

Thesis

Daniel Travaglia

September 7, 2020

1 A gentle introduction to credit risk

This section aims to give an overview of the topic that is at the heart of this paper: **credit risk**. Rather than jumping straight into formal definitions, which might create difficulties in grasping with the content, the reader will be first provided with real-life examples characterized by the presence of such risk. Moving on, some formal definitions will be given along with a description of the agents that usually take upon this risk and how they act to assess it and quantify it and finally take proper actions to mitigate it. To conclude, the chapter ends with an introduction to credit risk modelling, a topic that will be further explored later in the empirical analysis

1.1 Introduction

In a recently published online course on credit analysis [**creditanalysis**], the author provides a few real-life situations that occur daily and where it is possible to encounter credit risks. The following examples are based on those situations and might serve as a great starting point to understand the topic.

- **Example 1.** Imagine yourself receiving a call from one of your friends that you have not heard for quite a while. You can hear from his voice that he is really upset, and apparently, the reason for this is that he has to pay a fine by the end of the day to avoid an additional charge, but he does not have the money to pay at that moment. He tells you that the reason for this is that he had to cover fo huge expenses lately and he still has not received his monthly salary. Given that, he then asks you, whether you would be kind enough to lend him the money, promising to repay you as soon as he receives his pay. What would you do? Would you lend him the money?
- **Example 2.** Imagine to be in the same situation as above, but instead of simply being one of your friends, the requestor is one of your colleagues at work? Would you do it in this case? Would it make a difference if it was someone you knew well, for instance, the colleague you always work with?
- **Example 3.** Imagine now the request to change, and instead of asking for a large sum, your colleague asks you to lend him money for

a pizza without mentioning you when and how is going to pay you back. Would this make a difference in your evaluation on whether to lend the money or not?

Although representing a very simplistic situation concerning what usually financial institutions have to deal with when they need to evaluate whether to grant the credit or not, the questions that need to be answered could be equivalent:

1. Who is borrowing?
2. How much are they borrowing?
3. When will the debt be fully repaid?
4. How much is there for me (either as an individual or an institution)?

Considering the last example, imagine instead that your colleague promises you to pay back the week after and on top of that, he is willing to take you out for lunch. This time, what you get back is more than what you lent. In financial terms, this is also known as *interest rate*. On the other hand, imagine he gives you his watch to keep until he pays you back to make the lending much safer for you. The watch in this case represents a *collateral* (or a *security*).

The procedure just introduced is usually called "**credit analysis**". According to [creditanalysis], the latter can be defined as: "*the process through which the lender elaborates if he believes the counterparty is going to honour its obligations or not*". Ultimately, this determines whether or not the agent will enter in that contract, along with the relative risk associated with it, or more precisely, the **credit risk** [creditriskprinciple].

1.1.1 Credit risk

It is possible to find many definitions of *credit risk* around. Investopedia provides us with the right one in this context:

Definition (Credit Risk): *the risk that a lender has to take into account due to the uncertainty related to the borrower either failing to repay a loan or to meet its obligations.* [investopediacreditrisk]

In simple terms, this definition suggests that the lender should consider the possibility that a borrower will not be able to pay back the principal and the interest rate according to the initial contractual agreement. The party granting the credit should then quantify this probability and based on the result, charge a coupon rate (i.e. interest rate) to protect itself against this risk. The methods to derive the likelihood a counterparty won't be able to meet the contractual agreements are part of the broad area of **Credit Risk Modelling** [bluhm2016introduction] which will be at the heart of the empirical analysis, and for this reason, it is introduced later in the paper.

