Urn models and other approaches to risk and tails, with applications in risk management and climatology

Cheng, D.

DOI
10.4233/uuid:a1908c73-2bc4-4947-944e-2bd3a177bfe6

Publication date
2022

Document Version
Final published version

Citation (APA)
Cheng, D. (2022). *Urn models and other approaches to risk and tails, with applications in risk management and climatology*. [Dissertation (TU Delft), Delft University of Technology].
https://doi.org/10.4233/uuid:a1908c73-2bc4-4947-944e-2bd3a177bfe6

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

URN MODELS AND OTHER APPROACHES TO RISK AND TAILS, WITH APPLICATIONS IN RISK MANAGEMENT AND CLIMATOLOGY

DAN CHENG

URN MODELS AND OTHER APPROACHES TO RISK AND TAILS, WITH APPLICATIONS IN RISK MANAGEMENT AND CLIMATOLOGY

Propositions

accompanying the dissertation

Urn models and other approaches to risk and tails, with applications in risk management and climatology

by

Dan CHENG

1. In credit risk, if we divide the credit migration into the potential to migrate and the final decision to move or not, the potential to migrate might be used to design early warning mechanisms. (Chapter 2)
2. The model used in mortgage loans can also be easily applied to other types of loan categories, such as campus loans and consumer loans. (Chapter 3)
3. A stochastic Poisson equation can better predict a region of deterministic but unknown temperatures, under a certain metric, than a deterministic Poisson equation. (Chapter 4)
4. Financial institutions do interact with each other; hence, the credit risk of one counterparty will affect that of another. The cascading bankruptcies in the 2007-2008 financial crisis are a good example.
5. Bayesian statistics relies upon three elements: prior knowledge, empirical evidence, and the posterior distribution. Our opinion is updated with more and more empirical evidence. This is similar to the way of thinking of the human brain.
6. The extreme is better than all the data to make inferences about the more extreme.
7. The prosecutor's fallacy tells us that a small probability event does not necessarily occur with a small probability. Take for example a traffic accident at a hospital in a given region: one reason could be that the driver did it on purpose, for revenge on society or any other motivation; the other reason could be that the driver had no intention of making it happen and it was just an error. Both are small probability events since, in such a region, car accidents in hospitals have rarely occurred in the past. However, the second reason is more plausible, hence it can be seen to happen with a higher probability.
8. Give yourself permission to be human.
9. Life is not perfect, but we need to be positive enough to see the positive side of life.
10. A researcher deals with the unknown; hence I only need to deal with myself.

These propositions are regarded as opposable and defendable, and have been approved as such by the promotor prof. dr. Frank Redig.

Stellingen

behorende bij het proefschrift

Urn models and other approaches to risk and tails, with applications in risk management and climatology

door

Dan CHENG

1. Als we in kredietrisico de kredietmigratie verdelen in de migratiepotentie en de uiteindelijke beslissing om wel of niet te verplaatsen, zou de migratiepotentie gebruikt kunnen worden om mechanismen voor vroegtijdige waarschuwing te ontwerpen. (Hoofdstuk 2)

2. Het model dat wordt gebruikt voor hypotheekleningen kan ook gemakkelijk worden toegepast op andere typen leningscategorieën, zoals campusleningen en consumentenleningen. (Hoofdstuk 3)

3. Onder een bepaalde metriek is een stochastische Poissonvergelijking beter in het voorspellen van een gebied van deterministische maar onbekende temperaturen dan een deterministische Poissonvergelijking. (Hoofdstuk 4)

4. Financiële instituten hebben interactie met elkaar; daardoor beïnvloedt het kredietrisico van de ene tegenpartij die van een andere. De opeenstapeling van faillissementen tijdens de financiële crisis van 2007-2008 is een goed voorbeeld.

5. Bayesiaanse statistiek steunt op drie elementen: a-priori-kennis, empirisch bewijs en de a-posteriori-verdeling. Onze mening wordt geüpdatet met meer en meer empirisch bewijs. Dit is vergelijkbaar met de manier van denken van het menselijk brein.

6. Het extreme is beter dan alle data in de gevolgtrekking over het meer extreme.

7. De aanklagersdrogreden vertelt ons dat een gebeurtenis met kleine kans niet noodzakelijk met kleine kans voorkomt. Neem bijvoorbeeld een verkeersongeluk bij een ziekenhuis in een gegeven regio: een reden zou kunnen zijn dat de bestuurder het expres deed, uit wraak op de samenleving of een willekeurige andere motivatie; de andere reden zou kunnen zijn dat het niet de intentie van de bestuurder was dat dit zou gebeuren en het slechts een fout was. Beide zijn gebeurtenissen met kleine kans, aangezien in een dergelijke regio in het verleden zelden ongelukken met auto's in ziekenhuizen hebben plaatsgevonden. Echter, de tweede reden is meer plausibel, dus die zien we met een grotere kans gebeuren.

8. Sta jezelf toe om mens te zijn.

9. Het leven is niet perfect, maar het is nodig dat we positief genoeg zijn om de positieve kant van het leven te zien.
10. Een onderzoeker gaat om met het onbekende; daarom hoef ik alleen om te gaan met mijzelf.

Deze stellingen worden opponeerbaar en verdedigbaar geacht en zijn als zodanig goedgekeurd door de promotor prof. dr. Frank Redig.

URN MODELS AND OTHER APPROACHES TO RISK AND TAILS, WITH APPLICATIONS IN RISK MANAGEMENT AND CLIMATOLOGY

Dissertation

For the purpose of obtaining the degree of doctor
at Delft University of Technology
by the authority of the Rector Magnificus Prof.dr.ir. T.H.J.J. van der Hagen
chair of the Board for Doctorates
to be defended publicly on
Monday 12 September 2022 at 10:00 o'clock

by

Dan CHENG

Master of Science in Probability,
University of Science and Technology of China, China,
born in Xianning, China

This dissertation has been approved by the promotors.

Composition of the doctoral committee:

Rector Magnificus,	chairperson
Prof. dr. F.H.J. Redig,	Delft University of Technology, promotor
Prof. dr. P. Cirillo,	ZHAW School of Management and Law, copromotor
Independent members:	
Prof. dr. U. Cherubini,	University of Bologna, Italy
Prof. dr. I. Monasterolo,	EDHEC Business School, France
Prof. dr. A. Papapantoleon,	Delft University of Technology
Prof. dr. ir. C.W. Oosterlee,	Delft University of Technology/Utrecht University
Reserve member:	
Prof. dr. ir. G. Jongbloed,	Delft University of Technology

This research was funded by Delft University of Technology and China Scholarship Council.

Keywords: Reinforced Urn Process, Credit Risk, Probability of Default, Loss Given Default, Extreme Value Theory, Stochastic Poisson Equation, Spatio-temporal Data

ISBN 000-00-0000-000-0

Printed in the Netherlands

An electronic version of this dissertation is available at
http://repository.tudelft.nl/.

CONTENTS

SUMMARY

This dissertation collects three scientific contributions, already published in international peer-reviewed journals, plus some extra considerations and work-in-progress.

First, we present a model based on reinforced urn processes, which conjugates to the right-censored recovery process, and empirically apply it to the time series of recovery rates. We perform a very thorough empirical study, including how different priors affect the posterior predictive distribution, how our model is updated with the empirical data during the global financial crisis, and we make predictions. Second, we apply a bivariate reinforced process derived from a Generalized Polya Urn scheme to model the linear dependence between the probability of default and the loss given default. Third, we offer a new perspective with Stochastic Poisson equation to deal with Spatio-temporal extremes.

As it will be clear, the leit motiv of this thesis is the analysis of risk using different tools, from urn models to extreme value theory. In particular, we have focused on two risk applications: the modelling of credit risk in some of its declinations, and the prediction of the joint tail behavior of extreme sea surface temperature (SST) anomalies for the Red Sea.

Almost every financial contract is affected by credit risk, that is the risk of changes in the creditworthiness of a counterparty. Financial economists, market participants, bank supervisors, and regulators have all paid close attention to credit risk measurement, pricing, and management. The probability of default, the recovery rate, and their dependence are fundamental aspects of the credit risk. Measuring credit risk accurately is pivotal for four reasons. First, for financial economists, credit risk measures are very important for pricing credit risk portfolios, credit derivatives, etc. The importance of credit risk in the pricing of financial contracts has been underlined by the global financial crisis. Second, during the management process of credit risk for companies, the accurate credit risk measure can help the management team better determine their risk appetite. Third, the well-known Basel capital requirements are calculated using credit risk measure. Fourth, the accurate estimation of the credit risk can help a manager improve decisions. For example, in the recovery activities after default, more effort will be put on the individual with a high estimated LGD to reduce the large loss.

During my PhD studies I have also took part in several conferences, among which the 11th international conference on Extreme Value Analysis (EVA 2019). In attending this conference, I decided to participate in one of the proposed challenges for young scholars, something that led to the writing of one of the contributions of this work, which also won the first prize in the competition.

SAMENVATTING

Dit proefschrift is een verzameling van drie wetenschappelijke bijdragen, die al gepubliceerd zijn in internationale gepeerreviewde tijdschriften, plus een aantal extra beschouwingen en lopend onderzoek.

Ten eerste presenteren we een model dat is gebaseerd op versterkte urnprocessen, die geconjungeerd zijn aan rechts-gecensureerde terugvorderingsprocessen, en passen we die empirisch toe op de tijdreeksen van terugvorderingspercentages. We doen een zeer grondige empirische analyse, waaronder hoe verschillende a-priori-kansverdelingen de a-posteriori voorspellende verdeling beïnvloeden en hoe ons model wordt geüpdatet met de empirische data tijdens de wereldwijde financiële crisis, en we doen voorspellingen. Ten tweede passen we een bivariaat versterkt proces dat is afgeleid van een Gegeneraliseerd Polya Urnschema toe om de lineaire afhankelijkheid tussen de kans op wanbetaling en het verlies bij wanbetaling te modelleren. Ten derde bieden we met de Stochastische Poissonvergelijking een nieuw perspectief voor het omgaan met tijdruimtelijke extremen.

Zoals duidelijk wordt, is het leidmotief van dit proefschrift het analyseren van risico met behulp van verschillende gereedschappen, van urnmodellen tot de theorie van extreme waarden. We hebben ons in het bijzonder gericht op twee risicotoepassingen: het modelleren van het wanbetalingsrisico en herstelrisico die horen bij kredietrisico en het voorspellen van het gezamenlijke staartgedrag van extreme afwijkingen van zeewateroppervlaktetemperaturen (SST) van de Rode Zee.

Vrijwel elk financieel contract wordt beïnvloed door kredietrisico, dat is het risico op verandering van de kredietwaardigheid van een tegenpartij. Financiële economen, marktdeelnemers, banktoezichthouders en regelgevers hebben allemaal veel aandacht gegeven aan het meten, de prijsbepaling en het beheren van kredietrisico. De kans op wanbetaling, het terugvorderingspercentage en hun afhankelijkheid zijn fundamentele aspecten van kredietrisico. Het nauwkeurig meten van kredietrisico is cruciaal om vier redenen. Ten eerste zijn maten van kredietrisico erg belangrijk voor financiële economen voor prijsbepaling van kredietrisicoportfolio's, kredietderivaten, etc. Het belang van kredietrisico voor de prijsbepaling van financiële contracten is onderstreept door de wereldwijde financiële crisis. Ten tweede kunnen nauwkeurige maten van kredietrisico tijdens het managementproces van kredietrisico voor bedrijven het managementteam helpen om hun risicobereidheid beter te bepalen. Ten derde worden de bekende kapitaaleisen van Basel berekend met behulp van kredietrisicomaten. Ten vierde kan het nauwkeurig schatten van kredietrisico een manager helpen om betere beslissingen te nemen. Bijvoorbeeld zal bij terugvorderingsactiviteiten na wanbetaling meer aandacht worden besteed aan individuen met een hoog geschat LGD om grote verliezen te verminderen.

Tijdens mijn promotietijd heb ik ook deelgenomen aan verschillende conferenties, waaronder de elfde internationale conferentie over de analyse van extreme waarden

(EVA 2019). Bij mijn deelname aan deze conferentie besloot ik om mee te doen aan een van de voorgestelde uitdagingen voor jonge wetenschappers, wat geleid heeft tot het schrijven van een van de bijdragen aan dit proefschrift en wat ook de eerste prijs heeft opgeleverd in de competitie.

NOTATION

CHAPTER 3

Here below we collect all the symbols we use in the paper. By convention, we first order Latin letters alphabetically, and then the Greek ones.

c	A element in C
C	The set of colors
l	A recovery level l which could take a value from 0 to m
t	A nonnegative integer t indicates a time which could be the sojourning time or the jump time of a counterparty
L	The set of recovery levels
F^l	Beta-stacy process for recovery level l
$\tilde{F}^l$	Posterior beta-Stacy process for recovery level l
d_t^l	Strength of belief in the definition of the a priori
G	The random semi-Markov kernel of the process $\{Y\}_p$
$\mathscr{G}$	The space of all semi-Markov kernels
$\tilde{G}$	Posterior semi-Markov kernel of the process $\{Y\}_p$
$\hat{G}$	Conditional semi-Markov kernel for H_{k+1}
i	Index value for counterparties
H_i	The observed recovery history for the i-th defaulted counterparty
$\boldsymbol{H_k}$	The observed recovery histories for the first k defaulted counterparties
$\boldsymbol{H_k^l}$	The observed sojourning time τ_i^l at level l, the next recovery level δ_i^l, and the right censoring indicator ρ_i^l in the first k counterparties
J_n	The n-th jump time of process $\{Y_p\}$
k_l	The number of counterparties visiting recovery level l in the first k counterparties
L_n	The level it jumps to at time J_n of process $\{Y_p\}$
m	A termination level
$m-1$	The full recovery level which corresponds to the recovery rate 100%
M_n	The sojourn time at level L_{n-1} before jumping to level L_n for the process $\{Y_p\}$
$N_s(c)$	The number of balls of color c in urn $U(s)$
q	A law of motion to indicate how the sampling of the different urns drives the process $\{X\}_n$
$Q(p)$	The number of jumps before time p for the process $\{Y_p\}$
r	The reinforcement, that is the additional number of balls, with the color same as the sampled ball, added into the urn in each sample
R	The random transition matrix for the process $\{X_n\}$
R_i	The recovery trajectory for the i-th counterparty

s	A element in S
S	A countable state space in which each element is a couple of the time and the recovery level
S^*	The countable space of the all admissible 0-blocks of $\{X\}_n$
s_0	An initial state for RUP
$s_{(t,l)}(j)$	The number of counterparties sojourning at level l in t and then jumping to level $j \in L$ in the first k counterparties
T	The total recovery time
T_i	The total recovery time for the i-th counterparty
$T^{\max}$	The maximum total recovery time under legal provisions
U	A function of s, describing the original composition of the urn associated to any state s
v_t^l	The number of counterparties whose sojourning time at level l is right censored in t in the first k counterparties
$\{V_t^l\}_{t\geq 0}$	The beta component of a beta-Stacy process for recovery level l
w_t^l	The number of counterparties whose exact sojourning time at level l exceeds time t in the first k counterparties
W^l	A series of Dirichlet distributions for recovery level l
$\tilde{W}^l$	A series of posterior Dirichlet distributions for recovery level l
$\{X_n\}$	The reinforced urn process
$\{Y_p\}$	The recovery reinforced urn process
$\{\alpha_t^l, \beta_t^l\}_{t\geq 0}$	The parameters of a beta-Stacy process for recovery level l
$\{\tilde{\alpha}_t^l, \tilde{\beta}_t^l\}_{t\geq 0}$	The parameters of a posterior beta-Stacy process for recovery level l
δ_n^l	The next recovery level after the n-th visit at level l for the process $\{X_n\}$
η_k^0	The k-th time the process $\{X\}_n$ reaches the point (0,0)
η_n^l	The n-th time the process $\{X_n\}$ touches $(0, l)$
$\{\gamma_t^l\}_{t\geq 0}$	The parameters of a series of Dirichlet distributions for recovery level l
$\{\tilde{\gamma}_t^l\}_{t\geq 0}$	The parameters of a series of posterior Dirichlet distributions for recovery level l
κ	The mixing measure of the process $\{Y_p\}$
ω	The parameters for R
ψ	A function projecting the element in S into its recovery level
ψ	A function projecting the element in S^* into corresponded recovery levels component-wisely
ρ_i	A binary variable indicating right censoring or not for the i-th 0-block
ρ_i^l	A binary variable indicating right censoring or not for the sojourning time at level l at i-th visit
τ_n^l	The sojourning time at recovery level l at n-th visit for the process $\{X_n\}$

CHAPTER 4

RR	the recovery rate

LGD	the loss given default
F	A random distribution function
$j \in \mathbb{N}_0$	j is the element in the set of natural numbers
$\{\alpha_j, \beta_j\}$	parameters of a beta-stacy process
$\{V_j\}$	beta distributed with parameters (α_j, β_j)
$U(j)$	a Polya urn
$n = 0, 1, 2, ...$	the discrete time
$\{Z_n\}_{n\geq 0}$	A two-color reinforced urn process
$B_1, B_2, ..., B_m$	the first m blocks generated by a two-color RUP $\{Z_n\}_{n\geq 0}$
T_i	the last state visited by $\{Z_n\}$ within block i
r_j	the times that $T_1, \ldots, T_m$ equal to j
s_j	the times that $T_1, \ldots, T_m$ bigger than j
c_j	the strength of belief, and it represents how confident we are in our a priori
$G(\{j\})$	the expectation of $F(\{j\})$
X_i	the PD
Y_i	the LGD
$\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$	three independent LS sequences generated by three independent two-color RUPs $\{Z_j^A\}_{j\geq 0}$, $\{Z_j^B\}_{j\geq 0}$, $\{Z_j^C\}_{j\geq 0}$
F_{XY}	a bivariate random distribution
$\mathbf{A}_{-j}$	$(A_1, ..., A_{j-1}, A_{j+1}, ..., A_m)$
$d_1^v, ..., d_9^v$	the deciles for the quantity $v \in \{PD, LGD\}$
σ_A	the standard deviation of the variable A
$Poi(\lambda)$	the possion distribution with the parameter λ

CHAPTER 5

The following notation is used: Italic lower-case letters refer to scalar values (x) or scalar-valued functions ($f(\cdot)$). Lower-case bold letters denote vectors ($\mathbf{x}$) or vector-valued functions ($\mathbf{f}(\cdot)$). Finally, upper-case bold letters represent matrices ($\mathbf{X}$).

(s, t)	space-time validation points
$\hat{A}(\cdot, \cdot)$	the predicted SST anomalies
$X(\cdot, \cdot)$	
x	scalar values
$f(\cdot)$	scalar-valued functions
$\mathbf{x}$	vectors
$\mathbf{f}(\cdot)$	vector-valued functions
$\mathbf{X}$	matrices
(x_0, x_1)	interval from x_0 to x_1
$u(x, y)$, $(x, y) \in \mathscr{S}$	
$\mathscr{S}$	the set of the region representing the red sea
$\mathscr{M}$	the subset of $\mathscr{S}$, representing the masked region
$\mathscr{T}$	the set of all days considered in this paper
$u(x, y; t')$	the temperature data of someday t at the location (x, y)
$S = 16703$	the total number of grid cells

Ω	the mesh grid
$\mathscr{G}$	an undirected graph with its nodes and edges
the Laplacian matrix $\mathbf{L}$	the Laplacian matrix of graph $\mathscr{G}$
$\mathbf{I}$	the identity matrix

1

INTRODUCTION

1

In this PhD thesis we deal with the modeling of risk and uncertainty, with applications in risk management, mainly credit risk and its declinations, and climatology, with a particular attention to the modeling of extreme and missing temperatures. We collect here results that have been published in international peer-reviewed journals [1], and which constitute the main chapters of these pages.

Even if, at a first glance, these topics may appear disconnected, the *fil rouge* is represented by the consistent use of a class of probabilistic tools–as we shall see–and the desire to play with tail risk events (in line with my interest for extreme value theory, EVT), like the default of a company, the recovery of a defaulted bond or mortgage, or the modeling of extreme temperatures over space and time. The presence of missing observations is also something I have dealt with in the different settings.

In particular, as it will be clearer later, the application in climatology is linked to a competition I have taken part to, during the 2019 Extreme Value Analysis Conference (EVA) in Zagreb. The approach I proposed together with Dr. Z. Liu was considered the best and we won the competition, being now an integral part of this work.

1.1. DEBT AND CREDIT RISK

Debt exists: our societies rely on debt [1].

Debt can surely be a burden leading to default, but it can also be an opportunity and generate profits for both the lender and the borrower. The former may easily earn a certain amount of interest, while the latter gets the capital to finance business ventures or to satisfy personal needs.

Naturally intertwined with debt we find credit risk (CR). According to the BCBS [2, 3], credit risk is defined as the risk of changes in the creditworthiness of a counterparty. Such a change can substantiate itself in an increase in the Probability of Default (PD), thus defining default risk; in a change of ratings for the worse, also known as migration risk; or in more specific situations we will investigate in the thesis, like for example recovery risk, when it becomes more difficult to recover a credit from an already defaulted counterparty.

From a historical perspective, credit risk was initially something local. Most debt bearing transactions used to happen among relatives, or at the scale of a village. It was during the Renaissance that debt started becoming a broader phenomenon, and so credit risk [1], because of the increasing number of international commercial travels and explorations. And it was during the mid-19th century that debt really became a global phenomenon, thanks to the large-scale investments needed in railroads and new industries.

Issuing corporate bonds became a common way to raise capital and bond ratings started appearing, as a way to lure potential investors. The first Rating agency, Dun & Bradstreet, was set up in New York City in 1841, and it rated the creditworthiness of several bonds, making such an information available to the public [4]. In a sense, we can consider this as a first step towards proper credit risk management [5].

The credit risk management process usually presents three steps: the specification of the risk appetite of a client, the quantification of credit risk in all its manifestations,

[1] We refer to each single chapter for the detailed bibliographic references.

and the hedging and management of such a risk.

As said above, default risk (DR) and recovery risk (RR) represent two fundamental elements of credit risk.

Following [6], we define default risk is as the risk that a counterparty is unable to fulfill its obligations, declaring bankruptcy and possibly going into liquidation. A key measure of default risk is the probability of default (PD), which is the probability that a counterparty will not meet its obligations in due time.

In [7], recovery risk (RR) is defined as the risk that the recovered amount actually obtained after the liquidation of the insolvent counterparty's assets is smaller than the estimated exposure, but also the the temporal length of the recovery process does not satisfy our expectations in terms of speed. We shall see that recovery risk arises from the randomness of the following four variables: the amount to be recovered, the workout administrative expenses (AC), the discount rate (r), and the duration (T) of the recovery process. Regarding the amount to be recovered, two main quantities are of interest, the Exposure at Default (EAD) and the Loss Given Default (LGD), which we will specify later [3].

Usually a portfolio consists of different assets, belonging to different counterparties. This means that credit risk tends to have a multivariate nature. Moreover the different counterparties may belong to the same industrial sector, the same economy, they could be tied by commercial relations, or be shocked by the same exogenous phenomena. this tells us that modeling the dependence among different exposures is pivotal, and that independence is not a credible assumption [5, 7].

1.2. EXTREME TEMPERATURES

My interest for the modeling of extreme temperatures comes from my desire to deepen my understanding of EVT and tail risk–something that we can find also in the modeling of credit risk–but also from the participation in the challenge proposed by [8] during the 2019 EVA Conference in Zagreb. Challenge I was very happy to win, together with my co-author Dr. Liu.

The extreme temperature events can cause devastating ecological and environmental degradations, such as volcanoes and earthquakes. The EVA 2019 data competition is for predicting the extreme temperature events at unobserved locations and times. Under the evaluation criterion–the threshold-weighted continuous ranked probability score (twCRPS), our Stochastic Partial Differential Equation outperforms other methods on this problem.

Generally speaking, there are two ways to model spatio-temporal extremes in the field of extreme value analysis. One is to define extremes as block maxima with max-stable processes [9–17]. The other is to model high threshold exceedances with asymptotic or sub-asymptotic extreme-value models [18–21]. Despite the fact that the field of SPDEs is one of the most rapidly evolving areas of mathematics, with a wide range of current and potential applications, few works use SPDEs in the field of extreme value analysis. Bayesian hierarchical models can be especially efficient for high-dimensional problems and large datasets. In [22], they developed a Bayesian method for this competition with the SPDE approach. Their method is a two-step approach: first, they remove temporal trends to make the marginal distribution more similar to the standard

Gaussian distribution, especially for tails; second, a latent Gaussian model (LGM)—a special type of hierarchical model—is fitted locally in space and time to the transformed data. Bayesian inference for LGMs can be implemented through the integrated nested Laplace approximation approach (INLA) [23] and its implementation, R-INLA. The INLA methodology aims to approximate the posterior marginals of the model effects and hyperparameters, by exploiting the computational properties of the Gaussian Markov random field (GMRF). The GMRF representation can be constructed explicitly by using a certain SPDE whose solution are Gaussian fields with Mátern covariance function [24]. The logic of the above model is that first we fit the data to a model we assume and then sample the statistics of interest from the posterior distribution. Both the model misspecification and the error arising from the simulation exist. We just sample from the data and then assume a model to get the statistics of interest. At least the error of simulation does not exist in this case.

Motivated by a highly cited paper in computer graphics [25], our method is to borrow N realizations $f_1, \ldots, f_N$ of f from other days (see more detail in Chapter 4). And then, for each realization f_i, we assume that the anamolies follow a Poisson equation with the right hand side f_i. This is a joint work with my collaborator Dr. Zishun Liu from the field of computer graphics. We use the finite difference method to numerically approximate the solution to this equation. All in all, we can see this as a process to solve the Stochastic Poisson equation with the finite difference method (FDM) numerically.

1.3. PLAN OF THE THESIS

This thesis aims to provide an answer to a few research questions that arise in credit risk management and in the modeling of extreme temperatures, from a spatio-temporal point of view. These questions represent the skeleton of our work.

As said, we have tried to provide our answers in a series of papers we have published in peer-reviewed journals, on the basis of which we have built the different chapters, apart from this introduction and the first two methodological chapters, in which we collect some of the tools that are needed to understand the rest of the work.

As it will be evident in reading the present thesis, the approach we have proposed to deal with the different research questions is mainly Bayesian nonparametric in nature, and it relies on the use of a special class of urn processes.

CHAPTER 2: ALL-IN MODELLING OF RECOVERY RISK

Chapter 2 introduces a RUP-based model we have called RRUP, where the first R stands for recovery. This is one of the fewest models dealing with the joint modeling of the recovery rates (or LGD if you prefer to change the perspective) and the recovery time. In fact, most approaches in the literature either consider the recovered amount, or the time duration of the recovery process. This, in our view, may introduce an underestimation of the overall recovery risk. Recall: time is money.

Our approach integrates ideas from standard RUPs [26] with semi Markov chains, to better deal with the problem of righ-censoring, so common in the modeling of recovery risk. Censoring, in fact, may occur because of regulations imposing a maximum time length for the recovery process, or because of a simple write-off decision from the creditor [5].

We assume that the time series of the recovery rate follows a semi-Markov chain, a generalization of the more commonly used Markov chains [27].

For a Markov chain, the future only depends on the current state even if when past information is also available. In a semi-Markov chain, the future depends not only on the current state, but also partly on the time spent in the current state. The distribution of a semi-Markov chain is determined by the initial state and the semi-Markov kernel (similar to the role of transition matrix for a Markov chain) [28].

An advantage of our method is that we can use the right-censored data to predict the semi-Markov kernel. Based on the predicted kernel, a lot of parameters such as LGD and recovery time can then be derived.

Furthermore, exploiting some classical RUP properties, our model can improve its performances over time, learning from data and updating its own parameters. All this while combining empirical evidence with experts' judgements, which can be used to deal with missing observations, known unknowns [29], tail risk underestimation [30], and historical bias among the others [31, 32].

CHAPTER 3: MODELLING "WRONG-WAY RISK"

Chapter 3 explicitly deals with the first research question of this work: how to effectively model the dependence between PD and LGD, which is empirically observed when studying mortgages and other defaultable assets [33]. In particular we focus our attention on wrong-way risk, i.e. when PD and LGD are comonotonic and show positive dependence [5, 6].

Starting from some previous work of [34], we propose a one-factor model built on RUPs, in which the PD and the LGD are affected by a common component plus some idiosyncratic factors. Clearly the dependence arises from the common factor.

Thanks to the Bayesian nature of RUPs, we assign a priori to the different factors, thus being able to use experts' judgements if available.

A Gibbs sampler method is then used to obtain the predictive distribution of the PD and the LGD for a yet-to-default counterparty, given the observed PD and LGD for comparable and already defaulted counterparties, which we assume to be exchangeable.

This is the most applied work in the thesis, and to the best of our knowledge one of the few dealing with a Bayesian nonparametric modeling of wrong-way risk.

CHAPTER 4: MODELLING (EXTREME) MISSING TEMPERATURES

This is the chapter dealing with the problem proposed by [8] during the EVA Conference in Zagreb.

The goal was to predict Red Sea surface temperature extremes over space and time, especially in the presence of missing observations.

To achieve this, we have used a method based on stochastic partial differential equations, and the Poisson equation in particular. We have used a special regularization to penalize certain solutions.

The approach has shown to be successful according to the competition's evaluation criterion, that is a threshold-weighted continuous ranked probability score.

Interestingly, the numerical method we have proposed is computationally efficient in dealing with the problem of dimensionality.

1

CHAPTER 5: CONCLUSIONS AND FUTURE WORK
This last chapter concludes the thesis, presents and discusses some open questions related to our work, and suggests some possible research directions for future work.

1.4. SOME FINAL THOUGHTS BEFORE WE START

When writing a scientific paper to tackle a research question, we first search the literature and try to fill out the gap we find there in. This can hopefully be interpreted as a novelty or innovation in the proposed research.

A quote I like comes from [35]:

> *There are two main types of innovation in research: re-exploring old topics with new evidence, viewpoints or more rigorous methods, to find new discoveries (old bottles of new wine); or using existing methods with moderate modification to explore new research fields, topics, and new discoveries (new bottles of old wine)".*

According to [35], we offer "old bottles of new wine" type.
Let us open one of these bottles and start our discussion with a glass.

REFERENCES

[1] S. Hakemy, *Debt: The First 5000 Years* (Penguin UK, 2017) pp. 1–78.

[2] Basel Committee on Banking Supervision, *Bcbs*, July (2005) p. 6.

[3] Bank for International Settlements, *Basel III: A global regulatory framework for more resilient banks and banking systems,* in *Bank for International Settlements*, Vol. 2010 (2010) p. 69.

[4] A. Miglionico, *The credit rating industry,* The Governance of Credit Rating Agencies **5**, 2 (2019).

[5] J. C. Hull, *Risk management and financial institutions*, 4th ed. (Wiley, New York, 2012).

[6] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools* (Princeton university press, 2005).

[7] A. Resti and A. Sironi, *Risk Management and Shareholders' Value in Banking*, edited by A. Resti and A. Sironi (John Wiley Sons, Inc., Hoboken, NJ, USA, 2012).

[8] R. Huser, *Editorial: EVA 2019 data competition on spatio-temporal prediction of Red Sea surface temperature extremes,* Extremes **24**, 91 (2021), arXiv:1912.00694 .

[9] A. Davison, R. Huser, and E. Thibaud, *Geostatistics of dependent and asymptotically independent extremes,* (2013), 10.1007/s11004-013-9469-y.

[10] S. Castruccio, R. Huser, and M. Genton, *High-order composite likelihood inference for max-stable distributions and processes,* (2014), 10.1080/10618600.2015.1086656.

