Risk Assessment of Credit Securities: The Notion and the Issues

Stelvio Cimato, Ernesto Damiani, Gabriele Gianini

Dipartimento di Tecnologie dell’Informazione
Università degli Studi di Milano
via Bramante 65, 26013 - Crema (CR), Italy
{cimato,damiani,gianini} at dti.unimi.it

Abstract

Assessing risk associated to structured financial products requires processing information about diverse risk factors. Some information comes from the different sources directly in aggregated form, so that it is not possible to estimate the correlation among different risk components. In this paper we argue that Secure Multiparty Computation could be used to merge the information from the different sources so as to compute more accurately the overall risk profile of securitized assets, thanks to the use of reliable information from the individual sources, but without disclosing the information from each source. Our simple model for securitization can solve in many cases information asymmetries between lenders and borrowers.

1 Introduction

Banks and other credit institutions provide loans to a wide range of subjects, thus facing the risk that – for a variety of reasons, related to different risk sources – a number of borrowers will not, or will only partially, meet their obligations. This is indicated as credit risk. As a consequence, the banks’ ex-ante assessment of the riskiness of loan applicants and the resulting decision to grant credit (or not) at some risk-adjusted interest rate, is of great importance: for instance under the regulations of the international Basel Accords [7] assessment is used to set aside an overall minimum regulatory buffer capital. The term securitization [9] describes the financial practice of pooling diverse contractual debts (including, to name but a few, mortgages, auto loans and credit card debt obligations) to create a ”security” sold to various investors, who then receive the principal and interest on the debt underlying the security. This way, credit risk is transferred by the originating bank to investors by pooling the loans into a portfolio which is then used to create suitably structured credit securities. In this case credit risk assessment of the
overall loan portfolio backing the security becomes the key component of the security’s rating (the other risk components being related to actors of the securitization process other than the borrower). Securitization has evolved from its tentative beginnings in the late 1970s to a vital funding source with an estimated outstanding of 10.24 trillion in the United States and 2.25 trillion in Europe as of the 2nd quarter of 2008. In this position paper, expanding on our previous work [34] we argue that obfuscation due to securitization pooling process can limit investors’ ability to monitor risk, and discuss how secure computation techniques could be used to restore this ability. According to many observers, securitization has played an important role in the U.S. subprime mortgage crisis [2, 1]. Credit risk assessment bears many similarities to reliability analysis in that, for the elementary component (e.g. a borrower’s loan), the default can be modeled as an arrival process, albeit normally more sophisticated than those used to model system reliability. Such a process can be driven by a number of factors – e.g. the state of a market, a region or an industry – modeling the failure of a composite system as the interplay of the failure of its individual components. In order to quantify – in terms of a loss probability distribution – the risk associated to a security one needs first to model the uncertainty related to the individual factors and then to model the relationships among them, known as linkages, which are a source of correlation. Linkages are very important, inasmuch mean loss and loss variance of different individual risk factors add linearly when the factors are uncorrelated, whereas variance can be substantially higher if some factors are positively correlated. Information about some of these factors, along with the correlations among them are normally available to any bank’s credit analysts. This is the case of publicly available historical data, known as systemic risks: macro-economical data, such as the economical situation of a borrowing company’s home country, region or the situation of the market (e.g., industry), in which the company has its main activities. This is also the case of some borrower-specific risk data, which, though not publicly available, are provided directly by the borrower during the application: debt, short- and long-term liabilities, and financial obligations, capital structure, liquidity of the firm’s assets, and so on. Generally speaking, defaults can be successfully predicted: firms headed to default are less profitable, show weaker sales and investment growth, have lower liquidity ratios and are more dependent on external funding sources. However, some important information, referring to the borrower and to the relation to other subjects, may not be available to the bank: their associated risks are normally grouped into a risk component indicated as idiosyncratic risk (and modeled quantitatively by a normal or a gamma distribution). We argue that lack of information can become the source of severe flaws in the credit assessment. Underestimating the idiosyncratic credit risk can lead to severe consequences, as can be seen from historical records [6]: typically not considering existing (but undetected) borrower-to-borrower link-
ages leads to underestimating the size of the so called contagion effect, making a security riskier than it appears. The impact of underestimating contagion may be amplified by the presence of derivative products. A derivative – a.k.a. contingent claim – is a financial product whose price is a function of the price of the underlying asset. Credit derivatives insure and protect against so called credit events, i.e. adverse movements in the credit quality of the borrower. Credit Default Swaps (CDSs), for instance, are used by lenders to insure against the borrower’s default or downgrade: the insurer, called protection seller, takes up the credit risk in exchange of a risk adjusted-fee. The risk of credit derivatives as well is sensitive to the correlation between risk components of the underlying credit. On the other hand, conservatively overestimating the idiosyncratic risk may lead a bank not to accept an applicant’s loan request, with a damage for both parties and for the economy as a whole. The issue of idiosyncratic risk is especially relevant at the light of the Basel II Accords on Capital Adequacy, published in 2004 [7], about how much capital banks need to put aside to guard against the types of financial and operational risks banks face. The accord proposes the use of credit risk models to determine banks’ capital requirements: banks can use internal (or external) rating models to classify borrowers according to their risk; capital requirements can then be determined based on such credit exposure, instead of being constant per credit type, as under the previous accord.

