_0-}^i) = (t_0 - t^*i) \left(\bar{D} - \frac{1}{2}(t_0 - t^*i) \right), \quad (5.35)$$

$$Leb_2(A_{t_0+}^i) = \frac{1}{2}(\bar{D} - t_0 + t^*i)^2, \quad (5.36)$$

$$Leb_1(C_{t_0+}^i) = \sqrt{2}(\bar{D} - t_0 + t^*i), \quad (5.37)$$

where sets $C_{t_0-}^i, A_{t_0-}^i, A_{t_0+}^i,$ and $C_{t_0+}^i$ are defined by (4.2), (2.18), (2.17), and (4.1).

The proof of this lemma will be given at the end of this section. Having disposed of this preliminary step, we can now prove the Proposition 4.2. We denote:

$$\begin{aligned} Z_i &= N_i(C_{t_0-}^i) \text{ and } z_i = v_i, \\ Z'_i &= N_i(A_{t_0-}^i) \text{ and } z'_i = n_i, \\ Z''_i &= N_i(A_{t_0+}^i \cup C_{t_0+}^i) \text{ and } z''_i = b_i - n_i - v_i, \\ \theta &= (\theta_1, \theta_2, \theta_3) = (\lambda_1, \lambda_2, \bar{D}). \end{aligned}$$

Using expressions (3.9), (4.3), and Lemma 5.3 we can deduce that (5.32) is true and we have:

$$f_i(\theta) = \lambda_1 Leb_1(C_{t_0-}^i) = \lambda_1 \sqrt{2}(t_0 - t^*i),$$

$$\begin{aligned}
g_i(\theta) &= \lambda_2 \text{Leb}_2(A_{t_0-}^i) = \lambda_2(t_0 - t^*i) \left(\bar{D} - \frac{1}{2}(t_0 - t^*i) \right), \\
h_i(\theta) &= \lambda_1 \text{Leb}_1(C_{t_0+}^i) + \lambda_2 \text{Leb}_2(A_{t_0+}^i) = \lambda_1 \sqrt{2}(\bar{D} - t_0 + t^*i) + \frac{\lambda_2}{2}(\bar{D} - t_0 + t^*i)^2, \\
f_i(\theta) + g_i(\theta) + h_i(\theta) &= \lambda_1 \sqrt{2} \bar{D} + \frac{\lambda_2}{2}(\bar{D})^2.
\end{aligned}$$

Variables $Z_i = N_i(C_{t_0-}^i)$, $Z'_i = N_i(A_{t_0-}^i)$, and $Z''_i = N_i(A_{t_0+}^i \cup C_{t_0+}^i)$ are independent so we can use Lemma 5.3. The likelihood equations, where z_i, z'_i , and z''_i are exchanged by v_i, n_i , and $b_i - n_i - v_i$ (see Section 4.2), lead to the following system:

$$\begin{aligned}
& \sum_{i=0}^{i_0} \left[v_i \frac{\frac{\partial \lambda_1 \text{Leb}_1(C_{t_0-}^i)}{\partial \theta_k}}{\lambda_1 \text{Leb}_1(C_{t_0-}^i)} + n_i \frac{\frac{\partial \lambda_2 \text{Leb}_2(A_{t_0-}^i)}{\partial \theta_k}}{\lambda_2 \text{Leb}_2(A_{t_0-}^i)} + (b_i - n_i - v_i) \frac{\frac{\partial (\lambda_1 \text{Leb}_1(C_{t_0+}^i) + \lambda_2 \text{Leb}_2(A_{t_0+}^i))}{\partial \theta_k}}{\lambda_1 \text{Leb}_1(C_{t_0+}^i) + \lambda_2 \text{Leb}_2(A_{t_0+}^i)} \right] \\
& - \sum_{i=0}^{i_0} \frac{\partial}{\partial \theta_k} \left[\lambda_1 \text{Leb}_1(C_{t_0-}^i) + \lambda_2 \text{Leb}_2(A_{t_0-}^i) + \lambda_1 \text{Leb}_1(C_{t_0+}^i) + \lambda_2 \text{Leb}_2(A_{t_0+}^i) \right] = 0,
\end{aligned}$$

where $k = 1, 2, 3$. Using Lemma 5.3 the likelihood equations become:

$$\begin{aligned}
& \sum_{i=0}^{i_0} \left[\frac{v_i}{\lambda_1} + \frac{(b_i - n_i - v_i) \sqrt{2}(\bar{D} - t_0 + t^*i)}{\lambda_1 \sqrt{2}(\bar{D} - t_0 + t^*i) + \frac{\lambda_2(\bar{D} - t_0 + t^*i)^2}{2}} \right] = \sqrt{2}(i_0 + 1) \bar{D}, \\
(S) : \sum_{i=0}^{i_0} \left[\frac{n_i}{\lambda_2} + \frac{(b_i - n_i - v_i)(\bar{D} - t_0 + t^*i)^2}{2\lambda_1 \sqrt{2}(\bar{D} - t_0 + t^*i) + \lambda_2(\bar{D} - t_0 + t^*i)^2} \right] &= \frac{(i_0 + 1)(\bar{D})^2}{2}, \\
\sum_{i=0}^{i_0} \left[\frac{n_i}{\left(\bar{D} - \frac{1}{2}(t_0 - t^*i)\right)} + \frac{(b_i - n_i - v_i) \left(\sqrt{2}\lambda_1 + \lambda_2(\bar{D} - t_0 + t^*i)\right)}{\lambda_1 \sqrt{2}(\bar{D} - t_0 + t^*i) + \frac{\lambda_2(\bar{D} - t_0 + t^*i)^2}{2}} \right] &= (\lambda_1 \sqrt{2} + \lambda_2 \bar{D})(i_0 + 1).
\end{aligned}$$

We denote $(\hat{\lambda}_1, \hat{\lambda}_2, \hat{\bar{D}})$ the solution of this system which corresponds to the likelihood maximum estimators of parameters λ_1, λ_2 , and \bar{D} . This achieves the proof of Proposition 4.2.

Proof of Lemma 5.3.

1. The likelihood of $(Z_i, Z'_i, Z''_i)_{0 \leq i \leq i_0}$ is denoted by L_{i_0} . Using the independence of variables $\{Z_i, Z'_i, Z''_i, 0 \leq i \leq i_0\}$ we obtain:

$$L_{i_0} = \frac{\prod_{i=0}^{i_0} [f_i(\theta)]^{z_i} [g_i(\theta)]^{z'_i} [h_i(\theta)]^{z''_i}}{\prod_{i=0}^{i_0} z_i! z'_i! z''_i!} \exp \left[- \sum_{i=0}^{i_0} (f_i(\theta) + g_i(\theta) + h_i(\theta)) \right],$$

where z_i, z'_i , and z''_i are contained in \mathbb{N} . So:

$$\ln(L_{i_0}) = - \ln \left(\prod_{i=0}^{i_0} z_i! z'_i! z''_i! \right) + \sum_{i=0}^{i_0} \left[z_i \ln [f_i(\theta)] + z'_i \ln [g_i(\theta)] + z''_i \ln [h_i(\theta)] \right] - \sum_{i=0}^{i_0} [f_i(\theta) + g_i(\theta) + h_i(\theta)]. \quad (5.38)$$

The likelihood equations are:

$$\frac{\partial}{\partial \theta_k} \left(\ln(L_{i_0}) \right) = 0, \quad k = 1, 2, 3. \quad (5.39)$$

Combining (5.38) and (5.39) the likelihood equations lead to this system:

$$\sum_{i=0}^{i_0} \left[z_i \frac{\frac{\partial f_i(\theta)}{\partial \theta_k}}{f_i(\theta)} + z'_i \frac{\frac{\partial g_i(\theta)}{\partial \theta_k}}{g_i(\theta)} + z''_i \frac{\frac{\partial h_i(\theta)}{\partial \theta_k}}{h_i(\theta)} \right] - \sum_{i=0}^{i_0} \frac{\partial}{\partial \theta_k} [f_i(\theta) + g_i(\theta) + h_i(\theta)] = 0, \quad k = 1, 2, 3.$$

2. The proof is straightforward with definitions of $C_{t_0-}^i, A_{t_0-}^i, A_{t_0+}^i$, and $C_{t_0+}^i$.

□

6 Appendix

6.1 Graphical illustration

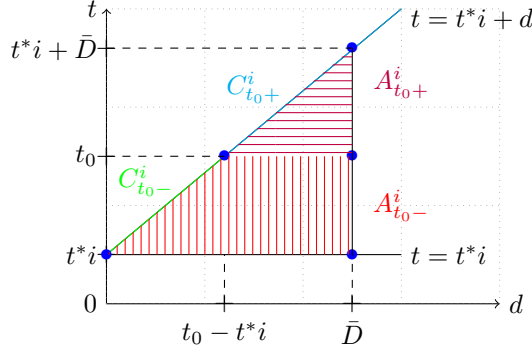


Illustration 5: Illustration of sets $A_{t_0-}^i, A_{t_0+}^i, C_{t_0-}^i$, and $C_{t_0+}^i$.

