Research Paper

A consumer credit risk structural model based on affordability: balance at risk

Marcelo S. Perlin, Marcelo B. Righi and Tiago P. Filomena

Escola de Administração, Universidade Federal do Rio Grande do Sul, Rua Washington Luiz, 855 – Centro Histórico de Porto Alegre – RS, 90010-460 Brazil; emails: marcelo.perlin@ufrgs.br, marcelo.righi@ufrgs.br, tiago.filomena@ufrgs.br

(Received August 3, 2017; revised June 11, 2018; accepted August 20, 2018)

ABSTRACT

This paper introduces an approach designed for personal credit risk, with possible applications in risk assessment and optimization of debt contracts. We define a structural model related to the financial balance of an individual, allowing for cashflow seasonality and deterministic trends in the process. Based on the proposed model, we develop risk measures associated with the probability of default rates conditional on time. This formulation is best suited to short-term loans, where the dynamics of individuals’ cashflow, such as seasonality and uncertainty, can significantly impact future default rates. In the empirical section of this paper, we illustrate an application by estimating risk measures using simulated data. We also present the specific case of optimization of a financial contract, where, based on an estimated model, we find the yield rate/time to maturity pair that maximizes the expected profit or minimizes the default risk of a short-term debt contract.

Keywords: credit risk; balance at risk (BaR); personal finance; cashflow.
1 INTRODUCTION

The assessment of consumer credit quality is an important aspect in banking and finance (Thomas 2000). The existence of a developed financial market allows an individual to sell future cashflows in exchange for present purchases of goods, such as a TV, car or real estate (Shiller 2013). Over the years, a higher demand for consumer credit products and competitive incentives for cost minimization in credit analysis have motivated a transition from the subjective evaluation of credit risk to the use of quantitative models for so-called credit and behavior scoring (Crook et al 2007; Hand and Henley 1997; Thomas 2000).

Traditional models for consumer credit risk assessment tend to rely on a broad range of techniques, such as artificial neural networks (Khashman 2010; Oreski et al 2012), support vector machines (Huang et al 2007), logistic regressions (Crook and Bellotti 2010; Wiginton 1980), decision trees (Matuszyk et al 2010) and mathematical programming (Crook et al 2007), to cite some examples. Another, less explored approach to consumer credit risk assessment is a structural model, which is more oriented toward obtaining the probability of default (PD) of a loan.

The first structural model for credit risk was an option-based approach proposed by Merton (1974) that focused on modeling the stochastic dynamics of a firm’s value. If the firm’s value falls below its debt value on its maturity date, then the firm is in default. This phenomenon is equivalent to stating that equity holders did not exercise their call option on a company’s assets: the option expired out-of-the-money. Thus, the PD on the loan equals the probability of the option being out-of-the-money on the expiration date (Allen et al 2004). Since Merton’s seminal work, other corporate structural credit risk models have appeared, such as Longstaff and Schwartz (1995), Leland and Toft (1996) and Collin-Dufresne and Goldstein (2001).

Despite the developments related to structural corporate credit risk models, not much attention has been paid to structural models for consumer credit risk. Perli and Nayda (2004) discuss an option-based model that is similar to one for corporate credit risk: if a consumer’s assets are lower than a threshold, then they will default on the loan. De Andrade and Thomas (2007) propose a structural model in which the PD is based on the consumer’s reputation. In their model, the consumer has a call option on the value of their reputation, and the strike price of this option is the debt value. The credit/behavior score is the proxy for the value of the reputation. Thomas (2009) classifies this model as a consumer structural model on the basis of reputation.

---

1 Reviews of other techniques can be found in Baesens et al (2003) and Lessmann et al (2013).
2 Reduced-form models, in which default is modeled exogenously, are also alternatives to structural models (see Jarrow et al 2004).
Our approach is a consumer structural model based on affordability (Thomas 2009), which differs not only from Perli and Nayda (2004) but also from De Andrade and Thomas (2007). We model directly the balance of the individual client by explicitly defining a stochastic process for their income and expenses. For the purposes of this paper, we define an applicant’s balance as the available funds in their bank account. This is not a strict definition; one can clearly define any asset as part of the balance, as long as it is liquid enough to be cashed without price restrictions. This or other definitions can be adapted to the use of our proposed model.

