Article

A urn-based nonparametric modeling of the dependence between PD and LGD with an application to mortgages

Dan Cheng¹,‡ and Pasquale Cirillo¹,†,‡

¹ Applied Probability Group, DIAM, Delft University of Technology
* Correspondence: P.Cirillo@tudelft.nl; Tel.: +31-15-2-782-589
† Current address: E1.260 Building 28, Van Mourik Broekmanweg 6, 2628 XE Delft, The Netherlands
‡ These authors contributed equally to this work.

Abstract: We propose an alternative approach to the modeling of the positive dependence between the probability of default and the loss given default in a portfolio of exposures, using a bivariate urn process. The model combines the power of Bayesian nonparametrics and statistical learning, allowing for the elicitation and the exploitation of experts' judgements, and for the constant update of this information over time, every time new data is available.

A real-world application on mortgages is described using the Single Family Loan-Level Dataset by Freddie Mac.

Keywords: Probability of default; Loss given default; Wrong-way risk; Dependence; Urn model

1. Introduction

The ambition of this paper is to present a new way of modeling the empirically-verified positive dependence between the Probability of Default (PD) and the Loss Given Default (LGD) using a Bayesian nonparametric approach, based on urns and beta-Stacy processes (Walker and Muliere 1997). The model is able to learn from the data and it improves its performances over time, compatibly with the machine learning paradigm. Similarly to the recent construction of Cheng and Cirillo (2018) for recovery times, this learning ability is mainly due to the underlying urn model.

The PD and the LGD are two fundamental quantities in modern credit risk management. The PD of a counterparty indicates the likelihood that counterparty will default and thus not be able to fulfil its debt obligations. The LGD represents the percent loss, in terms of the notional value of the exposure, known as the Exposure-at-Default or EAD, one actually experiences when a counterparty defaults and every possible recovery process is over. Within the Basel framework (BCBS 2000 2011), a set of international standards developed by the Basel Committee on Banking Supervision to harmonize the banking sector and improve the way banks manage risk, the PD and the LGD are considered pivotal risk parameters for the quantification of the minimum capital requirements for credit risk. In particular, under the so-called Internal Rating-Based (IRB) approaches, both the PD and the LGD are inputs of the main formulas for the computation of the risk-weighted assets (BCBS 2005 2006).

Surprisingly, in most theoretical models, and most of all in the formulas suggested by the BCBS, the PD and the LGD are assumed to be independent, even though several empirical studies have shown that there is a non-negligible positive dependence between them (Altman 2006; Altman et al. 2001; Frye 2005; Miu and Ozdemir 2006; Witzany 2011). Borrowing from the terminology developed in the field of credit valuation adjustment (CVA) (Hull 2015), this dependence is often referred to as wrong-way risk (WWR). Simply put, WWR is the risk that the possible loss generated by a counterparty increases

Electronic copy available at: https://ssrn.com/abstract=3360531
with the deterioration of its creditworthiness. Ignoring this WWR can easily lead to an unreliable estimation of credit risk for a given counterparty (Witzany 2011).

The link between the PD and the LGD is particularly important when dealing with mortgages, as shown by the 2007-2008 financial crisis, which was triggered by an avalanche of defaults in the US subprime mortgage market (Eichengreen et al. 2012), with a consequent drop in estate prices, and thus in the recovery rates of the defaulted exposures (Hull 2015). The financial crisis was therefore one of the main drivers for the rising interest in the joint modeling of PD and LGD, something long ignored both in the academia and in the practice—including regulators. In the paper we show how the model we propose can be used in the field of mortgages with a real-world application.

The first model implicitly dealing with the dependence between the PD and the LGD was Merton’s one. In his seminal work, Merton (1974) introduced the first structural model of default, developing what we can consider the Black and Scholes model of credit risk. In that model, the recovery rate $RR$ (remember $RR = 1 - LGD$), conditionally on the credit event, follows a lognormal distribution (Resti and Sironi 2007), and it is negatively correlated with the PD. While not consistent with later empirical findings (Altman et al. 2005; Jones et al. 1984) on WWR, Merton’s model remained for long time the only model actually dealing with this problem.

Many models have been derived from Merton’s original construction, both in the academic and in the industrial literature, think for example of Geske (1977), Kim et al. (1993), Longstaff and Schwartz (1995), Nielsen et al. (2001) and Vasicek (1984). During the 90’s, on the wake of the forthcoming financial regulations (Basel I and later Basel II), several new approaches were introduced, like the so-called reduced-form family (Duffie 1998; Duffie and Singleton 1999; Lando 1998) and the VaR models (JP Morgan 1997; Wilde 1997; Wilson 1998). However, and somehow surprisingly, the great majority of these models, while solving other problems of Merton’s original contribution—like the fact that default could only happen at predetermined times—often neglected the dependence between PD and LGD, sometimes even explicitly assuming independence, as in the Credit Risk Plus case (Wilde 1997), where LGD is mainly assumed deterministic.

In the new century, given the rising interest of regulators and investors, especially after the 2007 crisis, a lot of empirical research has definitely shown the positive dependence (in most cases, positive correlation) between PD and LGD. Frye (2000) proposed a standard one-factor model, becoming a pivotal reference in the modeling of the WWR between PD and LGD. In particular, the paper focused on the PD and the Market LGD of corporate bonds on a firm level, assuming dependence on a common systematic factor, plus some other independent idiosyncratic components to deal with marginal variability.

Witzany (2011) moved forward, proposing a two-factor model on retail data, looking at the overall economic environment as the source of dependence between PD and LGD. For the LGD only, then, a second component accounts for the conditions of the economy during the workout process. Other meaningful constructions in the common factor framework are Hamerle et al. (2011) and Yao et al. (2017). The latter is also interesting for the literature review, together with Altman et al. (2005).

Using a two-state latent variable, Bruche and Gonzalez-Aguado (2010) introduced another valuable model. The time series of the latent variable is referred to as credit cycle, and it is represented by a simple Markov chain. Interestingly, this credit cycle variable shows to be able to better capture the time variation in the joint distribution of the PD and the LGD, with respect to observable macroeconomic factors.

Notwithstanding the importance of the topic, and despite the notable efforts cited above, it seems that there is still a lot to do in the modeling of the PD/LGD dependence, which remains a yet-to-be-developed part of credit risk management, see Altman (2006), Maio (2017) and the references there cited.

Our contribution to this open problem is represented by the present paper, which finds its root in the recent literature about the use of urn models in credit risk management (Amerio et al. 2004; Cheng and Cirillo 2018; Cirillo et al. 2010; Peluso et al. 2015). The paper develops as follows: Section
2 is devoted to the description of the theoretical model and the introduction of all the necessary probabilistic tools. In Section 3 we briefly describe the Freddie Mac data we then use in Section 4 where we show the performances of the model, when studying the dependence between PD and LGD for residential US mortgages. Section 5 finally closes the paper.

2. Model

The model we propose was first presented in Bulla (2005) in the field of survival analysis, and it is based on powerful tools of Bayesian nonparametrics, like the beta-Stacy process of Walker and Muliere (1997). Here we give a slightly different representation in terms of Reinforced Urn Processes (RUP), to bridge towards some recent papers in the credit risk management literature like Cheng and Cirillo (2018) and Peluso et al. (2015).