In this paper we use the results of [34] to discuss the problem of assessing the credit risk associated to a borrower or to a security by privacy preserving methods. A solution to this problem comes from the applications of secure multiparty computation (SMC) techniques, which enable different parties to perform distributed computation on secret inputs. Following this paradigm, it is possible to compute any public function and share the result among the parties, preserving the privacy of the inputs such that each party learns anything more than the computed values after the execution of the protocol. In the present discussion we use the simplified model for securitization introduced in [34] for computing the overall credit risk and solve in many cases information asymmetries between lenders and borrowers.

2 An illustrative example: borrower-sets overlap

In this Section, we will use the simplified model of securitization introduced in [34] in order to capture the major critical elements of the problem and outline the role of Secure Multiparty Computation.

2.1 A simplified model of the Securitization process

In a simplified single-period model of securitization a number of Borrowers loans at a given time a unitary amount of money from a number of Lenders: after a fixed period of time they will have to pay back an amount of money $a$, comprehensive
of the original amount of money plus a suitable risk-adjusted interest. The Lenders
pass their assets to an Issuer, which pools the loans into a single security, where all
the loans have the same weight, so that an investor buying a share \( s \) of the security
buys an equal share of each individual loan.

Consider a single-period model with only two lenders, indicated by \( C \) and \( D \), in which, at time \( T = 0 \), \( m \) borrowers, \((B_1, B_2, \ldots, B_m)\), obtain a loan from
bank \( C \), whereas \( n \) borrowers, \((B_{m+1}, B_{m+2}, \ldots, B_{m+n})\), obtain a loan from bank
\( D \). We assume for sake of simplicity a that for each Borrower there are only two
possibilities: either at the end of the period the borrower can pays the full amount,
or at the end of the period the Borrower defaults completely and pays a null amount:
all the Borrowers have the same probability \( p \) of default, known to the Lender,
so that the loss of the individual investor will be proportional to the number of
individual loans’ defaults. The Rating Agency in charge for rating the security
is interested in computing the probability distribution of the loss, i.e. the amount
which is not payed back at the end of the period – synthesized by its mean and its
variance – under two different scenarios: in the first one the individual borrowers’
default events are completely uncorrelated, in the second one they are correlated in
a simple way, by the overlap of the two sets of borrowers. Note that, although the
expected value is often used as the sole indicator, criteria based on variance have
been demonstrated more efficient in achieving risk efficiency (see for instance, to
the point where more complex indicators become unnecessary [29]).

**Uncorrelated failures.** In the case there is no correlation at all, i.e. in the case
where the the individual loans’ default events are independent events, the total
numbers of degrees of freedom of the problem is \( n + m \) and the probability of a
number \( k \) of loans’ defaults is obviously a Binomial variable

\[
\text{Bin}(k | (m + n), p) = \binom{m + n}{k} p^k (1-p)^{n+m-k}
\]

with mean \( \mu = (n + m)p \) and variance \( \sigma^2 = (m + n)pq \), where \( q = (1 - p) \).