This approach allows for greater flexibility in simulating the scenarios and a more realistic assessment of the PD because it does not require any characteristic information from the applicant. The method is best suited to short-term loans, for which the distribution of cashflows over time can significantly impact the default rate of the debt. The model is easily justified because a better assessment of the applicants’ default rate is very important, given that short-term loans can improve the long-term relationship between a bank and its clients (Bodenhorn 2003) as well as the financial stability of the economy (Wagner and Marsh 2006).

Further, as discussed by Thomas (2010) and Crook and Bellotti (2010), introducing economics and market conditions into consumer risk assessment is still a challenge. Thus, we also contribute to the literature by directly considering the conditions of the economy, such as unemployment and wage levels. We call our model a balance-at-risk (BaR) model.

Note that there is a literature combining transaction data with credit bureau information based on a variety of machine learning techniques (Butaru et al 2016; Khandoani et al 2010; Oreski and Oreski 2014; Oreski et al 2012). These studies are reliant on finding patterns in different sets of data to make inferences about consumer credit ratings. Our BaR model, however, is organized around the consumer’s account balance, which means it is closely related to corporate cashflow-at-risk (CFaR) models. This type of model is a downside risk metric based on past cashflow distributions used to evaluate corporate risks (Andrén et al 2005; Oral and Akkaya 2015; Stein et al 2001).

The theory that supports this proposal, along with derivations of risk measures, is provided in this paper. Using an artificial data set, we illustrate the model’s usefulness by first estimating an empirical model and then using it to calculate forward PDs that are conditional on time. We also show how the model can be used in a loan application by finding the parameters of a debt contract that maximize the expected return of the financial transaction.

This paper is organized as follows. In Section 2, we discuss the theory that supports the empirical BaR model. In Section 3, an application model is examined. In Section 4, we finish the paper with some concluding remarks.
2 BALANCE AT RISK

In this section, we present our proposed approach, BaR, to measure and manage credit risk in a mathematical and theoretical framework. Unless otherwise stated, the content is based on the following notation. Consider the filtered atom-less probability space \( \mathcal{X}_T := \mathcal{X}(\Omega, \mathcal{F}, (\mathcal{F}_T)_{T \in T}, \mathbb{P}) \) of monetary values, where \( \Omega \) is the sample space; \( \mathcal{F} \) is the set of possible events in \( \Omega \); \( \mathcal{F}_T \) is a filtration with \( \mathcal{F}_0 = \{\emptyset, \Omega\} \), \( \mathcal{F} = \sigma(\bigcup_{T \geq 0} \mathcal{F}_T) \), \( T := \mathbb{R}_+ \cup \infty \) and the usual assumptions; and \( \mathbb{P} \) is a probability measure that is defined in \( \Omega \) of the events contained in \( \mathcal{F} \).

We consider adapted random processes \( X_T : \mathcal{X}_T \rightarrow \mathbb{R} \) to represent the variables in our approach. Thus, \( E[X_T] \) is the expected value of \( X_T \) under \( \mathbb{P} \). All equalities and inequalities are considered to be almost surely in \( \mathbb{P} \). \( F_{X_T} := F_{X|\mathcal{F}_T} \) is the probability function of \( X_T \), with inverse \( F_{X_T}^{-1} \) and density \( f_{X_T} \). At time \( T \), \( B_T, R_T, C_T, I_T, E_T \) and \( L_T \) represent, respectively, balance, risk-free rate, net cashflow, income, expense and loans to be paid. We assume that \( R_T \) and \( L_T \) are \( \mathcal{F}_{T-dt} \) measurable because their values are known by contract in the previous period.

The main idea is to consider the balance of an agent as a stochastic process, composed of their net cashflows, to measure the credit risk incurred for a financial institution when it loans money to this agent. The BaR approach is the analysis of risk measures through this stochastic process. Thus, in this section, we define and expose some properties of the stochastic process under analysis as well as the credit risk measures that are derived from it.