Ultimately, in the context of credit risk there are at least 3 additional points [investopediacreditrisk] worth to mention which should also be considered as key takeaways from this introductory chapter:

1. **Credit risk in every financial transaction:** although most of the times it might come naturally to associate this type of risk to transactions that occur exclusively between a party and a financial institution (e.g. mortgages, loans, credit cards, etc...), there are also situations where credit risk can be perceived between private parties (e.g. between companies and individuals: paying invoices, insurance coverage, etc..)
2. **Risk assessment:** before granting new credit (as often it is the case in the business world), the bank undergoes an assessment of the borrower takes into consideration his credit history, the capabilities to repay, the capital available, the loan's conditions and associated collateral. Such evaluation has the final goal of providing an accurate prior estimate of the credit risk, which will eventually tell whether the client should get or not the obligations, with an increasing interest rate for those who are perceived as riskier. In the case of bonds, this assessment is done by credit-rating agencies that assign a triple-A (i.e. *AAA*) for low-risk investments, all the way down to *C* for high-risk investments. Note that, although this process is almost exclusively conducted for bond products, the dataset used in the empirical analysis provides the same ratings for loans, perhaps as a result of internal procedures to assess credit risk.
3. **Credit risk vs. other risks:** so far it was assumed that the risk associated with the lending was purely driven by the borrower characteristics. However, in a typical lending transaction, usually, there are other types of risks kicking in: market risks (e.g. economic conditions, FX rates), country-specific risks (e.g. OECD country, emerging-markets

country, etc.), operational risk, issuer risk, and so on... To keep things simple, and given the nature of the data for the empirical analysis, although it might represent a far too simplistic assumption, we will stick with it and only deal with a particular type of credit risk: the *CCR* (*Counterparty Credit Risk*)

1.1.2 CCR - Counterparty Credit Risk

According to BIS - Bank for International Settlements, *CCR* (*Counterparty Credit Risk*) can be defined as:

Definition (Counterparty Credit Risk): *the measure of the **likelihood** that the counterparty to a transaction **might default** before meeting the contractual obligation.* [basel'CCR]

There is a subtle but fundamental difference between the two types of risk introduced above: when assessing the "credit risk" in general, the lender takes into account the chance that the borrower might not be able to fully repay the lender according to the contractual obligations, but do not exclude that the counterparty might cover for the remaining part in the future either. Hence, the bank is exposed (in terms of risk) only to the portion of the money lent that is perceived as being at risk.

On the other hand, the lender might also consider the possibility that the counterparty defaults on this loan with the inability to bear the debt neither in the present nor in the future. In such a context, the lender not only has to evaluate the risk associated with the loan, but it also has to take into account the risk associated with the counterparty itself. In conclusion, if such an event occurs, and the pool of transactions with the counterparty has a positive economic value, the result would be an economic loss for the lender.

As a last note, CCR is usually characterized by the bilateral risk of loss (i.e. either one of the parties involved in the transaction might default). Bilateral exposures usually factor in the uncertainty related to market movements, as part of the value of the transaction is driven by the underlying market value [basel'CCR]. Nevertheless, during the empirical analysis, we will assume unilateral exposures: the lender is the only one party to be exposed to the risk of the counterparty to default.

2 Basel Regulatory Framework

2.1 Introduction

This section aims to introduce the key regulatory framework in the context of credit risk and more specifically, counterparty credit risk: the Basel Accord. With a walk-through in history, the objective is to make the reader aware of the reasons why lenders and companies have become increasingly regulated from this standpoint.

Note: that the content provided here is for the most part based on the information provided by the following two references: [baselhistory'fu] and [baselhistory'bis]

2.2 Reasons for a more comprehensive framework to credit risk management

The needs for a more comprehensive and better approach to risk management, particularly for counterparty credit risk, has emerged quite significantly after the financial crisis of 2007-2009. Since then, regulators have sharpened their existing frameworks and applied more stringent controls to the stability of financial institutions. On this line of reasoning, this section provides an overview of how regulatory requirements in the context of counterparty credit risk have evolved, with a focus on the key regulatory framework in this environment: **The Basel Accords** [investopediabaselaccord]

After the well-known and aforementioned 2007-2009 financial crisis event, there has been an unprecedented revision of the global framework regulating the financial sector, culminating in what is known today as the **Basel III** regulatory framework. Before reaching this result, the banking sector has experienced a relevant number of systemic crisis usually driven by various factors, including the miscalculation of risk, represented by inadequate capital levels to carry out their business. Each of this crisis brought some contributions and changes to the previous framework, adding up to the first version, which consisted of 30 pages, more than 1500 pages of guidelines relating to the supervision of daily banking activities.

