UNIVERSAL APPROXIMATION OF CREDIT PORTFOLIO LOSSES USING RESTRICTED BOLTZMANN MACHINES

GIUSEPPE GENOVESE, ASHKAN NIKEGHBALI, NICOLA SERRA, AND GABRIELE VISENTIN

Abstract. We introduce a new portfolio credit risk model based on Restricted Boltzmann Machines (RBMs), which are stochastic neural networks capable of universal approximation of loss distributions.

We test the model on an empirical dataset of default probabilities of 30 investment-grade US companies and we show that it outperforms commonly used parametric factor copula models – such as the Gaussian or the t factor copula models – across several credit risk management tasks. In particular, the model leads to better out-of-sample fits for the empirical loss distribution and more accurate risk measure estimations.

We introduce an importance sampling procedure which allows risk measures to be estimated at high confidence levels in a computationally efficient way and which is a substantial improvement over the Monte Carlo techniques currently available for copula models.

Furthermore, the statistical factors extracted by the model admit an interpretation in terms of the underlying portfolio sector structure and provide practitioners with quantitative tools for the management of concentration risk.

Finally, we show how to use the model for stress testing by estimating stressed risk measures (e.g. stressed VaR) for our empirical portfolio under various macroeconomic stress test scenarios, such as those specified by the FRB's Dodd-Frank Act stress test.

1. Introduction

Portfolio credit risk models play a fundamental role within many financial institutions, in particular for the estimation of capital requirements and the pricing of credit derivatives.

Despite the variety of portfolio models that have been proposed over the years (see [41] and [28] for an overview), in practice they all rely on the same conditional independence assumption, also known as the *conditional independence framework* [41, 57, 31]. More specifically, in all such models obligors are assumed to default independently given a set of underlying factors, which can be either latent (i.e. statistically estimated) or observable (e.g. macroeconomic indicators). Models of this kind are most generally known as mixture models.

The rationale behind mixture models is that factors are expected to provide a low-dimensional parametrization of the complex dependence structure of the entire portfolio, which can then be exploited for dimensionality reduction and the simplification of otherwise computationally infeasible procedures. These latent factors can furthermore be interpreted by practitioners in terms of systematic and idiosyncratic risk, thus connecting credit risk practice with the neighboring fields of portfolio selection and asset pricing.

Date: February 20, 2022.

Key words and phrases. Credit risk, machine learning, Restricted Boltzmann Machine, universal approximation, stress testing, risk measures.

Apart from this common modelling assumption, each individual credit risk model relies on particular parametric specifications of the factor distributions and of the functional dependence between factors and obligors' default probabilities. Nevertheless, these parametric choices are almost never motivated on empirical grounds and are often influenced by considerations of mathematical tractability or statistical expediency.

For example, in the case of CreditRisk+, factor distributions are chosen purposefully in such a way as to obtain a closed-form formula for the probability generating function of the total portfolio losses [12][10], without addressing the arbitrariness of this distributional choice.

Similarly, in Moody's public-firm EDF™ model the distance-to-default values are first computed using a classical Merton model methodology – which is well-known to underestimate short-term credit spreads [95, 52, 53] – and are only subsequently readjusted by calibrating the term structure on an historical database of default frequencies [86], in what amounts to a methodological patchwork.

Another infamous example of blind reliance on parametric assumptions comes from the widespread use of the Gaussian copula model, which is easy to fit – since linear correlations are sufficient statistics for the Gaussian copula – and can be used to derive closed-form expressions for the price of some credit derivatives, under additional simplifying assumptions [89]. Notoriously the one-factor Gaussian copula gained broad acceptance among practitioners and was adopted as the standardized method in the Basel II capital adequacy framework [7]. Perhaps more damagingly, it also became the *de facto* standard model for the pricing of mortgage-backed asset securities, such as CDOs, in the early 2000s [14], up to the 2007-2008 financial crisis.