**Definition 2.1**  Let \( B_T, R_T, C_T, I_T, E_T : \mathcal{X}_T \rightarrow \mathbb{R} \) and \( C_T = I_T - E_T \). The closed form for the stochastic process that represents the balance is given by

\[
B_T = B_0 \exp \left( \int_0^T R_t \, dt \right) + \int_0^T C_t \exp \left( \int_t^T R_s \, ds \right) \, dt, \\
\frac{d B_T}{dt} = \left[ B_0 \exp \left( \int_0^T R_t \, dt \right) \right] \left( R_T + C_T \right) \, dt.
\]

**Remark 2.2**  The stochastic process in Definition 2.1 depends only on cashflows and can be considered information that summarizes the distinct dimensions (economic, social, geographic and others) of the agent’s behavior. Equations (2.1) and (2.2) can be recursively generalized for any time \( T^* \leq T \) and, respectively, conform to

\[
B_T = B_{T^*} \exp \left( \int_{T^*}^T R_t \, dt \right) + \int_{T^*}^T C_t \exp \left( \int_t^T R_s \, ds \right) \, dt, \\
\frac{d B_T}{dt} = \left[ B_{T^*} \exp \left( \int_{T^*}^T R_t \, dt \right) \right] \left( R_T + C_T \right) \, dt.
\]
Moreover, it is possible to obtain simpler formulations under a null initial balance, i.e., $B_0 = 0$.

Obviously, the properties of $B_T$ as a stochastic process directly depend on those possessed by $C_T$ and, consequently, by $I_T$ and $E_T$. At this point, we do not make assumptions about the stochastic process that governs such random variables. Our goal is to keep the model as general as possible. Nonetheless, for a special case, it is possible to analytically derive $F_{B_T}$ given $F_{I_T}$ and $F_{E_T}$. To do so, we use the concept of probability function convolution. Below, we present the definition of this concept and two well-known lemmas in probability theory. Given this background, we prove a theorem.

**Definition 2.3** The convolution between two probability densities $f_X$ and $f_Y$ is given by

$$f_X \Box f_Y(x) = \int_{-\infty}^{\infty} f_X(x-u) f_Y(u) \, du.$$  \hspace{1cm} (2.3)

**Lemma 2.4** Let $X$ and $Y$ be two independent random variables. Then, $f_{X+Y} = f_X \Box f_Y(x)$.

**Lemma 2.5** Let $k \in \mathbb{R}$. Then,

$$f(x) = \frac{1}{|k|} f_X\left(\frac{x}{k}\right).$$

**Proposition 2.6** Let $B_T$, $R_T$, $C_T$, $I_T$, $E_T : \mathcal{T} \rightarrow \mathbb{R}$, $C_T = I_T - E_T$ and $\Gamma_t = \int_t^T R_s \, ds$. If $I_T$, $-E_T$ are independent for all $T \in \mathbb{T}$, and $C_T$, $C_S$ are independent for all $T, S \in \mathbb{T}$, $T \neq S$, then

$$F_{B_T}(x) = \int_{-\infty}^x \left[ \lim_{dt \to 0} f_{B_0(t_0)}(s) \Box f_{C_t \Gamma_t}(s) \bigg|_0^T \right] \, ds,$$  \hspace{1cm} (2.4)

$$f_{C_t \Gamma_t}(x) = \left[ \frac{1}{|\Gamma_t|} f_{I_t}\left(\frac{x}{\Gamma_t}\right) \right] \Box \left[ \frac{1}{|\Gamma_t|} f_{-E_t}\left(\frac{x}{\Gamma_t}\right) \right] \quad \text{for all } t \in [0, T], \hspace{1cm} (2.5)$$

$$f_{C_t \Gamma_t}(x)|_0^T = f_{C_0 \Gamma_0}(x) \Box f_{C_{0+dt} \Gamma_{0+dt}}(x) \Box \cdots \Box f_{C_T \Gamma_T}(x). \hspace{1cm} (2.6)$$

**Proof** Because $R_T$ is $\mathcal{F}_{T-dt}$ measurable by assumption, $\Gamma_t \in \mathbb{R}$ for all $t \in [0, T]$. Because $C_t \Gamma_t = I_t \Gamma_t - E_t \Gamma_t$ for all $t \in [0, T]$, we have by Lemma 2.4 and Lemma 2.5 that

$$f_{C_t \Gamma_t}(x) = f_{I_t \Gamma_t - E_t \Gamma_t} = \left[ \frac{1}{|\Gamma_t|} f_{I_t}\left(\frac{x}{\Gamma_t}\right) \right] \Box \left[ \frac{1}{|\Gamma_t|} f_{-E_t}\left(\frac{x}{\Gamma_t}\right) \right]$$

for all $t \in [0, T]$.  