2.1. The two-color RUP

A RUP is a special type of combinatorial stochastic process, first introduced in Muliere et al. (2000). It can be seen as a reinforced random walk over a state space of urns, and depending on how its parameters are specified, it can generate a large number of interesting models. Essential references on the topic are Fortini and Petrone (2012) and Muliere et al. (2000 2003). In this paper we only need to specify a RUP able to generate a discrete beta-Stacy process, that is a particular random distribution over the space of discrete distributions.

Definition 1 (Discrete beta-Stacy Process (Walker and Muliere 1997)). A random distribution function $F$ is a beta-Stacy process with jumps at $j \in \mathbb{N}_0$ and parameters $\{\alpha_j, \beta_j\}_{j \in \mathbb{N}_0}$ if there exist mutually independent random variables $\{V_j\}_{j \in \mathbb{N}_0}$ each beta distributed with parameters $(\alpha_j, \beta_j)$, such that the random mass assigned by $F$ to $\{j\}$, written $F(\{j\})$, is given by $V_j \prod_{l<j}(1-V_l)$.

Consider the set of the natural numbers $\mathbb{N}_0$, including zero. On each $j \in \mathbb{N}_0$, place a Polya urn $U(j)$ containing balls of two colors, say blue and red, and reinforcement equal to 1. A Polya urn is the prototype of urn with reinforcement: every time a ball is sampled, its color is recorded, and the ball is put back into the urn together with an extra ball (i.e. the reinforcement) of the same color (Mahmoud 2008). This mechanism clearly increases the probability of picking the sampled color again in the future. We assume that all urns $U(j)$ contain at least a ball of each color, but their compositions can be actually different. As the only exception, the urn centered on 0, $U(0)$, only contains blue balls, so that red balls cannot be sampled.

Let $n = 0, 1, 2, ...$ represent time, which we take to be discrete. A two-color reinforced urn process $\{Z_n\}_{n \geq 0}$ is built as follows. Set $Z_0 = 0$ and sample urn $U(0)$: given our assumptions, the only color we can pick is blue. Put the ball back into $U(0)$, add an extra blue ball, and set $Z_1 = 1$. Now, sample $U(1)$: this urn contains both blue and red balls. If the sampled ball is blue, put it back, add an extra blue ball and set $Z_2 = 2$. If the sampled balls is red, put it back in $U(1)$, add an extra red ball and set $Z_2 = 0$. In general, the value of $Z_n$ will be decided by the sampling of the urn visited by $Z_{n-1}$. If a blue ball is selected, then $Z_n = Z_{n-1} + 1$, otherwise $Z_n = 0$. Blue balls thus make the process $\{Z_n\}_{n \geq 0}$ go further in the exploration of the natural numbers, while every red ball makes the process restart from 0. If one is interested in defining a two-color RUP only on a subset of $\mathbb{N}_0$, say $\{0, 1, ..., k\}$, it is sufficient to set to zero the number of blue balls in all urns $U(k), U(k+1), U(k+2), ...$

It is not difficult to verify that, if all urns $U(j)$, $j = 1, 2, ...$, contain a positive number of red balls, the process $\{Z_n\}_{n \geq 0}$ is recurrent, and it will visit 0 infinitely many times for $n \to \infty$.

A two-color recurrent RUP clearly generates sequences of nonnegative natural numbers, like the following

\[
\{Z_0, Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, ..., Z_{10}, Z_{11}, Z_{12}, ..., Z_{15}, ...\} = \{0, 1, 2, 0, 1, 2, 3, ..., 7, 0, 1, ..., 4, ...\}. \tag{1}
\]
These sequences are easily split into blocks, each starting with 0, as in Equation (1). In terms of sampling, those 0’s represent the times the process has been reset to 0 after the extraction of a red ball from the urn centered on the natural number preceding that 0. Notice that by construction no sequence can contain two contiguous 0’s. In the example above, the process is reset to 0 after the sampling of \( U(2) \) in \( n = 2 \), \( U(7) \) in \( n = 10 \), and \( U(4) \) in \( n = 15 \). Muliere et al. (2000) have shown that the blocks (they call them 0-blocks) generated by a two-color RUP are exchangeable, and that their de Finetti measure is a beta-Stacy process.

Exchangeability means that, if we take the joint distribution of a certain number of blocks, this distribution is immune to a reshuffling of the blocks themselves, i.e. we can permute them, changing their order of appearance, and yet the probability of the sequence containing them will be the same (for example \( P(\text{Block 1, Block 2, Block 3}) = P(\text{Block 1, Block 3, Block 2}) \) in Equation (1)). Notice that the exchangeability of the blocks does not imply the exchangeability of the single states visited, in fact one can easily check that each block constitutes a Markov chain. This implies that the RUP can be seen as a mixture of Markov chains, while the sequence of visited states is partially exchangeable in the sense of Diaconis and Freedman (1980).

Given their exchangeability, de Finetti’s representation theorem guarantees that there exists a random distribution function \( F \) such that, given \( F \), the blocks generated by the two-color RUP are i.i.d. with distribution \( F \). Theorem 3.26 in Muliere et al. (2000) tells us that such a random distribution is a beta-Stacy process with parameters \( \{\alpha_j, \beta_j\}_{j \in \mathbb{N}_0} \) where \( \alpha_j \) and \( \beta_j \) are the initial numbers of red, respectively blue, balls in urn \( U(j) \), prior to any sampling. In other words, \( F(0) = 0 \) with probability 1 and, for \( j \geq 1 \), the increment \( F(j) - F(j-1) \) has the same distribution as \( V_j \prod_{i=1}^{j-1} (1-V_i) \), where \( \{V_j\} \) is a sequence of independent random variables, such that \( V_j \sim \beta(\alpha_j, \beta_j) \). For a reader familiar with urn processes, each beta distributed \( V_j \) is clearly the result of the corresponding Polya urn \( U(j) \) (Mahmoud 2008).

Let \( B_1, B_2, ..., B_m \) be the first \( m \) blocks generated by a two-color RUP \( \{Z_n\}_{n\geq0} \). With \( T_i \) we indicate the last state visited by \( \{Z_n\} \) within block \( i \). Coming back to Equation (1), we would have \( T_1 = 2 \), \( T_2 = 7 \) and \( T_3 = 4 \). Since the random variables \( T_1, ..., T_m \) are measurable functions of the exchangeable blocks, they are exchangeable as well, and their de Finetti measure is the same beta-Stacy governing \( B_1, B_2, ..., B_m \). In what follows a sequence \( \{T_i\}_{i=1}^m \) is called LS (Last State) sequence. In terms of probabilities, for \( T_1, ..., T_m \), one can easily observe that

\[
P[T_1 = j] = \frac{\alpha_j}{\alpha_j + \beta_j} \prod_{i=0}^{j-1} \frac{\beta_i}{\alpha_i + \beta_i},
\]

and, for \( m \geq 1 \),

\[
P[T_{m+1} = j | T_1, T_2, ..., T_m] = \frac{\alpha_j + r_j}{\alpha_j + \beta_j + r_j + s_j} \prod_{i=0}^{j-1} \frac{\beta_i + s_i}{\alpha_i + \beta_i + r_i + s_i},
\]

while, trivially,

\[
P[T_{m+1} \geq j | T_1, T_2, ..., T_m] = \prod_{i=0}^{j} \frac{\beta_i + s_i}{\alpha_i + \beta_i + r_i + s_i},
\]