[11] R. Huser and A. C. Davison, *Space-time modelling of extreme events,* Journal of the Royal Statistical Society. Series B: Statistical Methodology **76**, 439 (2014), arXiv:1201.3245 .

[12] E. Thibaud, J. Aalto, D. Cooley, A. Davison, and J. Heikkinen, *Bayesian inference for the brown-resnick process, with an application to extreme low temperatures,* (2015), 10.1214/16-AOAS980.

[13] C. Dombry, S. Engelke, and M. Oesting, *Bayesian inference for multivariate extreme value distributions,* (2017), 10.1214/17-EJS1367.

[14] S. Buhl and C. Klüppelberg, *Generalised least squares estimation of regularly varying space-time processes based on flexible observation schemes,* (2019), 10.1007/S10687-018-0340-X.

[15] R. Huser, C. Dombry, M. Ribatet, and M. Genton, *Full likelihood inference for max-stable data,* (2019), 10.1002/STA4.218.

[16] S. Vettori, R. Huser, and M. G. Genton, *Bayesian modeling of air pollution extremes using nested multivariate max-stable processes,* Biometrics **75**, 831 (2019), arXiv:1804.04588 .

[17] S. Vettori, R. Huser, J. Segers, and M. Genton, *Bayesian model averaging over tree-based dependence structures for multivariate extremes,* (2019), 10.1080/10618600.2019.1647847.

[18] R. de Fondeville and A. Davison, *High-dimensional peaks-over-threshold inference,* (2016), 10.1093/BIOMET/ASY026.

[19] A. Kiriliouk, *Modelling extreme-value dependence in high dimensions using threshold exceedances,* (2016).

[20] S. A. Morris, B. Reich, E. Thibaud, and D. Cooley, *A space-time skew-t model for threshold exceedances.* (2017), 10.1111/biom.12644.

[21] R. Yadav, R. Huser, and T. Opitz, *Spatial hierarchical modeling of threshold exceedances using rate mixtures.* (2019), 10.1002/env.2662.

[22] D. Castro-Camilo, L. Mhalla, and T. Opitz, *Bayesian space-time gap filling for inference on extreme hot-spots: an application to red sea surface temperatures,* Extremes **24**, 105 (2020).

[23] H. Rue, S. Martino, and N. Chopin, *Approximate bayesian inference for latent gaussian models by using integrated nested laplace approximations,* (2009), 10.1111/J.1467-9868.2008.00700.X.

[24] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach,* Journal of the Royal Statistical Society. Series B: Statistical Methodology **73**, 423 (2011).

[25] P. Pérez, M. Gangnet, and A. Blake, *Poisson image editing,* ACM Transactions on Graphics **22**, 313 (2003).

[26] P. Muliere, P. Secchi, and S. G. Walker, *Urn schemes and reinforced random walks,* Stochastic Processes and their Applications **88**, 59 (2000).

[27] D. Cheng and P. Cirillo, *A reinforced urn process modeling of recovery rates and recovery times,* Journal of Banking and Finance **96**, 1 (2018).

[28] S. M. Ross, J. J. Kelly, R. J. Sullivan, W. J. Perry, D. Mercer, R. M. Davis, T. D. Washburn, E. V. Sager, J. B. Boyce, and V. L. Bristow, *Stochastic processes*, Vol. 2 (Wiley New York, 1996).

[29] H. Dickson and G. L. S. Shackle, *Ekonomisk Tidskrift*, Vol. 58 (Cambridge University Press, Cambridge, 1956) p. 250.

[30] E. Lybeck, *The Black Swan: The Impact of the Highly Improbable* (Random House, 2017) pp. 1–82.

[31] J. Derbyshire, *The siren call of probability: Dangers associated with using probability for consideration of the future,* Futures **88**, 43 (2017).

[32] F. H. Knight, *Risk, uncertainty and profit (reprint of the 1921 edition)* (Augustus M. Kelley, Bookseller).

[33] E. I. Altman, B. Brady, A. Resti, and A. Sironi, *The link between default and recovery rates: Theory, empirical evidence, and implications,* Journal of Business **78**, 2203 (2005).

[34] P. Bulla, P. Muliere, and S. Walker, *Bayesian nonparametric estimation of a bivariate survival function,* Statistica Sinica **17**, 427 (2007).

[35] M. Peng, *The complete survival handbook for graduate students: Methods, secrets, unspoken rules (in Traditional Chinese),* 1st ed. (Linking Publishing Company, 2017).

2

ALL-IN MODELLING OF RECOVERY RISK

Answering a major demand in modern credit risk management, we propose a nonparametric survival approach for the modeling of the recovery rate and the recovery time of a defaulted counterparty, by introducing what we call the Recovery Reinforced Urn Process, a special type of combinatorial stochastic process.

The new model allows for the elicitation and exploitation of prior knowledge and experts' judgements, and for the constant update of this information over time, as soon as new data become available. We show how to use it to perform Bayesian nonparametric prediction about the recovered amounts and the (total) recovery time of a series of defaulted exposures.

An application to real data is provided using the Single Family Loan-Level Dataset by Freddie Mac.

Keywords: *Loss-Given-Default, Recovery Rate, Reinforced Urn Process (RUP), Survival Analysis, Machine Learning, Bayesian Methods.*

Parts of this chapter have been published in D. Cheng and P. Cirillo, A Reinforced Urn Process Modeling of Recovery Rates and Recovery Times, Journal of Banking Finance 96, 1 (2018) [1].

2.1. RECOVERY MODELING: INTRODUCTION AND MAIN ISSUES

In the wake of the latest financial regulations for the banking sector [2–4], we provide an answer to the increasing demand of models for the recovery process of defaulted exposures. Differently from most contributions in the literature that mainly focus on the recovered amounts, here we propose a way to also study the duration of the recovery process, a common "known unknown" [5] in risk management, with the final goal of predicting the possible recovery trajectory of a counterparty, not only on the basis of the available data, but also with the possibility of using experts' judgements and other a priori knowledge to mitigate historical bias [6, 7], at least partially [8, 9]. The model we propose is able to learn, improving its performances over time, using the mechanism of Bayesian update, or machine learning in computer science language.

When a counterparty defaults, the corresponding loss is not necessarily equal to the Exposure-at-Default (EAD), that is the nominal value of the exposure at the time of default. In fact, thanks to the recovery process, i.e. the set of all the procedures that can be put in place to collect the amount due, one may be able to recover at least a part of the outstanding exposure [10]. But these procedures are costly and may require a substantial amount of time, a very important variable for correct risk management.

The recovery rate (RR) is the amount of principal and accrued interest on a defaulted exposure that can be recovered, expressed as a percentage of its face value, once again the EAD. In what follows, to ease notation, we will consider $RR \in [0\%, 100\%]$, even if the case $RR > 100\%$ is actually possible, thanks to fees and interests, as frequent among leasing contracts [11, 12]. The recovery rate is naturally linked to the Loss Given Default, or LGD, that is the percent loss experienced when a counterparty defaults and no further recovery is possible. For $RR \in [0\%, 100\%]$, we have $LGD = 100 - RR$. When $RR \geq 100\%$, we set $LGD = 0\%$.

In the last decade, because of the Basel Accords [2, 13, 14], and the new International Financial Reporting Standards [3], recovery modeling has become a major concern for banks and regulators. In particular, the new IFRS 9 regulations–just entered in full force–define LGD assessment and back-testing as a major task for banks and financial institutions [15]. Recovery risk, defined as the risk associated with the recovery process, in terms of time length and quantification of the actual loss, is officially one of the new challenges in credit risk management [15–17].

The recovery process is not at all a simple object [18], given that the entire process is influenced by a series of important factors, which affect its success and duration. As [19] point out, an effective recovery depends on the characteristics of the exposure (presence of collateral, degree of effectiveness), of the counterparty (e.g. industry, country and legal framework), on macroeconomic factors like the state of the economy, and on internal factors like the efficiency of a bank in recovering its money, for instance in dealing with out-of-Court settlements.

The complexity of the recovery process is evident if we take into account its possible outcomes, and the fact that its actual duration, the recovery time T, is not known until the very end. We can clearly distinguish three main scenarios: 1) the past due receivable is fully collected, so that, at some random time T, the recovery process terminates with $RR = 100\%$; 2) the uncollected receivable is fully or partially written-down or written-off, because it has no residual market value nor it can be monetized, so that at time T

we have $RR < 100\%$; 3) the recovery time exceeds a given maximum time $T^{\max}$, legally imposed, therefore $RR < 100\%$ and $T = T^{\max}$. From a statistical point of view, this last case involves censored data [20], something that needs to be taken into consideration [19].

The quantification of the recovery rate is a further problem, because also the recovered amount is precisely known only at the end of the recovery process. And while it is true that for the bond issues of large corporations one can typically rely on the so-called market recoveries (computed as the ratio of the price at which the defaulted bond is traded some days after default, provided the market is not illiquid, to the price of that asset at the time of default), it is also true that for smaller counterparties, or for other products like loans, the information about prices is either not available nor reliable. [21] have shown that post-default prices tend not to be rational, but rather biased and inefficient. Even when prices are available, the discounted value of cash flows is the preferred measures by both analysts [22] and regulators [2, 3].

The paper is organized as follows. Section 2.2 provides a brief overview of the state-of-the-art in recovery modeling. Section 2.3 gives an intuitive representation of the recovery process, which is propaedeutical to the the model we introduce in Section 2.4. In Section 2.5 we provide an extensive application of the model to real data, showing how to deal with the recovery trajectories of defaulted loans in the Single Family Loan-Level Dataset by [23]. All the mathematical contents related to the new model, from theorems to proofs and simulation details, are collected in the Appendix.

2.2. THE STATE-OF-THE-ART OF RECOVERY MODELING

The literature on recovery modeling is not as extensive as the one related to the probability of default (PD), but especially for the recovery rates it is nevertheless rich and varied, from the pioneering works on hazard functions for loss data [24], to the very recent contributions about the methodological implications of the new regulations on the estimation of LGD [15].

For good reviews, we refer to [18, 25, 26] and [27]. In particular, the first paper also contains details about some less used methods like those based on utility theory.

Statistically, the modeling of LGD includes parametric and nonparametric approaches [28]. In the first class, we find the different flavors of generalized linear models, from the convenient beta regression of [29] to the logit, probit and tobit regressions described in [30], or in [31], where interesting comparisons are discussed. In a very recent paper, [32] deal with quantile regression, and they are able to deal with both bulk and tail events, that is with the entire LGD spectrum. Regarding nonparametric methods, NP regression and model trees [28] are common tools, together with mixtures [33] and beta kernels [34].

Recently, a series of survival analysis models have also been proposed, e.g. [35], according to a trend in PD modeling, but also following the new regulations [2, 3] asking for both point-in-time (PIT) and through-the-cycle (TTC) estimates [15, 36]. We can cite [37] and [38] as examples of pure survival approaches. In particular, in the latter, different models are considered, including interesting semi-parametric and pseudo-survival constructions. For a comparison between regression and survival methods, we refer to [12].

Most of the contributions in the literature, a notable exception being [39], do not present models that are able to learn from data, updating their parameters, and improving their performances over time, whenever new pieces of information become available. This is actually our ambition: using a specially conceived reinforced urn process [40], we propose a nonparametric survival approach that allows for the elicitation and exploitation of prior information, and its automatic updating and correction over time using data. A model that can be seen as a mixture of combinatorial stochastic processes and machine learning, but that, differently from traditional machine learning [41], gives us full control of its probabilistic features.

2.3. VISUALIZING RECOVERY

Without loss of generality, we can discretize the recovery rate of a counterparty in terms of a scale of recovery levels from 0 to m. Level 0 corresponds to no recovery ($RR = 0\%$), levels 1 to $m-2$ represent intermediate stages of recovery ($0\% < RR < 100\%$), while level $m-1$ is full recovery ($RR = 100\%$). Level m is a special stage, not associated to any specific recovery rate, but representing the termination of the recovery process. We can read it as the situation in which the recovery process has reached full recovery or a write-off, and we need some little extra time for closing the bureaucratic procedures. As we shall see, the termination level m is a little artifice needed to fully develop the model, guaranteeing the property of recurrence (Appendix: Lemma 1).

Let us consider a simple example. Set $m = 4$ and define 4 possible levels for the recovery rate:

$$0 : 0\%,\ 1 : (0\%, 50\%],\ 2 : (50\%, 100\%),\ 3 : 100\%, \tag{2.1}$$

plus the extra level 4 representing the termination of the recovery process.

The larger m, the finer the partition for the discretized recovery rate. Notice that it is not required that each recovery level guarantees the same additional recovery. In other words, in the previous example, we could define a scale in which level 0 corresponds to 0%, level 1 to (0%,25 %], level 2 to (25%, 100%), and level 3 to 100%, according to our needs. The only constraints are that level 0 is equal to 0%, no recovery, level $m-1$ corresponds to 100%, full recovery, and level m is the termination level.

Following intuition, we can take the recovery level to be a non-decreasing quantity: once we have recovered 20% of the outstanding exposure, we can further increase our recovery level or not, but we cannot go back to 10%.

We set time t to be a nonnegative integer ($t \in \mathbb{N}_0$), representing a particular time unit, like days or months. For the recovery process, discrete time is not a limitation: according to [18], it is actually more realistic than a continuous time framework.

With the couple (t, l), we can therefore indicate that a given counterparty at time t is at recovery level $l \in \{0, 1, ..., m\}$.

By definition, the recovery process of a counterparty starts in $(0,0)$, at default, where the recovery clock is yet to start and the recovered amount is null. From (0,0), time starts running and we can spend several time units at a given recovery level, before being able to reach a higher one.

Let us consider a counterparty A, whose recovery trajectory R_A looks like

$$R_A = \{(0,0), (1,0), (2,0), (0,1), (1,1), (0,3), (1,3), (0,4), (1,4)\}. \tag{2.2}$$

This means that, after default in $(0,0)$, counterparty A spends 2 time units in level 0, where no amount is recovered, visiting $(1,0)$ and $(2,0)$. It then jumps to recovery level 1, where it stays 1 time unit, recovering up to 50% of its EAD in the scale of Equation (2.1). Finally, for one time unit, A reaches level 3, full recovery, before jumping to termination level $(0,4)$, which closes the recovery process. For the termination level, in what follows, we always assume a fixed fictitious permanence of 1 time unit (more later). The total recovery time for counterparty A is therefore $T_A = 2+1+1 = 4$. The time spent in level 4, the termination level, is not taken into consideration for the computation of the total recovery time.

2.4. RECOVERY MODELING AND REINFORCED URN PROCESSES

Consider a portfolio of k defaulted exposures, which we assume exchangeable[1]. Exchangeability is a relaxation of the stronger assumption of independence: the order in which we observe our counterparties is for us irrelevant, and the joint distribution of their recovery times and levels is immune to permutations in the order of appearance. Exchangeability is a common assumption in credit risk modeling, for instance in the class of mixture models [43], where it probabilistically represents the idea that we can split a group of counterparties into subgroups that are homogeneous in terms of risk, that is exchangeable within.

In this paper, the exchangeability of counterparties consists in the exchangeability of their recovery trajectories: it is not important if the recovery process of counterparty i is observed before or after that of counterparty j, because the joint distribution of their recoveries is unaffected, and so our estimates. We assume that the recovery trajectory R_i of counterparty i is representable as in the example of Equation (2.2), for $i = 1, ..., k$.

2.4.1. THE GENERAL RUP CONSTRUCTION

To develop our model we make use of Reinforced Urn Processes (RUPs) that are a class of combinatorial stochastic processes introduced by [40].

A RUP $\{X_n\}_{n\geq 0}$ is characterized by the following elements:

- A countable state space S of all the possible states the process $\{X_n\}$ can visit with positive probability.
- Every element $s \in S$ is associated with a Pólya urn, i.e. an urn characterized by sampling with reinforcement [44], containing colors belonging to a set C, with cardinality $\#C > 0$.
- For each urn we define a function $U\colon s \in S \to \mathbb{R}^{\#C}$ describing the composition of the urn itself. In other words, for every $s \in S$, urn $U(s)$ contains $N_s(c) \geq 0$ balls of color $c \in C$. To avoid degenerate cases, we assume $\sum_{c\in C} N_s(c) > 0$, so that every urn contains a positive amount of at least one color[2].

[1] Given two random variables X_1 and X_2, they are exchangeable whenever $P(X_1 \leq x_1, X_2 \leq x_2) = P(X_1 \leq x_2, X_2 \leq x_1)$, that is when their joint distribution is immune to permutations in the order of the variables. The concept is easily extended to higher dimensions [42].

[2] Notice that $N_s(c)$ is a real number, so that we can have 1.3 balls of color c. A real number of balls can be

- A law of motion $q : S \times C \to S$ indicating how the sampling of the different urns drives the process $\{X_n\}$. Given a color $c \in C$ and two states $s, w \in S$, so that w can be visited from s with positive probability after sampling c from $U(s)$, the function q is such that $w = q(s,c)$. Clearly the definition of q allows for the construction of very different processes, all falling within the general RUP framework of [40] to which we refer for details.

Without any loss of generality, fix an initial state s_0. A RUP $\{X_n\}$ on S with initial state s_0 is defined recursively as follows: set $X_0 = s_0$, and for all $n \geq 1$, if $X_{n-1} = s_{n-1} \in S$, sample a ball from the urn $U(s_{n-1})$ associated with s_{n-1}. Now, register the color of the ball, say $c \in C$, put it back in $U(s_{n-1})$, and Pólya-reinforce the urn with $r > 0$ balls of the same color. This increases the probability of picking again that color in the future, if the urn is sampled again: the higher r the stronger the update. Finally, using the rule of motion q, set $X_n = q(s_{n-1}, c)$. The sequence $\{X_n\}$ is a reinforced urn process with initial state s_0 and reinforcement r.

A RUP is just a reinforced random walk on a state space of urns. It is a Bayesian nonparametric model, in which the initial composition of the urns defines the a priori (the way in which the a priori is elicited is given in Appendix), which we can update over time thanks to the sampling of the urns, using reinforcement and other technical conditions we shall discuss later. This possibility of embedding prior knowledge and to learn from data is one of the points of strength of RUPs that, in the last years, have been used to develop some interesting models in finance [45, 46], biostatistics [47] and other fields.

2.4.2. THE RECOVERY RUP (R-RUP)

A generic RUP can be adapted to model recovery rates and recovery times, we call this special process the Recovery RUP or R-RUP.

We better specify the space S by setting $S = \mathbb{N}_0 \times L$, so that it contains all the couples (t, l) of recovery times t and recovery levels l the process $\{X_n\}$ can visit. As per Section 2.3, levels 0 to $m-1$ are a discretization of the recovery rates from 0% to 100%, while level m represents the termination level.

The set of colors is now $C = \{c_0, c_1, ..., c_m\}$, where each color from c_0 to c_m corresponds to a given recovery level, as if we color them.

For the fundamental rule of motion q, we define three behaviors:

A. $q((t,l), c_i) = (t+1, l)$, for $i \leq l \leq m$. In words: if, from the urn centered in (t,l), we extract a ball whose color corresponds to level l or lower, the process moves to the next time unit $t+1$, but stays at the same recovery level l. From (t,l) we move to $(t+1, l)$.
The rule can be further simplified if we assume that urns at level l only contain balls of colors $(c_l, c_{l+1}, ..., c_m)$, so that we can set $q((t,l), c_l) = (t+1, l)$. This is actually the version we adopt from now on.

represented using balls of different radius: the bigger the ball, the easier to sample it. Naturally we can choose $N_s(c)$ to be an integer: this will not affect the model, but it can help to have a simpler intuition of the sampling scheme.

B. $q((t,l),c_i) = (0,i)$, for $l < i \leq m$. The process jumps to level $i > l$ while time is reset to 0. Resetting time at every jump helps in counting the time units the R-RUP spends in each recovery level it visits.
When a c_m ball is extracted from the urn in state (t,l), the recovery process jumps to the termination level. Clearly, if c_m is extracted in level $m-1$, the recovery process has reached full recovery, otherwise, for $l = 0,...,m-2$, we are experiencing a write-off.

C. $q((1,m),c_m) = (0,0)$. The process $\{X_n\}$ can stay in level m only for one time unit thanks to rule A[3], after which it reaches $(1,m)$ and it is reset to $(0,0)$. This last movement allows us, using Lemma 1 in the Appendix, to define a recurrent R-RUP.

To avoid degenerate situations and to facilitate the application to real data, we can impose extra restrictions on the urn compositions. From one side, we can set $N_{(0,l)}(c_i) = 0$ for $i = l+1,...,m$, and all $l \in L$, guaranteeing that we cannot visit two recovery levels at the same (discrete) time, because when the process touches level l in $(0,l)$, it must spend at least one time unit at this level. From the other side, for all $t \geq 1$, we can impose $N_{(t,m-1)}(c_i) = 0$ for $i < m$, and $N_{(t,m-1)}(c_m) > 0$, so that the process necessarily moves to the termination level after visiting full recovery for one time unit.

In the R-RUP construction, the recovery process of counterparty i, with all its intermediate stages, can be represented as a sequence of points (t,l) in the space S. For a first counterparty we might for example observe

$$\begin{aligned} R_1 = \quad & \left[X_0, X_1, \ldots, X_{n_1}\right] = \\ & [(0,0),(1,0),\ldots,(t_0,0),(0,l_1),(1,l_1),\ldots,(t_1,l_1), \\ & \ldots,(0,l_{\max}),\ldots,(t_{\max},l_{\max}),(0,m),(1,m)], \end{aligned}$$

where $l_1, l_2, \ldots, l_{\max}$ are the levels visited by the counterparty ($l_{\max} < m$ being the maximum recovery level reached), and $t_1, t_2, \ldots, t_{\max}$ are the corresponding sojourn times, whose sum represents the total recovery time T_1. The quantities $l_{\max}$ and T_1 fully summarize the recovery trajectory (and risk) of exposure 1.

Similarly to R_1, the recovery process of each counterparty i is thus represented as a block R_i of visited states starting with $(0,0)$. If we consider k exposures, we will have k blocks. Every time we observe $(0,0)$ in the sequence generated by $\{X_n\}$, it means that we are looking at a new counterparty. Since we assume the recovery trajectories of the exposures to be exchangeable, the blocks R_i are exchangeable. The R-RUP is therefore able to model any number of counterparties if, every time a recovery process is over, we reset $\{X_n\}$ to (0,0), so that we can start a new block. In probabilistic terms, $\{X_n\}$ needs to be a recurrent, as we discuss in the Appendix.

In general, after n time steps, for $\{X_n\}$ we have

$$\{X_0, X_1, X_2, ..., X_n\} = \{R_1, R_2, ..., R_k\},$$

so that the first n states visited by $\{X_n\}$ can be collected into k 0-blocks, where the number k depends on the number of $(0,0)$ in the sequence of states generated by $\{X_n\}$. A group of k defaulted counterparties will thus be represented by k 0-blocks: $R_1, ..., R_k$.

[3] We are using the simplified version of rule A, otherwise it is sufficient to set $q((1,m),c_i) = (0,0)$ for every i.

Naturally the last recovery trajectory R_k could be incomplete at step n, depending on n and k.

2.4.3. A CLARIFYING EXAMPLE

The maximum recovery level is $m = 4$, so that each urn in S is characterized by a set of 5 colors $C = (c_0, c_1, c_2, c_3, c_4)$. According to Subsection 2.4.2, every urn at recovery level $l \in \{0,1,2,3,4\}$ only contains balls of color $(c_l, ..., c_4)$.

We start with the default of exposure 1, which begins its recovery process in $X_0 = (0,0)$. We sample urn $U((0,0))$ and extract a c_0 ball, so that we move to $X_1 = q((0,0), c_0) = (1,0)$, that is exposure 1 stays at the recovery level 0 for one time unit. Then, we sample urn $U((1,0))$ and we extract c_2, therefore we jump to $X_2 = (0,2)$, i.e. after staying at recovery level 0 for 1 time unit, the counterparty jumps immediately to level 2, without touching level 1. In urn $U((0,2))$, we sample again a c_2 ball, we set $X_3 = (1,2)$, and we stay at recovery level 2 another time unit. Finally, we extract a c_3 ball from urn $U((1,2))$ and the exposure reaches the full recovery level 3. The recovery process goes back to state $(0,0)$ after we stay one more time in the full recovery level 3, and in the termination level 4 for "bureaucratic" reasons. Summarizing, for exposure 1 we observe:

$$R_1 = \{X_0, ..., X_7\} = \{(0,0), (1,0), (0,2), (1,2), (0,3), (1,3), (0,4), (1,4)\}. \tag{2.3}$$

The total recovery time for exposure 1 is $T_1 = 1 + 1 + 1 = 3$, given that the one-period permanence in level 4 is not counted. The recovery trajectory is visible in Figure 2.1(a).

All the urns in S are Pólya, therefore every ball sampled from an urn is reinforced with r extra balls of the same color, thus changing the composition of that urn. Hence, after R_1 has been observed, the probability that a new counterparty follows the same trajectory increases. The R-RUP thus remembers and learns from what happens.

Let us now continue our sampling from urn $U((0,0))$, where the process $\{X_n\}$ has landed after the resetting due to the rule of motion q applied in $(1,4) \in R_1$. Imagine that we obtain two further recovery trajectories for exposures 2 and 3, i.e.

$$R_2 = \{X_8, ..., X_{17}\} = \{(0,0), (1,0), (2,0), (3,0), (0,1), (1,1), (0,2), (1,2), (0,4), (1,4)\},$$

and

$$\begin{aligned} R_3 = \{X_{18}, ..., X_{33}\} = \quad & \{(0,0), (1,0), (2,0), (3,0), (4,0), (0,1), (1,1), (0,2), (1,2), \\ & (2,2), (3,2), (4,2), (5,2), (6,2), (0,4), (1,4)\}. \end{aligned}$$

Notice that for exposure 2 (and also for 3), because of the recovery trajectory of exposure 1 and the Pólya-reinforcement mechanism, the probability of picking a c_2 ball in (1,0) is higher than what originally experienced by exposure 1 (there are r extra c_2 balls now), and this is true for all the already visited states.

The recovery processes for exposures 1, 2, and 3 can thus be represented as follows:

$$\begin{aligned} &\{R_1, R_2, R_3\} = \{X_0, X_1, ..., X_{33}\} = \\ &= \left\{ \overbrace{(0,0), ..., (1,3), (0,4), (1,4)}^{exposure\,1}, \overbrace{(0,0), ..., (1,2), (0,4), (1,4)}^{exposure\,2}, \overbrace{(0,0), ..., (6,2), (0,4), (1,4)}^{exposure\,3} \right\}. \end{aligned}$$

Using reinforcement, the R-RUP modifies its transition probabilities at every cycle. In the R-RUP, the initial compositions of the urns represent our a priori, which we modify through sampling, in order to get what looks like a Bayesian posterior, as clear from Theorem 2 in the Appendix.

2.4.4. THE INTRODUCTION OF CENSORING

To make the model more realistic, assume there are legal provisions on the market such that there exists a maximum recovery time $T^{\max}$, above which all recovery is exogenously stopped, notwithstanding the possibility to further recover later on. From a statistical point of view, the existence of $T^{\max}$ introduces the problem of right-censoring [20]. In looking at real data, it means that we have no idea about what happens after $T^{\max}$: an exposure could have reached full recovery or not, but that remains unknown to us.

As an example, set $T^{\max} = 9$, so that the recovery process of a given counterparty cannot last more than 9 time units. Figure 2.1 gives a graphical representation of the recovery processes of exposures 1, 2 and 3 of Section 2.4.3 in the case of censoring. For counterparty 1 and 2, the recovery process ends at time $T_1 = 1+1+1 = 3 < 9$ and $T_2 = 3+1+1 = 5 < 9$ respectively, well before the maximum time limit, therefore no censoring takes place. For exposure 3, on the contrary, $T_3 = 4+1+6 = 11 > 9$, hence its recovery is forcedly interrupted and right-censored at time $T^{\max} = 9$. The recovery trajectory for counterparty 3 in case of censoring is thus given by

$$R_3^{cens} = \{X_{18}, ..., X_{29}\} = \begin{array}{l}\{(0,0),(1,0),(2,0),(3,0),(4,0),(0,1),(1,1),(0,2),(1,2),\\ (2,2),(3,2),(4+,2)\},\end{array}$$

where $(4+,2)$ indicates that the time spent in level 2 is censored in 4, given that the global $T^{\max}$ has been reached. Again, in case of censoring, we do not know what happens to the recovery trajectory of exposure 3 after $T^{\max}$.

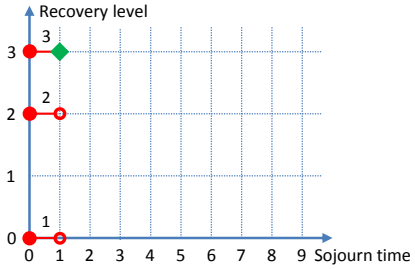
For every counterparty i, we will make use of a dummy variable ρ_i to indicate whether its recovery trajectory is censored (1) or not (0). More in the Appendix.

2.4.5. MAIN PROPERTIES

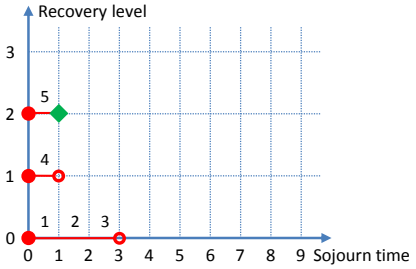
The R-RUP is characterized by several probabilistic properties, which make it powerful and flexible, and able to update its a priori knowledge with the information coming from actual recovery trajectories, even when these are censored, as very common in real business life [19].

When recurrent, the R-RUP may be represented as a mixture of semi-Markov chains, and this mixture is characterizable as a new bivariate random distribution given by a particular interaction of Dirichlet and beta-Stacy processes (Theorem 1). The knowledge of this random distribution allows for the Bayesian prediction of the possible recovery trajectories of new counterparties (Theorem 2 and Corollary 1).

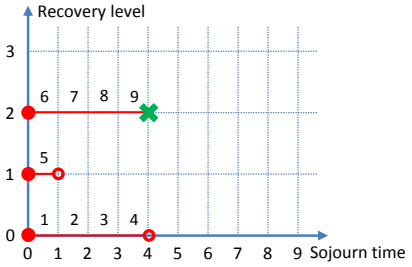
All the theorems, the proofs and the mathematical details related to the R-RUP are collected in the Appendix at the end of the paper.



(a) Exposure 1



(b) Exposure 2



(c) Exposure 3

Figure 2.1: The graphical representation of the recovery trajectories of exposures 1, 2, and 3 with maximum recovery time $T^{\max} = 9$. For 1 and 2, the recovery process ends at times 3 and 5 (green diamonds, no censoring), and the passing of time can be easily read on the recovery trajectories. Notice that counterparty 2 never reaches the maximum recovery level $m = 3$, but it stops in $(1,2)$ with a write-off. For 3, the recovery process is right-censored at time $T^{\max} = 9$ (green cross, censoring).

2.5. APPLICATION: RECOVERY MODELING OF FAMILY LOANS WITH US DATA

We show how the R-RUP can be used to jointly model recovery times and recovery levels, using US data from [23]. The results we get are encouraging, as they show the ability of the R-RUP to learn from data and to reach interesting out-of-sample performances, also being able to correct possible mistakes in the elicitation of our prior beliefs.