www.risk.net/journals  
Journal of Credit Risk
Repeating the argument for $C_t \Gamma_t, t \in [0, T]$, which are independent by assumption, we obtain

$$f_{\int_0^T C_t \Gamma_t \, dt}(x) = \lim_{dt \to 0} f_{C_t \Gamma_t}(x) \bigg|_{0}^{T}. $$

Because $f_{B_0 \Gamma_0}$ is a Dirac measure, which is independent of any probability function, and $F_X(x) = \int_{-\infty}^{x} f_X(s) \, ds$, we have

$$F_{B_T}(x) = \int_{-\infty}^{x} \left[ \lim_{dt \to 0} f_{B_0 \Gamma_0}(s) \square f_{C_t \Gamma_t}(s) \bigg|_{0}^{T} \right] \, ds. $$

This concludes the proof. 

**Remark 2.7** Naturally, such assumptions of independence in Proposition 2.6 are too restrictive and can be questioned in real life. However, even with such constraints, the degree of complexity becomes significant and leads to an analytical expression that is difficult to compute. Moreover, obtaining $F_{B_T}$ from cashflow distributions is at least an alternative. One can eliminate the assumption that $I_T$ and $-E_T$ are independent if the distribution for $C_T$ is directly known. Nonetheless, in practical situations, it is also possible to directly estimate $F_{B_T}$ from balance data using numerical simulations.

Having explained the stochastic process, we turn our focus to the next step in the BaR approach, which is measuring credit risk. We concentrate on two risk measures: the PD and the loss given default (LGD). As both PD and LGD are very well-known risk measures for credit risk, we shall not pursue a debate of their theoretical properties in this paper. Nonetheless, we formally define them and briefly discuss their financial meaning.

**Definition 2.8** Let $B_T, L_T : \mathcal{X}_T \to \mathbb{R}$. The probability of default, $PD_T : \mathcal{X}_T \to [0, 1]$, and the loss given default, $LGD_T : \mathcal{X}_T \to \mathbb{R}$, are defined, respectively, as follows:

$$PD_T := \mathbb{P}(B_T \leq L_T) = F_{B_T}(L_T), $$

$$LGD_T := -E[B_T - L_T \mid B_T \leq L_T = F_{B_T}^{-1}(PD_T)].$$

**Remark 2.9** The PD is the probability of an agent not having enough balance to pay their loan at time $T$. It is a very simple and intuitive credit risk measure. However, PD alone does not give the entire picture of a situation, because it does not consider the loss magnitudes in the event of a default. This shortcoming is handled by the LGD, which indicates how much money is necessary to fulfill the loss in the event of a default. This measure has similarities to the well-known market risk measure.
expected shortfall (ES), but for LGD the expectation is truncated by $L_T$ instead of a quantile of interest. Moreover, because (2.7) and (2.8) both depend directly on $F_{B_T}$, under the hypothesis of Proposition 2.6 one can obtain analytical formulations for these two risk measures.

3 AN EMPIRICAL BaR MODEL

Now, we consider a specific BaR model nested in the previous formulations. From this point forward, we change the notation so that the time index is represented by $t = 1, \ldots, T$, which is the usual notation in empirical time series models.

We use a discrete version of the model by setting the income and the expense equation of an individual to separate stochastic processes. We expect income to be far more predictable than expenses, because most of society is bound by work contracts that explicitly define the amount of financial reward that one receives per unit of time. However, when looking at expenses, we expect a higher quantity of noise because the individual decision to purchase goods and services is a function of diverse economic, social and personal factors. The heterogeneous stochastic properties of the inflow and outflow of cash separate the process of income and expenses.

Another important aspect of empirically modeling the inflow and outflow of financial resources is recognizing the existence of seasonality. An individual might receive more cash in particular periods than others. For instance, a worker in Brazil is entitled to an extra paycheck and an additional cash-based holiday premium throughout the year. These amounts are usually paid in June and December. Identifying these particular months is essential to developing a realistic credit risk model, particularly in the case of short-term loans. The expectation of receiving more cash in particular months affects the balance and adds a time dependency to the dynamic of cashflows and, therefore, the forward PD.

A realistic empirical model should also consider the effect of unemployment. If an individual loses their work contract for any reason, their cash income ceases, thus affecting the PD. When an individual loses their job, all that is left as support for expenses is the current value of the balance.