where \( r_j = \sum_{i=1}^{m} 1_{\{T_i = j\}} \) and \( s_j = \sum_{i=1}^{m} 1_{\{T_i > j\}} \). This means that, every time it is reset to 0, creating a new block, a RUP remembers what happened in the past thanks to the Polya reinforcement mechanism of each visited urn. The process thus learns, combining its initial knowledge, as represented by the quantities \( \{r(t), b(t)\}_{t \in \mathbb{N}_0} \), with the additional balls that are introduced in the system, to obtain the predictive in Equation (3).
Regarding the initial knowledge, notice that, when we choose the quantities \( a_j \) and \( \beta_j \) for \( j = 0, 1, 2, \ldots \), from a Bayesian point of view we are eliciting a prior. In fact, by setting

\[
a_j = c_j(G(\{j\}) \quad \text{and} \quad \beta_j = c_j \left( 1 - \sum_{i=0}^{j} (G(\{i\})) \right), \quad j \in \mathbb{N}_0,
\]

we are just requiring \( E[F(\{j\})] = G(\{j\}) \), so that the beta-Stacy process \( F \)– a random distribution on discrete distributions– is centered on the discrete distribution \( G \), which we guess may correctly describe the phenomenon we are modeling. The quantity \( c_j \geq 0 \) is called strength of belief, and it represents how confident we are in our a priori. Given a constant reinforcement, as the one we are using here (+1 ball of the same color), a \( c_j > 1 \) reduces the speed of learning of the RUP, making the evidence emerging from sampling less relevant in updating the initial compositions. In other terms, \( c_j \) helps in controlling the stickiness of \( F(\{j\}) \) to \( G(\{j\}) \). For more details, we refer to Muliere et al. (2000 2003).

\section*{2.2. Modeling dependence}
Consider a portfolio \( P \) containing \( m \) exposures. For \( i = 1, \ldots, m \), let \( X_i \) and \( Y_i \) represent the PD, respectively the LGD, of the \( i \)-th counterparty, when discretised and transformed into levels, in a way similar to what Cheng and Cirillo (2018) propose in their work. In other terms, one can split the PD and the LGD into \( l = 0, \ldots, L \) levels, such that for example \( l = 0 \) indicates a PD or LGD of 0\%, \( l = 1 \) something between 0\% and 5\%, \( l = 2 \) a quantity in (5\%, 17\%), and so on until the last level \( L \). The levels do not need to correspond to equally spaced intervals, and this gives flexibility to the modeling.

Clearly, the larger \( L \) the finer the partition we obtain. As we will see in Section 4, a convenient way of defining levels is through quantiles.

As observed in Altman et al. (2005), discretisation is a common and useful procedure in risk management, as it reduces the noise in the data. The unavoidable loss of information is more than compensated by the gain in interpretability, if levels are chosen in the correct way. From now on, the bivariate sequence \( \{(X_i, Y_i)\}_{i=1}^{m} \) is therefore our object of interest in studying the dependence between the PD and the LGD. Both \( X_i \) and \( Y_i \) will take values in \( \{0, 1, \ldots, L\} \).

Let \( \{A_i\}_{i=1}^{m}, \{B_i\}_{i=1}^{m}, \{C_i\}_{i=1}^{m} \) be three independent LS sequences generated by three independent two-color RUPs \( \{Z^A_j\}_{j \geq 0}, \{Z^B_j\}_{j \geq 0}, \{Z^C_j\}_{j \geq 0} \). As we know, the sequence \( \{A_i\}_{i=1}^{m} \) is exchangeable, and its de Finetti measure is a beta-Stacy process \( F_A \) of parameters \( \{a^A_j, \beta^A_j\}_{j \in \mathbb{N}_0} \). Similarly, for \( \{B_i\}_{i=1}^{m} \) and \( \{C_i\}_{i=1}^{m} \) we have \( F_B \) with \( \{a^B_j, \beta^B_j\}_{j \in \mathbb{N}_0} \) and \( F_C \) with \( \{a^C_j, \beta^C_j\}_{j \in \mathbb{N}_0} \).

Now, following Bulla et al. (2007), let us assume that, for each exposure \( i = 1, \ldots, m \), we have

\[
X_i = A_i + B_i, \quad Y_i = A_i + C_i.
\]

This construction builds a special dependence between the discretised PD and the discretised LGD of our exposures. In fact, we are assuming that for each counterparty \( i \) there exists a common factor \( A_i \) influencing both the PD and the LGD, while \( B_i \) and \( C_i \) can be seen as idiosyncratic components (conditionally on \( A_i \), \( X_i \) and \( Y_i \) are clearly independent). Since both \( X \) and \( Y \) are between 0 and 100\%\(^1\), the compositions of the urns defining the processes \( \{Z^A_j\}_{j \geq 0}, \{Z^B_j\}_{j \geq 0}, \{Z^C_j\}_{j \geq 0} \) can be tuned so that it is not possible to observe values larger than 100\%.

\(^1\) In reality, as observed in Zhang and Thomas (2012), the LGD can be slightly negative or slightly above 100\%, because of fees and interests, however we exclude that situation here. In terms of applications, all negative values can be set to 0, and all values above 100 can be rounded to 100.
From Equation (6) we can derive important features of the model we are proposing. Since we can write \( Y_i = X_i - B_i + C_i \), it is clear that we are assuming a linear dependence between \( X_i \) and \( Y_i \). This is compatible with several empirical findings, like Altman et al. (2005) or Miu and Ozdemir (2006).

Furthermore, given Equation (6) and the properties of the sequences \( \{ A_i \}_{i=1}^m \), \( \{ B_i \}_{i=1}^m \) and \( \{ C_i \}_{i=1}^m \), we can immediately observe that

\[
\begin{align*}
\text{Cov}(X_1, Y_1) &= \text{Var}(A_1) \geq 0, \\
\text{Cov}(X_{m+1}, Y_{m+1}|A_{m}, B_{m}, C_{m}) &= \text{Var}(A_{m+1}|A_m) \geq 0, \text{ for } m \geq 2,
\end{align*}
\]

where \( \text{Var} \) is the variance, \( \text{Cov} \) the covariance, \( A_m = [A_1, \ldots, A_m] \), and similarly \( B_m, C_m \). Therefore, with the bivariate urn model we can only model positive dependence. Again, this is totally in line with the purpose of our analysis—the study of wrong-way risk that by definition is a positive dependence between PD and LGD—and with the empirical literature, as discussed in Section 1. However, it is important to stress that the model in Equation (6) cannot be used when negative dependence is possible.

Always from Equation (6), we can verify that the sequence \( \{(X_i, Y_i)\}_{i=1}^m \) is exchangeable\(^2\). This comes directly from the exchangeability of \( \{A_i\}_{i=1}^m \), \( \{B_i\}_{i=1}^m \) and \( \{C_i\}_{i=1}^m \) and the fact that \( (X_i, Y_i) \) is a measurable function of \( \{A_i, B_i, C_i\} \). An implicit assumption of our model is therefore that the \( m \) counterparties in \( \mathcal{P} \) are exchangeable. As observed in McNeil et al. (2015), exchangeability is a common assumption in credit risk, as it is seen as a relaxation of the stronger hypothesis of independence (think about Bernoulli mixtures and the beta-binomial model). All in all, what we ask is that the order in which we observe our counterparties is irrelevant to study the joint distribution of their PDs and LGDs, which is therefore immune to changes in the order of appearance of each exposure. Exchangeability and the fact that \( X_i \) and \( Y_i \) are conditionally independent given \( A_i \) suggest that the methodology we are proposing falls under the larger umbrella of mixture models (Duffie and Singleton 2003; McNeil et al. 2015).