6.2 Variance of the total provision

We want to calculate the total provision variance.

$$\mathbb{V}[P_{t_0}] = \sum_{i=0}^{\lfloor \frac{t_0}{t^*} \rfloor} \mathbb{V}[P_{t_0}^i] + 2 \sum_{0 \leq i < i' \leq \lfloor \frac{t_0}{t^*} \rfloor} \left\{ \mathbb{E}[P_{t_0}^i P_{t_0}^{i'}] - \mathbb{E}[P_{t_0}^i] \mathbb{E}[P_{t_0}^{i'}] \right\}. \quad (6.1)$$

We calculate $\mathbb{E}[P_{t_0}^i P_{t_0}^{i'}]$.

$$\begin{aligned} \mathbb{E}[P_{t_0}^i P_{t_0}^{i'}] &= \mathbb{E} \left[\left\{ \sum_{j \geq 1} \mathbb{1}_{\{(T_{i,j}, D^j) \in A_{t_0+}^i\}} S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j} - t^*i}) \right\} \right. \\ &\quad \times \left. \left\{ \sum_{l \geq 1} \mathbb{1}_{\{(T_{i',l}, D^l) \in A_{t_0+}^{i'}\}} S(T_{i',l} - t^*i', D^l, M^l, R_{T_{i',l} - t^*i'}) \right\} \right] \\ &= \sum_{k \geq 0, h \geq 0} \left\{ \mathbb{P}[N_i(A_{t_0+}^i) = k] \mathbb{P}[N_{i'}(A_{t_0+}^{i'}) = h] \right\} \\ &\quad \times \sum_{1 \leq j \leq k, 1 \leq l \leq h} \mathbb{E} \left[S(T_{i,j} - t^*i, D^j, M^j, R_{T_{i,j} - t^*i}) S(T_{i',l} - t^*i', D^l, M^l, R_{T_{i',l} - t^*i'}) \right] \\ &= \sum_{k \geq 0, h \geq 0} \left\{ \mathbb{P}[N_i(A_{t_0+}^i) = k] \mathbb{P}[N_{i'}(A_{t_0+}^{i'}) = h] \right\} C^{i,i'} k h = C^{i,i'} \mathbb{E}[N_i(A_{t_0+}^i)] \mathbb{E}[N_{i'}(A_{t_0+}^{i'})]. \quad (6.2) \end{aligned}$$

Combining (6.1), (6.2), (5.18), and (3.2) we have:

$$\begin{aligned} \mathbb{V}[P_{t_0}] &= \sum_{i=0}^{\lfloor \frac{t_0}{t^*} \rfloor} \mathbb{V}[P_{t_0}^i] + 2 \sum_{0 \leq i < i' \leq \lfloor \frac{t_0}{t^*} \rfloor} \gamma^2 [\mathbb{E}[M^1]]^2 \left\{ \int_{A_{t_0+}^i \times A_{t_0+}^{i'}} \left\{ \int_{-\infty}^{\tilde{d}(\frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma, t-t^*i)} \right. \right. \\ &\quad \left. \left. F_3 \left(t - t^*i, d, t' - t^*i', d', x \right) dx \right\} d\Lambda_i(t, d) d\Lambda_{i'}(t', d') - \int_{A_{t_0+}^i} F_1 \left(t - t^*i, \frac{\phi_{t-t^*i}(d)}{\gamma}, \mu, \sigma \right) d\Lambda_i(t, d) \right. \\ &\quad \left. \times \int_{A_{t_0+}^{i'}} F_1 \left(t - t^*i', \frac{\phi_{t-t^*i'}(d)}{\gamma}, \mu, \sigma \right) d\Lambda_{i'}(t, d) \right\}, \end{aligned}$$

where $\mathbb{V}[P_{t_0}^i]$ is obtained by (3.7).

6.3 Simulation algorithm

Algorithm 1 Simulation of a sample of length l for the r.v. $P_{t_0}^i$.