2.5.1. DATA AND MODEL INITIALIZATION

We use a subset of the larger Single Family Loan-Level Dataset of [23], freely available online[4], covering around 23 million fixed-rate mortgages originated between January 1, 1999 and March 31, 2016. For each loan several covariates are available, the most relevant ones being the loan size, the loan-to-value, the FICO score (as indicator of the borrower's creditworthiness) and several measures of credit performance. For each single exposure the monthly EAD is registered together with all the repayments and, in case

[4] http://www.freddiemac.com/news/finance/sf_loanlevel_dataset.html

of a default, the information about the recovered amounts, the losses, the recovery procedure and the possible REO (real estate owned) dispositions is available.

Using Freddie Mac classification, a loan experiencing more than 180 days of delinquency is considered defaulted and the credit event is registered. From that very moment the recovery process starts, therefore delinquency day 180 corresponds to state $(0,0)$ in the R-RUP construction. It is important to stress that we do not take into consideration those loans that have been liquidated prior to a delinquency of 180 days because of a short sale, a repurchase or a foreclosure. This choice is due to the fact that we need a clear definition of credit event to correctly define the initial state of our reinforced urn process. Consistently with the literature [18], we also exclude all loans that, despite being formally defaulted after 180 days, have cured after the credit event and then prepaid or repurchased.

To train our model, that is to update the initial urn compositions (more in Section 2.5.2 and the Appendix), we start with a set of 20113 defaulted loans originated in the first quarter of 2006. These loans correspond to roughly 10% of all the loans originated in the months of January, February and March 2006. Given our definition of credit event, these loans start defaulting in the third quarter of 2006, but many of them later. Almost all loans in the Freddie Mac Dataset are not right-censored: a complete recovery process is indeed observed, probably thanks to the long time window over which the loans are followed. As a consequence, in what follows, $\rho_i = 0$ for $i = 1,...,20113$, where ρ_i is the dummy variable indicating censoring.

In our training set almost 93% of the defaulted loans have experienced a write-off, and only 7% a full recovery.

The performances of the model have then been tested in-sample and, most importantly, out-of-sample, making inference and prediction about the recovery times and the recovery levels of loans originated in subsequent periods, like the 20038 defaulted loans generated in the second quarter of 2006.

The 20113 loans of the training sample and the 20038 of the validation sample have been divided into different groups, to better model their recovery processes. In particular, we have considered 4 classes in terms of FICO score: ≤ 650, $(650,685]$, $(685,725]$, > 725; and 5 classes in terms of size of the exposure: ≤ 100, $(100,150]$, $(150,200]$, $(200,250]$ and > 250 in thousand dollars. Table 2.1 gives the number of observations in each class in the training and in the validation sets according to the two classifications.

Splitting the data into groups guarantees that the assumption of exchangeability can be reasonably made for the counterparties in the different classes. Within each group, we thus assume that all the defaulted exposures are homogeneous and exchangeable in terms of risk [43]. While it is plausible to assume that counterparties with similar FICO scores (or loan sizes) are exchangeable, it is not at all safe to assume that all loans together are. Each class will be modeled with a different R-RUP.

For each defaulted loan in the Freddie Mac Dataset the monthly recovered amounts are collected. By dividing these numbers by the corresponding Exposure-at-Default, we easily obtain the monthly recovery rates. In order to be used in our model, these recovery

rates are discretized into recovery levels. We have defined 13 levels:

$$L = \left\{ \begin{array}{l} 0:0\%,\ 1:(0\%,10\%),\ 2:[10\%,20\%),\ 3:[20\%,30\%),\ 4:[30\%,40\%), \\ 5:[40\%,50\%),\ 6:[50\%,60\%),\ 7:[60\%,70\%),\ 8:[70\%,80\%), \\ 9:[80\%,90\%),\ 10:[90\%,100\%),\ 11:100+\%,\ 12:\text{termination} \end{array} \right\}. \tag{2.4}$$

In level 11, corresponding to full recovery, we have also included the few cases (in the order of tens) of recovery rates above 100%, while level 0 includes the few (again in the order of tens) situations in which the actual recovery rate is slightly negative, because the defaulted exposure has generated fees that the debtor has not paid. For more details about the technicalities of these events we refer to [23].

Given the range of variation of the recovery times in the data, we have chosen $t = 0,1,...,100$, where each time step represents 1 month. The R-RUP is thus representable as a 13×101 matrix of urns, with levels of recovery on the rows, and times on the columns.

FICO score	≤ 650	(650,685]	(685,725]	> 725	
Training	4691	4936	5057	5509	
Validation	4851	4747	4999	5441	
Loan size	≤ 100	(100,150]	(150,200]	(200,250]	> 250
Training	3346	4386	4333	3198	4850
Validation	3317	4608	4407	3111	4595

Table 2.1: Repartitions of the training set (in-sample: 20113 loans originated in the first quarter of 2006) and of the validation set (out-of-sample: 20038 loans originated in the second quarter of 2006) using the FICO scores, and the size of the loans (in thousand dollars).

2.5.2. PRIOR ELICITATION AND UPDATE

A fundamental aspect of the R-RUP initialization is the definition of the starting urn compositions, which represent our prior beliefs about recovery times and recovery levels. To embed our a priori in the R-RUP, we need to intervene on the number of balls in each urn. At the end of the Appendix, we give all the details about the process of prior elicitation using the properties of the bivariate random distribution that characterizes the R-RUP. Here it is sufficient to say that we have used three different prior sets to test the model. In all sets, the priors on the risk levels are discrete uniforms, with the empirically-based assumption that the higher the recovery level an exposure reaches, the higher the chance of further recovery [48]. Regarding the recovery times, conversely, we use the empirical cumulative distribution function in Prior Set 1, a discrete uniform with empirically determined support in Prior Set 2, and a uniform on [0,100] in Prior Set 3.

As in all Bayesian models, the prior elicitation step is very important in the R-RUP, in that it influences the speed at which the model converges to the true posterior distribution, and thus all estimates and predictions. But there are good news: even a totally wrong prior can be corrected with a sufficient number of observations, provided it is not degenerate, in accordance to Cromwell's rule [49].

The urn compositions and our beliefs are then updated using the 20113 recovery trajectories of the training sample. This is done by translating the recovery process of each observed counterparty into a sequence of samplings from the urns of the R-RUP,

according to the rule of motion we have described in Section 2.4. For instance, when in the data we observe that counterparty i moves from state (t, l) to state $(t+1, l)$, we read this as an extraction of a c_l ball from the urn in (t, l). Therefore the composition of that urn is changed by adding r balls of color c_l. And so on for all the transitions we observe in the data, one counterparty at a time, under the hypothesis of exchangeability.

The parameter r thus plays the role of learning parameter. The bigger r, the quicker the model learns and adapts to the empirical data. The smaller r, the longer it will take to update our a priori. The calibration of r is therefore one of the ways in which we can show how confident we are about our beliefs, and how much we accept to modify them. Once again the importance of r diminishes with the number of available observations. With $r > 0$, thousands of observations will always be able to modify our a priori, shifting it towards the empirical reality that emerges from data. In the following we choose four different values for the reinforcement, $r \in \{0, 0.01, 1, 100\}$, where $r = 0$ means that we do not update our a priori, as if we do not trust data, while $r = 100$ indicates that empirical evidence is able to quickly modify our prior beliefs. For more details, once again we refer to the Appendix.

A natural question then arises: why do we need to elicit any a priori if we will always converge towards the empirical distribution of the data, with a sufficient number of observations? The answer is that: first, our prior beliefs become extremely important when the number of observations is not large, as in low default portfolios for instance [2]; second, a correct prior may compensate for the lack of information in the data and the problems of historical bias, as in the case of extreme and rare events [7, 8]; third, with the right prior elicitation we can also embed meaningful ideas about not-yet-observed trends and future developments. In the limit, in the totally ideal case in which our priors were true, we would need not a single observation in order to get perfect predictions.

2.5.3. RESULTS

We now discuss the performances of the different R-RUP models, when initialized with one of the Priors Sets of Subsection 2.5.2, trained using actual data like the training set in Table 2.1, and then used to obtain the posterior distributions of the total recovery times, and of the recovery levels, for the different classes of interest.

Following [50], we start by testing the performances of our model in-sample, comparing the posterior distributions with the empirical distributions of the training data. This comparison takes the name of posterior consistency check [51].

Tables 2.2 and 2.3 show the p-values of several Kolmogorov-Smirnov goodness-of-fit tests (KS-test), between the posteriors and the empirical distributions of the total recovery times in the training sample, for the different groups of Table 2.1. We show the KS-tests as they tend to be more conservative, hence in principle against our model. Compatible results hold using other tests like the χ^2, omitted for the sake of space.

Let us consider Table 2.2, where the p-values for the posterior consistency check are provided for the loan sizes. Looking at loan size class $(200K, 250K]$, the R-RUP successfully passes the KS-test for most combinations of Prior Sets and reinforcement $r > 0$, at the 95% or 99% confidence level. As anticipated, the relative irrelevance of the value of r is due to the large number of observations available, together with the rather regular behavior of this class of exposures, which help the R-RUP in quickly learning from data. Conversely, as expected, if we do not let the R-RUP learn from the data, setting $r = 0$, and

we just compare our naive prior beliefs with the empirical distributions of the recovery times, the null hypothesis of same distribution is definitely rejected.

Loan size	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	0.13	0.08	0.15	0.00
	2	0.16	0.09	0.06	0.00
	3	0.09	0.01	0.23	0.00
(100K,150K]	1	0.01	0.00	0.00	0.00
	2	0.01	0.00	0.04	0.00
	3	0.02	0.07	0.02	0.00
(150K,200K]	1	0.10	0.05	0.11	0.00
	2	0.09	0.03	0.06	0.00
	3	0.02	0.06	0.00	0.00
(200K,250K]	1	0.90	0.58	0.26	0.00
	2	0.92	0.45	0.47	0.00
	3	0.39	0.22	0.22	0.00
> 250K	1	0.31	0.86	0.64	0.00
	2	0.46	0.74	0.07	0.00
	3	0.60	0.75	0.68	0.00

Table 2.2: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times in the training sample (20113 defaulted exposures originated in the first quarter of 2006), comparing the posterior R-RUP distributions and the empirical ones, for different reinforcements r. The training sample has been divided into 5 groups in terms of loan size, as per Table 2.1.

FICO score	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 650	1	0.01	0.02	0.03	0.00
	2	0.06	0.04	0.02	0.00
	3	0.04	0.06	0.00	0.00
(650,685]	1	0.02	0.04	0.02	0.00
	2	0.07	0.28	0.02	0.00
	3	0.04	0.13	0.07	0.00
(685,725]	1	0.21	0.16	0.03	0.00
	2	0.08	0.26	0.05	0.00
	3	0.08	0.06	0.21	0.00
> 725	1	0.03	0.09	0.06	0.00
	2	0.06	0.03	0.08	0.00
	3	0.03	0.17	0.06	0.00

Table 2.3: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times in the training sample (20113 defaulted exposures originated in the first quarter of 2006), comparing the posterior R-RUP distributions and the empirical ones, for different reinforcements r. The training sample has been divided into 4 groups in terms of FICO scores, as per Table 2.1.

Interestingly, in the $(100K, 150K]$ class of Table 2.2, we need the strong reinforcement $r = 100$ in order not to reject the null hypothesis at the 1% significance level, indicating that this class probably contains more peculiar behaviors with respect to our prior beliefs (and the other classes), and the R-RUP thus requires a stronger reinforcement (or possibly more data) in order to deal with them. A check of the data confirms our suspects: in this class there is a very large variability in the total recovery times, almost twice the other classes. For all the other loan classes, results are in line with what we have just said.

In Table 2.3 the same type of analysis is performed using the FICO scores. In this case the R-RUP still shows interesting performances, and the best results are apparently obtained setting $r = 1$, that is a mild reinforcement, which averages our prior beliefs and empirical evidence. This is due to the fact that a strong reinforcement like $r = 100$

quickly amplifies recurring patterns in the data, so that the few unusual ones appear farther away from the bulk of the distribution, and this is known to have effects on the KS statistic, which is based on the supremum distance.

All in all, the R-RUP satisfactorily passes the posterior consistency check, when we deal with the total recovery times.

In Table 2.4 we show the p-values of the KS-tests for the posterior check when looking at the recovery levels per loan size. For the FICO score the results are completely comparable and they are omitted. The performances of the R-RUP on the recovery levels are extremely good, for most reinforcements, indicating that our prior beliefs are probably not too far from reality [48], and only minor modifications are necessary. This holds also for the $(100K, 150K]$ class, which is no longer as problematic as for the recovery times. In a nutshell: all combinations of Prior Sets and positive reinforcements give very good results. This makes sense, if we remember that the prior sets mainly differ for the time part, while the levels part is always the same (Subsection 2.5.2).

As further evidence of the good in-sample performances of the R-RUP, in Figure **??** we show the comparison between the empirical cumulative distribution function (ecdf) of the actual recovery levels in the 1st Quarter of 2016, for loan size class ≤ 100 (see Table 2.1), with a purposely wrongly elicited a priori, and with the updated posterior with reinforcement $r = 100$. As expected, given the good results with the KS tests, the trained model posterior very well approximates the actual distribution from the data, showing the ability of learning and improving, even when starting from a wrong set of beliefs.

Loan size	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	0.99	1.00	0.97	0.00
	2	1.00	0.98	0.98	0.00
	3	1.00	0.99	1.00	0.00
(100K,150K]	1	0.80	0.31	0.88	0.00
	2	0.82	0.80	0.72	0.00
	3	0.82	0.51	0.52	0.00
(150K,200K]	1	0.38	0.83	0.46	0.00
	2	0.49	0.32	0.25	0.00
	3	0.55	0.40	0.46	0.00
(200K,250K]	1	1.00	1.00	1.00	0.00
	2	1.00	1.00	1.00	0.00
	3	1.00	0.99	1.00	0.00
> 250K	1	0.99	1.00	0.97	0.00
	2	1.00	1.00	1.00	0.00
	3	1.00	0.99	1.00	0.00

Table 2.4: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the recovery levels in the training sample (20113 defaulted exposures originated in the first quarter of 2006), comparing the posterior R-RUP distributions and the empirical ones, for different reinforcements r. The training sample has been divided into 5 groups in terms of loan size, as per Table 2.1.

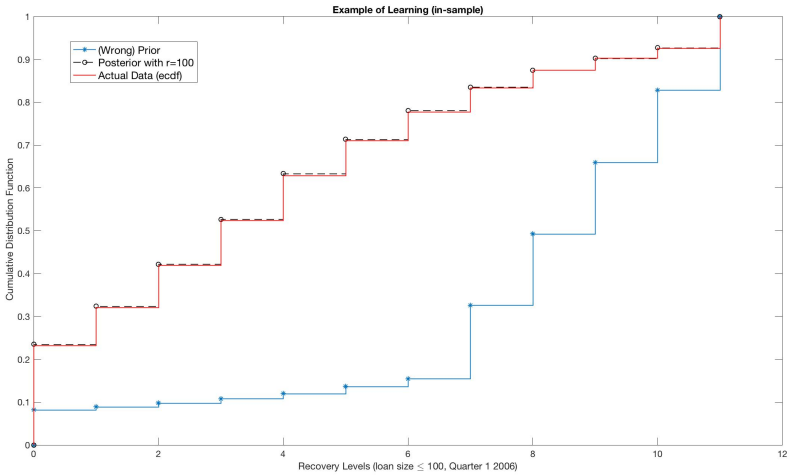


Figure 2.2: Example of (in-sample) learning, showing the ecdf of the actual recovery levels in the first quarter 2016, a wrong prior and the posterior after the training of the model with $r = 100$. As expected, the model posterior very well approximates the actual distribution from the data.

Satisfied with the in-sample performances of the R-RUP, we can move to the more interesting out-of-sample validation. Is the trained R-RUP able to predict the recovery times and levels of future companies?

Following again [50], we can compare the posterior marginal distributions generated by the trained R-RUP (on the first quarter of 2006) with the empirical distributions of the validation sample, for example the one containing the 20038 defaulted exposures originated in the second quarter of 2006. Theorem 2 in the Appendix proves extremely helpful in this situation. In Table 2.5, it is nice to see how the trained R-RUP is able to well approximate the distribution of the total recovery times in the validation sample. Similar results are obtainable using the loan size classes and/or the recovery levels, and they are available upon request.

Let us now consider some predictions made by the trained R-RUP with respect to the validation set, using the methodology we explain in the Appendix. Tables 2.6, 2.7, 2.8 and 2.9 compare the predicted median recovery times and levels for the defaulted exposures originated in the second quarter of 2006 with the actual medians from the validation set. The choice of the median as quantity of interest for credit risk purposes is consistent with works like that of [46], where medians are chosen for their statistical robustness.

Looking at Table 2.6, for example, we see that in the validation dataset, the actual median total recovery time is 15 months in loan size class $(100K, 150K]$, and 13 months in class $> 250K$. In both situations the R-RUP is able to provide predicted median values that are very close to the actual ones. The equivalence is also supported statistically by Mann-Whitney tests on medians. As before, the null reinforcement $r = 0$ is the one giving the worst results, and we find this comforting, as it underlines, once again, the ability of the R-RUP of learning and updating itself.

Upon request, we are glad to share the comparable results we obtain using other validation samples, like the defaulted exposures originated in the different quarters of 2007.

FICO score	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 650	1	0.37	0.27	0.71	0.00
	2	0.79	0.24	0.99	0.00
	3	0.83	0.19	0.20	0.00
(650,685]	1	0.20	0.21	0.38	0.00
	2	0.48	0.10	0.21	0.00
	3	0.12	0.05	0.21	0.00
(685,725]	1	0.21	0.21	0.06	0.00
	2	0.03	0.17	0.26	0.00
	3	0.31	0.20	0.06	0.00
> 725	1	0.54	0.69	0.46	0.00
	2	0.40	0.96	0.76	0.00
	3	0.60	0.86	0.30	0.00

Table 2.5: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests between the posterior distributions of the trained R-RUP and the empirical distributions of the validation sample (20038 defaulted exposures originated in the second quarter of 2006) for the total recovery times in the 4 FICO score classes.

Loan size	Prior Set	Actual	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	13	13	13	13	21
	2	13	13	13	13	18
	3	13	13	13	13	123
(100K,150K]	1	15	14	14	14	16
	2	15	14	14	15	9
	3	15	14	14	15	123
(150K,200K]	1	15	15	14	15	17
	2	15	14	15	15	9
	3	15	14	14	14	123
(200K,250K]	1	14	14	14	13	15
	2	14	14	13	13	9
	3	14	13	13	13	123
> 250K	1	13	13	13	13	17
	2	13	13	13	13	10
	3	13	13	13	13	124

Table 2.6: Median total recovery times per loan size class as predicted by the trained R-RUP against the actual values in the validation sample, for different configurations of priors and reinforcements r. Time expressed in months. Data in Table 2.1.

FICO score	Prior Set	Actual	r = 100	r = 1	r = 0.01	r = 0
≤ 650	1	16	15	15	15	25
	2	16	15	15	15	15
	3	16	15	15	15	123
(650,685]	1	15	14	14	14	16
	2	15	14	14	14	9
	3	15	14	14	14	123
(685,725]	1	14	14	14	14	21
	2	14	14	14	13	15
	3	14	14	14	14	124
> 725	1	13	12	13	12	17
	2	13	12	12	12	7
	3	13	13	13	12	124

Table 2.7: Median total recovery times per FICO scores class as predicted by the trained R-RUP against the actual values in the validation sample, for different configurations of priors and reinforcements r. Time expressed in months. Data in Table 2.1.

Loan size	Prior Set	Actual	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	3	3	3	3	9
	2	3	3	3	3	10
	3	3	3	3	3	11
(100K,150K]	1	5	5	5	5	7
	2	5	5	5	5	10
	3	5	5	5	5	10
(150K,200K]	1	5	5	5	5	7
	2	5	5	5	5	10
	3	5	5	5	5	10
(200K,250K]	1	5	5	5	5	7
	2	5	5	5	5	11
	3	5	5	5	5	10
> 250K	1	6	6	6	7	8
	2	6	6	6	7	10
	3	6	6	6	6	11

Table 2.8: Median recovery levels per loan size class as predicted by the trained R-RUP against the actual values in the validation sample, for different configurations of priors and reinforcements r. Levels follow the classification in Equation (2.4). Data in Table 2.1.

FICO score	Prior Set	Actual	r = 100	r = 1	r = 0.01	r = 0
≤ 650	1	5	5	5	5	9
	2	5	5	5	5	11
	3	5	5	5	5	11
(650,685]	1	5	5	5	5	7
	2	5	5	5	5	11
	3	5	5	5	5	11
(685,725]	1	5	5	5	6	9
	2	5	5	5	6	10
	3	5	5	5	5	11
> 725	1	5	5	5	5	7
	2	5	5	5	7	11
	3	5	5	5	7	11

Table 2.9: Median recovery levels per FICO scores class as predicted by the trained R-RUP against the actual values in the validation sample, for different configurations of priors and reinforcements r. Levels follow the classification in Equation (2.4). Data in Table 2.1.

In line with expectations, bad performances are obtained if we move further into the

FICO score	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 650	1	0.01	0.01	0.01	0.00
	2	0.01	0.03	0.00	0.00
	3	0.02	0.02	0.00	0.00
(650, 685]	1	0.00	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00
(685, 725]	1	0.00	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00
> 725	1	0.00	0.00	0.00	0.00
	2	0.00	0.00	0.00	0.00
	3	0.00	0.00	0.00	0.00

Table 2.10: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the recovery levels, when comparing the posterior distributions of the R-RUP trained on 2006 data, with the empirical ones from the fourth quarter of 2009 (1252 data points), for different reinforcements r and the usual Prior Sets, according to the four FICO score classes.

future, still using only data from 2006 to train the model. The behaviors of the total recovery times and levels in 2008 and 2009 are not correctly reproduced, notwithstanding the r value and the Prior Set. For instance, in Table 2.10 we show the bad predictive power of the R-RUP trained on 2006 data, when dealing with the recovery levels of the defaulted exposures originated in the fourth quarter of 2009, according to the FICO classes. The only exception is represented by the class ≤ 650, consisting of the least reliable loans. This makes sense to us: being the worst class, the possibility of getting worse is limited, therefore it is more stable and easier to predict.

The decrease in the goodness of fit and in the predictive power for 2008 and 2009 is probably due to the impact of the 2008 economic crisis, which has substantially changed the dynamics of the recovery processes, given the larger number of defaults and the higher uncertainty on the markets [52, 53]. To satisfactorily capture the dynamics of late 2008 or 2009, the R-RUP thus needs to be re-trained on 2007 and 2008 data[5]. Given its probabilist properties, and in particular conjugacy (Theorem 2), updating a R-RUP is rather simple, and it does not require to perform all computations anew. It is in fact sufficient to update its parameters using the new observations, as we show in the Appendix.

In Tables 2.11 and 2.12 we show the performances of the R-RUP for the total recovery times, if we complement its initial training on 2006 data with additional observations from 2007 and 2008. In particular, always focusing on the fourth quarter [6], we add the recovery trajectories of 20118 defaulted exposures originated in 2007, and 3900 trajectories from 2008. We then use the re-trained model to make predictions with respect to the 1252 defaulted exposures originated in the fourth quarter of 2009. The p-values in the tables clearly show the ability of the R-RUP to correctly predict the total recovery times. Interestingly, we can observe relatively poorer performances in two cases: class > 725 for FICO scores, and class ≤ 100 in terms of loan size. Our explanation is that these two classes are those historically experiencing lower default rates, higher recovery levels and shorter recovery times. A sudden change in the recovery trajectories, as that observed

[5] Or better priors have to be used. In a little experiment, on the basis of what we know happened in 2008 and 2009, we have modified our priors, and the R-RUP actually works well, because the a priori compensate for the lack of empirical information.

[6] Similar results hold for the other quarters, and combinations of quarters.

during the crisis, is therefore more difficult to grasp. Nevertheless we consider the overall results quite satisfactory. Again, notice that as expected the naive Prior Set 3 is the one with the worst performances.

FICO score	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 650	1	0.99	0.96	0.95	0.10
	2	0.96	0.97	0.72	0.00
	3	0.96	0.96	0.03	0.00
(650, 685]	1	0.22	0.14	0.17	0.00
	2	0.16	0.23	0.02	0.00
	3	0.14	0.17	0.00	0.00
(685,725]	1	0.27	0.19	0.22	0.00
	2	0.29	0.29	0.21	0.00
	3	0.24	0.21	0.00	0.00
> 725	1	0.04	0.01	0.00	0.00
	2	0.02	0.01	0.11	0.00
	3	0.00	0.00	0.00	0.00

Table 2.11: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times, when comparing the posterior distributions of the R-RUP trained on loans originated in the first quarter of 2006 and the fourth quarters of 2007 and 2008, with the empirical ones from the fourth quarter of 2009, for different configurations of priors, reinforcements r, and FICO score classes.

Loan size	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	0.02	0.02	0.00	0.00
	2	0.03	0.01	0.01	0.00
	3	0.01	0.01	0.00	0.00
(100K,150K]	1	0.17	0.12	0.18	0.00
	2	0.18	0.18	0.13	0.00
	3	0.15	0.17	0.00	0.00
(150K,200K]	1	0.15	0.12	0.15	0.00
	2	0.16	0.19	0.14	0.00
	3	0.17	0.10	0.00	0.00
(200K,250K]	1	0.75	0.77	0.73	0.03
	2	0.79	0.82	0.02	0.00
	3	0.77	0.78	0.00	0.00
> 250K	1	0.41	0.46	0.30	0.00
	2	0.38	0.40	0.93	0.00
	3	0.39	0.42	0.00	0.00

Table 2.12: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times, when comparing the posterior distributions of the R-RUP trained on loans originated in the first quarter of 2006 and the fourth quarters of 2007 and 2008, with the empirical ones from the fourth quarter of 2009, for different configurations of priors, reinforcements r, and loan size classes.

If we move from recovery times to recovery levels, we get similarly good results: the re-trained R-RUP is able to improve its predictive power, obtaining interesting performances, well above those of the old R-RUP trained on 2006 data only. All tables, omitted for the sake of space, are available upon request, also for the subsequent years up to the end of 2015. Here we only show Table 2.13, which contains the actual and the predicted median recovery levels in the 4th quarter of 2009, using data from 2006, 2007 and 2008. It is interesting to notice how the R-RUP, in this specific case, slightly underestimates the ultimate recovery level, on average by one level, showing a conservative behavior[7].

[7] Given the recent discussions about the margins of conservatism, and the preference of regulators for con-

The reason are once again unusual observations at the end of 2009, and, as before, our already good predictions could be improved by inputing more data in the R-RUP, for instance using those from the first half of 2009. It goes without saying that better priors, i.e. better experts' judgements, or a different scale (finer for instance) for the recovery levels could also be viable solutions.

Loan size	Prior Set	Actual	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	6	4	4	4	7
	2	6	4	4	4	10
	3	6	4	4	4	11
(100K,150K]	1	7	5	5	5	8
	2	7	5	5	6	10
	3	7	5	5	6	11
(150K,200K]	1	7	6	6	6	7
	2	7	6	6	6	10
	3	7	6	6	6	10
(200K,250K]	1	8	6	6	6	7
	2	8	6	6	7	11
	3	8	6	6	7	11
> 250K	1	8	7	7	7	8
	2	8	7	7	7	10
	3	8	7	7	7	10

Table 2.13: Median recovery levels as predicted by the trained R-RUP, using data from 2006 to 2008, against the actual values in the validation sample (4th quarter of 2009), for different configurations of priors, reinforcements r, and loan size classes.

FICO score	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 650	1	0.93	0.94	0.92	0.66
	2	0.95	0.75	0.00	0.00
	3	0.95	0.97	0.00	0.00
(650,685]	1	0.77	0.76	0.89	0.97
	2	0.74	0.63	0.00	0.00
	3	0.79	0.89	0.00	0.00
(685, 725]	1	0.97	0.94	0.87	0.78
	2	0.96	0.99	0.00	0.00
	3	0.97	0.83	0.00	0.00
> 725	1	0.88	0.89	0.98	0.19
	2	0.93	0.66	0.00	0.00
	3	0.80	0.97	0.00	0.00

Table 2.14: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times, when comparing the posterior distributions of the R-RUP trained on yearly data from 2008, with the empirical ones from 2009, for different configurations of priors, reinforcements r, and FICO score classes.

The good results we obtain for quarters also hold if we consider yearly observations. In Tables 2.14 and 2.15 we show the comparisons at the yearly level for the recovery times, using the data up to the end of 2008 to predict the times in 2009. Once again the R-RUP performs well, learning from the data, even after a major economic crisis. Please notice that, in order to lower the computational burden, for the yearly comparisons we have used a smaller dataset, always provided by Freddie Mac[8].

servative estimates, slightly underestimating the actual recovery level would be something acceptable in the Basel framework [4]. Unfortunately, we cannot guarantee that the R-RUP is always conservative, for all parameter choices. For sure a conservative prior set could be used as a standard, and even proposed by regulators.

[8] Citing [23]: "The sample dataset is a simple random sample of 50000 loans selected from each full vintage

Loan size	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	0.96	0.95	0.96	0.66
	2	0.96	0.77	0.00	0.00
	3	0.97	0.99	0.00	0.00
(100K, 150K]	1	0.38	0.48	0.49	0.52
	2	0.48	0.27	0.00	0.00
	3	0.42	0.51	0.00	0.00
(150K,200K]	1	0.84	0.80	0.70	0.29
	2	0.88	0.95	0.00	0.00
	3	0.92	0.65	0.00	0.00
(200K, 250K]	1	0.57	0.59	0.42	0.31
	2	0.61	0.77	0.00	0.00
	3	0.59	0.33	0.00	0.00
> 250K	1	0.32	0.34	0.42	0.87
	2	0.36	0.37	0.00	0.00
	3	0.34	0.32	0.00	0.00

Table 2.15: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times, when comparing the posterior distributions of the R-RUP trained on yearly data from 2008, with the empirical ones from 2009, for different configurations of priors, reinforcements r, and loan size classes.

All in all, the R-RUP construction is clearly able to model recovery trajectories in a satisfactory way, being capable of adapting to periods of crisis, by constantly updating its performances. And that is actually the idea: to use the model under continuous reinforcement, every time new data become available. As said, updating the R-RUP does not require to perform all computations anew, but only to add the new information.

It is important to stress that our R-RUP has been trained using a lot of observations, which naturally correct wrong priors via reinforcement. In case of smaller datasets, the reliability of the prior beliefs becomes fundamental in order to obtain satisfactory results. In case of no prior knowledge, as common in Bayesian statistics [54], we naturally suggest the use of the empirical distribution of the training set as starting point. If used in the IRB setting [2], specific prior sets could also be imposed by the regulator, which could also define restrictions for the reinforcement parameter.

2.5.4. THE IMPACT OF DISCRETIZATION

By construction, the R-RUP relies on the discretization of recovery times and rates. While the discretization of time is not a problem, given that observations are often taken at predetermined time points and, as discussed in [18], discrete time may even be an advantage in dealing with recovery risk, it may be worth analyzing the impact of discretizing recovery rates into recovery levels.

In Section 2.3, we said that one needs to define $m+1$ recovery levels, where level 0 accounts for no recovery, levels 1 to $m-1$ represent intermediate stages of recovery up to full recovery, and level m is the official termination level guaranteeing the technical condition of recurrence (Appendix: Lemma 1). The larger m, the finer the partition for the discretized recovery rates.

In the application to the Freddie Mac data, we have used the recovery levels of Equation (2.4). In that partition, each level between 1 and 10 accounted for an extra 10% recovery. Do our results change if we modify the interpretation of each level, or if we

year and a proportionate number of loans from each partial vintage year of the full Single Family Loan-Level Dataset."

chose a different number of levels?