**Simply correlated failures.** We now assume the loans’ defaults can be correlated
only in a simple way: a number \( r \) of borrowers obtained two loans: a loan from
bank \( C \) and one from bank \( D \): a default of one within those \( r \) borrowers will trigger
the default of two loans (this criticality is know as overlap problem).

Note that this assumption reduces the effective number of degrees of freedom
(number of independent actors) of the problem from \( n + m \) to \( n + m - r \), of which
\( r \) have a double weight in the loss computation, whereas \( m + n - 2r \) have the
usual unitary weight. In this case the distribution of the number \( k \) of loan defaults
consists in the sum of two independent variables

\[ k = k' + 2k'' \]
the variable $k'$ follows the binomial

$$Bin(k'|(m + n - 2)r), p)$$

whereas the variable $k''$ follows the binomial

$$Bin(k''|r, p)$$

The mean of the resulting variable is $\mu = (n + m - 2r)p + 2rp = (n + m)p$ i.e. equal to the uncorrelated case, however the variance is now

$$\sigma^2 = (m + n - 2r)pq + 4rpq = (m + n + 2r)pq$$

i.e. higher of an amount $2rpq$ with respect to the uncorrelated case. The risk for the investor, which are proportional to $\sigma$, are consequently higher. Furthermore if the risk-adjusted interest has been computed on the hypothesis of independent defaults investors are not being rewarded adequately for the risk actually taken.

2.2 Role of SMC in the securitization process

Should lenders know, through a common loan registry, the amount $r$ of borrowers who obtained a loan from both banks, they would be able to convey this information to the Issuer and in turn to the Rating Agency, which would then be able to compute correctly the risk associated to the security, and possibly to tranche the risk classes accordingly. When lenders do not have the availability of a common loan registry, and do not want to communicate to each other the identities of the respective borrowers, they would not be able to know of the amount $r$ of borrowers who obtained a loan from both banks, i.e. the degree of bank assets inter correlation. In this case the Rating Agency, by assuming any fixed correlation value (for instance uncorrelated defaults, $r = 0$), would issue an incorrect risk estimate. Assuming some prior probability distribution over $r$ would yield a variance estimate itself associated to some uncertainty and would be exposed to the prior-selection
bias. Now Secure Multiparty Computation comes into play: through SMC the parties C and D can find out the superposition between the two borrowers pools, i.e., to find the value of the parameter $r$, without disclosing one bank’s borrowers’ identities to the other bank. We will show in Section 3 how this task can be accomplished.

3 Secure Multiparty Computation

The paradigm of secure two-party computation was introduced by Yao in 1982 [33], and successively generalized to include multi participants. The resulting protocol assumes the representation of the function $F$ to be computed via a boolean circuit and enable the parties to engage in a protocol to securely compute the output of each gate till the completion of the result. While the theoretical result is strong (any function can be computed in a secure way), the resulting protocol is not very practical, since its complexity depends on both the number of the inputs and the size of the combinatorial circuit representing the function $F$. Different applications of SMC have been presented in the literature, however in many cases, the main results in this area remain of theoretical interests, since simple computations involving few parties still require several rounds of interaction. SMC can be efficiently used to perform the requested computations in the different scenarios. Basically, a protocol to compute set intersection and secure sum is requested, in which the rating agency executes a SMC round using the inputs provided by the banks (for the sake of simplicity we consider here computations involving two banks, but the protocols can be easily extended to take into account a large number of participants). Indeed, in the first example it is sufficient to compute the cardinality of the intersection of the borrower sets belonging to each of the involved banks. In the second and third scenarios it is requested to securely compute the sum of the default probability or the sum of the expected loss assigned to each of the borrowers, respectively. Finally in the last scenario the sum of the product of the default probability and of the expected loss is requested. In these three cases, the borrowers belonging to overlapping sets should be recognized in order to compute correctly the expectation and the variance values.