This over-reliance on parametric assumptions and lack of model validation in credit risk has been repeatedly shown to result in severe model misspecification across a range of financial applications. For instance, in CDO pricing it was pointed out very early that the one-factor Gaussian copula is unable to reproduce the observed market prices for tranches of different seniority, leading to the so called *correlation smile* (see [14] for an assessment of the problem for various parametric portfolio models). Model misspecification is particularly pernicious in credit risk management, in which the choice of the wrong copula may lead to the underestimation of risk measures and thus adversely impact risk management [41, 28, 33].

A substantial amount of research has focused on addressing these concerns by identifying alternative parametric families that are believed to be better suited for credit risk applications. So, for instance, the one-factor Gaussian copula in [61] has been superseded by stochastic correlations models [3, 80], t copula models [4, 21, 62, 32, 43, 66, 81], double t copula models [54, 17], Clayton copula models [83, 82, 78, 64, 34], Marshall–Olkin copula models [23, 61, 90, 27, 37, 63], more general Archimedean copulas [72, 36], and combinations of all the aforementioned models, via very general constructions in which pair copulas are glued together, as done in vine copulas [8, 58, 1, 18].

Unsurprisingly, the question of how to choose the right copula has become a pressing practical problem for many practitioners and common recipes include trying to fit as many models as possible and to select the one with highest likelihood on any given dataset [28].

In this paper we want to address these modelling questions by pursuing another approach and investigating instead the use of non-parametric, universal approximators from the field of unsupervised machine learning. These models are able to fit any dependence structure and

therefore provide a way of mitigating the intractable - and ultimately data-dependent - issue of choosing the right parametric copula, by letting the data speak for itself. Crucially these models maintain appealing mathematical features of traditional credit risk model – namely their conditional independence structure – which can be exploited in numerous applications and substantially improve interpretability, as we will see.

While our approach can be successfully applied to many stochastic neural networks, such as Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs), in this paper we focus on a single model, the Restricted Boltzmann Machines (RBMs), in order to showcase as many applications to credit risk as possible.

Section 2 is a self-contained introduction to RBMs, with particular emphasis on their conditional independence structure and issues pertaining to the training of the model.

In Section 3 we introduce the credit RBM model and we see how its universal approximation property allows us to fit an empirical dataset of default probabilities in a non-parametric way, with substantial improvements over common copula models.

Section 4 is devoted to the fundamental issue of importance sampling for credit risk models, which is pivotal for the accurate estimation of risk measures. We exploit the fact that the distribution of the credit RBM model admits an explicit formula and develop an efficient importance sampling procedure for the estimation of large losses. This result represents a substantial improvement over current standards, since no importance sampling procedure is known for portfolios other than the Gaussian copula model.

In Section 5 we deal with the issue of the interpretability of the latent factors extracted from data by the RBM model. In particular we show how the receptive field of the hidden units can be used to identify sector-level dependencies within a credit portfolio. The dependencies can be graphically represented using a weighted graph, in which hidden units induce cliques of locally dependent obligors, and the sector structure can be extracted via modularity maximization.

Finally, in Section 6, we see how the RBM model can be used to perform stress tests on credit portfolios, by training on a joint dataset of default probabilities and macroeconomic factors. The impact of stress test scenarios on the total loss distribution can then be assessed by performing conditional blocked Gibbs sampling. As an illustrative example, we implement the Federal Reserve Board's (FRB) Dodd-Frank Act stress test of June 2020 on a real portfolio. More generally, this methodology allows a portfolio manager to quantify the impact of any macroeconomic scenarios on portfolio performanceand to compute stressed risk measures (e.g. stressed VaR) for use in executive settings.

2. Introduction to RBMs

2.1. **Model description.** Restricted Boltzmann Machines (RBMs) are undirected probabilistic graphical models used for learning probability distributions. They are an instance of energy based models introduced in the 1980s [2, 46, 45, 85] which can be trained efficiently [47] and played a fundamental role in the early stages of the deep learning revolution [48, 51]. What follows is a short introduction to the model, but we refer the interested reader to [49, 9, 29, 70] for more detailed treatments.