2.3 Reasons for an independent Committee

Why was the Basel Committee ever needed? To answer this question, it is necessary to go back to 1970s, when the Herstatt Bank collapsed and was put under liquidation due to enormous trades on the foreign exchange market that did not go as planned. The license was withdrawn in 1974, as losses have reached an amount equal to 10 times the liquidity of the bank. However, there is more to the story: US counterparties engaging in multiple transactions with Herstatt Bank released "Deutschmark" in exchange of dollars. These lenders did never see their money, essentially because of time differences: the US was still in morning trades when the bank was revoked its license. Although this is purely related to FX activities, and consequently involves also FX and market risk, this event highlighted the necessity to create a central forum for banking supervision concerning matters related also to other types of risks, such as "credit risk". To enhance the financial stability and quality of banking supervision, in 1974 multiple central banks gave rise to a centralized committee which later on took the name of "Basel Committee on Banking Supervision". The latter expanded quite significantly and as of now, it includes 45 central banks worldwide.

2.4 Basel Frameworks

The need for a regulatory framework for risk management was further strengthened during the 70s-80s period, when the surge in debt in the Latin American countries, combined with the rise of interest rates in the US and Europe, led the way to a series of critical debt restructuring efforts for many countries worldwide. The need for a clearer and comprehensive framework for the banking sector forced the Basel committee to issue guidelines on weighted approach to risk management. Such need was satisfied with the release of the Basel 1 framework in 1988 when for the first time in history, banks were required to weigh the capital they held against the credit risk they took.

2.4.1 Basel 1

In 1988, the first regulatory framework, which later took the name of "Basel I", was released by the Basel Committee. The latter had the objective of setting up common international and shared regulations on capital adequacy supervision. The relevant aspects of the regulation are summarized in the following points:

- Credit institutions are required to classify assets into 5 different categories based on their underlying risk: from 0% for most secured assets (e.g. cash) to 100% for low-quality assets (e.g. private sector debt). These assets are also identified as **RWA**” (risk-weighted assets).
- Definition of the so-called **Minimum Capital Requirement**: the minimum ratio of capital to risk-weighted assets (RWA). The threshold was initially set at 8%, equally spread between most absorbing assets (i.e. **Tier-1 Capital**: examples are equity and retained earnings) and supplementary assets (**Tier-2 Capital**: examples are evaluated assets and subordinated debt). The latter includes assets that are more difficult to liquidate, and therefore less secure. This level was introduced to ensure that financial institutions had enough standalone capabilities to absorb potential losses resulting from defaulting clients.

Despite the effort, the first Basel Regulation had some shortcomings. These were mainly related to 3 factors: the duration of the service, the market risk and, most importantly, the counterparty risk. The introduction of amendments helped in assessing some of these issues (such as the Market Risk Amendment). Nevertheless, the complexity introduced by some financial products (Credit default swaps, Complex derivatives, etc..) was drastically incrementing the risk taken by financial institutions. This situation required further adjustments with the regulation and shined a light on the ever-increasing importance of accurate methodologies to assess exposures to risk.

2.4.2 Basel 2

The proposal for a replacement of the previous accord (i.e. Basel I) came in 1999 and was finally released in 2004 with the name **Basel II**. This comprised some of the previously mentioned and much-needed adjustments, which are summarized here:

1. **Minimum capital requirement**: it was developed and expanded and a new capital tier (i.e. **Tier-3**) was introduced to cover also for the market risk.
2. **Market discipline**: financial institutions are obliged to periodically disclose information on their risk exposures to enable much more informed decisions and constant monitoring.