In considering these effects, we propose a discrete version of (2.1) and (2.2) for our empirical example:

$$B_t = B_{t-1}(1 + r_t) + I_{i,t} - E_{i,t},$$

$$I_{i,t} = \begin{cases} 0 & \text{if } S_t = 1 \text{ (unemployed)}, \\ FI + MI_{i,t} & \text{if } S_t = 2 \text{ (employed)}, \end{cases}$$

$$E_{i,t} = FE + ME_{i,t}.$$
where \( B_t \) is the balance at time \( t \) \((t = 1, \ldots, T)\); \( r_t \) is the risk-free yield (monthly), assumed constant \((r_t = r)\); \( I_{i,t} \) is the total income for month \( i \) and time \( t \); \( E_{i,t} \) is the total expense for month \( i \) and time \( t \); \( FI \) is the fixed monthly income; \( MI \) is the monthly extra income for month \( i \); \( FE \) is the fixed monthly expense; and \( ME \) is the monthly extra expense for month \( i \).

Following the ideas described in Huh et al (2010) and Malik and Thomas (2012), the states of employment and unemployment \((S_t)\) follow a discrete Markov chain, given by transition matrix \( P \):

\[
P = \begin{bmatrix} p_{11} & 1 - p_{22} \\ 1 - p_{11} & p_{22} \end{bmatrix}
\]

(3.4)

Because only two states and two transition probabilities exist, we estimate the transition probabilities by inverting the expected duration of the states \( E(\text{Dur}(S_t = k)) = 1/(1 - p_{kk}) \):

\[
p_{11} = 1 - 1/E(\text{time employed}),
\]

(3.5)

\[
p_{22} = 1 - 1/E(\text{time unemployed}).
\]

(3.6)

Thus, depending on how much time it takes for a worker in the same profession as the applicant to get a job or lose their job, transforming this information into transition probabilities using (3.5) and (3.6) is easy. Given this setup, we allow for economic conditions in the job market to affect the applicant’s PD. If the unemployment duration increases, so does the likelihood of a lower income. Therefore, we can expect an increase in the PD on a loan. Note that the balance in (3.1) changes over time, indicating that the PD should be recalculated as time moves forward, in a manner similar to the behavioral models (Thomas 2000). If the effective balance level increases with an applicant’s good financial status, the likelihood of a future default decreases, and the individual should receive the reward of paying lower interest rates.

### 3.1 Simulation of the income and expense data

As an example of an empirical application, we assume the existence of a history of inflows and outflows of a balance account over five years. That is, in the data-generating stage, we also assume the individual was employed or had passive earnings throughout the period. Individual banking statements should be easily available within a commercial bank. In commercial applications of the model, it might be interesting to consider the creation of a central organization using applicants’ aggregate bank statements. On the one hand, banks would benefit from gaining a larger amount of information from applications. On the other hand, clients would benefit from competition between banks, which should create a downward pressure on the interest rates of debt.
To illustrate the use of this model, we simulate the inflows and outflows of money as random normal variables, with different means and standard deviations according to the month of the year. This simulation should emulate the expected noise from empirical banking data.

In the simulation, we set an initial balance of $1000, where, on the income side, the applicant makes an average of $3050 every month. In the months of June and December, they receive a bonus of 50%. The income in each month is a random normal variable, with its mean defined using the previous rules and a standard deviation of $50. On the expense side, the applicant expends an average of $3000, with 40% increases in June and December. The instability, however, is higher for expenses, with a standard deviation of $250. Given the previous information, we can simulate a banking history for any period of time.

In Figure 1, we present the time series plot of the simulated income and expense data. Note that the behavior of income is far more stable than that of expenses, with clear seasonality in June and December. The balance of this individual has a clear upward trend because their income is usually higher than their expenditure. Further, we assume that, generally, the applicant is employed for three years, and they can acquire a new job within three months when unemployed. This information can be retrieved from the applicant’s job records or other sources. For example, the Bureau of Labor Statistics reports that the average duration of employment in the United States for March 2016 was twenty-nine weeks (7.25 months).3

### 3.2 Estimation of the coefficients

Given the general mathematical formulation discussed in the previous section, the empirical model can be formulated in different ways. For simplicity, we assume that income and expenses follow a linear model conditional on time. However, it is important to point out that more sophisticated methods, such as cubic splines, can be used to model the seasonality of the banking data. Such an approach has been used with success in finance to model volatility and term structures (Audrino and Bühlmann 2009; Engle and Rangel 2008; Jarrow et al 2004).