Since \( \{(X_i, Y_i)\}_{i=1}^m \) is exchangeable, de Finetti’s representation theorem guarantees the existence of a bivariate random distribution \( F_{XY} \), conditionally on which the couples are i.i.d. with distribution \( F_{XY} \). The properties of \( F_{XY} \) have been studied in detail in Bulla (2005) and Bulla et al. (2007).

Let \( F_X \) and \( F_Y \) be the marginal distributions of \( X_i \) and \( Y_i \). Clearly we have

\[
\begin{align*}
F_X &= F_A \times F_B, \\
F_Y &= F_A \times F_C,
\end{align*}
\]

so that both \( F_X \) and \( F_Y \) are convolutions of beta-Stacy processes. The dependence between \( X \) and \( Y \), given \( F_{XY} \) and \( F_A \) is thus simply

\[
\text{Cov}_{F_{XY}}(X, Y) = \text{Var}_{F_A}(A) = \sigma_A^2.
\]

Furthermore, if \( P \) is the probability function corresponding to \( F \), one then has

\[
P_{XY}(x, y) = \sum_{a=0}^{\min(x,y)} P_A(a) P_B(x-a) P_C(y-a), \quad \forall x, y \in \mathbb{N}_0^2.
\]

Assume now that we have observed \( m \) exposures, and we have registered their actual PD and LGD, which we have discretised to get \( \{(X_i, Y_i)\}_{i=1}^m \). The construction of Equation (6), together with the properties of the beta-Stacy processes involved, allows for a nice derivation of the predictive

\(^2\) Please observe that exchangeability only applies among the couples \( \{(X_i, Y_i)\}_{i=1}^m \), while within each couple there is a clear dependence, so that \( X_i \) and \( Y_i \) are not exchangeable.
distribution for a new exposure \((X_{m+1}, Y_{m+1})\), given the observed couples \((X_{m} = x_{m}, Y_{m} = y_{m})\). This can be extremely useful in applications, when one is interested in making inference about the PD, the LGD and their relation.

In fact

\[
P[X_{m+1} = x, Y_{m+1} = y | X_{m} = x_{m}, Y_{m} = y_{m}] = \frac{P[X_{m+1} = x, Y_{m+1} = y, X_{m} = x_{m}, Y_{m} = y_{m}]}{P[X_{m} = x_{m}, Y_{m} = y_{m}]}.
\]

Given Equation (6), Equation (9) can be rewritten as follows

\[
P[X_{m+1} = x, Y_{m+1} = y | X_{m} = x_{m}, Y_{m} = y_{m}] = \sum_{a_{m}} P[X_{m+1} = x, Y_{m+1} = y | A_{m} = a_{m}, B_{m} = b_{m}, C_{m} = c_{m}]
\]

\[
\times P[A_{m} = a_{m} | X_{m} = x_{m}, Y_{m} = y_{m}],
\]

where \(b_{m} = x_{m} - a_{m}\) and \(c_{m} = y_{m} - a_{m}\).

From a theoretical point of view, computing Equation (10) just requires to count the number balls in the urns behind \(\{A_{1}\}_{i=1}^{m}, \{B_{1}\}_{i=1}^{m}\) and \(\{C_{1}\}_{i=1}^{m}\) and use formulas like those in Equations (2) and (3); something that for a small portfolio can be done explicitly. However, when \(m\) is large, it becomes numerically unfeasible to perform all those sums and products.

Luckily, developing an alternative Markov Chain Monte Carlo algorithm is simple and effective. It is sufficient to go through the following steps.

1. Given the observations \(X_{m} = x_{m}\) and \(Y_{m} = y_{m}\), the sequence \(A_{m} = (A_{1}, \ldots, A_{m})\) is generated via a Gibbs sampling. The full conditional of \(A_{m}, P[A_{m} = a_{m} | A_{m-1} = a_{m-1}, X_{m} = x_{m}, Y_{m} = y_{m}]\) is such that

\[
P[A_{m} = a_{m} | A_{m-1} = a_{m-1}, X_{m} = x_{m}, Y_{m} = y_{m}] \propto P[A_{m} = a_{m} | A_{m-1} = a_{m-1}]
\]

\[
\times P[X_{m} - A_{m} = x_{m} - a_{m} | B_{m-1} = b_{m-1}]
\]

\[
\times P[Y_{m} - A_{m} = y_{m} - a_{m} | C_{m-1} = c_{m-1}].
\]

Since \(\{A_{1}\}_{i=1}^{m}\) is exchangeable, all the other full conditionals, \(P[A_{j} = a_{j} | A_{-j} = a_{-j}, X_{m} = x_{m}, Y_{m} = y_{m}]\), where \(A_{-j} = (A_{1}, \ldots, A_{j-1}, A_{j+1}, \ldots, A_{m})\), have an analogous form.

2. Once \(A_{m}\) is obtained, compute \(B_{m} = X_{m} - A_{m}\) and \(C_{m} = Y_{m} - A_{m}\).

3. The quantities \(A_{m+1}, B_{m+1}\), and \(C_{m+1}\) are then sampled according to their beta-Stacy predictive distributions \(P(A_{m+1} | A_{m}), P(B_{m+1} | B_{m}),\) and \(P(C_{m+1} | C_{m})\) as per Equation (3).

4. Finally, set \(X_{m+1} = A_{m+1} + B_{m+1}\) and \(Y_{m+1} = A_{m+1} + C_{m+1}\).

3. Data

We want to show the potentialities of the bivariate urn construction of Section 2 in modeling the dependence between PD and LGD in a large portfolio of residential mortgages. The data we use come from Maio (2017). The author kindly allowed us to share them, and they are now available at (here link to repository of the journal).

Maio’s dataset is the result of cleaning operations (treatment of NA’s, inconsistent data, etc.) on the well-known Single Family Loan-Level Dataset Sample by Freddie Mac, freely available at http://www.freddiemac.com/research/datasets/sf_loanlevel_dataset.page. Freddie Mac’s sample contains 50’000 observations per year over the period 1999-2017. Observations are randomly selected, on a yearly basis, from the much larger Single Family Loan-Level Dataset, covering approximately 26.6 million fixed-rate mortgages. Extensive documentation about Freddie Mac’s data collections can be found in Freddie Mac (2019a).
Maio’s dataset contains 383465 loans over the period 2002-2016. Each loan is uniquely identified by an alphanumeric code, which can be used to match the data with the original Freddie Mac’s source.3

For each loan, several interesting pieces of information are available, like its origination date, the loan age in months, the geographical location (ZIP code) within the US, the FICO score of the subscriber, the presence of some form of insurance, the loan to value, the combined loan to value, the debt-to-income ratio, and many others. In terms of credit performance, quantities like the unpaid principal balance and the delinquency status up to termination date are known. Clearly, termination can be due to several reasons, from voluntary prepayment to foreclosure, and this information is also recorded, following Freddie Mac (2019a). A loan is considered defaulted when it is delinquent for more than 180 days, even if it is later repurchased (Freddie Mac 2019b). In addition to the information also available in the Freddie Mac’s collection, Maio’s dataset is enriched with estimates of the PD and the LGD for each loan, obtained via survival analysis. It is worth noticing that both the PD and the LGD are always contained in the interval [0, 1].