- 1: **for** $k = 1$ to l **do**
 - 2: Simulate a r.v. with distribution $\mathcal{P}\left(\hat{\lambda}_2 Leb_2(\hat{A}_{t_0+}^i)\right)$. Its realization is denoted by n .
 - 3: Simulate n r.v.'s i.i.d. uniformly distributed on $\hat{A}_{t_0+}^i$. The realizations are denoted by $(t_j, d^j)_{1 \leq j \leq n}$. Order dates $(t_j)_{1 \leq j \leq n}$: $t_{(1)} \leq \dots \leq t_{(n)}$.
 - 4: The variables $(M^j)_{1 \leq j \leq n}$ are simulated by reversing the empirical c.d.f.. The realizations are denoted by $(\tilde{m}^j)_{1 \leq j \leq n}$.
 - 5: Simulate n r.v.'s i.i.d. with standard normal distribution. The realizations are denoted by $(g^j)_{1 \leq j \leq n}$.
 - 6: Calculate: $p_k^i = \sum_{i=1}^n \tilde{m}^j \left\{ \phi_{t_{(j)} - t^* i}(d^{(j)}) - \gamma \prod_{k=1}^j r^k \right\}_+$, where ϕ and r^k are defined by (2.14) and (3.15).
 - 7: **end for**
-

6.4 Scatter plot

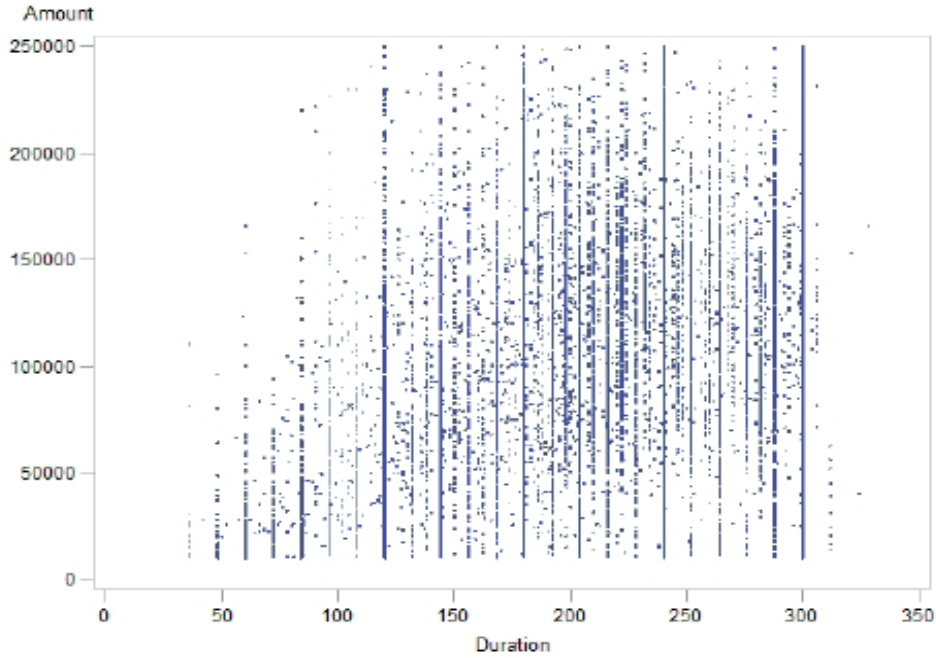


Illustration 6: Scatter plot of the borrowed amount and the loan duration.

References

- [1] Elja Arjas. The claims reserving problem in non-life insurance: some structural ideas. *Astin Bulletin*, Vol. 19:139 – 152, 1989.
- [2] Erhan Çinlar. *Introduction to stochastic processes*. Prentice-Hall Inc., Englewood Cliffs, N.J., 1975.
- [3] Anne Claude and Gérard Chaigneau. *Mathématiques financières*. Ellipses Marketing, 2010.
- [4] Christiane Coccozza-Thivent. *Processus stochastiques et fiabilité des systèmes*, volume 28 of *Mathématiques & Applications (Berlin) [Mathematics & Applications]*. Springer-Verlag, Berlin, 1997.
- [5] ACPR Banque de France. Housing finance in france in 2013. *ACPR - Banque de France - Analyses et Synthèses*, 2014.
- [6] John Hull. *Options, Futures and Other Derivatives*. Prentice Hall finance series. Pearson/Prentice Hall, 2009.