Let us consider the following three alternatives:

$$L_1 = \left\{ \begin{array}{l} 0:0\%,\ 1:(0\%,2\%),\ 2:[2\%,18\%),\ 3:[18\%,28\%),\ 4:[28\%,36\%), \\ 5:[36\%,44\%),\ 6:[44\%,51\%),\ 7:[51\%,60\%),\ 8:[60\%,71\%), \\ 9:[71\%,88\%),\ 10:[88\%,100\%),\ 11:100+\%,\ 12:\text{termination} \end{array} \right\},$$

$$L_2 = \left\{ \begin{array}{l} 0:0\%,\ 1:(0\%,h\%),\ 2:[h\%,2h\%),\ 3:[2h\%,3h\%),\ 4:[3h\%,4h\%), \\ 5:[4h\%,5h\%),\ 6:[5h\%,6h\%),\ 7:[6h\%,7h\%),\ 8:[7h\%,8h\%), \\ 9:[8h\%,9h\%),\ 10:[9h\%,10h\%),\ 11:[10h\%,11h\%),\ 12:[11h\%,100\%), \\ 13:100+\%,\ 14:\text{termination} \end{array} \right\},$$

where $h = 8.\bar{3}$, and

$$L_3 = \left\{ \begin{array}{l} 0:0\%,\ 1:(0\%,12.5\%),\ 2:[12.5\%,25\%),\ 3:[25\%,37.5\%),\ 4:[37.5\%,50\%), \\ 5:[50\%,62.5\%),\ 6:[62.5\%,75\%),\ 7:[75\%,87.5\%),\ 8:[87.5\%,100\%), \\ 9:100+\%,\ 10:\text{termination} \end{array} \right\}.$$

In L_1 we use the same number of levels of L in Equation (2.4), but we change their interpretation: they are no longer equally-spaced. The intervals are empirically chosen so that each of them contains the same amount of observations, about 10%. In L_2 we increase the number of levels to 14, with a finer partition, in which each level between 1 and 12 accounts for an extra recovery of $8.\bar{3}$%. Finally, in L_3 we decrease the number levels, and each intermediate level accounts for an extra recovery of 12.5%.

Table 2.16 contains the same information of Table 2.2, when we substitute L with L_1. Clearly the use of L_1 improves the goodness-of-fit. Since the intervals in L_1 are chosen empirically, to guarantee the same number of observations per level, this partition optimizes the information in the dataset in terms of number of updates per level, and this improves the fitting. Table 2.17 shows the results of the goodness-of-fit for L_2: in this case no dramatic difference is observed with respect to Table 2.2. The same holds when using L_3.

Regarding the goodness-of-fit for the ultimate recovery rate and the predictive medians of the total recovery times and the ultimate recovery rate, the choice of L_1, L_2 and L_3 does not have any particular influence. The R-RUP shows to be quite robust with respect to alternative choices of the recovery level partition. This is probably due to the large number of observations available, and–to a minor extent–to the fact that the partitions above are not dramatically different.

Loan size	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	0.84	0.76	0.86	0.00
	2	0.78	0.66	0.00	0.00
	3	0.89	0.88	0.00	0.00
(100K, 150K]	1	0.85	0.97	0.64	0.00
	2	0.48	0.96	0.00	0.00
	3	0.39	0.91	0.00	0.00
(150K, 200K]	1	0.92	0.30	0.62	0.00
	2	0.36	0.96	0.00	0.00
	3	0.97	1.00	0.00	0.00
(200K, 250K]	1	1.00	0.91	1.00	0.00
	2	0.96	0.54	0.00	0.00
	3	0.83	1.00	0.00	0.00
> 250K	1	0.58	0.86	0.94	0.00
	2	0.87	1.00	0.00	0.00
	3	0.74	0.93	0.00	0.00

Table 2.16: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times in the basic training sample (20113 defaulted exposures originated in the first quarter of 2006), comparing the posterior R-RUP distributions and the empirical ones, for different reinforcements r, under L_1.

Loan size	Prior Set	r = 100	r = 1	r = 0.01	r = 0
≤ 100K	1	0.13	0.20	0.23	0.00
	2	0.51	0.09	0.00	0.00
	3	0.49	0.18	0.00	0.00
(100K,150K]	1	0.07	0.02	0.03	0.00
	2	0.01	0.00	0.00	0.00
	3	0.02	0.05	0.00	0.00
(150K,200K]	1	0.19	0.23	0.36	0.00
	2	0.36	0.07	0.00	0.00
	3	0.11	0.23	0.00	0.00
(200K,250K]	1	0.28	0.70	0.68	0.00
	2	0.55	0.45	0.00	0.00
	3	0.29	0.62	0.00	0.00
> 250K	1	0.77	0.37	0.55	0.00
	2	0.37	0.67	0.00	0.00
	3	0.90	0.53	0.00	0.00

Table 2.17: P-values for two sample Kolmogorov-Smirnov Goodness-of-Fit Tests for the total recovery times in the basic training sample (20113 defaulted exposures originated in the first quarter of 2006), comparing the posterior R-RUP distributions and the empirical ones, for different reinforcements r, under L_2.

On average, choosing a smaller number of levels improves fitting, because the number of observations and transitions per level increases, reinforcing the Bayesian learning process, ceteris paribus. Similarly, as seen above, an improvement is observed when intervals are chosen to guarantee more or less the same number of observations per level. Conversely, increasing m too much may lead to the situation in which for a specific level no transition is observed, so that our a priori is not changed, and, if our beliefs are wrong (or not meant to compensate an alleged lack of information in the data), this has naturally an impact on the goodness of fit.

As natural trade-off, changing m has also effects in terms of precision. In the limit, we could consider a partition with just 4 levels, in which level 2 corresponds to a recovery rate in the interval $(0\%, 100\%)$. This would dramatically increase the goodness of fit, but we all agree it would be useless, as we would end up saying that most counterparties reach some recovery rate between 0% and 100%, without distinction. On the opposite

side we could be very precise and have thousands of levels, each representing an additional recovery of just a few decimals, but our a priori would be almost never updated for many levels (apart from the difficulty of eliciting it), and the definition of the strength of the reinforcement would become extremely complicated.

The best way to decide the number of levels is to find 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, but in any case each level should be characterized by a minimum number of transitions in order to exploit the Bayesian learning mechanism. To improve fitting, rarely visited levels should be aggregated, or, at least, the corresponding reinforcement calibrated to maximize the available information. Once again, experts' judgements could represent a viable solution.

APPENDIX

We here collect all the mathematics related to the R-RUP construction, including definitions, theorems, proofs and technical details. Please notice that, not to overload notation, vector quantities are not expressed in bold.

NOTATION, THE PROCESS $\{Y_p\}$ AND RECURRENCE

A process is said recurrent when, in infinite time, it will visit a given state infinitely often with probability one. For us this point is the origin $(0,0)$.

From [40] we know that a recurrent RUP is partially exchangeable in the sense of [55], defining a sequence of visited states that can be collected into exchangeable blocks, each one representing a Markov chain. Therefore, a recurrent RUP gives rise to a mixture of Markov chains with a given de Finetti (mixing) measure [40].

Can we obtain something similar for the R-RUP? The answer is yes.

Fix $\eta_0^0 = 0$, and let the random variable $\eta_i^0 = \inf\{j > \eta_{i-1}^0 : X_j = (0,0)\}$ represent the i-th time the R-RUP process visits the point $(0,0)$ for all nonnegative integer i, defining what [40] call a 0-block, i.e. a sequence of states in S starting with (0,0) and containing no further (0,0)[9]. Thus $R_i = \{X_{\eta_{i-1}^0}, \ldots, X_{\eta_i^0 - 1}\}$ will be the recovery history, with all the intermediate stages of recovery, of the i-th defaulted counterparty, until its recovery process stops, because of a write-off, a full recovery or censoring. Let $\{R_i\}_{i=1}^k$ be the sequence of successive 0-blocks in $\{X_n\}$, representing the recovery histories of the first k defaulted exposures.

Let ψ be a mapping projecting all the finite sequences of elements of S, starting with an initial state $(0,0)$ and ending up with $(1,m)$, with no $(0,0)$ appearing in between, into a sequence of recovery levels belonging to L. For $k \in \mathbb{N}_0$, any $t_0, \ldots, t_k \in \mathbb{N}_0$ and $l_1, l_2, \ldots, l_k \in L$, we have

$$\psi((0,0),\ldots,(t_0,0),\ldots,(0,l_k),\ldots,(t_k,l_k),(0,m),(1,m)) = \left\{ \overbrace{0,\ldots,0}^{t_0 \text{ times}}, \overbrace{l_1,\ldots,l_1}^{t_1 \text{ times}}, \ldots, \overbrace{l_k,\ldots,l_k}^{t_k \text{ times}}, m \right\}.$$

[9] Notice that, given our assumptions on the urns, every 0-block necessarily ends with the state $(1,m)$.

Notice that $\psi(R_k)$ is just another way of representing the recovery process in block R_k, just focusing on the recovery levels.

The one-to-one correspondence between R_k and $\psi(R_k)$ is revealed by a simple example with termination level $m = 4$. For a block

$$R_k = \{(0,0),(1,0),(0,2),(1,2),(0,3),(1,3),(0,4),(1,4)\},$$

we have $\psi(R_k) = (0,2,3,4)$. In this new representation, to recognize a 0-block, we have to look for the first 0 following termination level m.

Since the map ψ is measurable and bijective, and, in case of recurrence, the blocks $\{R_k\}_{k\geq 1}$ are exchangeable, the sequence $\{\psi(R_k)\}_{k\geq 1}$ is also exchangeable [55], and we can use the sequence of levels $\{\psi(R_k)\}$ to build a new process $\{Y_p\}_{p\geq 0}$.

Assume that the first two recovery trajectories we observe are

$$R_1 = \{(0,0),(1,0),(0,2),(1,2),(0,3),(1,3),(0,4),(1,4)\}$$

and

$$R_2 = \{(0,0),(1,0),(2,0),(3,0),(0,1),(1,1),(0,2),(1,2),(0,4),(1,4)\}.$$

Then, the corresponding realization of the process $\{Y_p\}$ is

$$\{Y_0,Y_1,Y_2,Y_3,Y_4,Y_5,Y_6,Y_7,Y_8,Y_9\} = \{\psi(R_1),\psi(R_2)\} = \{0,2,3,4,0,0,0,1,2,4\}.$$

We define all the notations we need to use afterwards. For any $l \in L$, set $\xi_0^l = 0$, and for the positive integer n, define recursively

$$\xi_n^l = \inf\{t > \xi_{n-1}^l : Y_{t-1} \neq l,\, Y_t = l\},$$

as the n-th time the process $\{Y_p\}_{p\geq 0}$ touches l. For every $l \in L$, define a sequence $\{\tau_n^l\}_{n\geq 1}$ as follows:

$$\tau_n^l = \inf\{\, t - \xi_{n-1}^l : t > \xi_{n-1}^l,\, Y_t \neq l\}. \tag{2.5}$$

If $\{X_n\}$ is recurrent, for any $l \in L$ and $t \geq 0$, define

$$\delta_n^l = Y_{\xi_{n-1}^l + \tau_n^l}.$$

All in all, τ_n^l summarizes the sojourn time at recovery level l during the n-th visit, while δ_n^l is the next recovery level touched by the R-RUP after the n-th visit at level l.
Set the initial level and the initial jump time as $L_0 = J_0 = 0$. For any positive integer n, define then

$$J_n = \inf\{t > J_{n-1} : Y_t \neq Y_{t-1}\}.$$

Set also $L_n = Y_{J_n}$ and $M_n = J_n - J_{n-1}$. Clearly J_n is the n-th jump time of process $\{Y_p\}$, L_n is the level the process jumps to at time J_n, and M_n is the sojourn time at level L_{n-1} before jumping to level L_n.

From a probabilistic point of view, a necessary and sufficient condition for the R-RUP to be recurrent, so that state $(0,0)$ is visited infinitely many times, is provided by the following Lemma.

Lemma 1 (Recurrence of the R-RUP) *The R-RUP $\{X_n\}$ is recurrent if and only if, for all $l \in L$,*

$$\lim_{n\to\infty} \prod_{t=0}^{n} \frac{N_{(t,l)}(c_l)}{\sum_{i\geq l} N_{(t,l)}(c_i)} = 0, \tag{2.6}$$

where $N_{(t,l)}(c_i)$ is the number of balls of color $c_i \in C$ in urn $U((t,l))$.

Proof 1 *From Lemma 2.13 and Lemma 3.23 in [40], it is clear that Equation (2.6) is a necessary condition. Then, to prove that the process is recurrent, it is sufficient to show*

$$\mathbb{P}\left[\bigcap_{u=1}^{\infty} \{\eta_u^0 < \infty\}\right] = 1.$$

First, $\mathbb{P}\left[\eta_1^0 < \infty\right] = 1$ holds since

$$\mathbb{P}[\eta_1^0 < \infty] \geq \mathbb{P}\left[\bigcap_{l=0}^{m} \{\tau_1^l < \infty\}\right] = 1, \tag{2.7}$$

where the equality is obtained from Lemma 3.23 in [40] and the finiteness of the maximum level m. Then, by induction, we can prove on n that

$$\mathbb{P}\left[\bigcap_{u=1}^{n+1} \{\eta_u^0 < \infty\}\right] = 1,$$

if $\mathbb{P}\left[\cap_{u=1}^{n} \{\eta_u^0 < \infty\}\right] = 1$. Since

$$\mathbb{P}[\eta_{n+1}^0 < \infty] \geq \mathbb{P}\left[\bigcap_{l=0}^{m}\bigcap_{u=1}^{n+1} \{\tau_u^l < \infty\}\right] = 1,$$

we get that

$$\mathbb{P}\left[\bigcap_{u=1}^{n+1} \{\eta_u^0 < \infty\}\right] = 1,$$

and the result follows.

Lemma 1 plays with the rule of motion q, by requiring that the probability for the R-RUP to stay at a given level l for infinite time is zero. This forces the R-RUP starting in $(0,0)$ to jump to higher levels, reaching either full recovery (sampling of a c_m ball at recovery level $m-1$) or write-off (sampling of a c_m ball at recovery level $l < m-1$), before visiting the termination level m for one time unit and then restarting from $(0,0)$. Clearly, when $\{X_n\}$ is recurrent, $\{Y_p\}$ is recurrent as well.

In what follows, all the definitions, propositions, lemmas, and theorems are based on the assumption of recurrence of the process $\{X_n\}$, or equivalently of the process $\{Y_p\}$.

THE R-RUP AS MIXTURE OF SEMI-MARKOV CHAINS

Let us give some important definitions.

Definition 1 (beta-Stacy Process) *The random distribution function F is a beta-Stacy process with jumps at $t \in \mathbb{N}_0$ and parameters $\{\alpha_t, \beta_t\}_{t\in\mathbb{N}_0}$, if there exist mutually independent random variables $\{V_t\}_{t\in\mathbb{N}_0}$, each beta distributed with parameters (α_t, β_t), such that the random mass assigned by F to $\{t\}$, written $F(\{t\})$, is given by $V_t \prod_{u<t}(1-V_u)$.*

Introduced by [56], the beta-Stacy process can be seen as a generalization of the well-known Dirichlet process, a pivotal random distribution in Bayesian nonparametrics [54]. Its RUP representation was first given in [40].

Definition 2 (beta-Stacy Dirichlet Process) *The random probability mass function Q on $L\times\mathbb{N}_0$ is called beta-Stacy Dirichlet (BSD) process with parameters $\{\alpha_t, \beta_t\}_{t\geq 0}$ and $\{\gamma_t\}_{t\geq 0}$, if there exist mutually independent Dirichlet processes $\{W_t\}_{t\geq 0}$ of parameters $\{\gamma_t\}$, and a beta-Stacy process F with jumps at $t \in \mathbb{N}_0$ and parameters $\{\alpha_t, \beta_t\}_{t\in\mathbb{N}_0}$, independent of the sequence $\{W_t\}$, such that the random mass assigned to the element (j,t) in $L\times\mathbb{N}_0$ is $Q(j,t) = F(\{t\})W_t(j)$.*

The BSD process is the bivariate random distribution that characterizes the R-RUP, as stated by Theorem 1 below. However, before considering that fundamental result, we first need to introduce the concept of semi-Markov chain.

Definition 3 *A process $\{Y_p\}_{p\geq 0}$ is a discrete time semi-Markov chain on $L\times L\times\mathbb{N}_0$, if the values Y_{J_n} at its jump times form a Markov chain. Moreover, conditionally on Y_{J_n} and all the previous information, the sojourn time in state Y_{J_n}, and the next state the process jumps to, only depend on Y_{J_n}.*

If $\{Y_p\}_{p\geq 0}$ is a discrete time semi-Markov chain, let $G_l(j,t)$ be the probability of staying at level l for a time t and then jumping to another state j, i.e.

$$G_l(j,t) = \mathbb{P}\left[L_{n+1} = j, M_{n+1} = t \mid L_n = l\right].$$

Then G is the semi-Markov kernel of $\{Y_p\}$ on $L\times L\times\mathbb{N}_0$.

Consider the set of semi-Markov kernels $\mathcal{G}$ on $L\times L\times\mathbb{N}_0$ in the topology of coordinate convergence. If the process $\{Y_p\}$ is a mixture of semi-Markov chains, there exists a probability measure κ, also called the mixing measure, on the Borel subset of the space $\mathcal{G}$, such that, for any $l_1,\ldots,l_n \in L$, $t_1,\ldots,t_n \geq 0$, $n\geq 1$ and $l_0 = 0$, we have

$$\mathbb{P}\left[(L_1,M_1) = (l_1,t_1),\ldots,(L_n,M_n) = (l_n,t_n)\right] = \int_{\mathcal{G}} \prod_{u=1}^{n} G_{l_{u-1}}(l_u,t_u)\kappa(\mathrm{d}G). \tag{2.8}$$

Proposition 1 *For every $l \in L$, the sub-sequence $\{(\delta_n^l, \tau_n^l)\}$ is exchangeable. Moreover, the sub-sequences*

$$\{(\delta_n^0,\tau_n^0)\},\{(\delta_n^1,\tau_n^1)\},\ldots,\{(\delta_n^{m-1},\tau_n^{m-1})\},\{(\delta_n^m,\tau_n^m)\} \tag{2.9}$$

are mutually independent.

Proof 2 *The exchangeability of the sequence $\{(\delta_n^l,\tau_n^l)\}$ simply derives from the fact that both δ_n^l and τ_n^l are measurable functions of the 0-blocks, which are exchangeable by construction. The independence, conversely, simply derives from the R-RUP construction.*

The following theorem shows that the R-RUP $\{Y_p\}$ is a mixture of semi-Markov chains, i.e. the blocks $\psi(R_i)$ constituting $\{Y_p\}_{p\geq 0}$ are exchangeable and each of them is a semi-Markov chain. The mixing measure κ can be obtained by characterizing G. Knowing the functional form of the semi-Markov kernel G will prove essential to use the R-RUP in practice.

Theorem 1 *If recurrent, the process $\{Y_p\}_{p\geq 0}$ is a mixture of semi-Markov chains, and its mixing measure κ is unique. More precisely, there exists a unique random kernel G such that, conditionally on G, Y_p is a semi-Markov chain with semi-Markov kernel G.*

For each level of risk $l \in L$, the kernel G can be explicitly characterized as the product $G_l(j,t) = W_{(t,l)}(j)F^l(\{t\})$, for any $l, j \in L$ and $t \in \mathbb{N}_0$, where:

- *F^l is a beta-Stacy process with jumps at $t \in \mathbb{N}_0$, and parameters*

$$\alpha_t^l = \sum_{h\neq l} \frac{N_{(t,l)}(c_h)}{r} \qquad \text{and} \qquad \beta_t^l = \frac{N_{(t,l)}(c_l)}{r},$$

  with r the reinforcement of the different Pólya urns.

- *$\{W_{(t,l)}\}_{t\in\mathbb{N}_0, l\in L}$ are mutually independent Dirichlet processes of parameter $\gamma_{(t,l)}(\cdot)$, all independent of $\mathbf{F} = \{F^l, l \in L\}$, assigning mass $\frac{N_{(t,l)}(c_j)}{r}$ to the j-th component for $l+1 \leq j \leq m$, and mass 0 for $0 \leq j \leq l$.*

Proof 3 *[40] show that a recurrent RUP $\{X_n\}$ defines a mixture of Markov chains. This means that there exists a random transition matrix R, which is a random element of the set of all stochastic matrices, such that for all finite sequences $(s_0, s_1, \ldots, s_n)$ of elements of S,*

$$\mathbb{P}[X_0 = s_0, X_1 = s_1, \ldots, X_n = s_n \mid R] = \prod_{u=1}^{n} R(s_{u-1}, s_u) a.s. \tag{2.10}$$

Theorem 2.16 in [40] characterizes R by showing that its rows are mutually independent random probability masses on S, and for all $s \in S$, $R(s)$ follows a Dirichlet process with parameter $\omega(s)$, which assigns $\frac{N_s(c)}{r}$ to state $q(s,c) \in S$ with $c \in C$.

For any $l, j \in L$ and $t \in \mathbb{N}_0$, define

$$\begin{aligned} W_{(t,l)}(j) &= \frac{R[(t,l),(0,j)]}{1 - R[(t,l),(t+1,l)]}, \\ V_t^l &= 1 - R[(t,l),(t+1,l)], \\ F^l(\{t\}) &= V_t^l \textstyle\prod_{u<t}(1 - V_u^l). \end{aligned}$$

Because of the tail free property of the Dirichlet process [54],

$$W_{(t,l)}(\cdot) = \left[W_{(t,l)}(0), \ldots, W_{(t,l)}(m)\right]$$

follows a Dirichlet process with measure $\gamma_{(t,l)}(\cdot)$ assigning $N_{(t,l)}(c_j)/r$ to $j > l$ and 0 to the rest. Moreover, $\{V_t^l\}_{t\geq 0}$ are a series of independent beta distributed random variables with parameters (α_t^l, β_t^l) where $\alpha_t^l = \sum_{h\neq l} N_{(t,l)}(c_h)/r$ and $\beta_t^l = N_{(t,l)}(c_l)/r$.

From [56], under the recurrence condition of Lemma 1, F^l is a discrete-time beta-Stacy process with parameters $\{\alpha_t^l, \beta_t^l\}$. This said, the goal is to characterize the mixing measure of the component $\{(\delta_n^l, \tau_n^l)\}$ of the semi-Markov chain $\{Y_p\}$.

According to the de Finetti representation theorem, the exchangeability of the sub-sequence $\{(\delta_n^l, \tau_n^l)\}$ guarantees the existence and uniqueness of a random probability measure Q_l on $L \times \mathbb{N}_0$, conditionally on which the elements of the sub-sequence are i.i.d. with a bivariate mass function Q_l [42]. The following lemma characterizes this de Finetti measure as a new combination of well-known random distributions, namely the Dirichlet and the beta-Stacy processes.

Lemma 2 *If recurrent, for any $l \in L$, the unique random probability measure Q_l on $L \times \mathbb{N}_0$ is a BSD process with parameters $\{\alpha_t^l, \beta_t^l\}_{t\geq 0}$ and $\{\gamma_{(t,l)}\}_{t\geq 0}$, where, for any $t \in \mathbb{N}_0$, $\alpha_t^l = \sum_{h\neq l} \frac{N_{(t,l)}(c_h)}{r}$, $\beta_t^l = \frac{N_{(t,l)}(c_l)}{r}$, and $\gamma_{(t,l)}(\cdot)$ assigns mass $\frac{N_{(t,l)}(c_j)}{r}$ to the j-th component for $l < j \leq m$, and mass 0 to all the other components.*

Proof 4 *For all $t_1, \ldots, t_n \in \mathbb{N}_0$ and $l_1, \ldots, l_n \in L$, we have*

$$
\begin{aligned}
& \mathbb{P}[\delta_1^l = l_1, \tau_1^l = t_1, \delta_2^l = l_2, \tau_2^l = t_2, \ldots, \delta_n^l = l_n, \tau_n^l = t_n \mid R] \\
= \; & \prod_{k=1}^n \prod_{u<t_k} R[(u,l),(u+1,l)] \times R[(t_k,l),(0,l_k)]) \\
= \; & \prod_{k=1}^n \prod_{u<t_k} R[(u,l),(u+1,l)] \times \\
& (1 - R[(t_k,l),(t_k+1,l)]) \times \frac{R[(t_k,l),(0,l_k)]}{1 - R[(t_k,l),(t_k+1,l)]} \\
= \; & \prod_{k=1}^n [W_{(t_k,l)}(l_k) V_{t_k}^l \prod_{u<t_k} (1 - V_u^l)] \\
= \; & \prod_{k=1}^n [W_{(t_k,l)}(l_k) F^l(\{t_k\})] \qquad a.s.
\end{aligned}
$$

Define $Q_l(j,t) = W_{(t,l)}(j) F^l(\{t\})$, for any $l, j \in L$ and $t \in \mathbb{N}_0$. Since the bivariate measure is measurable with respect to the $\mathbb{P}$-completion of the Borel σ-algebra of R, the above relation remains valid after replacing R with Q_l. Hence, for any $l_1, \ldots, l_n \in L$ and $t_1, \ldots, t_n \in \mathbb{N}_0$,

$$
\mathbb{P}[\delta_1^l = l_1, \tau_1^l = t_1, \delta_2^l = l_2, \tau_2^l = t_2, \ldots, \delta_n^l = l_n, \tau_n^l = t_n \mid Q_l] = \prod_{k=1}^n Q_l(l_k, t_k) \;\; a.s.
$$

The result follows.

Lemma 3 *$F^0, \ldots, F^m$ are mutually independent. The elements $\{W_{(t,l)}\}_{t\in\mathbb{N}_0, l\in L}$ are mutually independent, and independent from $\mathbf{F} = \{F^l, l \in L\}$.*

Proof 5 *It is sufficient to show that for any $l \in L$ and any $t, u \in \mathbb{N}_0$, $W_{(t,l)}$ and $F^l(\{u\})$ are independent. We have already seen that for every $l \in L$ and $u \in \mathbb{N}_0$, $\{W_{(t,l)} : t \neq u, t \in \mathbb{N}_0\}$ and V_u^l are independent thanks to the R-RUP construction. We only need to prove that $W_{(t,l)}$ and V_t^l are independent, which can be obtained from the tail free property of the Dirichlet process. The result then follows.*

Define $G_l(\cdot,\cdot) = Q_l$, for any $l \in L$. Thanks to Lemmas 2 and 3, for any $l_1,\dots,l_n \in L$ and $t_1,\dots,t_n \in \mathbb{N}_0$, we have that

$$\mathbb{P}[(L_1,M_1)=(l_1,t_1),\dots,(L_n,M_n)=(l_n,t_n)\mid G] = \prod_{u=1}^{n} G_{l_{u-1}}(l_u,t_u) \qquad a.s. \tag{2.11}$$

Theorem 1 is thus finally proved.

THE POSTERIOR PREDICTIVE DISTRIBUTION WITH AND WITHOUT RIGHT-CENSORING

The results we have obtained with recurrence and semi-Markovianity are extremely useful to derive the posterior predictive distribution of the R-RUP in terms of the BSD process, which proves to be conjugate [54].

Here below we give an important theorem, in which we show that, given the information about the recovery processes of k counterparties, the recovery trajectory of the $k+1$-th one is still a semi-Markov chain with updated kernel. With updated kernel we indicate the one resulting from the combination of the prior knowledge, embedded into the initial composition of the urns of the R-RUP, and the empirical evidence we can collect from data, i.e. the k observed recovery trajectories.

The result we provide deals with the general case of possibly censored observations, as per Subsection 2.4.4. Naturally it is completely valid even when all recovery trajectories are fully observed, being this just a special case.

Let ρ_i, $i = 1,...,k$, be a dummy variable indicating censoring. Each recovery block R_i, $i = 1,...,k$, can then be accompanied by a corresponding ρ_i, so that $(R_i,\rho_i = 0)$ indicates that the recovery trajectory of counterparty i is not censored, while $(R_i,\rho_i = 1)$ tells us that the recovery has been censored because of some $T^{\max}$.

The information in (R_i,ρ_i) can be further summarized by $H_i = (\psi(R_i),\rho_i)$, where H_i represents the recovery history of counterparty $i = 1,...,k$, including the information about censoring. With $\boldsymbol{H_k} = [H_1,H_2,\dots,H_k]$ we indicate the histories of the first k counterparties in our portfolio. Let $s_{(t,l)}(j)$ be the number of counterparties sojourning at level l in t and then jumping to another level $j \in L$, w_t^l be the number of counterparties whose exact sojourning time at level l exceeds time t, and v_t^l be the number of counterparties whose sojourning time at level l is right-censored in t.

Theorem 2 *Conditionally on $\boldsymbol{H_k}$, possibly with right-censoring, the element H_{k+1}, representing the recovery history of the $(k+1)$-th counterparty, is (still) a semi-Markov chain with semi-Markov kernel $\hat{G}$, such that $\hat{G}_l(j,t) = E\left[\tilde{G}_l(j,t)\right]$, for $l,j \in L$ and $t \in \mathbb{N}_0$, where $\tilde{G}_l(j,t) = \tilde{F}^l(\{t\})\tilde{W}_{(t,l)}(j)$ and*

- *$\tilde{F}^l$ is a beta-Stacy process with jumps at $t \in \mathbb{N}_0$ and updated parameters*

$$\begin{aligned} \tilde{\alpha}_t^l &= \alpha_t^l + \textstyle\sum_{h>l} s_{(t,l)}(h), \\ \tilde{\beta}_t^l &= \beta_t^l + w_t^l + v_t^l. \end{aligned}$$

- *$\{\tilde{W}_{(t,l)}\}$ are mutually independent Dirichlet with parameters $\tilde{\gamma}_{(t,l)}$ where*

$$\tilde{\gamma}_{(t,l)}(j) = \gamma_{(t,l)}(j) + s_{(t,l)}(j),$$

for $t \in \mathbb{N}_0$ and $j \in L$.
As in Theorem 1, $\{\tilde{W}_{(t,l)}\}$ is independent from $\tilde{F}^l$ for every $l \in L$.

2

Proof 6 *Let $\rho_i^l, i = 1, \ldots, k_l$ be a binary variable indicating local right-censoring, where k_l is the number of counterparties sojourning at recovery level l. The recovery histories of the first k defaulted counterparties $\boldsymbol{H_k}$ can be split into several sets $\boldsymbol{H_k^l}$, $l \in L$, each of which contains the observed sojourning time τ_i^l at level l, and the next recovery level δ_i^l, when $\rho_i^l = 0$, or just the observed sojourning time if censoring has occurred with $\rho_i^l = 1$.*

Lemma 4 *For every $l \in L$, given $\boldsymbol{H_k^l}$, the posterior predictive probability mass function for the couple $(\delta_{k_l+1}^l, \tau_{k_l+1}^l)$ is given by*

$$\mathbb{P}\left[\delta_{k_l+1}^l = j, \tau_{k_l+1}^l = t \mid \boldsymbol{H_k^l}\right] = \frac{\tilde{\alpha}_t^l}{\tilde{\alpha}_t^l + \tilde{\beta}_t^l} \prod_{u<t} \frac{\tilde{\beta}_u^l}{\tilde{\alpha}_u^l + \tilde{\beta}_u^l} \times \frac{\tilde{\gamma}_{(t,l)}(j)}{\tilde{\alpha}_t^l},$$

for $j \in L$ and all nonnegative integer t, with

$$\begin{aligned} \tilde{\alpha}_t^l &= \alpha_t^l + v_t^l, \\ \tilde{\beta}_t^l &= \beta_t^l + w_t^l, \\ \tilde{\gamma}_{(t,l)}(j) &= \gamma_{(t,l)}(j) + s_{(t,l)}(j), \end{aligned} \tag{2.12}$$

where $s_{(t,l)}(j)$ is the number of counterparties sojourning at level l in t and then jumping to level j, where w_t^l is the number of counterparties whose exact sojourning time at level l exceeds time t, and where v_t^l is the number of counterparties whose sojourning time at level l is right-censored in t.