If one considers the pooled data, the correlation between the PD and the LGD is 0.2556. The average PD is 0.0121 with a standard deviation of 0.0157, while for LGD we have 0.1494 and 0.0937 respectively. Regarding the minima and the maxima, we have $2.14 \times 10^{-5}$ and 0.3723 for PD, and 0 and 0.5361 for LGD.

In the parametric approach proposed by Maio (2017), the covariates which affect both the PD and LGD, possibly justifying their positive dependence, are the unpaid principal balance (UPB) and the debt-to-income ratio (DTI). Other covariates, from the age of the loan to the ZIP code, are then relevant in explaining the marginal behaviour of either the PD or the LGD. In modeling both the PD and the LGD, Maio (2017) proposes the use of two Weibull accelerated failure time (AFT) models, following a recent trend in modern credit risk management (Narain 1992). The dependence between PD and LGD is then modelled parametrically using copulas and a brand new approach involving a bivariate beta distribution. For more details, we naturally refer to Maio (2017).

A correlation around 0.26 clearly indicates a positive dependence between PD and LGD, and it is in line with the empirical literature we have mentioned in Section 1. However, considering all mortgages together may not be the correct approach, as we are pooling together very different counterparties, possibly watering down more meaningful areas of dependence.

Given the richness of the dataset, there are many ways of disaggregating the data. For example, one can compute the correlation between PD and LGD for different FICO score classes. The FICO score, originally developed by the Fair Isaac Corporation (https://www.fico.com), is a leading credit score in the US, and one of the significant covariates for the estimation of PD in Maio’s survival model (Maio 2017). In the original Freddie Mac’s sample it ranges from 301 to 850.

Following a common classification (Experian 2019), we can define five classes of creditworthiness:

- Very poor, for a FICO below 579;
- Fair, for 580-669;
- Good, for 670-739;
- Very good, for 740-799; and
- Exceptional, with a score above 800.

Table 1 contains the number of loans in the different classes, the corresponding average PD and LGD, and naturally the correlation $\rho$ between PD and LGD. As expected the average PD is higher when the FICO score is lower, while the opposite is observable for the average LGD. This is probably due to the fact that, for less creditworthy counterparties, stronger insurances are generally required, as compensation for the higher risk of default. Moreover, in terms of recovery, in putting the collateral on the market, it is probably easier to sell a house with a lower value than a very expensive property, for which discounts on the price are quite common. Interestingly, in disaggregating the data, we see that the PD-LGD correlation is always above 0.3, with the only exception of the most reliable FICO class (0.19).

3 To avoid any copyright problem with Freddie Mac, which already shares its data online, from Maio’s dataset we only provide the PD and the LGD estimates, together with the unique alphanumeric identifier. In this way, merging the data sources is straightforward.
Table 1. Some descriptive information about the data used in the analysis. Loans are collected in terms of FICO score.

| Class       | Number of Loans | Avg. PD  | Avg. LGD | \( \rho \)  |
|-------------|-----------------|----------|----------|-------------|
| Very Poor   | 1627            | 0.0378   | 0.1013   | 0.3370      |
| Fair        | 46720           | 0.0238   | 0.1237   | 0.4346      |
| Good        | 124824          | 0.0138   | 0.1409   | 0.3159      |
| Very good   | 177891          | 0.0083   | 0.1574   | 0.3599      |
| Exceptional | 32403           | 0.0080   | 0.1777   | 0.1858      |

As an example, Figure 1 shows two plots of the relation between PD and LGD for the "Very poor" class. On the left, a simple scatter plot of PD vs LGD. On the right, to deal with the large number of overlapping points, thus improving interpretability, we provide a hexagonal heatmap with counts. To obtain this plot, the plane is divided into regular hexagons (20 for each dimension), the number of cases in each hexagon is counted, and it is then mapped to a color scale. Both plots tell us that most of the PD-LGD couples lie in the square \([0,0.1]^2\).

![Scatter plot and heatmap](image-url)

Figure 1. Plots of PD vs LGD for mortgages in the "very poor" (FICO score below 579) class. On the left, a simple scatter plot. On the right, a hexagonal heatmap with counts.

In Figure 2, the two histograms of the marginal distributions of PD and LGD in the "Very poor" rating class are shown. While for PD the distribution is unimodal, for LGD we can clearly see bimodality (a second bump is visible around 0.25). These behaviours are present among the different classes and at the pooled level.

4. Results

In this section we discuss the performances of the bivariate urn model on the mortgage data described in Section 3. For the sake of space, we will mainly focus our attention on the "Very poor" and the "Exceptional" FICO score classes, as per Table 1.

In order to use the model, we need 1) to transform and discretise both the PD and the LGD into levels, and 2) to define an a priori for the different beta-Stacy processes involved in the construction of Equation (6).

The results we obtain are promising and suggest that the bivariate urn model can represent an interesting way of modeling PD and LGD dependence.

4.1. Discretisation

In order to discretise the PD and LGD into \( X \) and \( Y \), it is necessary to choose the appropriate levels \( l = 0, 1, \ldots, L \). In the absence of specific ranges, possibly arising from a bank's business practice...
or imposed by a regulator, a convenient way for defining the levels is through quantiles (Cheng and Cirillo 2018).

For example, let $d_{v1}, \ldots, d_{v9}$ be the deciles for the quantity $v \in \{PD, LGD\}$. We can set levels $l = 0, 1, \ldots, 10$, where

$$L1 := \{0 : 0; 1 : (0, d_{v1}]; 2 : (d_{v1}, d_{v2}]; \ldots; 9 : (d_{v8}, d_{v9}]; 10 : > d_{v9}\}.$$  \hspace{1cm} (11)

Such a partition guarantees that each level, apart from $l = 0$ contains 10% of the values for both PD and LGD. Notice that the thresholds are not the same, for example, when focusing on the "Very poor" class, $d_{PD}^{1} = 0.0099$, while $d_{LGD}^{1} = 0.0030$. A finer partition can be obtained by choosing other percentiles. Clearly, in using a similar approach, one should remember that she is imposing a uniform behaviour on $X$ and $Y$, in a way similar to copulas (Nelsen 2006). However, differently from copulas, the dependence between $X$ and $Y$ is not restricted to any particular parametric form: dependence will emerge from the combination of the a priori and the data.

Another simple way for defining levels is to round the raw observations to the nearest largest integer (ceiling) or to some other value. For instance, we can consider:

$$L2 := \{0 : 0; 1 : (0, 1\%]; 2 : (1, 2\%]; \ldots; 100 : (99, 100\%]\}.$$  \hspace{1cm} (12)

Even if it is not a strong requirement, given the meaning of the value 0 in a RUP (recall the 0-blocks), we recommend to use 0 as a special level, not mapping to an interval.

Notice that, if correctly applied, discretisation maintains the dependence structure between the variables. For the "Very poor" class, using the levels in Equation (11) in defining $X$ and $Y$, we find that $cor(X, Y) = 0.3348$, in line with the value 0.3370 in Table 1. For the "Exceptional" group, using the levels in Equation (12), we obtain 0.2080, still comparable.

In what follows, we discuss the results mainly using the levels in Equation (12). However, our findings are robust to different choices of the partition. In general, choosing a smaller number of levels improves fitting, because the number of observations per level increases, reinforcing the Polya learning process, ceteris paribus. Moreover–and this is why partitions based on quantiles give nice performances–better results are obtained when intervals guarantee more or less the same number of observations per level.