3. **Supervisory review**: providing clear guidances on periodic assessments of an institution's capital adequacy and in case of capital pressure, it granted intervention powers

Despite the efforts to bring more stability, the accord required time and effort before it was released. Most financial institutions had already started taking full advantage of the subprime mortgages - lending money to low credit profile - given their higher expected returns. Moreover, these institutions entered the financial crisis with too much leverage supported by the later favourable economic conditions, despite the regulatory environment that was set in place with Basel II. These events were part of a much broader series of dysfunctionality (e.g. poor governance, bad risk management, a combination of excessive credit growth and credit risk mispricing) that served as key lessons for the Basel Committee to bring substantial updates to the baseline framework. Eventually, the Committee agreed to design a reform package to outline new ways of handling credit and liquidity risk.

2.4.3 Basel 3

The agreement culminated with the publication in 2010 of what is today known as the **Basel III** framework. The aim was to improve on top of the 3 pillars of the previous accord and extend them to also other areas. The key changes can be summarized in the following points:

1. **Capital conservation buffer**: part of the common equity that, if ever breached, restricts payouts (e.g. dividends) to be used to meet the minimum common equity requirements.
2. **New leverage ratio**: based on loss-absorbing capital and total institution's assets and takes into account also off-balance sheet exposures (regardless of RWA).
3. **Liquidity Coverage Ratio (LCR)**: a minimum short-term liquidity ratio to provide banks with enough liquidity to cover for at least 30-days stress period.
4. **Net Stable Funding Ratio (NSFR)**: a minimum long-term liquidity ratio to address maturity mismatches over the balance sheet.

The Basel Committee realized that the approach of "one fits all" is hard to meet the expectations outlined in the framework. For this reason, it

started a series of reform to condition the requirements on the institutions' size, complexity and systematic importance. This culminated in a series of on-going reform packages that are still happening today and should be published under the name of **Basel IV**.

3 Credit Risk - Empirical Analysis

The objective of this section is to give a high-level overview of the methodologies that the financial industry usually adopt to forecast credit losses and later leverage the data available to develop a machine-learning approach and estimate the probability that a counterparty will default on its obligation.

3.1 Credit risk modeling

3.1.1 Introduction

The motivations for a financial institution to develop credit risk models are much due to regulatory requirements. However, as highlighted in [bluhm2016introduction], other reasons explain the importance of internal credit risk modelling. To better understand this, let us provide a simplified, but not too unrealistic example based on the example provided in [bluhm2016introduction]:

Example 1.

Imagine a bank is asked from a major tech company to provide a very huge loan in the size of \$5B. As already discussed, a credit analyst, (who for ease the example we will call Ben) will have to go through the request and see whether it is the case to grant or not this loan based on the information that he can gather. Let us assume that Ben already knows that the CEO of this company has a great friendly relationship with the CFO of the bank, and, from recent studies, he discovers that the specific sector of the company has recently experienced a steady drop in sales and the bank-internal rating system suggests this company is on its way down to a sub-investment grade (i.e. classified as a risky investment). What should Ben do in this situation? Usually, the analyst has 2 options at this stage:

1. *Reject* the deal based on the information he gathered on the company and the relative market
2. *Accept* the deal BUT, protect the bank against potential losses thanks to **credit risk management instruments** (such as credit derivatives) and transfer the risk to a third-party in exchange of a fee. [investopediacreditderivative]

Hence, financial institutions have also a way out: as individuals would do through health insurance, banks can protect themselves from the underlying

exposures to risk when lending to particularly risky clients. In particular, banks had already designed ways to get loan insurance in the past and usually comprises the whole bank's credit risk portfolio, and not only some critical positions. This brings directly to what in [bluhm2016introduction] is perceived as the building block of credit risk modelling within a financial institution: the **expected loss**.