Some of the requested computation can be straightforwardly obtained after the application of some of the SMC protocols presented in literature. In some cases however they must be adapted to return the expected result.

The protocols for SMC used in this paper rely on a semantically secure [23], additively homomorphic public-key crypto-system. If $E_{pk}(X)$ denotes the encryption function with public key $pk$, the homomorphic crypto-system allows the computation of the following operations, performed without knowledge of the private key. If $E_{pk}(a)$ and $E_{pk}(b)$, are the encryptions of $a$ and $b$, respectively, it is possible to efficiently compute the
\[ E_{pk}(a + b) = E_{pk}(a) + E_{pk}(b) \]

\[ E_{pk}(c \cdot a) = c \cdot E_{pk}(a) \]

where \( c \) is a constant. When such operations are performed, we require that the resulting cipher texts be re-randomized for security, i.e. the ciphertext is transformed so as to form an encryption of the same plaintext, under a different random string than the one originally used. The Paillier’s crypto-system [28] satisfies the above requirements, since it is additively homomorphic and supports ciphertexts re-randomization. In the remainder of this paper, we simply use \( E(X) \) to denote the encryption function of the homomorphic crypto-system which satisfies all the aforementioned properties.

4 Computing overlapping sets

Let us consider the SMC protocol presented in [20] and successively refined in [25] to privately compute the size of the intersection set of two parties. As far as adversaries are concerned, we consider a semi-honest behaviour. Such an approach rely on a homomorphic crypto-system allowing the parties to perform some basic operations on (encrypted) coefficients of a polynomial. We refer the interested reader to their papers for the details, and for a deep discussion of the security of the protocol.

In our case, we consider two banks \( B_1 \) and \( B_2 \), each owning its own set of borrowers, \( U = \{u_1, \ldots, u_m\} \) for \( B_1 \) and \( V = \{v_1, \ldots, v_n\} \) for \( B_2 \), and wanting to compute the size of the overlapping identities of their borrowers, without disclosing any other detail. We assume that \( B_1 \) and \( B_2 \) have a common approach to map identities of the borrowers to a given domain set of identifiers, we denote with \( ID \). \( B_1 \) should define a new polynomial \( P \) whose roots are the identities of the borrowers belonging to its dataset:

\[ P(y) = (u_1 - y)(u_2 - y) \cdots (u_m - y) = \sum_{i=0}^{m} \alpha_i y^i \]

She then sends to \( B_2 \) the homomorphic encryption of the \( m + 1 \) coefficients \( \alpha_i \). \( B_2 \) can use the homomorphic properties to evaluate the polynomial at a given point knowing only the encryptions of the coefficients. She then chooses a random value \( r \) and computes \( E(r \cdot P(v_i) + 0^+) \) for each \( v_i \in V \), where \( 0^+ \) is a string of 0’s. Then she randomly permutes the obtained results, and sends back them to \( B_1 \). \( B_1 \) has now to decrypt the returned ciphertexts and count the number of 0 strings, which is equivalent to the number of borrowers contained in the intersection set. The obtained parameter is sufficient to compute the expectation and the variance.
4.1 Overlapping sets with exposure

The protocol needed to compute the expectation and variance in the remaining scenarios is a composition of the basic intersection protocol and of a secure sum protocol. In this case, each of the borrower belonging to the borrower set $U$ and $V$, has a default probability $p_i$ associated. The protocol should output the summation of such default probabilities considering the overlapping identities.