Choosing a very large $L$ may lead to the situation in which for a specific level no transition is observed, so that for that level no learning via reinforcement is possible, and, if our beliefs are wrong–or not meant to compensate some lack of information in the data, this has naturally an impact on the goodness of fit. Therefore, in choosing $L$ there is a trade-off between fitting and precision. The higher $L$, the lower the impact of discretisation and noise reduction.
The right way to define the number of levels is therefore to look for a compromise between precision, as required by internal procedures or the regulator, and the quantity and the quality of the empirical data. The more and the better the observations, the more precise the partition can be, because higher is the chance of not observing empty levels. In any case each level should guarantee a minimum number of observations in order to exploit the Bayesian learning mechanism. To improve fitting, rarely visited levels should be aggregated. Once again, experts’ judgements could represent a possible solution.

4.2. Prior elicitation

In order to use the bivariate urn model, it is necessary to elicit an a priori for all its components, namely \( \{A_i\}_{i=1}^m \), \( \{B_i\}_{i=1}^m \) and \( \{C_i\}_{i=1}^m \). This is fundamental to initialise the Markov Chain Monte Carlo algorithm discussed in Subsection 2.2.

Prior elicitation can be performed in a completely subjective way, on the basis of one’s knowledge of the phenomenon under scrutiny, or by looking at the data. Recalling Equation (8), it seems natural to choose the prior for \( \{A_i\}_{i=1}^m \) so that its variance coincides with the empirical covariance between \( X \) and \( Y \). When looking at the "Very poor" and "Exceptional" classes, this covariance is approximately 3 (i.e. 2.7345 and 3.2066 respectively), using the levels in Equation (11). For the situation in Equation (12), conversely, we have a covariance of 10.28 for the "Very poor" class, and 1.46 for the other. In general, apart from the constraint on the variance—which does make sense—the choice of the distribution for \( A \) can be completely free.

Given the ranges of variation of \( X \) and \( Y \), one can choose the appropriate supports for the three beta-Stacy processes. Considering the levels in Equation (11), for \( \{A_i\}_{i=1}^m \), \( \{B_i\}_{i=1}^m \) and \( \{C_i\}_{i=1}^m \), it is not rational to choose priors putting a positive mass above 10. Being the levels defined via deciles, it is impossible to observe a level equal to 11 or more. Similarly, using the levels in Equation (12) and recalling Figure 1, where neither the PD nor the LGD reach values above 40% (0.4), once can decide not to allow large values of \( X \) and \( Y \), initialising the RUP’s urns only until level 40.

Since we do not have any specific experts’ knowledge to exploit, we have tried different prior combinations. Two possibilities are:

- Independent discrete uniforms for \( \{A_i\}_{i=1}^m \), \( \{B_i\}_{i=1}^m \) and \( \{C_i\}_{i=1}^m \), where the range of variation for \( B \) and \( C \) is simply inherited from \( X \) and \( Y \) (but extra conditions can be applied, if needed), while for \( A \) the range is chosen to guarantee \( \sigma_A^2 = \text{Cov}(X, Y) \). For instance, if the covariance between \( X \) and \( Y \) is approximately 3, the interval \([0, 5]\) guarantees that \( \sigma_A^2 \approx 3 \) as well. We can simply use the formula for the variance of a discrete uniform, i.e.

\[
\sigma^2 = \frac{(b - a + 1)^2 - 1}{12}.
\]

- Independent Poisson distributions, such that \( A \sim \text{Poi}(\lambda_A = \text{Cov}(X, Y)) \), while for \( B \) and \( C \) one sets \( \text{Poi}(\bar{X} - \lambda_A) \) and \( \text{Poi}(\bar{Y} - \lambda_A) \), where \( \bar{X} \) is the empirical mean of \( X \). This guarantees, for example, that \( X \sim \text{Poi}(\bar{X}) \). Recall in fact that in a Poisson random variable the mean and the variance are both equal to the intensity parameter. Given our data, where the empirical variances of \( X \) and \( Y \) are not at all equal to the empirical means, but definitely larger, the Poisson prior can be seen as an example of wrong prior.

Once the priors have been decided, the urn compositions can be derived via Equation (5). Different values of \( c_j \) are tried in our experiments. Here we discuss the cases \( c_j = 1 \) and \( c_j = 100 \) for all values of \( j \). The former indicates a moderate trust in our a priori, while the latter shows a strong confidence in our beliefs. Cheng and Cirillo (2018) have observed that, in constructions involving the use of reinforced urn processes, if the number of observations used to train the model is large, then priors become asymptotically irrelevant, for the empirical data prevail. However, when the number of data points is not very large, having a strong prior does make a difference. Given the set cardinalities, we
shall see that a strong prior has a clear impact for the "Very poor" rating class ("only" 1627 observations), while no appreciable effect is observable for the "Exceptional" one (32403 data points).

As a final remark, it is worth noticing that one could take all the urns behind \( \{A_i\}_{i=1}^m \) to be empty. This would correspond to assuming that no dependence is actually possible between PD and LGD, so that \( X_i = B_i \), \( Y_i = C_i \) and \( \text{Cov}(X, Y) = 0 \). As discussed in Bulla et al. (2007), it can be shown that when \( A \) is degenerate on 0–and no learning on dependence is thus possible, given the infringement of Cromwell’s rule (Lindley 1991)–the bivariate urn model simply corresponds to playing with the bivariate empirical distribution and the bivariate Kaplan-Meier estimator. While certainly not useful to model a dependence we know is present, this setting further props the flexibility of the bivariate urn model up.

4.3. Fitting

Figure 3 shows, for the exposures in the "Very Poor" group, the very good fitting performances of the model for the marginal distributions of \( X \) and \( Y \), the discretised PD and LGD respectively. Each subfigure shows the elicited prior, the empirical cumulative distribution function (ecdf) and the posterior, as obtained via learning and reinforcement. In the figure shown, the levels are given by Equation (12), while the priors are discrete uniforms with \( c_j = 1 \). Since the covariance between \( X \) and \( Y \) is 10.28, the discrete uniform on \( A \) is on \([0, 10]\). For both \( B \) and \( C \) the support is \([0, 40]\). A two sample Kolmogorov-Smirnov (KS) test does not reject the null hypothesis of same distribution between the ecdfs and the relative posteriors (p-values stably above 0.05).

Figure 4 is generated in the same way of Figure 3. The only difference is that now \( c_j = 100 \), indicating a strong belief on the uniform priors. In this case, it is more difficult for the bivariate urn process to update the prior, and in fact the posterior distribution is not as close as before to the ecdf. The effect is more visible for PD than for LGD, however in both cases the KS test rejects the null (p-value 0.005 and 0.038). Please notice that this is not necessarily a problem, if one really believes that the available data do not contain all the necessary information, or she want to incorporate specific knowledge about future trends and so on.

As anticipated, the number of observations available in the data plays a major role in updating—and, in case of a wrongly elicited belief, correcting—the prior. Focusing on the PD, Figure 5 shows that when the "Exceptional" rating class is considered, no big difference can be observed in the obtained posterior, no matter the strength of belief \( c_j \). In fact, for both cases, a KS test does not reject the null hypothesis with respect to the ecdf (p-values: 0.74 and 0.58). The reason is simple: with more than thirty thousand observations, the model necessarily converges towards the ecdf, even if the prior distribution is clearly wrong and we put a strong belief on it. Similar results hold for the LGD. In producing the figure, \( A \) has a uniform prior on \([0, 3]\), while \( B \) and \( C \) have support \([0, 45]\).