The model’s estimation is straightforward. We define a statistical model for income and expenses as

\[
I_{i,t} = \alpha_I + \sum_{i=2}^{12} \beta_i D_i + \varepsilon_{i,t},
\]

\[
E_{i,t} = \alpha_E + \sum_{i=2}^{12} \phi_i D_i + \eta_{i,t},
\]

where $D_i$ is a dummy variable that takes the value 1 if the current month is $i$ and 0 otherwise. We exclude the dummy for the first month, $i = 1$, to avoid identification issues in the regression model. Using the simulated data set as input, we estimate with least squares the empirical model defined in (3.7) and (3.8). The resulting coefficients from the estimation are omitted. Note that they capture the seasonality of the data. For reproducibility, the R code that generates the data and estimates the model is available at https://bit.ly/2X9eaS1.

As an illustration of the use of this model, we set the expected time of employment to unemployment as three years (thirty-six months) and the expected time
of unemployment to employment as three months. These are used to define the transition probabilities of (3.4) using (3.5) and (3.6).

3.3 Methodology for and results of simulating an individual’s future balance and the associated PDs

Given information on the regression models, transition matrix and an initial balance of $1000, we simulate an individual’s future balance by sampling the estimated residuals from the regression, $\varepsilon_{i,t}$ and $\eta_{i,t}$, conditional on the month of the year. This simulation allowed for seasonality not only with expected income and expenses but also with their corresponding volatility. The estimated transition probabilities will drive the Markov chain forward. In the simulation, we first assume state 1 at time 0 ($S_0 = 1$): employed. Then, we sample a random number from 0 to 1. If this number is lower than the probability of staying in regime 1, we set $S_1 = 1$. If this random number is higher than the probability of staying in regime 1, we switch the state to $S_1 = 2$: unemployed. Likewise, following (3.2), if $S_t = 1$, then $I_{i,t} = 0$; else, $I_{i,t} = FI + MI_{i,t}$. We use the same algorithm to drive the Markov chain for $t = 2, 3, 4, \ldots, T$. We perform 10 000 simulations with a time horizon of twenty-four months.

Once the balance is simulated, calculating the PD of the individual is straightforward: it can be done by simply looking at the number of simulated scenarios in which the resulting balance was negative for each forward point in time (see (2.7)). Because our first example had no loans, this figure represents the probability of the applicant falling short on their expenses. We call this figure the benchmark default rate curve.

In Figure 2, we note that the PD is also seasonal, with a small drop in month 12, which is when a larger cash income occurs. We also see that the PDs generally increase over time, indicating that this applicant is more likely to run out of cash as time passes. While not an intuitive result, this is the effect of the chances of unemployment, driven by the transition matrix. Losing income puts pressure on the balance of the individual. Even if the overall chance of unemployment is small, it will still have a great impact on the forward PD.

We now illustrate a common practical case in which the same applicant asks for a loan that can be paid through monthly installments of $400 for two years or monthly installments of $200 for four years. Given the previous regression for the expenses, implementing this information in the model is easy by defining

$$E_{i,t} = \alpha E + \Delta + \sum_{i=2}^{12} \phi_i D_i + \eta_{i,t}.$$
where Δ is the expected change in expenses represented as the monthly loan payment value. Figure 3 presents the impact on the PD.

Based on Figure 3, the payment of the loan has an explicit impact on the forward PD. For monthly payments of $400, the applicant is very likely to default on their loan in month 4. However, when setting a monthly payment of $200, the PD decreases significantly, indicating a better financial contract for both sides. This illustration shows the flexibility of the empirical model and could accommodate different practical scenarios when evaluating credit risk. Next, in Table 1, we present the calculation of the LGD for each scenario.

As is observed, the first scenario leads to an LGD of 3346.6, which is significantly higher than the second case with longer maturity and lower monthly installments. As suspected, the second setup presents a lower expected cost for the bank.
FIGURE 3  Probabilities of default given a new loan.

Here, “pm” denotes “per month”.

TABLE 1  Risk measures for empirical BaR model.

<table>
<thead>
<tr>
<th>Case</th>
<th>LGD</th>
</tr>
</thead>
<tbody>
<tr>
<td>Installments of $400 per month</td>
<td>3346.6</td>
</tr>
<tr>
<td>Installments of $200 per month</td>
<td>475.2</td>
</tr>
</tbody>
</table>

3.4 Optimizing parameters of a loan using a BaR model

In practice, the bank receives data from the applicant and must decide on the structure of the offered debt contract, including its maturity and rate of return. If the contract
is perceived as risky, the bank is likely to charge a higher yield rate and set a lower maturity range. In this section, we present an illustration of how the BaR formulation can be used to optimize the parameters from the point of view of the bank.