Proof 7 *If right-censoring occurs, from Lemma 2, we have that,*

$$\mathbb{P}[\tau_1^l = t_1, \rho_1^l = 1 \mid G] = 1 - \textstyle\sum_{u \le t_1} F^l(\{u\}) = 1 - F^l(t_1).$$

Otherwise,

$$\mathbb{P}[\delta_1^l = l_1, \tau_1^l = t_1, \rho_1^l = 0 \mid G] = W_{(t_1,l)}(l_1) F^l(\{t_1\}).$$

Hence,

$$
\begin{aligned}
&\mathbb{P}\left[(\tau^l_{k_l+1},\delta^l_{k_l+1})=(t,j)\mid \boldsymbol{H^l_k}\right]\\
&=\frac{\mathbb{P}[(\tau^l_{k_l+1},\delta^l_{k_l+1})=(t,j),\boldsymbol{H^l_k}]}{\mathbb{P}[\boldsymbol{H^l_k}]}\\
&=\frac{E\left[W_{(t,l)}(j)F^l(\{t\})\prod_{i=1}^{k_l}\left[W_{(t_i,l)}(l_i)F^l(\{t_i\})\mathbb{1}[\rho^l_i=0]+[1-F^l(t_i)]\mathbb{1}[\rho^l_i=1]\right]\right]}{E\left[\prod_{i=1}^{k_l}\left[W_{(t_i,l)}(l_i)F^l(\{t_i\})\mathbb{1}[\rho^l_i=0]+[1-F^l(t_i)]\mathbb{1}[\rho^l_i=1]\right]\right]}\\
&=\frac{E\left[W_{(t,l)}(j)F^l(\{t\})\prod_{\rho^l_i=0}\left[W_{(t_i,l)}(l_i)F^l(\{t_i\})\right]\prod_{\rho^l_i=1}\left[1-F^l(t_i)\right]\right]}{E\left[\prod_{\rho^l_i=0}\left[W_{(t_i,l)}(l_i)F^l(\{t_i\})\right]\prod_{\rho^l_i=1}\left[1-F^l(t_i)\right]\right]}\\
&=\frac{E\left[W_{(t,l)}(j)\prod_{\rho^l_i=0}W_{(t_i,l)}(l_i)\right]E\left[F^l(\{t\})\prod_{\rho^l_i=0}F^l(\{t_i\})\prod_{\rho^l_i=1}\left[1-F^l(t_i)\right]\right]}{E\left[\prod_{\rho^l_i=0}W_{(t_i,l)}(l_i)\right]E\left[\prod_{\rho^l_i=0}F^l(\{t_i\})\prod_{\rho^l_i=1}\left[1-F^l(t_i)\right]\right]}\\
&=\frac{\tilde\alpha^l_t}{\tilde\alpha^l_t+\tilde\beta^l_t}\prod_{u<t}\frac{\tilde\beta^l_u}{\tilde\alpha^l_u+\tilde\beta^l_u}\times\frac{\tilde\gamma_{(t,l)}(j)}{\tilde\alpha^l_t},
\end{aligned}
$$

with the second last equality coming from Lemma 3, and where, for $j\in L$ and all nonnegative integer t,

$$
\begin{aligned}
\tilde\alpha^l_t &= \alpha^l_t + \nu^l_t,\\
\tilde\beta^l_t &= \beta^l_t + w^l_t,\\
\tilde\gamma_{(t,l)}(j) &= \gamma_{(t,l)}(j) + s_{(t,l)}(j).
\end{aligned}
\tag{2.13}
$$

Thanks to Proposition 1, Theorem 2 is then proved.

Corollary 1 *The $(k+n)$-th recovery history H_{k+n} for any positive integer n, conditionally on $\boldsymbol{H_k}$, possibly with right-censoring, is a semi-Markov chain with a common semi-Markov kernel $\hat G$ defined as in Theorem 2.*

Exploiting Theorem 2 and Corollary 1, we can perform Bayesian prediction about the recovery trajectories (levels and times) of future counterparties, given our a priori–as expressed by the initial compositions of the urns in the R-RUP–and the collected empirical evidence in $\boldsymbol{H_k}$, with or without censoring. In fact, Theorem 2 tells us that the semi-Markov kernel governing the future recovery histories is the updated BSD process combining our a priori with the data, and whose parameters we know. In statistical terms, the BSD process is conjugate. The probability of every feasible future recovery history can therefore be computed explicitly, given the available information.

The more our priori is close to the truth, the more reliable our predictions from the very beginning, and the better we can overcome the problem of censoring, by compensating the lack of information in the data with our beliefs. In case of a wrong a priori, however, in a way similar to what happens with machine learning, a sufficient amount of recovery histories can compensate unrealistic beliefs, as we see in Section 2.5.

PRIOR ELICITATION

Notice that G is a random kernel on $L \times L \times \mathbb{N}_0$ and for every $l \in L$, G_l is a bivariate random probability measures taking values in the space of bivariate discrete probability measures, which can be obtained as the product of beta-Stacy and Dirichlet processes. Therefore, if we want to center G on a given semi-Markov kernel $\bar{G}$, we can exploit the following relationship between the prior guess $\bar{G}$ and the initial composition of the Pólya urns constituting the R-RUP. Generalizing results in [40] and [47], we have

$$E[G_l(j,t)] = \bar{G}_l(j,t) = \frac{\alpha_t^l}{\alpha_t^l + \beta_t^l} \prod_{u<t} \frac{\beta_u^l}{\alpha_u^l + \beta_u^l} \times \frac{\gamma_{(t,l)}(j)}{\alpha_t^l},$$

for $j, l \in L$, $j > l$, $t \in \mathbb{N}$, and where $\alpha_t^l = \sum_{h=l+1}^{m} \frac{N_{(t,l)}(c_h)}{r}$, $\beta_t^l = \frac{N_{(t,l)}(c_l)}{r}$, and $\gamma_{(t,l)}(j) = \frac{N_{(t,l)}(c_j)}{r}$.

From this we derive that, for every state $(t,l) \in S$, and every color c_j, with $j = 0, ..., m$, the initial urn composition is given by

$$N_{(t,l)}(c_j) = \begin{cases} d_t^l \times \bar{G}_l(j,t), & \text{if } j > l; \\ d_t^l \times \left[1 - \sum_{k=0}^{t} \sum_{h>l} \bar{G}_l(h,k)\right], & \text{if } j = l; \\ 0 & \text{if } j < l. \end{cases} \tag{2.14}$$

From Definition 3, the semi Markov kernel $\bar{G}$ is controlled by $\bar{F}^l$, charactering the distribution of the sojourn times in level l, and $\bar{W}_{(t,l)}$, expressing the probability of jumping to other levels from level l, given the time t spent in l. Hence, to elicit prior guesses easily and flexibly, we can just specify the one-dimensional distributions $\bar{F}^l$ and $\bar{W}_{(t,l)}$, replacing $\bar{G}_l$ in Equation (2.14) with their product.

The parameter $d_t^l > 0$ in Equation (2.14) is what [40] call the strength of belief. It is a value representing how confident we are in our a priori. The larger $d_t^l > 0$, the more we are convinced that our prior beliefs are correct, so that the R-RUP will need more time and more observations, or a stronger reinforcement r, in order to correct them if wrong. In this paper we set $d_t^l = 1$ for all $t \in \mathbb{N}_0$ and $l \in L$.

Regarding the reinforcement parameter r, which in the R-RUP represents the strength of updating, i.e. how empirical observations affect and modify our a priori, we have chosen four different values to compare, i.e. $r \in \{0, 0.01, 1, 100\}$. A value of $r = 0.01$ says that we do not trust the empirical evidence too much, and that we prefer to stick to our a priori, only allowing for very slow updates. A value of $r = 100$, conversely, indicates that we trust the empirical evidence and we are ready to update our beliefs quickly. Naturally for $r = 0$ the process never learns from data, and our a priori is never changed.

The following list collects the prior beliefs we have elicited to test our model, for the different classes (size of the exposure and FICO score) described in Subsection 2.5.1:

Prior Set 1: For every $l \in L$ and $t \in \mathbb{N}_0$, $\bar{W}_{(t,l)}$ assigns mass 0 to $\{0, \ldots, l\}$ and equal mass to $\{l+1, ..., m\}$, while $\bar{F}^l$ is the empirical cumulative distribution function (ecdf) of the sojourn times, as obtainable from the data, for each recovery level l.

Prior Set 2: For every $l \in L$ and $t \in \mathbb{N}_0$, $\bar{W}_{(t,l)}$ assigns mass 0 to $\{0, \ldots, l\}$ and equal mass to $\{l+1, ..., m\}$, while $\bar{F}^l$ is a discrete uniform distribution on $\{0, 1, \ldots, T_{emp}^l\}$, with

T^l_{emp} indicating the maximum empirical sojourn time at level l as observable in the training data.

Prior Set 3: For every $l \in L$ and $t \in \mathbb{N}_0$, $\bar{W}_{(t,l)}$ assigns mass 0 to $\{0,\dots,l\}$ and equal mass to $\{l+1,...,m\}$, while $\bar{F}^l$ a discrete uniform distribution on $\{0,1,\dots,100\}$.

Remember that $\bar{W}_{(t,l)}$ essentially governs the probability of jumping to higher recovery levels from (t,l), while $\bar{F}^l$ accounts for the probability of the permanence time at level l. Please notice that the a priori elicited by $\bar{W}_{(t,l)}$, with the support $\{l+1,...,m\}$ depending on the level l, corresponds to assuming that the higher the recovery level an exposure reaches, the higher the chance of further recovery. This is consistent with the empirical evidence about recovery rates in most countries [18, 19].

Given Priors Sets 1 to 3, Equation (2.14) is used to define the initial composition of all the urns of the R-RUPs in the 13×101 matrix of visitable states. Naturally, the urn compositions' limitations of Subsection 2.4.2 still apply.

REFERENCES

[1] D. Cheng and P. Cirillo, *A reinforced urn process modeling of recovery rates and recovery times,* Journal of Banking and Finance **96**, 1 (2018).

[2] Bank for International Settlements, *Basel III: A global regulatory framework for more resilient banks and banking systems,* in *Bank for International Settlements*, Vol. 2010 (2010) p. 69.

[3] IFRS9, *Financial instruments. IFRS Foundation* (2014).

[4] European Banking Authority, *Guidelines on PD estimation, LGD estimation and the treatment of defaulted exposures*, December (2016) p. 200.

[5] P. Jorion, *Risk management lessons from the credit crisis,* European Financial Management **15**, 923 (2009).

[6] F. H. Knight, *Risk, uncertainty and profit (reprint of the 1921 edition)* (Augustus M. Kelley, Bookseller).

[7] E. Lybeck, *The Black Swan: The Impact of the Highly Improbable* (Random House, 2017) pp. 1–82.

[8] H. Dickson and G. L. S. Shackle, *Ekonomisk Tidskrift*, Vol. 58 (Cambridge University Press, Cambridge, 1956) p. 250.

[9] J. Derbyshire, *The siren call of probability: Dangers associated with using probability for consideration of the future,* Futures **88**, 43 (2017).

[10] J. C. Hull, *Risk management and financial institutions,* 4th ed. (Wiley, New York, 2012).

[11] M. Schmit, *Credit risk in the Leasing Business-A case study of low probability of default,* Journal of Banking and Finance **28**, 811 (2004).

[12] J. Zhang and L. C. Thomas, *Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD,* International Journal of Forecasting **28**, 204 (2012).

[13] Basel Committee on Banking Supervision, *Bcbs*, July (2005) p. 6.

[14] BCBS, *International convergence of capital measurement and capital standards for banks,* in *Reserve Bank of New Zealand Bulletin*, Vol. 52 (1989) p. 285.

[15] W. Reitgruber, *Methodological thoughts on expected loss estimation for IFRS 9 impairment: Hidden reserves, cyclical loss predictions and LGD backtesting,* Credit Technology **92**, 1 (2015).

[16] T. Schuermann, *What do we know about loss given default?* SSRN Electronic Journal , 1 (2011).

[17] B. Bade, D. Rösch, and H. Scheule, *Default and recovery risk dependencies in a simple credit risk model,* European Financial Management **17**, 120 (2011).

[18] E. I. Altman, A. C. Resti, and A. Sironi, *Recovery risk: The next challenge in credit risk management* (Risk Books, London, 2005).

[19] A. Resti and A. Sironi, *Risk Management and Shareholders' Value in Banking*, edited by A. Resti and A. Sironi (John Wiley Sons, Inc., Hoboken, NJ, USA, 2012).

[20] P. D. Sasieni and A. R. Brentnall, *Handbook of Epidemiology: Second Edition* (Springer, 2014) pp. 1195–1239.

[21] H. D. Khieu, D. J. Mullineaux, and H. C. Yi, *The determinants of bank loan recovery rates,* Journal of Banking and Finance **36**, 923 (2012).

[22] J. Frye, *The effects of systematic credit risk: A false sense of security,* Recovery Risk: The Next Challenge in Credit Risk Management , 187 (2005).

[23] F. Mac, *Single-family loan-level dataset general user guide March 2013*, Freddie Mac `http://www.freddiemac.com/fmac-resources/research/pdf/user_guide.pdf` (2013).

[24] N. Kiefer and N. Kiefer, *Economic duration data and hazard functions,* Journal of economic literature **26**, 646 (1988).

[25] M. Höchstötter and A. Nazemi, *Analysis of loss given default,* Investment Management and Financial Innovations **10**, 70 (2013).

[26] M. Gürtler and M. Hibbeln, *Improvements in loss given default forecasts for bank loans,* Journal of Banking and Finance **37**, 2354 (2013).

[27] O. Yashkir and Y. Yashkir, *Loss given default modeling: A comparative analysis,* Journal of Risk Model Validation **7**, 25 (2013).

[28] T. Hartmann-Wendels, P. Miller, and E. Töws, *Loss given default for leasing: Parametric and nonparametric estimations,* Journal of Banking and Finance **40**, 364 (2014).

[29] X. Huang and C. W. Oosterlee, *Generalized beta regression models for random loss given default,* Journal of Credit Risk **7**, 45 (2011).

[30] T. Bellotti and J. Crook, *Loss given default models incorporating macroeconomic variables for credit cards,* International Journal of Forecasting **28**, 171 (2012).

[31] G. Loterman, I. Brown, D. Martens, C. Mues, and B. Baesens, *Benchmarking regression algorithms for loss given default modeling,* International Journal of Forecasting **28**, 161 (2012).

[32] S. Krüger and D. Rösch, *Downturn LGD modeling using quantile regression,* Journal of Banking and Finance **79**, 42 (2017).

[33] R. Calabrese and M. Zenga, *Bank loan recovery rates: Measuring and nonparametric density estimation,* Journal of Banking and Finance **34**, 903 (2010).

[34] O. Renault and O. Scaillet, *On the way to recovery: A nonparametric bias free estimation of recovery rate densities,* Journal of Banking and Finance **28**, 2915 (2004).

[35] J. K. Im, D. W. Apley, C. Qi, and X. Shan, *A time-dependent proportional hazards survival model for credit risk analysis,* Journal of the Operational Research Society **63**, 306 (2012).

[36] G. Chawla, *Point-In-Time (PIT) LGD and EAD models for IFRS9 / CECL and stress testing,* Journal of Risk Management in Financial Institutions **9**, 1 (2016).

[37] S. Bonini and G. Caivano, *The survival analysis approach in basel II credit risk management: Modeling danger rates in the loss given default parameter,* Journal of Credit Risk **9**, 101 (2013).

[38] J. Witzany, M. Rychnovsky, and P. Charamza, *Survival analysis in LGD modeling,* SSRN Electronic Journal , 1 (2012).

[39] J. A. Bastos, *Forecasting bank loans loss-given-default,* Journal of Banking and Finance **34**, 2510 (2010).

[40] P. Muliere, P. Secchi, and S. G. Walker, *Urn schemes and reinforced random walks,* Stochastic Processes and their Applications **88**, 59 (2000).

[41] C. Robert, *Chance*, Vol. 27 (The MIT Press, Cambridge MA, 2014) pp. 62–63.

[42] B. De Finetti, *Community Care*, 1638 (Wiley, 2006) pp. 26–27.

[43] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools* (Princeton university press, 2005).

[44] H. M. Mahmoud, *Polya Urn Models* (CRC Press, Boca Raton, 2008) pp. 1–291.

[45] P. Cirillo, J. Hüsler, and P. Muliere, *A nonparametric urn-based approach to interacting failing systems with an application to credit risk modeling,* International Journal of Theoretical and Applied Finance **13**, 1223 (2010), arXiv:arXiv:1010.3586v1 .

[46] S. Peluso, A. Mira, and P. Muliere, *Reinforced urn processes for credit risk models,* Journal of Econometrics **184**, 1 (2015).

[47] M. Mezzetti, P. Muliere, and P. Bulla, *An application of reinforced urn processes to determining maximum tolerated dose,* Statistics and Probability Letters **77**, 740 (2007).

[48] E. I. Altman, B. Brady, A. Resti, and A. Sironi, *The link between default and recovery rates: Theory, empirical evidence, and implications,* Journal of Business **78**, 2203 (2005).

[49] S. Jackman, *Bayesian Analysis for the Social Sciences* (Wiley, New York, 2009) pp. 1–558.

[50] J. K. Kruschke, *Wiley Interdisciplinary Reviews: Cognitive Science*, 2nd ed., Vol. 1 (Chapman and Hall/CRC, 2010) pp. 658–676.

[51] X. L. Meng, *Multiple-imputation inferences with uncongenial sources of input,* Statistical Science **9**, 538 (1994).

[52] S. Caselli, S. Gatti, and F. Querci, *The sensitivity of the loss given default rate to systematic risk: New empirical evidence on bank loans,* **34**, 1.

[53] R. Jankowitsch, F. Nagler, and M. G. Subrahmanyam, *The determinants of recovery rates in the US corporate bond market,* Journal of Financial Economics **114**, 155 (2014).

[54] E. Ziegel, *Technometrics*, Handbook of Statistics, Vol. 48 (Elsevier Science, 2006) pp. 576–577.

[55] P. Diaconis and D. Freedman, *De Finetti's theorem for Markov chains,* The Annals of Probability **8**, 115 (2007).

[56] S. Walker and P. Muliere, *Beta-stacy processes and a generalization of the polya-urn scheme,* Annals of Statistics **25**, 1762 (1997).

3

MODELLING "WRONG-WAY RISK"

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.

3.1. 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 [2]. The model is able to learn from the data, and it improves its performances over time, compatibly with the machine/deep learning paradigm. Similarly to the recent construction of [3] 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 such a counterparty defaults, thus not fulfilling 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 [4, 5], a set of international standards developed by the Basel Committee on Banking Supervision (BCBS) 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.

Parts of this chapter have been published in D. Cheng and P. Cirillo, An urn-based nonparametric modeling of the dependence between pd and lgd with an application to mortgages, Risks 7, 76 (2019) [1].

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 [6, 7].

Surprisingly, in most theoretical models in the literature, 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 [8–12]. Borrowing from the terminology developed in the field of credit valuation adjustment (CVA) [13], 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 with the deterioration of its creditworthiness. Ignoring this WWR can easily lead to an unreliable estimation of credit risk for a given counterparty [12].

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 market [14], with a consequent drop in estate prices, and thus in the recovery rates of the defaulted exposures [13]. 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, [15] introduced the first structural model of default, developing what we can consider the Black and Scholes model for credit risk. In that model, the recovery rate RR (with $RR = 1 - LGD$), conditionally on the credit event, follows a lognormal distribution [16], and it is negatively correlated with the PD. While not consistent with later empirical findings [17, 18] 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 [19], [20], [21], [22] and [23]. During the 1990's, on the wake of the–at the time–forthcoming financial regulations (Basel I and later Basel II), several new approaches were introduced, like the so-called reduced-form family [24–26] and the VaR methodology [27–29]. 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 [27], 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, hence linear dependence–between PD and LGD. For example, [30] proposed a standard one-factor model, becoming a pivotal reference in the modeling of the WWR between PD and LGD. The paper focuses 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.

[12] moved forward, proposing a two-factor model on retail data, looking at the over-

all economic environment as the source of dependence between PD and LGD. Then, for the LGD only, a second component accounts for the conditions of the economy during the workout process. Other meaningful constructions in the common factor framework are [31] and [32]. The latter is also interesting for the literature review, together with [17].

Using a two-state latent variable construction, [33] 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 the discussions in [8] and [34], and the references therein.

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 [3, 35–37], and more in general in the Bayesian modeling of credit risk, see for example [38–42].

For the reader's convenience, we here summarise the main findings of the paper:

- An intuitive bivariate model is proposed for the joint modeling of PD and LGD. The construction exploits the power of Polya urns to generate a Bayesian nonparametric approach to wrong-way risk. The model can be interpreted as a mixture model, following the typical credit risk management classification [43].

- The proposed model is able to combine prior beliefs with empirical evidence and, exploiting the reinforcement mechanism embedded in Polya urns, it learns, thus improving its performances over time.

- The ability of learning and improving gives the model a machine/deep learning flavour. However, differently from the common machine/deep learning approaches, the behavior of the new model can be controlled and studied in a rigorous way from a probabilistic point of view. In other words, the common "black box" argument [44] of machine/deep learning does not apply.

- The possibility of eliciting an a priori allows for the exploitation of experts' judgements, which can be extremely useful when dealing with rare events, historical bias and data problems in general [3, 45, 46].

- The model we propose can only deal with positive dependence. Given the empirical literature, this is not a problem in WWR modeling, however it is important to be aware of this feature, if other applications are considered.

The paper develops as follows: Section 3.2 is devoted to the description of the theoretical framework and the introduction of all the necessary probabilistic tools. In Section 3.3 we briefly describe the Freddie Mac data we then use in Section 3.4, where we show how to use and the performances of the model, when studying the dependence between PD and LGD for residential US mortgages.

3.2. MODEL

The main idea of the model we propose was first presented in [47], in the field of survival analysis, and it is based on powerful tools of Bayesian nonparametrics, like the beta-Stacy process of [2]. Here we give a different representation in terms of Reinforced Urn Processes (RUP), to bridge towards some recent papers in the credit risk management literature like [37] and [3].

3.2.1. THE TWO-COLOR RUP

A RUP is a combinatorial stochastic process, first introduced in [48]. 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 [48, 49] and [50]. In this paper we need to specify a RUP able to generate a discrete beta-Stacy process, a particular random distribution over the space of discrete distributions.

Definition 4 (Discrete beta-Stacy Process [2]) *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 (α_j, β_j), such that the random mass assigned by F to $\{j\}$, written $F(\{j\})$, is given by $V_j \prod_{i<j}(1-V_i)$.*

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 [51]. 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 $n = 0, 1, 2, \ldots$ 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 at time $n-1$, 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 $\{0, 1, \ldots, k\}$, it is sufficient to set to zero the number of blue balls in all urns $U(k), U(k+1), U(k+2)$ and so on. For the rest, the mechanism stays unchanged.

It is not difficult to verify that, if all urns $U(j)$, $j = 1, 2, \ldots$, 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

$$\{Z_0, Z_1, Z_2, Z_3, Z_4, Z_5, Z_6, ..., Z_{10}, Z_{11}, Z_{12}, ..., Z_{15}, ...\} = \{\underbrace{0,1,2}_{Block\,1}, \underbrace{0,1,2,3,...,7}_{Block\,2}, \underbrace{0,1,...,4}_{Block\,3}, ...\}. \tag{3.1}$$

These sequences are easily split into blocks, each starting with 0, as in Equation (3.1). In terms of sampling, those 0's represent the times the process has been reset 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 $n = 15$. [48] have shown that the blocks–or the 0-blocks in their terminology– 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 instance $P(Block\,1, Block\,2, Block\,3) = P(Block\,1, Block\,3, Block\,2)$ in Equation (3.1). 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 [52].

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 [48] tells us that such a random distribution is a beta-Stacy process with parameters $\{\alpha_j, \beta_j\}_{j\in\mathbb{N}_0}$, where α_j and β_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 $U(j)$ [51].

Let $B_1, B_2, ..., B_m$ be the first m blocks generated by a two-color RUP $\{Z_n\}_{n\geq 0}$. With T_i we indicate the last state visited by $\{Z_n\}$ within block i. Coming back to Equation (3.1), 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

$$P[T_1 = j] = \frac{\alpha_j}{\alpha_j + \beta_j} \prod_{i=0}^{j-1} \frac{\beta_i}{\alpha_i + \beta_i}, \tag{3.2}$$

and, for $m \geq 1$,

$$P[T_{m+1} = j | T_1, T_2, ..., T_m] = \frac{\alpha_j + r_j}{\alpha_j + \beta_j + r_j + s_j} \prod_{i=0}^{j-1} \frac{\beta_i + s_i}{\alpha_i + \beta_i + r_i + s_i}, \tag{3.3}$$

while

$$P[T_{m+1} \geq j | T_1, T_2, ..., T_m] = \prod_{i=0}^{j} \frac{\beta_i + s_i}{\alpha_i + \beta_i + r_j + s_i}, \tag{3.4}$$

where $r_j = \sum_{i=1}^{m} 1_{\{T_i = j\}}$ and $s_j = \sum_{i=1}^{m} 1_{\{T_i > j\}}$. From Equations (3.3) and (3.4) we see 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 $\{\alpha_j, \beta_j\}_{j \in \mathbb{N}_0}$, with the additional balls that are introduced in the system, to obtain the predictives in Equations (3.3) and (3.4).

Regarding the initial knowledge, notice that, when we choose the quantities α_j and β_j, for $j = 0, 1, 2, ...$, from a Bayesian point of view we are eliciting a prior. In fact, by setting

$$\alpha_j = c_j(G(\{j\}) \qquad \text{and} \qquad \beta_j = c_j \left(1 - \sum_{i=0}^{j} (G(\{i\})\right), \qquad j \in \mathbb{N}_0, \tag{3.5}$$

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 [48, 49].

3.2.2. MODELING DEPENDENCE

Consider a portfolio $\mathscr{P}$ containing m exposures. For $i = 1, ..., m$, 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 [3] propose in their work. In other terms, one can split the PD and the LGD into $l = 0, ..., L$ 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 3.4, convenient ways of defining levels are through quantiles, via rounding and, when available, thanks to experts' judgements.

As observed in [17], 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 $\{0, 1, ..., L\}$

Let $\{A_i\}_{i=1}^{m}$, $\{B_i\}_{i=1}^{m}$ and $\{C_i\}_{i=1}^{m}$ be three independent LS sequences generated by three independent two-color RUPs $\{Z_j^A\}_{j \geq 0}$, $\{Z_j^B\}_{j \geq 0}$, $\{Z_j^C\}_{j \geq 0}$. 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 $\{\alpha_j^A, \beta_j^A\}_{j \in \mathbb{N}_0}$. Similarly, for $\{B_i\}_{i=1}^{m}$ and $\{C_i\}_{i=1}^{m}$ we have F_B with $\{\alpha_j^B, \beta_j^B\}_{j \in \mathbb{N}_0}$, and F_C with $\{\alpha_j^C, \beta_j^C\}_{j \in \mathbb{N}_0}$.

Now, as in [53], let us assume that, for each exposure $i = 1, ..., m$, we have

$$\begin{aligned} X_i &= A_i + B_i, \\ Y_i &= A_i + C_i. \end{aligned} \tag{3.6}$$

This construction builds a special dependence between the discretised PD and the discretised LGD: we are indeed assuming that for each counterparty i there exists a common factor A_i influencing both, while B_i and C_i can be seen as idiosyncratic components. Observe that, 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 $\{Z_j^A\}_{j\geq 0}$, $\{Z_j^B\}_{j\geq 0}$, $\{Z_j^C\}_{j\geq 0}$ can be tuned so that it is not possible to observe values larger than 100%.

From Equation (3.6) we can derive important features of the model we are proposing. Since we can write $Y_i = X_i - B_i + C_n$, it is clear that we are assuming a linear dependence between X_i and Y_i. This is compatible with several empirical findings, like [17] or [11].

Furthermore, given Equation (3.6) 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

$$\begin{aligned} \text{Cov}(X_1, Y_1) &= \text{Var}(A_1) \geq 0, \\ \text{Cov}(X_{m+1}, Y_{m+1} | \mathbf{A}_m, \mathbf{B}_m, \mathbf{C}_m) &= \text{Var}(A_{m+1} | \mathbf{A}_m) \geq 0, \text{ for } m \geq 2, \end{aligned} \tag{3.7}$$

where Var is the variance, Cov the covariance, $\mathbf{A}_m = [A_1, \ldots, A_m]$, and similarly $\mathbf{B}_m$, $\mathbf{C}_m$. Therefore, with the bivariate urn construction 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 3.1. However, it is important to stress that the model in Equation (3.6) cannot be used when negative dependence is possible.

Always from Equation (3.6), 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 $\mathscr{P}$ are exchangeable. As observed in [43], exchangeability is a common assumption in credit risk, for 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 [43, 55].

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 [47].

[1] In reality, as observed in [54], 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.

[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.

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

$$\begin{aligned} F_X &= F_A \times F_B, \\ F_Y &= F_A \times F_C, \end{aligned}$$

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

$$\mathrm{Cov}_{F_{XY}}(X,Y) = \mathrm{Var}_{F_A}(A) = \sigma_A^2. \tag{3.8}$$

3

Furthermore, if P is the probability function corresponding to F, one has

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

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 (3.6), together with the properties of the beta-Stacy processes involved, allows for a nice derivation of the predictive distribution for a new exposure (X_{m+1},Y_{m+1}), given the observed couples $(\mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m)$. This can be extremely useful in applications, when one is interested in making inference about the PD, the LGD and their relation.

In fact

$$P[X_{m+1}=x, Y_{m+1}=y | \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m] = \frac{P[X_{m+1}=x, Y_{m+1}=y, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]}{P[\mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]}. \tag{3.9}$$

Given Equation (3.6), Equation (3.9) can be rewritten as follows

$$\begin{aligned} & P[X_{m+1}=x, Y_{m+1}=y | \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m] \\ = & \sum_{\mathbf{a}_m} P[X_{m+1}=x, Y_{m+1}=y \mid \mathbf{A}_m = \mathbf{a}_m, \mathbf{B}_m = \mathbf{b}_m, \mathbf{C}_m = \mathbf{c}_m] \\ & \times P[\mathbf{A}_m = \mathbf{a}_m \mid \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m], \end{aligned} \tag{3.10}$$

where $\mathbf{b}_m = \mathbf{x}_m - \mathbf{a}_m$ and $\mathbf{c}_m = \mathbf{y}_m - \mathbf{a}_m$.