- [7] Christian Røhølte Larsen. An individual claims reserving model. Astin Bulletin, Vol. 37:95 – 115, 2007.
- [8] Thomas Mack. Distribution-free calculation of the standard error of chain ladder reserve estimates. Astin Bulletin, Vol. 23:213 – 25, 1993.
- [9] Thomas Mack. The standard error of chain ladder reserve estimates: Recursive calculation and inclusion of a tail factor. Astin Bulletin, Vol. 29(2):361 – 266, 1999.
- [10] Geoffrey Nichil. Provisionnement en assurance non vie pour des contrats à maturité longue et à prime unique - application à la réforme solvabilité 2. Thèse - Université de Lorraine, 2014.
- [11] Ragnar Norberg. Prediction of outstanding liabilities in non life insurance. Astin Bulletin, Vol. 23:95 – 115, 1989.
- [12] Mathieu Pigeon, Katrien Antonio, and Michel Denuit. Individual loss reserving with the multivariate skew normal distribution. Open access publications from katholieke universiteit leuven, Katholieke Universiteit Leuven, December 2012.
- [13] Tomasz Rolski, Hanspeter Schmidli, Volker Schmidt, and Jozef Teugels. Stochastic processes for insurance and finance. Wiley Series in Probability and Statistics. John Wiley & Sons Ltd., Chichester, 1999.
- [14] Gref Taylor, Grainne McGuire, James Sullivan, and University of Melbourne. Centre for Actuarial Studies. Individual Claim Loss Reserving Conditioned by Case Estimates. Research paper. Centre for Actuarial Studies, Department of Economics, University of Melbourne, 2007.
- [15] Robert Van Order. Modeling the credit risk of mortgage loans: a primer. 2007.
- [16] Mario V. Wüthrich and Michel Merz. Stochastic Claims Reserving Methods in Insurance. The Wiley Finance Series. Wiley, 2008.

List of notations

t_0	Analysis date	1
t^*	Time unit, $t^* = 1$ month	2
B	Borrower set	2
D^j	Loan duration of the borrower j	2
I^j	Drawdown date of the borrower j	2
M^j	Loan amount of borrower j	2
\bar{D}	Maximal loan duration	3
r	Exponential rate of the loans	3
$(B_t)_{t \geq 0}$	Standard Brownian motion	3
$(R_t)_{t \geq 0}$	Geometric Brownian motion	3
μ and σ	Parameters of the geometric Brownian motion $(R_t)_{t \geq 0}$	3
$S(t, d, m, r_t)$	Claim amount submitted at date t	4
$T_{i,j}$	Maturity of the loan drawn down the month i by the borrower j	4
$A_{t_0+}^i$ (resp. $A_{t_0-}^i$)	Set in which a claim can occur in the period of time $]t_0, t^*i + \bar{D}[$ (resp. $[t^*i, t_0]$)	4
Λ_i	Intensity of the PPP $(T_{i,j}, D^j)_{j \in B}$	5
$P_{t_0}^i$	Loss provision of month i	5
P_{t_0}	Total provision	5
F_1	Function participant to the provision mean calculation	7
Φ	C.d.f. of the standard normal distribution	7
F_2, F_3, \tilde{d}	Functions participant to the provision variance calculation	8
$N_i(A)$	Number of claims in a set A	8
$C_{t_0+}^i$ (resp. $C_{t_0-}^i$)	Set of loans which have matured after (resp. before) t_0	9
λ_1, λ_2	Parameters intervening in the measure Λ_i	10

d^j (resp. m^j)	Observation of the loan term (resp. amount) of borrower j	10
$t_{i,j}$	Observation of the loan maturity of borrower j	10
b	Observation of the number of loans drawn down prior to t_0	10
b_i	Observation of the number of loans drawn down the month i	10
n_i	Observation of the number of default submitted prior to t_0	10
v_i	Observation of the number of loans which have matured prior to t_0	10
p	Coefficient due to late payment interest.....	14

Author's current address:

1. Pierre Vallois, Institut Elie Cartan de Lorraine, UMR 7502, Université de Lorraine, B.P. 239, 54506 Vandoeuvre-lès-Nancy Cedex, France. Mail: Pierre.Vallois@univ-lorraine.fr.
2. Geoffrey Nichil, Institut Elie Cartan de Lorraine, UMR 7502, Université de Lorraine, B.P. 239, 54506 Vandoeuvre-lès-Nancy Cedex, France. Mail: Geoffrey.Nichil@gmail.com.