Figure 6 shows the bivariate distribution we obtain on for the discretised PD and LGD for the "Very poor" FICO score group, when \( c_j = 1 \) and we use the Poisson priors on the levels of Equation

Electronic copy available at: https://ssrn.com/abstract=3360531
Figure 4. Prior, posterior and ecdf for the discretised PD and LGD in the "Very poor" rating class, when priors are uniform and levels are like in Equation (12). The strength of belief is always set to 100.

Figure 5. Prior, posteriors \(c_j = 1\) and \(c_j=100\) and ecdf for the discretised PD in the "Exceptional" rating class, when priors are uniform.

(12). Figure 7 shows the case \(c_j = 100\). As one would expect, the strong prior provides a smoother joint distribution, while the weak one tends to make the empirical data prevail. In Figure 8 the equivalent of Figure 6 is given for the "Exceptional" rating class.

Regarding the numbers, all priors and level settings are able to model the dependence between \(X\) and \(Y\). The correlation is properly captured (the way in which the prior on \(A\) is defined surely helps), as well as the mean and the variances of the marginals, especially when the strength of beliefs is smaller. The only problem is represented by the Poisson priors used on the "Very poor" class. In this case, the number of observations is not sufficient to correct the error induced by the use of a Poisson distribution, i.e. same value for mean and variance, and the variance of both PD and LGD is underestimated, while the mean is correctly captured. In particular, while the estimated and actual means are 3.81 and 3.79 for \(X\), and 10.61 and 10.12 for \(Y\), in the case of the variances the actual values 10.26 and 92.97 are definitely larger than the predicted ones, i.e. 5.88 and 22.50. This is a clear signal that more data would be necessary to properly move away from the prior beliefs.

In terms of dependence, the model is always able to satisfactorily replicate the positive correlation between PD and LGD. For example, in the case of Figure 6, the actual correlation is 0.3370, and the model estimates 0.3366. However, under Poisson priors with strong degrees of belief a certain
overestimation is observable for the "Very poor" rating group, in line with the underestimation of the variances.

The results we have just commented are all in sample, that is we have used all the observations available to verify how the model fits the data, no out of sample performance was checked. Using the Bayesian terminology of Jackman (2009) and Meng (1994), we have indeed performed a posterior consistency check.

Luckily, the out of sample validation of the bivariate urn model is equally satisfactory. To perform it, we have used the Freddie Mac’s sampling year to create two samples. The first one includes all the loans (362104) sampled in the period 2002-2015 and we call it the training sample. The validation sample, conversely, includes all the loans sampled in 2016, for a total of 21321 data points. The samples have then been split into FICO score groups, as before.

For the "Very Poor" class, Figure 9 shows the predictive distribution of the PD as obtained by training the bivariate urn construction (uniform priors) on the training sample (1590 loans) against the ecdf of the corresponding validation sample (37 loans). The fit is definitely acceptable, especially if we consider the difference in the sample sizes. The mean (level) of the predictive distribution is 3.85, while that of the validation sample is 3.81. The standard deviations are 3.24 and 2.53. The median 3 in both cases. Regarding dependence, a correlation of 0.3419 is predicted by the model, while the realised one in the validation sample is smaller and equal to 0.1963. This difference is probably due to the fact that in the validation set the PD never exceeds level 13, while in the training one the maximum level reached is 31. When the validation set is larger, as in the case of the "Exceptional" class, the dependence is better predicted: one has a correlation of 0.1855 against 0.1799. Similar or better results are obtained for the other classes, using the different prior sets, once again with the only exception of the Poisson priors with large strength of belief, when applied to the "Very poor" class.

It is important to stress that this simple out of sample validation does not guarantee that the predictive distribution would be able to actually predict new data, in the case of a major change in the underlying phenomenon—something we have not observed in our data—like a structural break to use the econometric jargon. That would be exactly the case in which the clever use of the prior distribution—in the form of expert prior knowledge—could contribute in obtaining a better fit.
5. Conclusions

We have presented a bivariate urn construction to model the dependence between PD and LGD, also showing a promising application on mortgage data. Exploiting the reinforcement mechanism of Polya urns and the conjugacy of beta-Stacy processes, the Bayesian nonparametric model we propose is able to combine experts’ judgements, in the form of a priori knowledge, with the empirical evidence coming from data, learning and improving its performances every time new information becomes available.

The possibility of using prior knowledge is an important feature of the Bayesian approach. In fact, it can compensate for the lack of information in the data with the beliefs of experienced professionals about possible trends and/or rare events. For rare events in particular, the possibility of eliciting an a priori on something rarely or never observed before empirically can mitigate, at least partially, the relevant problem of historical bias (Derbyshire 2017; Shackle 1955).

One could argue that nothing guarantees the ability of eliciting a reliable a priori: experts could be easily wrong. The answer to such a relevant observation is that the bivariate urn model learns over time, at every interaction with actual data. A sufficient amount of data can easily compensate unrealistic beliefs. Moreover, as already observed in Cirillo et al. (2013), thanks to reinforcement, urns are able to learn hidden patterns in the data, casting light on previously ignored relations and features. Such a capability bridges towards the paradigm of machine learning (Murphy 2012). However, differently from standard machine learning techniques, the combinatorial stochastic nature of the bivariate urn model allows for a greater control of its probabilistic features. There is no black box (Knight 2017).

Clearly, if an a priori cannot be elicited nor be desired, one can still use the model in a totally data-driven way. As shown in Section 4, in the absence of any a priori on the dependence between PD and LGD, the bivariate urn can easily generate an empirical bivariate cumulative distribution function.

Further, thanks to its nonparametric nature, the model we propose does not require the development of parametric assumptions, with the subsequent problem of choosing among them, as one should for instance do in using copulas (Nelsen 2006).

As observed in Section 2, an important limitation of the bivariate urn model is the possibility of only modeling positive linear dependence. Its use when dependence can be negative, or strongly non linear, is not advised. But this seems not to be the case with mortgages.
Finally, the bivariate urn model can be computationally intensive. For large datasets and portfolios, a standard laptop may require several hours if not days to obtain the posterior distribution. Clearly the quality of coding, which was not our main investigation path, has a big impact on the final performances.

**Author Contributions:** D.C. and P.C. contributed equally to the present work.

**Funding:** This research received no external funding.

**Acknowledgments:** Empty for the moment.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**

Altman, Edward I.. 2006. Default Recovery Rates and LGD in Credit Risk Modeling and Practice: An Updated Review of the Literature and Empirical Evidence. *New York University, Stern School of Business.*

Altman, Edward I., Brooks Brady, Andrea Resti, and Andrea Sironi. 2005. The Link between Default and Recovery Rates: Theory, Empirical Evidence, and Implications. *The Journal of Business* 78(6), 2203–2227.

Altman, Edward I., Andrea Resti, and Andrea Sironi. 2001. Analyzing and Explaining Default Recovery Rates. A report submitted to the International Swaps & Derivatives Association, [http://people.stern.nyu.edu/ealtman/Review1.pdf](http://people.stern.nyu.edu/ealtman/Review1.pdf).