From a theoretical point of view, computing Equation (3.10) just requires to count the balls in the urns behind $\{A_i\}_{i=1}^m$, $\{B_i\}_{i=1}^m$ and $\{C_i\}_{i=1}^m$, and then to use formulas like those in Equations (3.2) and (3.3); 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 $\mathbf{X}_m = \mathbf{x}_m$ and $\mathbf{Y}_m = \mathbf{y}_m$, the sequence $\mathbf{A}_m = (A_1, \ldots, A_m)$ is generated via a Gibbs sampling. The full conditional of A_m, $P[A_m = a_m \mid \mathbf{A}_{m-1} = \mathbf{a}_{m-1}, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]$, is such that

$$\begin{aligned} & P[A_m = a_m \mid \mathbf{A}_{m-1} = \mathbf{a}_{m-1}, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m] \\ \propto \; & P[A_m = a_m \mid \mathbf{A}_{m-1} = \mathbf{a}_{m-1}] \\ & \times P[X_m - A_m = x_m - a_m \mid \mathbf{B}_{m-1} = \mathbf{b}_{m-1}] \\ & \times P[Y_m - A_m = y_m - a_m \mid \mathbf{C}_{m-1} = \mathbf{c}_{m-1}]. \end{aligned}$$

Since $\{A_j\}_{j=1}^m$ is exchangeable, all the other full conditionals, $P[A_j = a_j \mid \mathbf{A}_{-j} = \mathbf{a}_{-j}, \mathbf{X}_m = \mathbf{x}_m, \mathbf{Y}_m = \mathbf{y}_m]$, where $\mathbf{A}_{-j} = (A_1, ..., A_{j-1}, A_{j+1}, ..., A_m)$, have an analogous form.

2. Once $\mathbf{A}_m$ is obtained, compute $\mathbf{B}_m = \mathbf{X}_m - \mathbf{A}_m$ and $\mathbf{C}_m = \mathbf{Y}_m - \mathbf{A}_m$.

3. The quantities A_{m+1}, B_{m+1}, and C_{m+1} are then sampled according to their beta-Stacy predictive distributions $P(A_{m+1} \mid \mathbf{A}_m)$, $P(B_{m+1} \mid \mathbf{B}_m)$, and $P(C_{m+1} \mid \mathbf{C}_m)$ as per Equation (3.3).

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

3.3. DATA

We want to show the potentialities of the bivariate urn construction of Section 3.2 in modeling the dependence between PD and LGD in a large portfolio of residential mortgages. The data we use come from [34].

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 [56].

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.

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 [56]. A loan is considered defaulted when it is delinquent for more than 180 days, even if it is later repurchased [57]. 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 [3]. 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 2.14×10^{-5} and 0.3723 for PD, and 0 and 0.5361 for LGD.

[3] To avoid any copyright problem with Freddie Mac, which already freely shares its data online, from Maio's Freddie datastat we only provide the PD and the LGD estimates, together with the unique alphanumeric identifier. In this way, merging the data sources is straightforward.

In the parametric approach proposed by [34], 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, [34] proposes the use of two Weibull accelerated failure time (AFT) models, following a recent trend in modern credit risk management [58]. 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 refer to [34].

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 3.1. 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 [34]. In the original Freddie Mac's sample it ranges from 301 to 850.

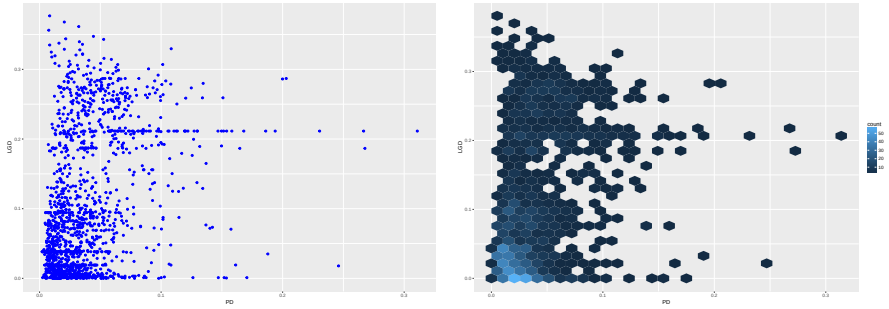
Following a common classification [59], we can define five classes of creditworthiness: Very poor, for a FICO below 579; Fair, for 580-669; Good, for 670-739; Very good, for 740-799; and Exceptional, with a score above 800. Table 3.1 contains the number of loans in the different classes, the corresponding average PD and LGD, and naturally their correlation ρ. 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 [14]. 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).

Table 3.1: 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	ρ
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 3.1 shows two plots of the relation between PD and LGD for

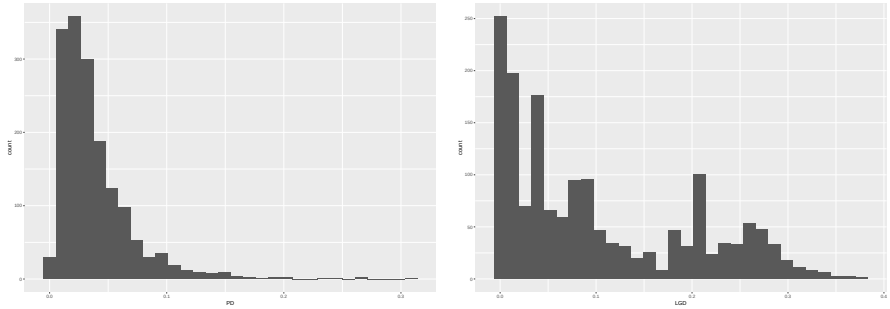
the "Very poor" class. On the left, a simple scatter plot of PD vs LGD. On the right, to deal with the large number of overlapping points, thus improving interpretability, we provide a hexagonal heatmap with counts. To obtain this plot, the plane is divided into regular hexagons (20 for each dimension),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



(a) Scatter plot (b) Hexagonal heatmap

Figure 3.1: 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 3.2, 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.



(a) PD (b) LGD

Figure 3.2: Histograms of the marginal distributions of PD and LGD in the "Very poor" rating class.

3.4. RESULTS

In this section we discuss the performances of the bivariate urn model on the mortgage data described in Section 3.3. For the sake of space, we show the results for "Very poor"

and the "Exceptional" FICO score classes, as per Table 3.1.

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 (3.6).

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

Please notice that our purpose is to show that the bivariate urn model actually works. The present section has no ambition of being a complete empirical study on the PD-LGD dependence in the Freddie Mac's or Maio's datasets.

3.4.1. DISCRETISATION

In order to discretise the PD and LGD into X and Y, it is necessary to choose the appropriate levels $l = 0, 1, ..., L$. 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 [3].

For example, let $d_1^v, ..., d_9^v$ be the deciles for the quantity $v \in \{PD, LGD\}$. We can set levels $l = 0, 1, ..., 10$, where

$$L1 := \{0 : 0;\ 1 : (0, d_1^v];\ 2 : (d_1^v, d_2^v];\ ...;\ 9 : (d_8^v, d_9^v];\ 10 :> d_9^v\}. \tag{3.11}$$

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, $d_1^{PD} = 0.0099$, while $d_1^{LGD} = 0.0030$. 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 [60]. However, differently from copulas, the dependence between X and Y is not restricted to any particular parametric form (the copula function): 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:

$$L2 := \{0 : 0;\ 1 : (0, 1\%];\ 2 : (1, 2\%];\ ...;\ 100 : (99, 100\%]\}. \tag{3.12}$$

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. Moreover, as already observed, equally-spaced intervals are easier to implement and often to interpret, but again this is not a necessity: levels can represent intervals of different size.

Notice that, if correctly applied, discretisation maintains the dependence structure between the variables. For the "Very poor" class, using the levels in Equation (3.11) in defining X and Y, we find that $cor(X, Y) = 0.3348$, in line with the value 0.3370 in Table 3.1. For the "Exceptional" group, using the levels in Equation (3.12), we obtain 0.2080, $still comparable$.

In what follows, we discuss the results mainly using the levels in Equation (3.12). 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 fully exploit the Bayesian learning mechanism. To improve fitting, rarely visited levels should be aggregated. Once again, experts' judgements could represent a possible solution. We refer to [3] for further discussion on level setting.

3.4.2. 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 for any computation and to initialise the Markov Chain Monte Carlo algorithm discussed in Subsection 3.2.2.

Prior elicitation can be performed 1) in a completely subjective way, on the basis of one's knowledge of the phenomenon under scrutiny, 2) by just looking at the data in a fully empirical approach, or 3) by combining data and beliefs, as suggested for example by [61]. In reality, a fully data-driven approach based on the sole use of the empirical distribution functions, as for example the one suggested in [3], is not advisable for the bivariate urn construction. While there is no problem in observing X and Y, thus exploiting their empirical distributions for prior elicitation, it is not immediate to do the same for the unobserved quantity A, necessary to obtain both B and C. A possibility could be to use the empirical Kendall's function [62], but then one would end up with a model not different from the standard empirical copula approach [63], especially if the number of observations used for the definition of the a priori is large. A compromise could thus be to elicit a subjective prior for A only, and to combine this information with the empirical distributions of X and Y, so to obtain B and C. As common in Bayesian nonparametrics [64], no unique best way exists: everything depends on personal preferences–this is the unavoidable subjectivity of every statistical model [65]–and on exogenous constraints that, in credit risk management, are usually represented by the actual regulatory framework [4, 13].

Recalling Equation (3.8), 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 (3.11). For the situation in Equation (3.12), conversely, we have a covariance of 10.28 for the "Very poor" class, and 1.46 for the other. Apart from the reasonable constraint on the variance, 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 (3.11), 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 (3.12) and recalling Figure 3.1, 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. Here we discuss two possibilities:

- 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 $\sigma_A^2 = Cov(X,Y)$. 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 $A \sim Poi(\lambda_A = Cov(X,Y))$, while for B and C one sets $Poi(\bar{X} - \lambda_A)$ and $Poi(\bar{Y} - \lambda_A)$, where $\bar{X}$ is the empirical mean of X. This guarantees, for example, that $X \sim Poi(\bar{X})$. Recall in fact that in a Poisson random variable the mean and the variance are both equal to the intensity parameter, and that independent Poissons are closed under convolution. 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.

Notice that the possibility of eliciting an a priori is an extremely useful feature of the every urn construction [3, 35, 36]. In fact, a good prior–when available–can compensate for the lack of information in the data, and it can effectively deal with extremes and rare events, a relevant problem in modern risk management [13, 66, 67]. For example, if an expert believes that her data under-represent a given phenomenon, like for example some unusual combinations of PD and LGD, she could easily solve the problem by choosing a prior putting a relevant mass on those combinations, so that the posterior distribution will always take into account the possibility of those events, at least remotely. This can clearly correct for the common problem of historical bias [45, 46].

Once the priors have been decided, the urn compositions can be derived via Equation (3.5). Different values of c_j have been 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. [3] 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

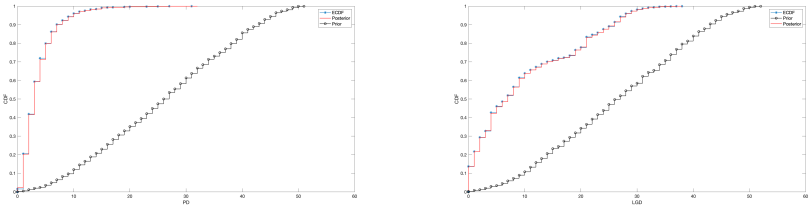
that a strong prior has a clear impact for the "Very poor" rating class ("only" 1627 observations), while no appreciable effect is observable for the "Exceptional" one (32403 data points).

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 [53], 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 [68]–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 possibility further props the flexibility of the bivariate urn model up.

3.4.3. FITTING

Figure 3.3 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 (3.12), 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).

It is worth noticing that, qualitatively, the results we discuss here are robust to different choices of the levels, until a sufficient number of observations is available for the update of the urns and hence learning, as already observed at the end of Subsection 3.4.1. For example we have tested the partitions $\{0:0;\ 1:(0,2\%];\ 2:(2,4\%];\ \ldots;\ 50:(98,100\%]\}$ and $\{0:0;\ 1:(0,5\%];\ 2:(5,10\%];\ \ldots;\ 20:(95,100\%]\}$: all findings were consistent to what we see in these pages. An example of partition which does not work is conversely the one based on increments equal to 0.1% per level, which proves to be too fine, so that several urns are never updated and the posteriors do not pass the relative KS tests.

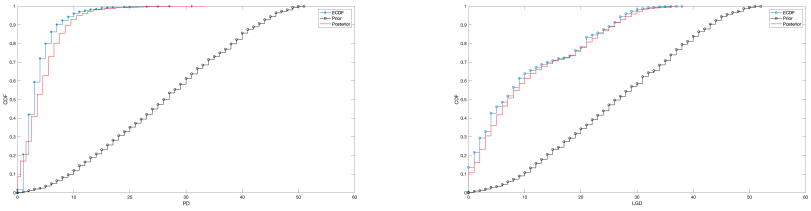


(a) PD (b) LGD

Figure 3.3: Prior, posterior and ecdf for the discretised PD and LGD in the "Very poor" rating class, when priors are uniform and levels are like those in Equation (3.12). The strength of belief is always set to 1.

Figure 3.4 is generated in the same way of Figure 3.3. 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, thus correcting for historical bias, or she wants to incorporate specific knowledge about future trends.



(a) PD (b) LGD

Figure 3.4: Prior, posterior and ecdf for the discretised PD and LGD in the "Very poor" rating class, when priors are uniform and levels are like those in Equation (3.12). The strength of belief is always set to 100.

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 3.5 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 (one needs $c_j > 500$ to see some difference). Similar results hold for the LGD. In producing the figure, A has a uniform prior on $[0,3]$, while B and C on $[0,45]$.

Figure 3.6 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 (3.12). Figure 3.7 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, with more peaks, bumps and holes. In Figure 3.8 the equivalent of Figure 3.6 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 small. 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 initial 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.

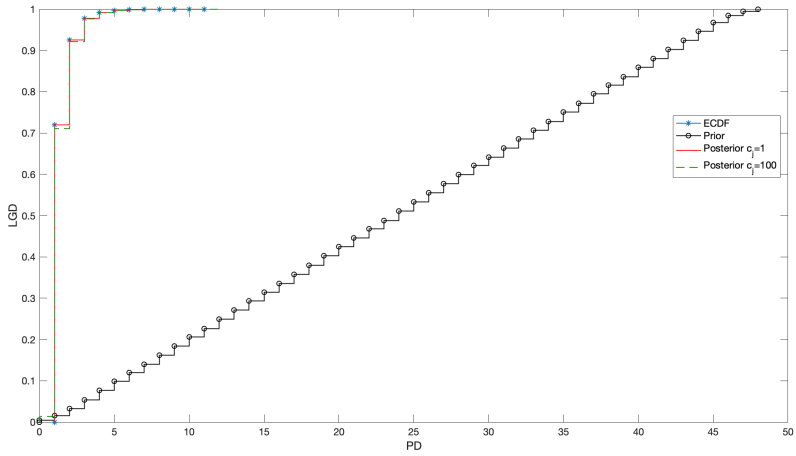


Figure 3.5: Prior, posteriors (c_j = 1 and c_j=100) and ecdf for the discretised PD in the "Exceptional" rating class, when priors are uniform.

5.88 and 22.50. This is a clear signal that more data would be necessary to properly move away from the wrong prior beliefs, forgetting the Poisson nature.

In the case of Figure 3.6, 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: we have indeed used all the observations available to verify how the model fits the data, and no out of sample performance was checked. Using the Bayesian terminology of [69] and [70], we have therefore 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 3.9 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

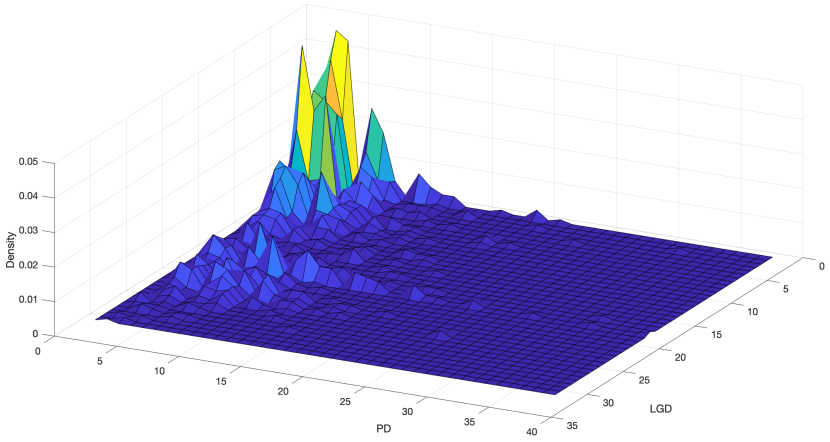


Figure 3.6: Bivariate density distribution of PD and LGD in the "Very poor" rating class, with Poisson priors and strength of belief equal to 1.

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. In Table 3.2 we show the PD and correlation results for all classes under uniform priors, while Table 3.3 deals with the LGD.

Table 3.2: PD values (%) and correlation. Some descriptive descriptive statistics (mean, median, standard deviation, correlation for the joint) for the predictive distribution (P) and the validation set (V) for the different FICO classes, under uniform priors.

Class	$mean_P$	$mean_V$	$median_P$	$median_V$	SD_P	SD_V	ρ_P	ρ_V
Very Poor	3.85	3.81	3.02	3.03	3.24	2.53	0.34	0.20
Fair	2.41	2.32	1.50	1.44	2.36	2.33	0.45	0.51
Good	1.43	0.49	0.83	0.25	0.71	0.66	0.34	0.38
Very good	0.36	0.32	0.62	0.64	0.14	0.16	0.38	0.42
Exceptional	0.16	0.18	0.60	0.54	0.82	0.83	0.19	0.18

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 not observed in Maio's dataset–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.

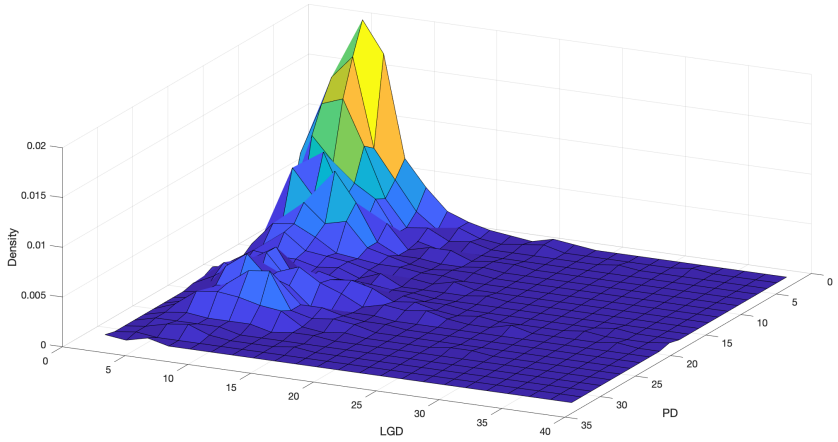


Figure 3.7: Bivariate density distribution of PD and LGD in the "Very poor" rating class, with Poisson priors and strength of belief equal to 100.

Table 3.3: LGD values (%). Some descriptive descriptive statistics (mean, median, standard deviation) for the predictive distribution (P) and the validation set (V) for the different FICO classes, under uniform priors.

Class	$mean_P$	$mean_V$	$median_P$	$median_V$	SD_P	SD_V
Very Poor	9.82	10.2	6.63	6.14	9.51	8.32
Fair	12.1	11.5	9.43	10.6	9.69	5.53
Good	13.8	19.8	14.5	19.2	5.70	5.26
Very good	15.6	18.5	18.7	18.0	5.27	5.13
Exceptional	17.8	17.2	20.4	19.7	8.00	4.99

3.4.4. WHAT ABOUT THE CRISIS?

In line with findings in the literature [12, 14, 71, 72], Maio's dataset–as well as the original one by Freddie Mac–shows that the financial crisis of 2007-2008 had a clear impact on the dependence between PD and LGD. In particular an increase in the strength of correlation is observable during the the crisis and in the years after. In fact we can observe an overall value of 0.11 before 2008, which grows up to 0.24 in 2011 and remains pretty stable after. A compatible behavior is observed for the different FICO classes.

When evaluated in-sample, the bivariate urn model, given the large number of observations is always able to perform satisfactorily: this holds true for example when we restrict our attention on the intervals [2002 – 2007], [2007-2010] or [2010-2016]. Performances are conversely not good when the model trained on pre-crisis data is used to predict the post-crisis period, given the substantial difference in the strength of the dependence, which is persistently underestimated. Clearly we are speaking about those situations in which the prior we elicited were not taking into consideration a strong in-

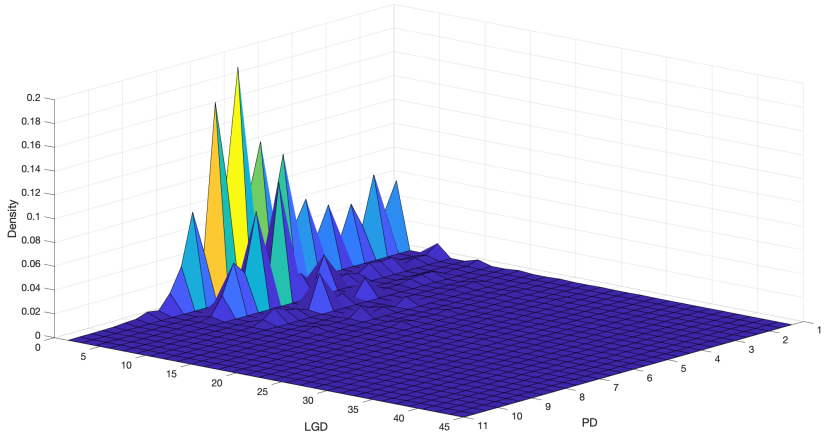


Figure 3.8: Bivariate density distribution of PD and LGD in the "Exceptional" rating class, with Poisson priors and strength of belief equal to 1.

crease in correlation. An ad-hoc modification of the support of A could represent a viable solution, provided we could be aware of this ex ante, and not just ex post.

An interesting thing, worth mentioning, happens when we focus our attention on the loans originated (not defaulted) during the crisis and follow them. Here the relation between PD and LGD tends to 0 or, for some FICO classes, it becomes slightly negative, probably because of the stricter selection the scared banks likely imposed on applicants, asking for more collateral and guarantees. While the model is still able of dealing with no dependence, a negative correlation cannot be studied.

REFERENCES

[1] D. Cheng and P. Cirillo, *An Urn-based nonparametric modeling of the dependence between PD and LGD with an application to mortgages,* Risks **7**, 76 (2019).

[2] S. Walker and P. Muliere, *Beta-stacy processes and a generalization of the polya-urn scheme,* Annals of Statistics **25**, 1762 (1997).

[3] D. Cheng and P. Cirillo, *A reinforced urn process modeling of recovery rates and recovery times,* Journal of Banking and Finance **96**, 1 (2018).

[4] The Basel Committe, *Principles for the management of credit risk,* IFAS Extension , 1 (2000).

[5] Bank for International Settlements, *Basel III: A global regulatory framework for more resilient banks and banking systems,* in *Bank for International Settlements*, Vol. 2010 (2010) p. 69.

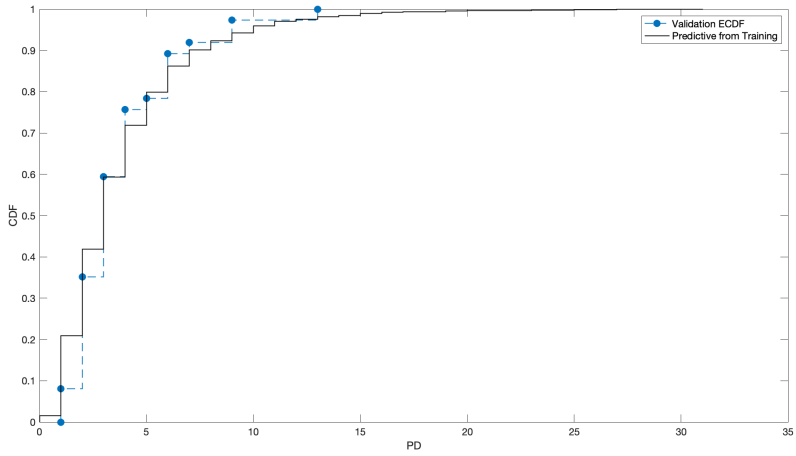


Figure 3.9: Predictive distribution generated by the bivariate urn model for the "Very poor" class when trained on the training sample (1590 loans), against the empirical ecdf from the validation set (37 loans).

[6] R. Douglas Fields, J. A. Black, and S. G. Waxman, *Filipin-cholesterol binding in CNS axons prior to myelination: Evidence for microheterogeneity in premyelinated axolemma,* Brain Research **404**, 21 (1987).

[7] BCBS, *International convergence of capital measurement and capital standards for banks,* in *Reserve Bank of New Zealand Bulletin*, Vol. 52 (1989) p. 285.

[8] E. I. Altman, *Default recovery rates and LGD in credit risk modelling and practice,* The Oxford Handbook of Credit Derivatives (2012), 10.1093/oxfordhb/9780199546787.013.0003.

[9] E. I. Altman, A. Resti, and A. Sironi, *Analyzing and explaining default recovery rates,* The International Swaps Dealers Association (2001).

[10] J. Frye, *The effects of systematic credit risk: A false sense of security,* Recovery Risk: The Next Challenge in Credit Risk Management , 187 (2005).

[11] P. Miu and B. Ozdemir, *Basel requirement of downturn LGD: modeling and estimating PD LGD correlations,* Journal of Credit Risk **2**, 43 (2006).

[12] J. Witzany, *A two-factor model for PD and LGD correlation,* SSRN Electronic Journal **18** (2011), 10.2139/ssrn.1476305.

[13] J. C. Hull, *Risk management and financial institutions*, 4th ed. (Wiley, New York, 2012).

[14] B. Eichengreen, A. Mody, M. Nedeljkovic, and L. Sarno, *How the Subprime Crisis went global: Evidence from bank credit default swap spreads,* Journal of International Money and Finance **31**, 1299 (2012).

[15] R. C. Merton, *On the pricing of corporate debt: The risk structure of interest rates,* The Journal of Finance **29**, 449 (1974).

[16] A. Resti and A. Sironi, *Risk Management and Shareholders' Value in Banking*, edited by A. Resti and A. Sironi (John Wiley Sons, Inc., Hoboken, NJ, USA, 2012).

[17] E. I. Altman, B. Brady, A. Resti, and A. Sironi, *The link between default and recovery rates: Theory, empirical evidence, and implications,* Journal of Business **78**, 2203 (2005).

[18] E. P. JONES, S. P. MASON, and E. ROSENFELD, *Contingent claims analysis of corporate capital structures: An empirical investigation,* The Journal of Finance **39**, 611 (1984).

[19] R. Geske and H. E. Johnson, *The valuation of corporate liabilities as compound options: A correction,* The Journal of Financial and Quantitative Analysis **19**, 231 (1984).

[20] I. J. Kim, K. Ramaswamy, and S. Sundaresan, *Does default risk in coupons affect the valuation of corporate bonds?: A contingent claims model,* Financial Management **22**, 117 (1993).

[21] F. A. LONGSTAFF and E. S. SCHWARTZ, *A simple approach to valuing risky fixed and floating rate debt,* The Journal of Finance **50**, 789 (1995).

[22] L. Nielsen, J. Saá-Requejo, and P. Santa-Clara, *unpublished, University of California, Los Angeles* (INSEAD, Paris, 1993).

[23] O. A. Vasicek, *Credit valuation,* (2015).

[24] D. Duffie, *Graduate School of Business, Stanford . . .*, Tech. Rep. (Graduate School of Business, Stanford University, 1998).

[25] D. Duffie and K. J. Singleton, *Modeling term structures of defaultable bonds,* Review of Financial Studies **12**, 687 (1999).

[26] D. Lando, *On cox processes and credit risky securities,* Review of Derivatives Research **2**, 99 (1998).

[27] C. S. F. Boston, *Credit Suisse First Boston*, Tech. Rep. (1997).

[28] J. P. Morgan, *Creditmetrics - Technical Document,* (1997).

[29] T. C. Wilson, *Portfolio credit risk,* SSRN Electronic Journal **4**, 71 (2011).

[30] J. Frye, *Depressing recoveries,* Policy Studies **Emerging I**, 108 (2000).

[31] A. Hamerle, M. Knapp, and N. Wildenauer, *Modelling loss given default: A "point in time"-approach,* in *The Basel II Risk Parameters: Estimation, Validation, and Stress Testing*, edited by B. Engelmann and R. Rauhmeier (Springer, Berlin, 2006) pp. 127–142.

[32] X. Yao, J. Crook, and G. Andreeva, *Is it obligor or instrument that explains recovery rate: Evidence from US corporate bond,* Journal of Financial Stability **28**, 1 (2017).

[33] M. Bruche and C. Gonzalez-Aguado, *Recovery rates, default probabilities, and the credit cycle,* Journal of Banking & Finance **34**, 754 (2010).

[34] V. Maio, *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 (2017).

[35] E. Amerio, P. Muliere, and P. Secchi, *Reinforced urn processes for modeling credit default distributions,* International Journal of Theoretical and Applied Finance **7**, 407 (2004).

[36] P. Cirillo, J. Hüsler, and P. Muliere, *A nonparametric urn-based approach to interacting failing systems with an application to credit risk modeling,* International Journal of Theoretical and Applied Finance **13**, 1223 (2010), arXiv:arXiv:1010.3586v1 .

[37] S. Peluso, A. Mira, and P. Muliere, *Reinforced urn processes for credit risk models,* Journal of Econometrics **184**, 1 (2015).

[38] B. Baesens, D. Roesch, and H. Scheule, *Credit risk analytics: Measurement techniques, applications, and examples in SAS* (Wiley, 2016).

[39] P. Giudici, *Bayesian data mining, with application to benchmarking and credit scoring,* Applied Stochastic Models in Business and Industry **17**, 69 (2001).

[40] P. Giudici, M. Mezzetti, and P. Muliere, *Mixtures of products of Dirichlet process for variable selection in survival analysis,* Journal of Statistical Planning and Inference **111**, 101 (2003).

[41] P. Cerchiello and P. Giudici, *Bayesian credit rating assessment,* Communications in Statistics: Theory and Methods **111**, 101 (2014).

[42] Y. Fong, H. Rue, and J. Wakefield, *Bayesian inference for generalized linear mixed models,* Biostatistics **11**, 397 (2010).

[43] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools* (Princeton university press, 2005).

[44] W. Knight, *The dark secret at the heart of AI,* Technology Review **120**, 54 (2017).

[45] J. Derbyshire, *The siren call of probability: Dangers associated with using probability for consideration of the future,* Futures **88**, 43 (2017).

[46] H. Dickson and G. L. S. Shackle, *Ekonomisk Tidskrift*, Vol. 58 (Cambridge University Press, Cambridge, 1956) p. 250.

[47] P. Bulla, *Application of reinforced urn processes to survival analysis*, Ph.D. thesis, Bocconi University (2005).

[48] P. Muliere, P. Secchi, and S. G. Walker, *Urn schemes and reinforced random walks,* Stochastic Processes and their Applications **88**, 59 (2000).

[49] P. Muliere, P. Secchi, and S. G. Walker, *Reinforced random processes in continuous time,* Stochastic Processes and their Applications **104**, 117 (2003).

[50] S. Fortini and S. Petrone, *Hierarchical reinforced urn processes,* Statistics and Probability Letters **82**, 1521 (2012).

[51] H. M. Mahmoud, *Polya Urn Models* (CRC Press, Boca Raton, 2008) pp. 1–291.

[52] P. Diaconis and D. Freedman, *De Finetti's theorem for Markov chains,* The Annals of Probability **8**, 115 (2007).

[53] P. Bulla, P. Muliere, and S. Walker, *Bayesian nonparametric estimation of a bivariate survival function,* Statistica Sinica **17**, 427 (2007).

[54] J. Zhang and L. C. Thomas, *Comparisons of linear regression and survival analysis using single and mixture distributions approaches in modelling LGD,* International Journal of Forecasting **28**, 204 (2012).

[55] A. Corelli, *Credit risk,* Understanding Financial Risk Management , 239 (2020).