Basically the first steps are equal to the previous case until $B_2$ evaluates the polynomial in the point $E(r \cdot P(v_i) + v_i||p_i)$ for each $v_i \in V$, where $p_i$ is the associated default probability of borrower $v_i$. When $B_1$ decrypts the ciphertexts returned by $B_2$, she can determine the borrowers who are in the intersection set (since the resulting decryption returns the same element $v_i$) and the associated default probability. Now she can compute the sum of the default probability of the borrowers belonging to the intersection set (which will count two times in the computation of the expectation) and the sum of the default probability of the users $u_i \in U \setminus (U \cap V)$, $PM = \sum_{i \in V \setminus (U \cap V)} p_i + 2 \sum_{j \in U \cap V} p_j$. Such value will be sent to $B_2$ in encrypted form as well as the identities of the borrowers belonging to the intersection set. $B_2$ can now compute the sum of the default probability of the users $v_j \in V \setminus (U \cap V)$, obtaining $PV = \sum_{v_j \in V \setminus (U \cap V)} p_j$ and returning the result $E(PM + PV)$ to $B_1$. $B_1$ can decrypt and publish the final value. In the last scenario each borrower is associated with both a default probability $p_i$ and expected loss $w_i$. The computations can be performed using the same protocol above described, with the modification that both $p_i$ and $w_i$ values should be disclosed for the participants belonging to the intersection set. Basically $B_2$ evaluates the polynomial in the point $E(r \cdot P(v_i) + v_i||p_i||w_i)$ for each $v_i \in V$.

4.2 Other roles of Secure Multiparty Computation in Loan Origination and in Securitization

In most countries the above presented stylized problem of borrower-sets overlap is not what occurs in practice. Rather than in the identity of the borrower, the problem resides in the correlation between borrowers, which not always can be spotted by querying loan databases (nor comprehensively modeled through factorial models). Suppose that, differently from what assumed above, the degree of correlation may assume values other than zero and one and allow different reasons for correlation other that two borrowers sharing the same identity: the borrowers could be correlated because associated in some entrepreneurial activity, or because they belong to the same family and are likely to be backed or backing up the same family members, or because they applied for the loan to support their activities as subcontractors of the same contractor. These linkages opens the possibility for hidden correlations among loans issued by the same bank (beside leaving open the pos-
sibility for the inter-bank asset correlations mentioned in the previous example), including correlations which general models cannot account for completely. Even in presence of a loan registry, those correlating conditions are not easy to be spotted by the bank: open queries to the register may be restricted so as to not allow a cross-check with other registries, the restriction could be enforced by the trusted authority in charge for each registry management. Now, from the point of view of the borrower there might be the need for asking for a loan without disclosing all the information relating them to other borrowers (this behavior is indicated as predatory borrowing). The borrower could even be not aware of the relevance of some information and of the existence of other potential correlated borrowers. For the bank on the other hand it would be important to know those conditions. In this context the Secure Multiparty computation could find applications in many ways.

- A round of secure computation over different registries could compute, _ex-ante_, the superposition of the attributes of the loan applicant with other already approved borrowers: in case the bank deems that the approval of the loan would introduce excessive correlation within the bank asset, (reducing too much the effective degrees of freedom), the applicant may be denied the loan or charged a congruent risk-adjusted rate. All this would take place without any particular disclosure from the applicant, and, for what matters, without the bank being implicitly forced to reveal to the applicant anything about its borrowers portfolio profile.

- After having accepted a set of loan applications – on a case-by-case basis by using the information available under the restricted case-by-case queries policies – the bank may wish to perform, _ex-post_, through secure computation, a the credit risk assessment also based on the degree of correlation of its own loan assets.

- At the moment of pooling the loan assets of more banks into a security, the rating agency may use a round of secure computation to assess the credit risk of the overall portfolio.

### 5 Conclusions and Outlook

The proliferation of personal information collection tools and the increasing power of analysis tools for advanced data mining are often seen as a serious threat to the privacy of individuals. The problem of performing different kind of analysis on confidential data, held by different parties which do not want to disclose their information, is emerging as interesting research field in the last years. In several contexts, there is a request for information accountability, meaning that the use of information should be transparent and misuse should be traceable [32]. One of
such contexts is the credit market either during the loan origination or during loan pool securitization process.

In our research work [34] we investigated role of Secure Multiparty Computation with respect to the process of securitization, showing how SMC can allow a fine grained control of the variables controlling the risk embedded in a security, without compromising data privacy.

References

[1] Adam B. Ashcraft, Til Schuermann, Understanding the Securitization of Sub-prime Mortgage Credit, Federal Reserve Bank of New York Staff Reports, Staff Report no. 318, March 2008, JEL classification: G24, G28