3.1.2 Expected Loss

For banks to protect against this risk, they would like to know the *cost* that arises from a particular client (a group of clients when dealing with a portfolio of loans) before granting the loan, so that they can charge a *premium* on the interest rate accordingly and add up to a reserve (also called *expected loss reserve*) to cover for potential losses due to defaulting loans. In this context, the bank is interested to know, *on average*, the potential loss that might occur when lending money to a client. This value is called **Expected Loss** and it is characterized by the 3 pillars of credit risk modelling [bankofenglandcreditrisk]:

1. **Probability of Default (PD)**: the *likelihood* of a default in a defined time-horizon associated to a client. Note however that this is a general definition that may vary depending on the method used to estimate such probability. For instance, under the Basel II IRB framework, the PD per rating rate is defined as the average percentage of obligators that will default over one year (i.e. the time-horizon specified in the regulation for each asset class).
2. **Exposure At Default (EAD)**: *estimated outstanding amount* given that a client defaulted
3. **Loss Given Default (LGD)**: *proportion of EAD* that is expected not to be recovered in case of default.

To be mathematically rigorous, the model describing the expected loss is defined on top of a probability space $(\Omega, \Phi, \mathcal{P})$, where Ω consists of the *sample space* defining all the possible events, Φ is the σ -*Algebra* containing all the measurable events of the sample space (such as the information on whether an obligator default or not) and lastly, \mathcal{P} , the probability space, which attaches a probability to each measurable event.

If we declare D_i as the event that the client ' i ' defaults within a time-horizon and $\mathcal{P}(D_i)$ as the probability associated to this event, then the loss related to the obligator can be then defined as:

$$L_i = I * EAD_i * LGD_i \quad \text{with} \quad I = \mathbf{1}_{D_i}, \quad \mathcal{P}(D_i) = PD_i \quad (1)$$

where I is equal to the *indicator function* of the underlying event D . The variable takes value 1 when the event occur (i.e. default) and 0 when it does not occur. Note that this is a Bernoulli R.V., and as such, the expectation is equal to the probability of the event to occur. In other words, it is equal to the *Probability of Default (PD)*. Before laying down the formula of *Expected Loss*, is necessary to list down some assumptions ([**bankofenglandcreditrisk**], [**bluhm2016introduction**]):

1. To keep things simple, we will assume that the 3 constituents of the loss formula (i.e. PD, EAD, LGD) are independent among each other. Despite being a more than questionable statement and far from being true in general, this setting will allow us to state the most simple formula for the *Expected loss*.
2. To be able to simply define the expected loss over the whole portfolio of obligations, we also assume the joint distribution to be composed of independent and identically distributed R.V., each describing the expected loss of a specific obligator (i).

The formula of the expected loss of an obligator now comes very naturally as it is simply the expectation of the loss function defined above:

$$E[L_i] = \mathcal{P}(D_i) * EAD_i * LGD_i \quad (2)$$

where $\mathcal{P}(D_i)$ represents the expectation of I .

At this point, it is possible to identify the expected loss of a portfolio as the proportion of obligators that might default over a time-horizon, multiplied by the expected portion of the total expected exposures that is assumed not to be recovered if the default event occurs. More formally:

$$E[L_P] = \sum_{i=1}^N \mathcal{P}(D_i) * EAD_i * LGD_i \quad (3)$$

3.1.3 Unexpected Loss

Note that the formula to calculate the expected losses aims at measuring the **expected** proportion of losses due to defaulting obligators based on historical default experiences, but it does not consider the **unexpected** losses that might come up as a deviation from the average experienced losses of the past. [bluhm2016introduction]. The capital held to sustain losses should then consider both, and for the sake of completeness of this chapter, the formula will be provided here:

$$UL_P = \sum_{i=1}^N \sigma_i * \varphi_i \quad (4)$$

where σ_i is the individual standard deviation of credit losses for the obligator i , while φ_i indicates the correlation between the credit losses of individual i and all the other obligators in the portfolio.