[56] F. Mac, *Single-family loan-level dataset general user guide March 2013*, Freddie Mac `http://www.freddiemac.com/fmac-resources/research/pdf/user_guide.pdf` (2013).

[57] Freddie Mac, *Single family loan-level dataset summary statistics December 2015*, Freddie Mac `http://www.freddiemac.com/fmac-resources/research/pdf/summary_statistics.pdf`.

[58] B. Narain, *Survival analysis and the credit granting decision,* in *Credit Scoring and Credit Control. Oxford University Press: Oxford*, edited by E. D. B. Thomas LC Crook JN (Oxford University Press, Oxford, 1992) pp. 109–121.

[59] Experian, *What are the different credit score ranges?* (2016).

[60] R. B. Nelsen, *Journal of the American Statistical Association*, Vol. 95 (Springer, New York, 2000) p. 334.

[61] S. Figini and P. Giudici, *Statistical merging of rating models,* Journal of the Operational Research Society **62**, 1067 (2011).

[62] C. Genest and L. P. Rivest, *Statistical inference procedures for bivariate Archimedean copulas,* Journal of the American Statistical Association **88**, 1034 (1993).

[63] L. Rüschendorf, *On the distributional transform, Sklar's theorem, and the empirical copula process,* Journal of Statistical Planning and Inference **139**, 3921 (2009).

[64] N. L. Hjort, C. Holmes, P. Mueller, and S. G. Walker, *Bayesian Nonparametrics* (Cambridge University Press, 2010).

[65] M. C. Galavotti, *Subjectivism, objectivism and objectivity in Bruno de Finetti's Bayesianism,* in *Foundations of Bayesianism*, edited by D. Corfield and J. Williamson (Springer, Dordrecht, 2001) pp. 161–174.

[66] R. Calabrese and P. Giudici, *Estimating bank default with generalised extreme value regression models,* Journal of the Operational Research Society **66**, 1783 (2015).

[67] E. Lybeck, *The Black Swan: The Impact of the Highly Improbable* (Random House, 2017) pp. 1–82.

[68] D. J. Gentleman, *Environmental Science and Technology*, 2nd ed., Vol. 45 (Wiley, New York, 2011) p. 5066.

[69] S. Jackman, *Bayesian Analysis for the Social Sciences* (Wiley, New York, 2009) pp. 1–558.

[70] X. L. Meng, *Multiple-imputation inferences with uncongenial sources of input,* Statistical Science **9**, 538 (1994).

[71] M. Turlakov, *Wrong-way risk, credit and funding,* AsiaRisk **26**, 46 (2013).

[72] V. Ivashina and D. Scharfstein, *Bank lending during the financial crisis of 2008,* Journal of Financial Economics **97**, 319 (2010).

4

MODELLING (EXTREME) MISSING TEMPERATURES

*This paper presents our winning entry for the EVA 2019 data competition, the aim of which is to predict Red Sea surface temperature extremes over space and time. To achieve this, we used a stochastic partial differential equation (*Poisson equation*) based method, improved through a regularization to penalize large magnitudes of solutions. This approach is shown to be successful according to the competition's evaluation criterion, i.e. a threshold-weighted continuous ranked probability score. Our stochastic Poisson equation and its boundary conditions resolve the data's non-stationarity naturally and effectively. Meanwhile, our numerical method is computationally efficient at dealing with the data's high dimensionality, without any parameter estimation. It demonstrates the usefulness of stochastic differential equations on spatio-temporal predictions, including the extremes of the process.*

4.1. INTRODUCTION

The aim of EVA 2019 data competition is to predict spatio-temporal extremes of Red Sea surface temperature [2]. The dataset of the competition consists of daily sea surface temperature (SST) anomalies for the entire Red Sea from 1985 to 2015 whilst part of the data is masked. The main goal of this data competition is to predict the distribution of

$$X(s,t) = \min_{(\tilde{s},\tilde{t})\in\mathscr{N}(s,t)} \hat{A}(\tilde{s},\tilde{t}), \tag{4.1}$$

for space-time validation points (s,t), where $\hat{A}(\cdot,\cdot)$ is the predicted SST anomalies. The competition evaluates the performance via a threshold-weighted continuous ranked probability score (twCRPS). The main challenge is the data's high dimensionality (16703 locations and 11315 days) and strong non-stationarity in space and time (Section 3.3 of

Parts of this chapter have been published in D. Cheng and Z. Liu, Spatio-temporal prediction of missing temperature with stochastic poisson equations, Extremes 24, 163 (2021) [1].

[2]). More details of the data competition, including the definition of the anomaly, plots of the data along with some exploratory analysis, as well as the competition's rules and goals, are available in the Editorial [2]. s To predict the extremes on the spatio-temporal domain, it is sufficient to predict the whole field of values on it. That is to say, we can predict $\hat{A}(\cdot,\cdot)$ instead of directly predicting $X(\cdot,\cdot)$ in Eq. (4.1). Given the partial data of $\hat{A}(\cdot,t)$ for someday t, the predicted data in the unobserved region should comply with some continuity constraints. According to Fourier's analytical theory of heat, it is reasonable to assume a temperature field has a second order derivative, implying that the anomaly field has its Laplacian. We formulate the missing value prediction problem as a Poisson equation based model. The Poisson equations are defined on the two-dimensional spatial domain, not the spatio-temporal domain. For each time step (day), Poisson equations are solved to interpolate the unobserved regions. To generate the stochasticity in predictions, our Poisson equation is not deterministic but stochastic. The stochastic components (*i.e.* the Laplace fields) in our SPDEs (stochastic partial differential equations) are sampled from the observed data, assuming temporal stationarity, which allows bypassing explicit estimation of parameters defining a stochastic model. The method is further improved with a regularization term to penalize large magnitudes of solutions. Experimental results show that it is effective and efficient to use the regularized stochastic Poisson equation to fill in the missing data in this problem.

A lot of approaches have been proposed on spatio-temporal extremes in the field of extreme value theory (EVT). One category of approaches is to model block maxima using max-stable processes [3–5]. The other category is to model the exceedances of a threshold [6, 7] using extreme distributions such as generalized Pareto distribution (GPD). Both categories of models have been used on climate data [8, 9]. SPDE-based approaches have been made popular by the INLA method in spatial and spatio-temporal statistics [10]. In [11], a type of SPDE similar to our Poisson equations is studied, which will be discussed in the last section of this paper. Comparing to EVT-based methods, our method is more efficient since we do not need any parameter estimation. The drawback is its performance will be inferior comparing to methods tailored to the prediction of the upper tail. Our approach is very similar to the Poisson image editing method [12] from the image processing community, which blends multiple images seamlessly with Poisson equations.

The remainder of this paper is structured as follows. In Section 4.2, we give the details about the stochastic Poisson equation based model. Its numerical method and the regularized version are given in Section 4.3. In Section 4.4, we explain our scheme to generate predictions, and the cross-validation dataset we prepared to evaluate our approach, followed by some concluding discussions in Section 4.5.

Notation The following notation is used: Italic lower-case letters refer to scalar values (x) or scalar-valued functions ($f(\cdot)$). Lower-case bold letters denote vectors ($\mathbf{x}$) or vector-valued functions ($\mathbf{f}(\cdot)$). Finally, upper-case bold letters represent matrices ($\mathbf{X}$).

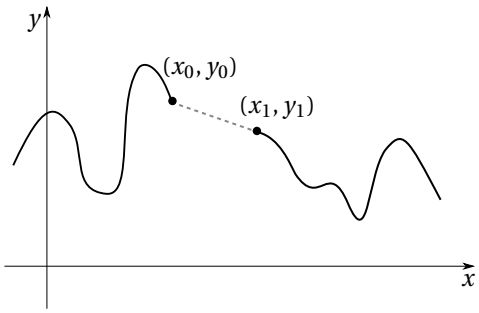


Figure 4.1: One-dimensional data interpolation.

4.2. METHOD

The missing data prediction problem can be studied from the perspective of data interpolation. By considering a 1D curve $y = f(x)$ with missing interval (x_0, x_1), as shown in Fig. 4.1, the interpolation result should have at least C^0 continuity. It is straightforward to connect two endpoints of the missing interval with a straight line segment, on which the second derivative is zero. Thus finding such a line segment is equivalent to solving the differential equation

$$\begin{cases} \dfrac{\mathrm{d}^2 f(x)}{\mathrm{d}x^2} = 0, \quad x \in (x_0, x_1); \\ \\ f(x_0) = y_0,\ f(x_1) = y_1. \end{cases} \tag{4.2}$$

The one-dimensional case above can be easily generalized to 2D, *i.e.* assuming the second derivative on the missing region is zero. Given a function $u(x, y)$, $(x, y) \in \mathscr{S}$ with masked region $\mathscr{M} \subset \mathscr{S}$, the unobserved region can be interpolated by solving the equation

$$\begin{cases} \Delta u = \dfrac{\partial^2 u}{\partial x^2} + \dfrac{\partial^2 u}{\partial y^2} = 0, \quad (x, y) \in \mathscr{M} \\ \\ u(x, y) \text{ takes known value in } \mathscr{S} \setminus \mathscr{M}. \end{cases} \tag{4.3}$$

Equations taking the form $\Delta u = 0$ are known as Laplace equations. An interpolation example is demonstrated in Fig. 4.2 (a–b). It is observed that the predicted region is oversmooth, lacking variations, compared to the known area, indicating that the right-hand side (RHS) in Eq. (4.3) should be some function f which is not zero. Thus the missing region can be predicted by solving the equation $\Delta u = f$, which is known as Poisson equation.

The physical meaning of Poisson equation $\Delta u = f$ here is that, given the partial temperature data u of someday $t \in \mathscr{T}$, the Laplacian of u (we call it Laplace field) in the unobserved region should follow some target field f. Immediately, it is necessary to model the target Laplace field f conditioning on the known data. We assume f is random since its exact value is unknown. Then the equation we are solving is actually a stochastic Poisson equation, instead of a deterministic one. Although we can construct some statistical

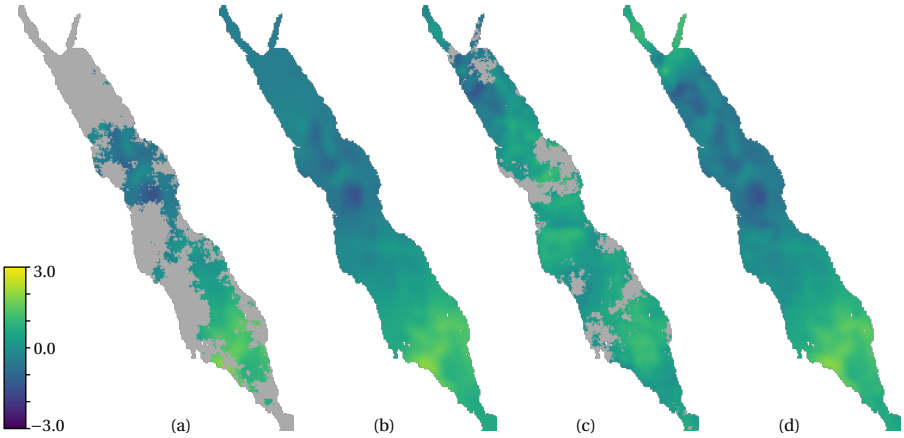


Figure 4.2: Given the partial data of Dec 22, 2015 (a), the unobserved regions (in gray) are interpolated with Laplace equation, which is over-smooth (b). With the data of Apr 1, 1985 as reference (c), the solution of Poisson equation is more realistic (d).

model for random f, it is more straightforward to sample existing fields from known data directly, with the assumption that the Laplace field is independent and identically distributed (i.i.d.) in time for the sake of simplicity. That is to say, the Laplace fields of day $t_1', t_2', \ldots$ are independent and identically distributed realizations of the random function f. The stochastic Poisson equation is solved by solving a series of deterministic ones. In terms of solving one of the Poisson equations, we take a reference day $t' \in \mathcal{T}$ as an example. Day t' is picked as a reference day, then Laplace field of day t' is a reference target Laplace field for day t. Therefore, missing data interpolation for day t can be accomplished by solving

$$\begin{cases} \Delta u(x,y;t) = \Delta u(x,y;t'), & (x,y) \in \mathscr{M}_t \\ u(x,y;t) = \text{known, in } \mathscr{S} \setminus \mathscr{M}_t, \end{cases} \tag{4.4}$$

in which $\mathscr{M}_t$ is the unobserved region of the day t. In general, this function does not always have a unique solution, but our numerical method gives a unique solution in our specific setting, which is described in Section 4.3. There are still missing data in $\Delta u(x,y;t')$ since $u(x,y;t')$ is incomplete for any t'. These values in $\Delta u(x,y;t')$ will be simply filled with *zeros*. An example is shown in Fig. 4.2 (c–d).

As indicated in [2], the temperature anomalies are not stationary in space and time. The Poisson equation $\Delta u = f$ can handle it effectively. In fact, the solution of Poisson equation can be decomposed as the sum of the solutions of the following two equations,

$$\begin{cases} \Delta u = 0, & (x,y) \in \mathscr{M} \\ u(x,y) = \text{known, in } \mathscr{S} \setminus \mathscr{M}, \end{cases} \quad \text{and} \quad \begin{cases} \Delta u = f, & (x,y) \in \mathscr{M} \\ u(x,y) = 0, \text{ in } \mathscr{S} \setminus \mathscr{M}. \end{cases} \tag{4.5}$$

In these two equations, the left one is the Laplace equation that imposes the continuity constraints, and the right one provides the randomness. The left one, which is

only related to the boundary conditions, ensures the solution is adaptive to the known data of the current day, tackling the non-stationarity in *time*. The right one has RHS $f = \Delta u(x, y; t')$, making the Laplace field be the same as some other day in a location-wise sense. Thus it resolves the temperature anomalies' non-stationarity in *space*. In conclusion, the sum of them can handle the non-stationarity in both space and time.

4.3. NUMERICAL SOLUTION

4.3.1. NUMERICAL SOLUTION OF THE POISSON EQUATION

The numerical solution of Poisson equation on a rectangular domain with Dirichlet boundary condition can be found with the finite difference method [13]. The key idea is discretizing the domain as a grid and approximating the Laplace operator with a linear combination of neighbor grid nodes. The differential equation is then converted to a linear system, of which the left-hand side (LHS) matrix is named the Laplacian matrix. In this paper, Poisson equation Eq. (4.4) is solved with the finite difference method.

First, the spatial domain is represented as a mesh grid Ω. The Red Sea $\mathscr{S}$ has been discretized into $S = 16703$ grid cells in the dataset of this competition. The grid cells are represented with their row/column indices (i, j), in which $i \in \{1, \ldots, 359\}$, $j \in \{1, \ldots, 233\}$. Each cell (i, j) has at most four neighbors $(i \pm 1, j)$ and $(i, j \pm 1)$. Then the mesh grid Ω is established after taking cells (i, j) as mesh nodes and the neighboring relation as mesh edges.

Then, the Laplacian matrix is determined after defining the mesh grid Ω. Noting that the mesh grid Ω is equivalent with an undirected graph $\mathscr{G}$ with its nodes and edges, the Laplacian matrix $\mathbf{L}$ of graph $\mathscr{G}$ is defined according to [14], as a generalization of the Laplacian matrix of a rectangular domain. The Poisson equation Eq. (4.4) is converted to the linear system

$$\mathbf{L}\mathbf{u} = \mathbf{L}\mathbf{u}', \tag{4.6}$$

in which matrix $\mathbf{L}$ has shape $S \times S$, $\mathbf{u}$ is an S-dimensional column vector containing the temperature anomalies of all locations on the day t, and $\mathbf{u}'$ is the temperature anomalies vector of some other day t'. In this equation, the LHS $\mathbf{L}\mathbf{u}$ is the Laplace field of day t, and the RHS $\mathbf{L}\mathbf{u}'$ is the reference Laplace field taken from day t'. Both of them are computed with the Laplacian matrix $\mathbf{L}$ defined on the mesh domain. As discussed above, some elements in $(\mathbf{L}\mathbf{u}')$ are not available since $\mathbf{u}'$ is incomplete (containing values represented with NaNs). These invalid values in $(\mathbf{L}\mathbf{u}')$ are filled with zeros.

Finally, unknown values of $\mathbf{u}$ are determined by solving the equation. Without loss of generality, it can be assumed that vector $\mathbf{u}$ is split into two parts $\mathbf{u}_0$ and $\mathbf{u}_1$, of which the first one $\mathbf{u}_0$ contains all unobserved locations while the known values appear in the second part $\mathbf{u}_1$. This form can be obtained by multiplying permutation matrices on both sides of Eq. (4.6). Then $\mathbf{L}$ and $\mathbf{u}'$ are partitioned accordingly, resulting in the linear system

$$\begin{pmatrix} \mathbf{L}_{00} & \mathbf{L}_{01} \\ \mathbf{L}_{10} & \mathbf{L}_{11} \end{pmatrix} \begin{pmatrix} \mathbf{u}_0 \\ \mathbf{u}_1 \end{pmatrix} = \begin{pmatrix} \mathbf{L}_{00} & \mathbf{L}_{01} \\ \mathbf{L}_{10} & \mathbf{L}_{11} \end{pmatrix} \begin{pmatrix} \mathbf{u}'_0 \\ \mathbf{u}'_1 \end{pmatrix}, \tag{4.7}$$

which implies that

$$\mathbf{L}_{00}\mathbf{u}_0 = \mathbf{L}_{00}\mathbf{u}'_0 + \mathbf{L}_{01}(\mathbf{u}'_1 - \mathbf{u}_1). \tag{4.8}$$

This equation has a unique solution since $\mathbf{L}_{00}$ is invertible, which will be proved in the next subsection. At last, Eq. (4.8) can be solved with LU decomposition of matrix $\mathbf{L}_{00}$.

4.3.2. AN OPTIMIZATION PERSPECTIVE: LEAST SQUARES

From the viewpoint of mathematical optimization, the Poisson equation Eq. (4.4)

$$\Delta u = \nabla \cdot \nabla u = \nabla \cdot \nabla u' \tag{4.9}$$

is the Euler-Lagrange equation of the type

$$\min_u \iint \|\nabla u - \nabla u'\|^2. \tag{4.10}$$

A discrete approximation of this optimization problem is

$$\min_{\mathbf{u}} \|\mathbf{G}\mathbf{u} - \mathbf{G}\mathbf{u}'\|^2, \tag{4.11}$$

in which $\mathbf{G}$ is the gradient matrix. The gradient matrix is defined as transpose of graph $\mathscr{G}$'s incidence matrix, when an arbitrary *orientation* is given to $\mathscr{G}$ [14]. Rows and columns of matrix $\mathbf{G}$ are the numbers of edges and nodes in graph $\mathscr{G}$ respectively. If $\mathbf{G}$ is written in the block form according to values' availability as before, *i.e.* $\mathbf{G} = \begin{pmatrix}\mathbf{G}_0 & \mathbf{G}_1\end{pmatrix}$, finding unknown values $\mathbf{u}_0$ is to solve the least squares problem

$$\min_{\mathbf{u}_0} \|\mathbf{G}_0\mathbf{u}_0 - \mathbf{h}\|^2, \tag{4.12}$$

in which $\mathbf{h} = \mathbf{G}_0\mathbf{u}_0' - \mathbf{G}_1\left(\mathbf{u}_1' - \mathbf{u}_1\right)$. Invalid values (NaNs) in $\mathbf{h}$ are also filled with zeros.

The least squares problem (4.12) is solved with the pseudo inverse theory, *i.e.*

$$\mathbf{u}_0 = (\mathbf{G}_0^\top \mathbf{G}_0)^{-1}\mathbf{G}_0^\top \mathbf{h}. \tag{4.13}$$

Noting that $\mathbf{L}_{00} = \mathbf{G}_0^\top \mathbf{G}_0$ since $\mathbf{L} = \mathbf{G}^\top \mathbf{G}$, solution of the least squares problem should be identical with that of Eq. (4.8) if there is no missing value on the reference day. When there is missing data, the only difference between them is that it is the invalid gradient values that are filled with zeros in Eq. (4.13), while it is the invalid Laplacian values that are replaced with zeros in Eq. (4.8).

To make $\mathbf{L}_{00} = \mathbf{G}_0^\top \mathbf{G}_0$ invertible, block $\mathbf{G}_0$ in the gradient matrix is required to have linearly independent columns. This is true because each unknown node on graph $\mathscr{G}$ is connected with at least one known node by a *path* [14], which can be concluded from the fact that the spatial domain Ω has only one connected component and all days have at least one known node.

4.3.3. LEAST SQUARES WITH REGULARIZATION

Solution of the least squares problem in Eq. (4.12) gives a prediction of the missing data that is smooth and vibrant (Fig. 4.3 (c)). However, it is suffering from numerical issues. Although the Laplacian matrix $\mathbf{L}_{00}$ is invertible, its *condition number* is large from the viewpoint of computing, making the numerical solution instable. If the Poisson equation has a closed domain with Dirichlet boundary condition, the solution will be well bounded by known values. However, this is not our case.

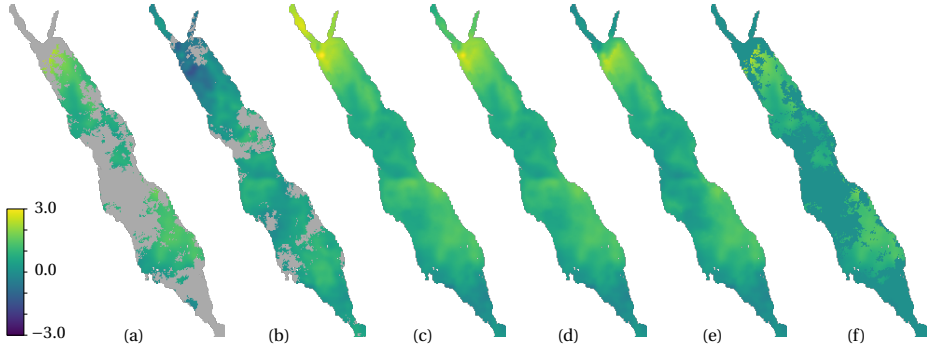


Figure 4.3: Results of the regularized least squares. Given the partial data of Sept 23, 2015 (a) and the reference data of Apr 1, 1985 (b), the spatial domain is completed with the regularized least squares method, using $\lambda = 0.0, 0.001, 0.01, +\infty$ respectively (c–f).

Regularization is commonly used to stabilize least squares, based on which Eq. (4.12) becomes

$$\min_{\mathbf{u}_0} \|\mathbf{G}_0\mathbf{u}_0 - \mathbf{h}\|^2 + \lambda\|\mathbf{u}_0\|^2, \tag{4.14}$$

with $\lambda \geq 0$. The regularized least squares problem takes the form of ridge regression [15], of which the solution is given by

$$(\mathbf{L}_{00} + \lambda\mathbf{I})\mathbf{u}_0 = \mathbf{f}. \tag{4.15}$$

Here, $\mathbf{f}$ is $\mathbf{G}_0^\top\mathbf{h}$, and $\mathbf{I}$ is the identity matrix. Please note that $\mathbf{f}$ can also take the value of RHS in Eq. (4.8), depending on how invalid values at the RHS are handled, as discussed in the last subsection. When $\mathbf{f}$ takes the RHS of Poisson equation, Eq. (4.15) actually solves $(\Delta + \lambda)u = f$, which is named *screened Poisson equation*. It is the Euler-Lagrange equation of the type

$$\min_u \iint \|\nabla u - \nabla u'\|^2 + \lambda\|u\|^2. \tag{4.16}$$

This minimization objective means the gradient field of the target day should be close to that of the reference day. At the same time, large values of u are penalized.

It is necessary to determine the value of λ in the regularized problem. Observing the results with different λ values in Fig. 4.3 (c–f), we can see that larger λ generates flatter result. By considering the limit situation when $\lambda \to +\infty$, the predicted temperature anomalies will degenerate to all zeros (Fig. 4.3 (f)). In this case, there is a step between observed and unobserved regions ("boundary step"), which can be evaluated as follows

$$\mathrm{BS}(\lambda; t) = \sum_{(p,q)\in P} \|\mathbf{u}_0(p) - \mathbf{u}_1(q)\|^2, \tag{4.17}$$

where P is the set of all neighboring location pairs (p, q) where p is unobserved and q is observed. Unavailable data $\mathbf{u}_0$ is uniquely determined once given $\lambda \geq 0$, thus $\mathbf{u}_0$ is a function of λ. So is the "boundary step" $\mathrm{BS}(\lambda; t)$. The "boundary step" is large when

the regularization parameter λ is too large or too small. However, finding the optimal λ is not trivial. We pick the value of λ that minimizes the "boundary step" among all candidates in a finite set Λ, *i.e.*

$$\lambda^*(t) = \arg\min_{\lambda\in\Lambda} \mathrm{BS}(\lambda; t). \tag{4.18}$$

The candidate set Λ will be given numerically in the next section. Please note that the best weight λ^* is determined for each day t separately. We do not estimate a fixed value of λ for all days.

4.4. INFERENCE

To predict the distribution of $X(s,t)$ (see Eq. (4.1)) for space-time validation point (s,t), we will predict the whole temperature anomaly field $\hat{A}(s,t)$ for all days that fall in the ± 3 days temporal neighborhoods of validations days first, then compute the spatio-temporal neighborhood minimums directly. And last, we use the empirical distribution of the minimums as the prediction.

4.4.1. INFERENCE PROCEDURE

Given a space-time validation point (s_0, t_0), the temperature anomaly field of that target day, $\hat{A}(s, t_0),\ s \in \mathscr{S}$, is predicted first. To achieve this, a guidance day t' is picked to provide a reference of Laplace or gradient field. The sampled day t' should be as complete as possible, thus it is sampled from the period 1985–2006 with only 20% missing data. After that, linear systems are constructed and solved following the method presented in the previous section. All space-time validation points from the day t_0 are computed together from the same linear system, to avoid redundant computing, which saves time, energy, and helps fight climate change.

After predicting the field of the target day t_0, we repeat the same procedure to predict all seven days in its temporal neighborhood, *i.e.* the days $t_0 + i,\ -3 \le i \le 3$. Hence seven reference days should be sampled. Instead of choosing them independently, it is better to sample seven successive days, *i.e.* the days $t' + i,\ -3 \le i \le 3$. The advantage is that the temporal dependency of the temperature anomalies is implicitly maintained in the data of sampled seven days. Until now, all $\hat{A}(\cdot,\cdot)$ values in $\mathscr{N}(s_0, t_0)$ are obtained. Then we calculate the minimum to get one realization of $X(s,t)$.

An empirical distribution of $X(s,t)$ is predicted if the whole procedure above is performed repeatedly for $t'_1, t'_2, \ldots, t'_N$. We repeated N times to generate N values of $X(s,t)$ for each validation point (s,t), and used the empirical cumulative distribution function (ECDF) of them as the prediction of that validation point. It is noted that the LHS Laplacian matrix in Eq. (4.6–4.8) only depends on the data availability, *i.e.* the binary mask, of the target day; its temperature anomaly values and that of the reference day only affect the RHS of the linear systems. We take advantage of this to pre-factorize the Laplacian matrix with LU decomposition only once. Then all of those days sharing the same data availability can be solved directly by forward and backward substitutions.

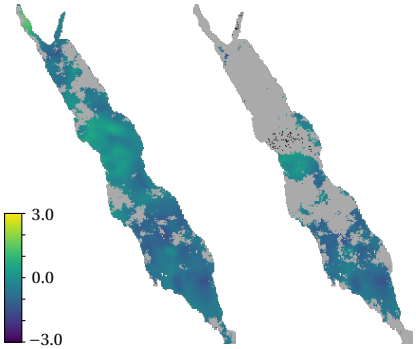


Figure 4.4: Cross-validation set is constructed by masking part of the training data. Left: Training data of Apr 5, 1985. Right: Gray regions are masked. Colored regions are the training data for cross-validation. Black dots are cross-validation points.

Table 4.1: Average twCRPS of different methods on the cross-validation set.

Method	$10^4 \times \overline{\text{twCRPS}}$
Benchmark	5.37
Poisson equation	2.46
Regularized least squares	2.33
Screened Poisson equation	2.35
Ensemble	2.19

4.4.2. LEAVE-ONE-OUT VALIDATION

Since the validation set of this competition is not available to the teams, a cross-validation set is prepared during the competition. Data in the range of 1985–2006 is used for 22-fold cross-validation. For the i-th fold ($i = 0, \ldots, 21$), data of the year $1985 + i$ is left out for validation while the other 21 years are used for training. For each day in the left-out year, some of the locations have already been unobserved in the original training set (gray regions in the left of Fig. 4.4) and more locations are masked out (gray regions in the right of Fig. 4.4) for cross-validation. For example, as demonstrated in Fig. 4.4, 1985 is left out and part of Jan 5's data is masked out (new gray regions in the right figure comparing to the left). Unmasked regions (colored regions in the right figure) are used as training data in the cross-validation stage. For some points in the gray region, they can be sampled as cross-validation points if their spatio-temporal neighborhood $\mathcal{N}(\cdot, \cdot)$ is completely available in the original training set, which are shown as black dots in the right of Fig. 4.4. Ground truth $X(\cdot, \cdot)$ is available on these points. Therefore, we have generated the training set and validation points for cross-validation. Since the average temperature anomalies of the first 22 years are lower than that of the last 9 years (where the competition's final predictions should be made), as indicated in [2], we translated the data by the differences of empirical means between the last 9 years and our cross-validation set in order to match the means.

We evaluated different configurations of our method on the generated cross-validation set, including Poisson equation Eq. (4.8), regularized least squares Eq. (4.15), and screened Poisson equation that takes Eq. (4.15)'s LHS and Eq. (4.8)'s RHS. For those configurations with regularizers, the candidate set of weight λ is $\Lambda = \{0.02 \times (0.5)^k : \ k = 0, 1, \ldots, 11\}$. For each day, we choose the λ value that gives the least "boundary step" as Eq. (4.18). Consequently, the Poisson equation generates one solution for each day; each of the regularized methods, *i.e.* the regularized least squares and the screened Poisson equation, chooses one solution from $|\Lambda| = 12$ candidate solutions for each day. At last, an ensemble is proposed, which pools all 25 resulting candidate $\mathbf{u}_0$s (one from the Poisson equation and 12 from each of the regularized methods) and picks the one with the least "boundary step" from them as Eq. (4.18). In all methods above, the sample size N is 1000. These methods are compared to the benchmark proposed in [2], as shown in Tab. 4.1. The ensemble gives the lowest error in terms of the average twCRPS on the cross-validation set, thus we used its result as the final submission for the competition.

After sampling $N = 1000$ times, the ECDF curves are already smooth enough. We have tried to fit the ECDFs with the generalized extreme value (GEV) distribution, while there is little difference according to the twCRPS.

4.4.3. FINAL TECHNICAL REMARK

Our approach is implemented with Python to perform parallel computing with an Intel® Core™ i7-9700K CPU on a desktop computer. In our final submission to the competition, we used the ensemble method with sample size $N = 1000$ and candidate weight set Λ as described above. The total computational cost is less than 70 minutes.

4.5. DISCUSSION

This approach is shown to be successful in the context of this competition, thanks to its advantages. First, the Laplacian matrix is a nice encoding of the spatial domain's geometry. It contains the links between locations and its neighbors, and helps to build the spatial dependence between locations. Second, the temporal dependence is considered implicitly via borrowing the Laplace fields from successive days, *i.e.* reference weeks, instead of independent reference days. Third, the non-stationarity over time and space is embedded in our model since the values in unobserved regions can be decomposed into two parts (Eq. (4.5)); one is associated only with the current day's information and the other considers the spatial variation. Last, this method is training-free. Training data are used directly as references without any time-consuming parameter estimation.