We consider a case in which the applicant requests a loan for 10,000 units of cash and has the same banking records as used in the previous example (see Figure 1). The debt is paid back by the applicant in fixed monthly installments that are defined for a given maturity and annual yield rate. This loan is equivalent to a fixed rate bond with monthly coupons (Fabozzi and Mann 2012). Using an annual yield rate, we define the installment value as:

\[
I = 10000 \frac{r_m}{1 - (1 + r_m)^{-T}},
\]

where \(r_m\) is the monthly yield of the debt contract, \(r_y\) is the yearly yield of the debt contract, \(T\) is the maturity of the contract (in months) and \(I\) is the fixed value of monthly installments.

Equation (3.10) indicates that the increase in the yield rate and the decrease in the maturity of a contract increase the value of the installments and the risk of the contract. In the BaR model, a higher installment leads to a higher level of uncertainty and, consequently, a higher PD.

From the bank side, a natural question to investigate is this: given the expected cashflow dynamic of the applicant, which values of yield and maturity maximize the profit of the contract? We analyze this question by defining the objective function as the expected return from the loan, which is calculated on the basis of the simulation procedure defined in the previous section. Note that the risk is also considered because higher installment values increase the likelihood of a default and, thus, decrease the expected profit.

We first create a vector of values for maturity (\(T = 1, 2, \ldots, 23, 24\)) and yields (\(r_y = 0.025, 0.05, \ldots, 0.475, 0.5\)). For each pair of maturity and yields, we simulated the BaR model and calculated the expected return from the contract. We performed a grid search procedure to find the pair of maturities and yields that maximizes the value of \(E(R)\), the expected cashflow divided by the total value of the loan. Our results are presented in Figure 4.

Figure 4 indicates that the pair of maturities and yields that maximizes the contract’s expected profit is twenty-three months and a 30% rate of return on the debt. The white spots in the figure show that a low maturity and a higher yield lead to a null return of the contract. This result is explained by the fact that a higher installment (see (3.10)) quickly drains the applicant’s balance, leading to a default on the contract. We also point out that the risk of the contract, LGD (2.8), is minimized with a maturity of eighteen months and a 5% yield. This exercise clearly shows how the
BaR model can be used to better design short-term financial contracts by taking into account the applicant’s banking history.

4 CONCLUSIONS

In this paper, we propose an alternative model to calculate personal financial risk by defining a stochastic process for an individual’s cash income and expenditures. Formal theoretical definitions for this approach along with the usual risk measures are presented. Using an artificial data set, we illustrate the model’s usage by estimating...
an empirical version of the general formulation and calculating the forward PDs on different loan scenarios. We also present an example of an optimization scenario in which we find the optimal maturity and yield rate of a short-term loan.

The main advantage of using our approach is its flexibility. By directly employing a stochastic process for income and expenses, we allow seasonality to play a role in the analysis of short-term credit risk. By using this model, one can verify the impact of several economic factors on an individual’s personal credit risk, e.g., loan payment increases, specific payment schedule adjustments or job market changes.

Empirical applications of the model are straightforward. On the basis of banking data and information from the job market, one can estimate and simulate an individual’s future income and expenses. As argued in this paper, the creation of a central organization to store and analyze banking data could facilitate the process and benefits for both parties in the transaction.

Because banks deal with portfolios of individual loans, one suggestion for a future study is to investigate a multivariate version of the model. Formulating a model that incorporates systematic dependencies within a pool of applicants’ cashflows, for which economic shocks such as an increase in general unemployment affect the overall PD on loans, would be interesting in the analysis of the general risk of a bank. Future investigations with real banking data are also suggested. Given individual records, one could study the dynamics of cashflows and test for the structure of an empirical model that efficiently represents the data set and, thus, provides realistic default probability values. Understanding and modeling how applicants react when faced with a low balance in their accounts is also suggested. A person is likely to place a hold on their expenses once their balance passes a particular threshold. An investigation using real data could shed light on this effect. These and other ideas are left for future development.

DECLARATION OF INTEREST

The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

REFERENCES


Journal of Credit Risk www.risk.net/journals