After this brief parenthesis, let us go back to one of the key constituents of expected loss, the **probability of default**.

3.1.4 Probability of Default

The probability of default will be the subject of our empirical analysis where, given a pool of counterparties and associated loans, we wish to estimate how likely a loan is to default based on the underlying characteristics of both the obligators and the obligation.

Predicting the probability of default is in itself a very challenging task as often it involves very complicated models that need to assign a probability to each of the clients in the portfolio. Banks have set up multiple departments (usually under the division "risk methodology") in charge of estimating the 3 main constituents of the expected loss as a way to keep the situation monitored and intervene when needed. Great importance is also given to the

PD.

In principle, there are many methods to estimate the probability of default, but we will list only 3 different approaches and, for the scope of our analysis we will implement the last two listed here:

1. *Calibration from the **market data***: calibrating default probabilities on market data is not an easy task, and for this reason, there are already ad-hoc models (such as the KVM-model) that take care of this. Another option is based on credit spreads of the product bearing credit risk.
2. *Calibration from the **ratings***: default probabilities are linked with ratings (e.g. AAA, AA, ..., BBB, ...) which can be provided either by internal-based methodologies or by external rating agencies (e.g. Moody's, S&P, Fitch, ...)
3. ***Model-based** calibration*: lastly, it is possible to construct a model based on the available information on the obligator and the obligation and try to predict whether the counterparty will likely default or not on that loan. Note that the dependent variable here is treated as a binary variable (i.e. taking values 0 and 1 only) and for this reason, we are approximating a probability towards either 0 or 1 for non-default and default events respectively.

The dataset on top of which the empirical analysis will be conducted provides a sample of counterparties and relative loans, along with some information on both. In particular, given the rating for each loan, one idea might be to estimate the probability of default of each loan and use that estimate as a benchmark to see how the model performs on the data.

3.2 Empirical work

Introduction

This section presents a credit risk analysis and modelling on a portfolio of loans applying machine learning techniques. The core objective is to inspect whether, given a set of characteristics on the loans and the counterparties, it is possible to accurately assess the probability of default of an obligator with the ultimate objective of decreasing the default rate of the portfolio as a whole with mutual short and long-term benefits for both the lender and the client.

Frame the problem

Having a clear idea of the problem at hand is fundamental, specially when dealing with empirical data. The aim here is first define the requests from a business perspective and then translate them to reflect how to optimally approach these questions from a data science point of view. Following this procedure, it is possible to use statistical techniques to explore the data and provide justifiable answers.

From a managerial point of view, as a financial institution, the focus will be on understanding whether a counterparty of a loan in our loan portfolio is going to default or not on its obligation. In other words, we want to focus on assessing the default probability of single loans, which, from a data science perspective, it represents a binary classification problem. The output of the model can take two possible values:

- **1**: the counterparty is predicted to **default**
- **0**: the counterparty is predicted to **NOT default**

To inspect how the model performed and punish it for mistakes, it is necessary to define an error metric to minimize. Choosing an error metric much depend on the business case as well as on the data at hand, but for a classification problem the set of possible options boils down to just 3 : accuracy, precision and recall. Our choice will be taken and justified later in the analysis

Data collection

The data used to conduct the analysis was collected from Kaggle, a popular data science platform., and it is collected into a panel dataset. The latter simulates a loan portfolio from a credit bureau, including characteristics on the loans and on the obligators. For convenience of the analysis, the dataset is first splitted into two subsets to separate counterparties from loans characteristics and it is then combined together before the model gets trained.

Analysis

Modelling - Probability of Default

After an extensive overview and analysis of the features characterizing the dataset, it is time to proceed into modelling taking into consideration the findings emerging from the previous sections. Multiple models will be implemented here: first, a logistic regression model will be fitted and optimized in order used to derive the

4 Blockchain and asymmetric information