In [11], the SPDE $(\lambda - \Delta)^{\alpha/2} u = f$ is studied, where $\alpha > 0$, $\lambda \geq 0$ and f corresponds to Gaussian white noise. First, it should be noted that the $-\Delta$ in [11] is actually the $+\Delta$ in this paper. The latter notation is used here to make it compatible with the discretization of Laplace operator in [14]. Therefore, the above SPDE coincides with the (screened) Poisson equation in this work, when $\alpha = 2$, $\lambda \geq 0$. The difference is that our model is derived from an interpolation perspective. Our regularization term with λ is motivated by penalizing large magnitudes of solutions. At the same time, our model is parameter-free since the value of λ is determined according to the "boundary step" on each day. Also, we do not assume the RHS f corresponds to Gaussian white noise.

However, there are still some aspects that are neglected, which will be further studied to improve the current approach. For example, this approach is not tailored specifically to the evaluation criterion twCRPS. The performance might be further improved by putting the emphasis on the upper tail, according to the weight function in twCRPS. Also, reference weeks are just sampled randomly, which means some of the sampled reference fields might not be compatible with the known data. It will be helpful to study which Laplace field is more likely to present subject to the known data.

Thanks to the efficiency of our method, we were able to simulate a large number of temperature anomaly fields and compute local minimums directly. That is why extreme value theory methods were not used in our competition contribution. In the future, it will be studied to improve this model with EVT methods. For example, the sample size N can be largely reduced if EVT methods are used to model the distributions of reference Laplace or gradient fields. This model can also be combined with the EVT method to further reduce its error on large temperature anomalies.

REFERENCES

[1] D. Cheng and Z. Liu, *Spatio-temporal prediction of missing temperature with stochastic Poisson equations: The LC2019 team winning entry for the EVA 2019 data competition,* Extremes **24**, 163 (2021).

[2] R. Huser, *Editorial: EVA 2019 data competition on spatio-temporal prediction of Red Sea surface temperature extremes,* Extremes **24**, 91 (2021), arXiv:1912.00694 .

[3] A. C. Davison, S. A. Padoan, and M. Ribatet, *Statistical modeling of spatial extremes,* Statistical Science **27**, 161 (2012).

[4] R. A. Davis, C. Kluppelberg, and C. Steinkohl, *Statistical inference for max-stable processes in space and time,* Journal of the Royal Statistical Society. Series B: Statistical Methodology **75**, 791 (2013).

[5] R. Huser and A. C. Davison, *Space-time modelling of extreme events,* Journal of the Royal Statistical Society. Series B: Statistical Methodology **76**, 439 (2014), arXiv:1201.3245 .

[6] J. P. III, *Statistical Inference Using Extreme Order Statistics,* The Annals of Statistics **3**, 119 (1975).

[7] J. N. Bacro, C. Gaetan, T. Opitz, and G. Toulemonde, *Hierarchical space-time modeling of asymptotically independent exceedances with an application to precipitation data,* Journal of the American Statistical Association **115**, 555 (2020).

[8] V. Chavez-Demoulin and A. C. Davison, *Generalized additive modelling of sample extremes,* Journal of the Royal Statistical Society. Series C: Applied Statistics **54**, 207 (2005).

[9] B. J. Reich and B. A. Shaby, *A hierarchical max-stable spatial model for extreme precipitation,* Ann. Appl. Stat. **6**, 1430 (2012).

[10] E. Krainski, V. Gómez-Rubio, H. Bakka, A. Lenzi, D. Castro-Camilo, D. Simpson, F. Lindgren, and H. Rue, *Advanced Spatial Modeling with Stochastic Partial Differential Equations Using R and INLA* (Chapman and Hall/CRC, 2018).

[11] F. Lindgren, H. Rue, and J. Lindström, *An explicit link between gaussian fields and gaussian markov random fields: The stochastic partial differential equation approach,* Journal of the Royal Statistical Society. Series B: Statistical Methodology **73**, 423 (2011).

[12] P. Pérez, M. Gangnet, and A. Blake, *Poisson image editing,* ACM Transactions on Graphics **22**, 313 (2003).

[13] P. Knabner and L. Angerman, *Numerical Methods for Elliptic and Parabolic Partial Differential Equations* (Springer-Verlag New York, 2003).

[14] U. Knauer and K. Knauer, *Algebraic Graph Theory* (Springer-Verlag New York, 2019).

[15] D. Ruppert, *Journal of the American Statistical Association*, Vol. 99 (Springer-Verlag New York, 2004) pp. 567–567.

5

CONCLUSIONS AND FUTURE WORK

In this last chapter I collect some thoughts about my work so far, and most of all about possible future work.

5.1. ABOUT CHAPTER 3

Chapter 3 applied a bivariate reinforced urn process–the R-RUP–to model the dependence between PD and LGD. The building block of this bivariate reinforced process is the beta-Stacy process, in its urn representation. The PD and LGD are affected by both a common component and some idiosyncratic shocks, following a one-factor model. The dependence between PD and LGD clearly arises from this common factor. It is natural to think that the product instead of the sum could also build a kind of dependence relation. But that would trivially correspond to a sum in the logs.

The R-RUP exhibits many interesting probabilistic properties, like partial exchangeability, semi-Markovianity and conjugacy. Its construction is rather intuitive and it allows for simple simulations. The Bayesian update mechanism embedded in the Pólya urns builds a model that is able to combine prior knowledge, for example in the form of expert judgements, with empirical evidence; a model that learns and adapts to new phenomena and trends emerging from data, even in the case of censoring. The rate of updating can be controlled by acting on the strength of the a priori, and on the reinforcement mechanism of the single Pólya urns.

An application on the Freddie Mac Single Family Loan-Level Dataset showed that the R-RUP actually provides interesting performances in terms of Bayesian prediction, with results that could be easily improved in case of more reliable a prioris. This is in fact a weak point at the moment. Further research is indeed needed to develop a more reliable set of priors for PD and LGD.

In Chapter 3, we have not dealt with wrong-way risk, another interesting component of recovery risk, that is the positive dependence between PD and LGD observed in the empirical literature, especially during financial crises [1–3]. We assume in fact that default is given and it corresponds to the beginning of the recovery process, so that the

modeling of PD is not relevant. An extension of the model to include wrong-way risk is in our future plans, and interesting starting points are the works of [4] and [5], notwithstanding our results from Chapter 4.

5.2. ABOUT CHAPTER 4

In Chapter 4, we have presented a one-factor model built with RUPs, 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 once again 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 rare events. For rare events in particular, the possibility of eliciting an a priori on something rarely or never observed before can mitigate, at least partially, the relevant problem of historical bias [6–8].

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 [9], 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/deep learning [10]. However, differently from standard machine/deep 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 [11].

Clearly, if an a priori cannot be elicited nor is desired, one can still use the model in a totally data-driven way, building priors based on ecdfs [12]. In such a case, however, the results of the bivariate urn construction would not substantially differ from those of a more standard empirical copula approach [13], notwithstanding the difficulty of dealing with non-directly (fully) observable quantities like the joint factor A. More interesting can therefore be the "merging" methodology developed by [14], in which the combination of quantitative and qualitative information can (at least partially) solve the lack of experts' priors.

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 [15].

As observed in Section 3.2, 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 (we estimate a correlation of about 0.25), and PD/LGD in general [3].

Finally, the bivariate urn model can be computationally intensive. For large datasets and portfolios, a standard laptop may require from a few hours to a full day 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, and it should be addressed more

seriously for final deployment.

5.3. ABOUT CHAPTER 5

In Chapter 5, some stochastic Poisson equations are proposed to make the spatio-temporal prediction of missing temperature. The ultimate goal is to predict the spatio-temporal extremes, which is the distribution of the minimum of the predicted SST anomalies over space and time (Eqn. (1)).

Instead of modeling the extremes directly, we chose to first predict the SST anomalies over space and time and then the prediction of the extremes based on that is made. In order to predict SST anomalies over space and time, our model is derived from a very simple setting, in order to better capture the features of the data.

The first method is based on Laplace equations. We use a five-point stencil finite-difference method to approximate the Laplacian. In this sense, modeling the missing SST anomalies with Laplace equations means that every value at (x, y) equals to the average of four neighboring values. Given the boundary condition, the missing values can be solved as shown in Fig.2 (a-b) of Chapter 5.

However, compared with the realistic data, the prediction via Laplacian equation is over-smooth. Therefore, we move forward towards Poisson equations, which means that the difference between every value and the average of its four neighbors can be nonzero. Mathematically, for some time $t \in \mathcal{T}$, we have $\Delta u(x, y; t) = f(x, y; t)$. To solve this equation and obtain the missing temperature $u(x, y; t)$, the right hand side $f(x, y; t)$ has to be given. The $f(x, y; t)$ is unavailable for location (x, y) and day t; however, at the same location and different day $t^{'}$, $\Delta u(x, y; t')$ is available in our dataset. As a result, we borrow $\Delta u(x, y; t')$ and replace the unknown $f(x, y; t)$ with it.

The most ideal situation is that the information we borrow from the day $t^{'}$ is just the same as that in the day t we want to predict. It is quite difficult to find the most ideal day $t^{'}$. This leads to the third method: stochastic Poisson equation. We assume that the unknown $f(x, y; t)$ is random, which can take different values with the same probability. These values are chosen from other days. In this way, the stochastic Poisson equation can be described as $\Delta u(,; t) = \Delta u(,; t')$, where t' takes N different values $t_1, \ldots, t_N$ where each value is assigned a probability $\frac{1}{N}$. This means that $\Delta u(,; t)$ is possible to be $\Delta u(,; t_1)$ with the probability $1/N$. In order to solve it, we solve $\Delta u(,; t) = \Delta u(,; t_1), \ldots, \Delta u(,; t) = \Delta u(,; t_N)$. Then, we obtain N numerical solutions of $u(,; t)$ and assign equal probability to each $u(,; t)$, as the solution of our stochastic Poisson equation.

Since the ultimate goal is to find the extremes over 7 days, we borrow the seven successive days' information to maintain the dependence over time. We use three ways to get the numerical solution. The first is to solve Poisson equation with the finite difference method. The second one is from another perspective-least squares-to numerically solve the Poisson equation.

The solution of Poisson's equation $-\Delta u = -\Delta u'$ in U with the Dirichlet boundary $u = g$ on ∂U is related to a minimization problem [16]:

$$\min_{w} \frac{1}{2} \int \int_{U} \|\nabla(w - u')\|^2 \text{ with } w = g \text{ on } \partial U,$$

where u' is known. Set $v = w - u'$ and define $E[v] = \frac{1}{2} \int \int_{U} \|\nabla v\|^2$.

Discretize the gradient $\nabla(w-u')$ along an edge $e=(i,j)$ where i,j represent the locations as the finite difference (v_i-v_j). Then $E[v]$ can be approximated by

$$E_d[v]=\sum_e h_{ij}(v_i-v_j)^2$$

where $h_{ij}=1$ when there is an edge between i and j; otherwise, $h_{ij}=0$. Then we can derive that

$$E_d[v]=\mathbf{v}^T L\mathbf{v}$$

where L is a Laplacian matrix of which the elements are

$$L_{i,j}:=\begin{cases} deg(v_i) & \text{if } i=j \\ -1 & \text{if } v_i \text{ is adjacent to } v_j \\ 0 & \text{otherwise} \end{cases} \tag{5.1}$$

where $\deg(v_i)$ is the degree of the vertex i. Write the boundary constraints $v=g-u'$ on ∂U in the form of matrix as $A\mathbf{v}=\mathbf{c}$. Then our goal is to

$$\begin{array}{ll} \text{minimize} & E_d[v]:=\mathbf{v}^T L\mathbf{v} \\ \text{subject to} & A\mathbf{v}=\mathbf{c} \end{array} \tag{5.2}$$

This is a quadratic programming with the equality constraints. The minimization problem is transferred to the following linear system

$$\begin{pmatrix} L & A^T \\ A & \mathbf{0} \end{pmatrix}\begin{pmatrix} \mathbf{v} \\ \lambda^* \end{pmatrix}=\begin{pmatrix} \mathbf{0} \\ \mathbf{c} \end{pmatrix} \tag{5.3}$$

where λ^* is the associated Lagrange multiplier.

5.4. FUTURE WORK

Here I collect some ideas for future work, mainly in the field of urn models. Interesting intermediate results have already been developed, but I have chosen not to include them in the main text of this thesis, because I still need time to further develop and validate them.

5.4.1. BERNSTEIN PRIORS FOR MODELING THE DEPENDENCE BETWEEN PD AND LGD

The prime interest in Chapter 3 are the PD and the LGD, usually taking values in a continuous space $[0,1]^2$. However, the bivariate survival function we used in [17] takes value in the natural numbers $\mathbb{N}_0^2$, which is a discrete space. In order to use this bivariate survival function, the discretization of $[0,1]^2$ is conducted; this can be seen as a limitation of our model. In order to overcome this, we construct a prior for the observations which takes values in the continuous space $[0,1]^2$.

Recall the bivariate survival function F_{XY} in Chapter 4, and its constituent independent discrete beta-Stacy processes F_A, F_B, and F_C, with parameters $\{\alpha_j^A,\beta_j^A\}$, $\{\alpha_j^B,\beta_j^B\}$,

and $\{\alpha_j^C, \beta_j^C\}$, $j \in \mathbb{N}_0$ respectively. Given P_A, P_B, and P_C which are the corresponding probability distribution of F_A, F_B, and F_C, define

$$P_{XY}(x,y) = \sum_{a=1}^{x \wedge y} P_A(a) P_B(x-a) P_C(y-a), \tag{5.4}$$

for any $x, y \in \mathbb{N}_0$. Denote the distribution function associated with P_{XY} by F_{XY}.

A Bernstein prior is a probability measure on the space of all the distribution functions on the unit interval $[0,1]$, based on Bernstein polynomials. Following [18], assuming a Bernstein prior is equivalent to the following hierarchical model:

- For any $n \in \mathbb{N}_0$, $(X_1, \ldots, X_n)$ are conditionally independent and identically distributed, given $\tilde{k} = k$ and a random probability distribution $\tilde{F} = F$, with a common distribution

$$B(x;k,F) = \begin{cases} 0 & x < 0, \\ \sum_{i=0}^{k} F(\frac{i}{k}) C_k^i x^i (1-x)^{k-i} & 0 \le x \le 1, \\ 1 & x > 1. \end{cases} \tag{5.5}$$

  It is easy to show that $B(\cdot;k,F)$ is a distribution function; in particular, if it has probability mass zero at point 0, it is almost surely absolutely continuous on $[0,1]$ with the density function

$$b(x;k,w_k) = \sum_{i=1}^{k} w_{i,k} \beta(x;i,k-i+1),$$

  i.e a mixture of beta densities with parameters $(i, k-i+1)$, with k components and mixing weights $w_k = (w_{1,k}, \ldots, w_{k,k})$, where $w_{i,k} = F(\frac{i}{k}) - F(\frac{i-1}{k})$.

- Suppose that $(\tilde{k}, \tilde{F})$ has a joint distribution $\mathbf{P}$.

Then the Bernstein polynomial $B(\cdot;\tilde{k},\tilde{F})$ is a random distribution function on $[0,1]$, whose probability law is induced by $\mathbf{P}$ and called a Bernstein prior with parameter $\mathbf{P}$ [18].

However, for two-dimentional data with dependence between them, a further generalization is necessary.

For any $n \in \mathbb{N}_0$, $[(X_1,Y_1), \ldots, (X_n,Y_n)]$ are conditionally independent and identically distributed, given $\tilde{k} = k$ and a random bivariate probability distribution on $\mathbb{N}_0^2$, $\tilde{F}_{WV} = F_{WV}$, with a common distribution

$$B(x,y;k,P_{WV}) = \begin{cases} 0 & x < 0 \text{ or } y < 0, \\ \int_0^x \int_0^y \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} P_{WV}(i,j) K(a,b;i,k-i+1,j,k-j+1) \mathrm{d}a\mathrm{d}b & 0 \le x,y \le 1, \\ 1 & x > 0 \text{ or } y > 1, \end{cases} \tag{5.6}$$

where for any $a, b \in [0,1]$, and any $i, j, k \in \mathbb{N}_0$,

$$K(a,b;i,k-i+1,j,k-j+1) = \begin{cases} \beta(a;i,k-i+1)\beta(b;j,k-j+1) & \text{if } 1 \le i,j \le k, \\ 1 & \text{otherwise.} \end{cases} \tag{5.7}$$

5

If P_{WV} is a probability measure on $\mathbb{N}_0^2$, it is easy to show that $B(\cdot,\cdot;k,P_{WV})$ is a distribution function. Moreover, it is almost surely absolutely continuous on $[0,1]^2$ with the the density function

$$b(x,y;k,P_{WV}) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} P_{WV}(i,j)K(x,y;i,k-i+1,j,k-j+1)$$

a mixture of beta densities and uniform densities. We can then derive that the marginal density for $B(\cdot,\cdot;k,P_{WV})$ is respectively

$$\sum_{i=1}^{k} P_W(i)\beta(x;i,k-i+1),$$

and

$$\sum_{j=1}^{k} P_V(j)\beta(x;j,k-j+1),$$

which is the density function associated with a Bernstein prior. For any $x,y\in[0,1]$, define $F_{XY}(x,y) = B(x,y;\tilde{k},\tilde{P}_{WV})$.

Building the connection between the bivariate Bernstein prior and continuous-state RUPs is the next step.

5.4.2. REALISTIC MODELLING OF CREDIT RATINGS TRANSITION MATRICES WITH URNS

A credit rating is a measure of the creditworthiness of a debtor as evaluated by credit rating agencies, the largest of which are Standard & Poor's, Moody's and Fitch. The time series of the ratings is usually assumed to be a Markov chain which is determined by an initial state and a transition matrix whose elements are the probabilities of moving between any two rating classes [19]. The PD is one of the element in this transition matrix, i.e. the probability of ending in the default class, which is also an absorbing state.

However, the more recent empirical literature has shown that ratings do not constitute standard homogeneous Markov chains [20, 21]. Ratings are indeed characterized by four main features: a downward momentum, time non-homogeneity, rating duration, and an aging effect.

The downward momentum indicates that a counterparty experiencing a downgrade is more likely to further downgrade, with respect to a counterparty that never downgraded before.

Time non-homogeneity tells us that the credit transition matrix varies with time [19, 20], suggesting that the use of homogeneous Markov chains is simply wrong.

Rating duration represents the intuitive situation in which the time spent in a given rating class influences the probability of jumps to other classes.

Finally, the aging effect means that the transition matrix at the current time relies on the issue date of the rating, so that past estimates become less and less reliable over time, also because of business cycles [20].

Ratings are therefore better modeled in a semi-Markov framework [22, 23]. A semi-Markov model assumes that the rating in the future depends not only on the current

state, but also on the time spent in the current state. Further time non-homogeneity can be also embedded in the semi-Markov framework, as in [24, 25]. For the ageing effect, a discrete time non-homogeneous semi Markov model with an age index is proposed for example in [26]. To address the downward rating momentum, [27] applied semi-Markov processes with an extended state space to account for downward rating momentum.

Recently, [28] proposed a Bayesian nonparametric method to estimate rating migration matrices. This method is a mixture of Markov chains based on RUPs. Taking this approach as an inspiration, we have been developing an urn model for rating migrations, which are decomposed into two parts: the potential to migrate, and the "decision" to accept this migration or just keep the same.

Credit rating agencies such as S&P, Moody's, and Fitch detect the potential change of the credit rating for companies as watch signals. After observing a permanent change in an issuer's creditworthiness without any substantial uncertainty, the rating is usually changed. We assume that the potential to move in one period is controlled by a common factor for all companies and the decision to accept whether to transfer from one rating to another one is determined by an idiosyncratic factor. All these factors are modeled by urns. In this way, we can consider the time non-homogeneity and the downward momentum in our model.

We consider companies divided into L rating groups. Withing each group companies are assumed to be exchangeable. Over time a company can stay in the same group or move to another one. One of the groups contains defaulted companies, which can no longer move away.

Let us model the credit transitions over T years. We have T common urns for T years. We construct a urn system consisting of an urn M, T common urns $U_1,\dots,U_T$, and $4L$ idiosyncratic urns $V_{11},V_{12},V_{13},V_{14},\dots,V_{L1},V_{L2},V_{L3},V_{L4}$. We specify the initial compositions for these urns. Let M be the urn with L different colors of balls inside: $b_1,b_2,\dots,b_L>0$. The number of balls of color b_l could be specified as the number of counterparties for the l-th group. For each group l, there are four additional standard Pólya urns $V_{l1},V_{l2},V_{l3},V_{l4}$. These urns are used to model possible moves for a company: improve rating, decrease rating, stay in the same rating class, jump to default. All individual urns V_{lj}, where $1\le l\le L$ and $1\le j\le 4$, are initialized in the same way; there are w white balls and b black balls and the total number in the urn is $\beta=w+b$. Let U represent a multicolored urn with 4 different types of balls: c_1,c_2,c_3 and c_4. Set the same urn composition in all urns $U_1,\dots,U_T$; call α_j the number of the balls of color c_j, for any $1\le j\le 4$, and let the total number in this urn be $A=\sum_{j=1}^{4}\alpha_j$.

We then generate a stochastic process $\{(X_{tn},\hat{Y}_{tn},Z_{tn})\}$, which will be used to model the credit rating transitions for the counterparties in those L groups.

The mechanism is the following:

- for $t\in[1,T]$ do
  - Sample the urn M with replacement. If the color is b_i, denote $X_{t1}=i$, so that in the following we deal with a firm belonging to rating $i\in\{1,\dots,L\}$;
  - Associated with each credit rating i, there are four urns $V_{i1},V_{i2},V_{i3},V_{i4}$ with black and white balls inside. If $X_{t1}=i$ and $\hat{Y}_{t1}=j$, we sample from the urn V_{ij} to ask for the approval of this trend. If we pick a white ball, denote $Z_{t1}=1$

and reinforce a white ball into urn V_{ij}. The value 1 is an approval for the trend. If the ball is black, denote $Z_{t1} = 0$ and a black ball is reinforced into the urn V_{ij}.

– Repeat this mechanism, and obtain $(X_{t2}, \hat{Y}_{t2}, Z_{t2}), (X_{t3}, \hat{Y}_{t3}, Z_{t3}), \ldots$.

At present I am working on a full characterization of this non-trivial process.

REFERENCES

[1] E. B. Johnston Ross and L. Shibut, *What drives loss given default? Evidence from commercial real estate loans at failed banks,* SSRN Electronic Journal (2015), 10.2139/ssrn.2608442.

[2] J. Dermine and C. N. de Carvalho, *Bank loan losses-given-default: A case study,* Journal of Banking and Finance **30**, 1219 (2006), arXiv:NIHMS150003 .

[3] E. I. Altman, B. Brady, A. Resti, and A. Sironi, *The link between default and recovery rates: Theory, empirical evidence, and implications,* Journal of Business **78**, 2203 (2005).

[4] C. Han, *Modeling severity risk under PD–LGD correlation,* European Journal of Finance **23**, 1572 (2017).

[5] B. Bade, D. Rösch, and H. Scheule, *Default and recovery risk dependencies in a simple credit risk model,* European Financial Management **17**, 120 (2011).

[6] J. Derbyshire, *The siren call of probability: Dangers associated with using probability for consideration of the future,* Futures **88**, 43 (2017).

[7] H. Dickson and G. L. S. Shackle, *Ekonomisk Tidskrift*, Vol. 58 (Cambridge University Press, Cambridge, 1956) p. 250.

[8] E. Lybeck, *The Black Swan: The Impact of the Highly Improbable* (Random House, 2017) pp. 1–82.

[9] P. Cirillo, J. Hüsler, and P. Muliere, *Alarm systems and catastrophes from a diverse point of view,* Methodology and Computing in Applied Probability **15**, 821 (2013).

[10] C. Robert, *Chance*, Vol. 27 (The MIT Press, Cambridge MA, 2014) pp. 62–63.

[11] W. Knight, *The dark secret at the heart of AI,* Technology Review **120**, 54 (2017).

[12] D. Cheng and P. Cirillo, *A reinforced urn process modeling of recovery rates and recovery times,* Journal of Banking and Finance **96**, 1 (2018).

[13] L. Rüschendorf, *On the distributional transform, Sklar's theorem, and the empirical copula process,* Journal of Statistical Planning and Inference **139**, 3921 (2009).

[14] S. Figini and P. Giudici, *Statistical merging of rating models,* Journal of the Operational Research Society **62**, 1067 (2011).

[15] R. B. Nelsen, *Journal of the American Statistical Association*, Vol. 95 (Springer, New York, 2000) p. 334.

[16] P. Bochev and M. Gunzburger, *Least-squares finite element methods*, **166** (2014).

[17] D. Cheng and P. Cirillo, *An Urn-based nonparametric modeling of the dependence between PD and LGD with an application to mortgages*, Risks **7**, 76 (2019).

[18] S. Petrone and L. Wasserman, *Consistency of Bernstein polynomial posteriors*, Journal of the Royal Statistical Society. Series B: Statistical Methodology **64**, 79 (2002).

[19] J. C. Hull, *Risk management and financial institutions*, 4th ed. (Wiley, New York, 2012).

[20] A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools* (Princeton university press, 2005).

[21] A. McNeil and J. Wendin, *Dependent credit migrations*, The Journal of Credit Risk **2**, 87 (2006).

[22] G. D'Amico, S. Dharmaraja, R. Manca, and P. Pasricha, *A review of non-Markovian models for the dynamics of credit ratings*, Reports on Economics and Finance **5**, 15 (2019).

[23] G. D'Amico, G. Di Biase, J. Janssen, and R. Manca, *Semi-Markov Migration Models for Credit Risk* (John Wiley & Sons, 2017) pp. 1–297.

[24] G. D'Amico, J. Janssen, and R. Manca, *Initial and final backward and forward discrete time non-homogeneous semi-markov credit risk models*, Methodology and Computing in Applied Probability **12**, 215 (2010).

[25] A. Vasileiou and P. C. Vassiliou, *An inhomogeneous semi-Markov model for the term structure of credit risk spreads*, Advances in Applied Probability **38**, 171 (2006).

[26] G. D'Amico, J. Janssen, and R. Manca, *Discrete time non-homogeneous semi-Markov reliability transition credit risk models and the default distribution functions*, Computational Economics **38**, 465 (2011).

[27] G. D'Amico, J. Janssen, and R. Manca, *Downward migration credit risk problem: A non-homogeneous backward semi-Markov reliability approach*, Journal of the Operational Research Society **67**, 393 (2016).

[28] S. Peluso, A. Mira, and P. Muliere, *Reinforced urn processes for credit risk models*, Journal of Econometrics **184**, 1 (2015).

ACKNOWLEDGEMENTS

Finishing my PhD is a really tough journey for me and luckily I arrive.

During my high school period, I worked hard and entered the University of our province, in order to get rid of the hard life. I want to change my fate by learning knowledge. I was born in a small county in China. My mother, with her diligence, blocked a lot of wind and rain for me and gave me enough love. Even though, the most profound thing in my memory is that my mother went to the vegetable market to pick up the leaves that others didn't want, and took them home to raise chickens. At that time, as a high school student, I felt very powerless about this tough life. I could only accompany her to pick up free vegetables but I firmly told myself in my heart that I don't want to live my mother's life any more. I have to change my fate by knowledge. So I worked hard and entered Hubei University.

After walking out of the mountain, God seems to have heard my inner cry, which made me meet good tutors one after another. They changed my destiny. After entering Hubei University, I was lucky to meet my mentor, Prof. Xu, who has actively guided me all my life. He is a professional tutor of Chucai College of Hubei University. I still remember Prof. Xu's strict requirements for my studies and his concern for the future of my students. When I was an undergraduate, he took me to a seminar on abstract algebra for postgraduates. After I got the letter of acceptance from the summer camp of USTC, he told the dean of Chucai College that I was a very good student. After successfully entering USTC, I met Prof. Zhang. I still remember what he said in a seminar. The mathematical formula should not only be proved, but also understand the intuitive meaning of the mathematical formula. These things are so precious to a student who comes out of the poor mountain area. Under their positive influence, I bravely applied to study abroad for a PhD.

The study abroad for a PhD has taught me how to think out of the box, upgrading my cognition, trying to find my inner love. After going abroad, the problems to be solved became more and more complex. I needed to figure out the solutions myself more and more, since to be an independent researcher is one of the goals in my PhD. This is a process to create, in which I can see the change of my thinking. When I'm stuck in a certain place, I need to get out of it. It's very difficult. At this time, you may need to ponder, discuss with your supervisor, and discuss with your colleagues. In a word, this process needs to be completed by ourselves. In the process of communicating with my supervisor Pasquale (Prof. dr. Cirillo, which I thank for his work over this years), I found that due to cultural differences, we often have totally different views on the same thing, which strongly upgraded my cognition and world view. To avoid misunderstanding and achieve the purpose of solving problems, I had to learn good communication skills. In the process of solving problems, I encountered many difficulties which bring about a lot of pain, but the sense of achievement after completion made me feel that everything was worth it and could bring real big pleasure.

After becoming a strong person, don't forget to be gentle with the world. Chinese government gives me the opportunity to study abroad. My family, teachers, colleagues and friends give me very valuable suggestions and support so that I can stick with what I do. My everything, including this PhD, is because of them. Hence, I need to serve the society afterwards.

After those incredible years at TUD, it's time for me to say goodbye and move on to something new. I feel very proud I have been a part of the applied probability group at Delft Institute of Applied Mathematics and it was such a pleasure working here. I am grateful to the people whom I have met over the years, who have shared their knowledge and skills with me and I'll miss you all terribly.

Thank you to all my colleagues, Martina, Eni, Fracesca, Jasper and Kailun, Bruno, Lixue with whom I shared evenly joys and doubts both professional and existential. In particular, thank my colleague Bart for helping translate my English summary and proposition into dutch. The discussion with Martina, Eni, Francesca have been precious moments that were necessary to counter the loneliness that a Ph.D. candidate can face; I have been lucky to meet you. Thanks to Prof.dr. Frank Redig for his co-supervision, and to Prof.dr.ir. Guert Jongbloed and dr. Jiao Chen with whom I played games in the DIAM trip which took place a few years ago, and that I recall with sympathy. Also, I would like to thank Juan (Prof. Cai) for her kindness and encouragement: I learned a lot from her course Nonparametric Statistics and the EVA seminars she tutored. Moreover, a lot of thanks to Chunyan Zhu, with whom I spent two years on studying swimming from scratch and some time in sports center doing BBB. We improve and enjoy together for being a better self.

I would like to thank my roommates Xuerui Wang, Mengshi Yang, Jun Nie, Mingxue Zheng, Nan Jiang, and Zheyi Zeng, who teach me how to live in a daily life and motivates me to enjoy the life instead of working as a machine. Also a lot of thanks to my best friends Yujiao Xu, Jie Hu, Meiyun He, Shuqiao Fang, Yanze Yang, Tianzi Wang, Xiangrong Wang who are role models for me about how to be a better self.

CURRICULUM VITÆ

Dan CHENG

02-09-1991 Born in Xianning, China.

EDUCATION

2015 – 2022 PhD in Applied Probability. Delft University of Technology, the Netherlands
Promotor: Prof. Dr. Frank Redig
Supervisors: Prof. Dr. Pasquale Cirillo

2013–2015 Master in Probability. University of Science and Technology of China, China
Thesis: 1-D Stochastic Wave Equation Driven by Poisson White Noise
Supervisor: Prof. Dr. Tusheng Zhang

2009 – 2013 Bachelor in Mathematics. Hubei University, China
Supervisor: Prof. Dr. Yunge Xu