Amerio, Emanuele, Pietro Muliere, and Piercesare Secchi. 2004. Reinforced Urn Processes for Modeling Credit Default Distributions. *International Journal of Theoretical and Applied Finance* 07(04), 407–423.

BCBS. 2000. Principles for the Management of Credit Risk. [https://www.bis.org/publ/bcbs75.pdf](https://www.bis.org/publ/bcbs75.pdf).

BCBS. 2005. An Explanatory Note on the Basel II IRB Risk Weight Functions. [https://www.bis.org/bcbs/irbriskweight.pdf](https://www.bis.org/bcbs/irbriskweight.pdf).

BCBS. 2006. International Convergence of Capital Measurement and Capital Standards. *In Bank for International Settlements, Number 30 June*, pp. 285.

BCBS. 2011. Basel III: A Global Regulatory Framework for More Resilient Banks and Banking Systems. *In Bank for International Settlements, Number 01 June*, pp. 69.

Bruche, Max and Carlos Gonzalez-Aguado. 2010. Recovery Rates, Default Probabilities, and the Credit Cycle. *Journal of Banking & Finance* 34(4), 754–764.

Bulla, Paolo. 2005. *Application of Reinforced Urn Processes to Survival Analysis*. Ph. D. thesis, Bocconi University.

![Figure 8. Bivariate density distribution of PD and LGD in the "Exceptional" rating class, with Poisson priors and strength of belief equal to 1.](image)
Figure 9. Predictive distribution generated by the bivariate urn model for the "Very poor" class when trained on the training sample (1590 loans), against the empirical ecdf from the validation set (37 loans).
Frye, Jon. 2005. The Effects of Systematic Credit Risk: A False Sense of Security. In E. Altman, A. C. Resti, and
A. Sironi (Eds.), Recovery risk, pp. 187–200. London: Risk books.

Geske, Robert. 1977. The Valuation of Corporate Liabilities as Compound Options. Journal of Financial and
Quantitative Analysis 12(4), 541–552.

Hamerle, Alfred, Michael Knapp, and Nicole Wildenauer. 2011. Modelling Loss Given Default: A “Point in
Time”-Approach. In B. Engelmann and R. Rauhmeier (Eds.), The Basel II Risk Parameters: Estimation, Validation,
Stress Testing - with Applications to Loan Risk Management, pp. 137–150. Berlin: Springer.

Hull, John C.. 2015. Risk Management and Financial Institutions (4th ed.). New York: Wiley.

Jackman, Simon. 2009. Bayesian Analysis for the Social Sciences. New York: Wiley.

Jones, Philip, Scott P. Mason, and Eric Rosenfeld. 1984. Contingent Claims Analysis of Corporate Capital
Structures: An Empirical Investigation. The Journal of Finance 39(3), 611–625.

JP Morgan. 1997. Creditmetrics - Technical Document, http://www.defaultrisk.com/_pdf6j4/creditmetrics_
techdoc.pdf.

Kim, In Joon, Krishna Ramaswamy, and Suresh Sundaresan. 1993. Does Default Risk in Coupons Affect
the Valuation of Corporate Bonds?: A Contingent Claims Model. Financial Management 22(3), 117–131.
doi:10.2307/3665932.

Knight, Will. 2017. The Dark Secret at the Heart of AI. Technology Review 120, 54–63.

Lando, David. 1998. On Cox Processes and Credit Risky Securities. Review of Derivatives Research 2(2), 99–120.

Lindley, Dennis. 1991. Making Decisions (2nd ed.). New York: Wiley.

Longstaff, Francis A. and Eduardo S. Schwartz. 1995. A Simple Approach to Valuing Risky Fixed and Floating
Rate Debt. The Journal of Finance 50(3), 789–819.

Mahmoud, Hosam. 2008. Polya Urn Models. Boca Raton: CRC Press.

Maio, Vittorio. 2017. Modelling the Dependence between PD and LGD. A New Regulatory Capital Calculation
with Empirical Analysis from the US Mortgage Market. Master’s thesis, Politecnico di Milano, https://www.
politesi.polimi.it/handle/10589/137281.

McNeil, Alexander J., Ruediger Frey, and Paul Embrechts. 2015. Quantitative Risk Management. Princeton:
Princeton University Press.

Meng, Xiao-Li. 1994. Multiple-Imputation Inferences with Uncongenial Sources of Input. Statistical Science 9(4),
538–558.

Merton, Robert C.. 1974. On the Pricing of Corporate Debt: The Risk Structure of Interest Rates. The Journal of
Finance 29(2), 449–470.

Miu, Peter and Bogie Ozdemir. 2006. Basel Requirement of Downturn LGD: Modeling and Estimating PD & LGD
Correlations. Journal of Credit Risk 2(2), 43–68.

Muliere, Pietro, Piercesare Secchi, and Stephen G Walker. 2000. Urn Schemes and Reinforced Random Walks.
Stochastic Processes and their Applications 88(1), 59–78.

Muliere, Pietro, Piercesare Secchi, and Stephen G Walker. 2003. Reinforced Random Processes in Continuous
Time. Stochastic Processes and their Applications 104(1), 117–130.

Murphy, Kevin P.. 2012. Machine Learning: A Probabilistic Perspective. Cambridge MA: The MIT Press.

Narain, B. 1992. Survival Analysis and the Credit Granting Decision. In E. D. B. Thomas LC, Crook JN (Ed.),
Credit Scoring and Credit Control, pp. 109–121. Oxford: Oxford University Press.

Nelsen, Roger B.. 2006. An Introduction to Copulas. New York: Springer.

Nielsen, Lars Tyge, Jesus Saá-Requejo, and Pedro Santa-Clara. 2001. Default Risk and Interest Rate Risk: The Term
Structure of Default Spreads. Paris: INSEAD.

Peluso, Stefano, Antonietta Mira, and Pietro Muliere. 2015. Reinforced Urn Processes for Credit Risk Models.
Journal of Econometrics 184(1), 1–12.

Resti, Andrea and Andrea Sironi. 2007. Risk Management and Shareholders’ Value in Banking. New York: Wiley.

Shackle, George Lennox Sharman. 1955. Uncertainty in Economics and Other Reflections. Cambridge: Cambridge
University Press.

Vasicek, Oldrich A. 1984. Credit valuation, http://www.ressources-actuarielles.net/EXT/ISFA/1226.nsf/0/
c181fb77ee99d464c125757a800505078/$FILE/Credit_Valuation.pdf.

Walker, Stephen and Pietro Muliere. 1997. Beta-Stacy Processes and a Generalization of the Polya-Urn Scheme.
The Annals of Statistics 25(4), 1762–1780.

Wilde, Tom. 1997. CreditRisk+: A Credit Risk Management Framework. Technical report.
Wilson, Thomas C. 1998. Portfolio credit risk. *Economic Policy Review* 4(3), 71–82.

Witzany, Jiří. 2011. A Two Factor Model for PD and LGD Correlation. *Bulletin of the Czech Econometric Society* 18(28).

Yao, Xiao, Jonathan Crook, and Galina Andreeva. 2017. Is It Obligor or Instrument That Explains Recovery Rate: Evidence from US Corporate Bond. *Journal of Financial Stability* 28, 1–15.

Zhang, Jie and Lyn C. Thomas. 2012. Comparisons of Linear Regression and Survival Analysis Using Single and Mixture Distributions Approaches in Modelling LGD. *International Journal of Forecasting* 28(1), 204–215.