The units of an RBM are partitioned into two layers: the layer of visible units, $V = (V_1, V_2, \dots, V_n)$, encoding the distribution of the training data, and the layer of ancillary hidden units, $H = (H_1, H_2, \dots, H_m)$. The two layers are fully connected, while intra-layer

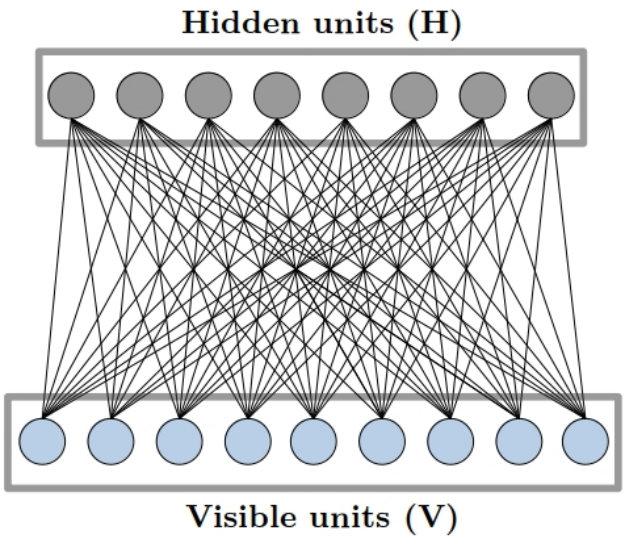


FIGURE 1. Graphical representation of a Restricted Boltzmann Machine.

connections are absent, thus leading to a bipartite network, as depicted in Figure 1. We associate to a RBM the following probability density:

$$p(v,h;\theta) = \frac{1}{Z} e^{-E(v,h;\theta)} p_V(v) p_H(h), \quad v \in \mathbb{R}^n, h \in \mathbb{R}^m, \tag{2.1}$$

where p_V and p_H denote the a priori distributions of the units belonging respectively to the visible and hidden layer and

$$E(v,h;\theta) = -h^T W v - b^T v - c^T h \tag{2.2}$$

is the so-called *energy function*. It is parametrized by a matrix of real-valued weights, $W = (W_{ji}) \in \mathbb{R}^{m \times n}$, and two bias vectors, b and c (for the visible and hidden units respectively), and we set $\theta = (W, b, c)$. The constant Z is simply a normalization constant called the *partition function* of the model.

One important remark is that the bipartite structure of an RBM implies a corresponding factorization of the distribution function, so that an RBM can be seen as a mixture model, in which visible units are conditionally independent given the values of all the hidden units, and viceversa. To see this, we specialize to the case of uniform a priori distributions and binary units, i.e. $v \in \{0,1\}^n$ and $h \in \{0,1\}^m$, as we will do for the rest of the paper, and compute the following conditional distributions:

$$\mathbb{P}(V_i = 1|H) = \sigma\left(\sum_{j=1}^{m} W_{ji} H_j + b_i\right), \quad \forall i = 1, \dots, n \tag{2.3}$$

$$\mathbb{P}(H_j = 1|V) = \sigma\left(\sum_{i=1}^{n} W_{ji} V_i + c_j\right), \quad \forall j = 1, \dots, m \tag{2.4}$$

where $\sigma(\cdot)$ denotes the sigmoid function.

We stress that the availability of an explicit formula for the joint distribution of the RBM's units is an important advantage over other generative models, such as Variational Autoencoders (VAEs) or generators trained via adversarial learning, as in Generative Adversarial Networks (GANs). In Section 4 and Section 6, we will see how to exploit this fact in the context of credit risk.

Nevertheless, sampling from the RBM distribution (2.1) can be computationally inefficient. Indeed an exact computation of the normalisation constant Z requires a number of operations that is exponential in the number of units, which makes exact sampling from the model distribution infeasible. One is thus led to the use of approximate methods, such as Markov chain Monte Carlo techniques.

Indeed Equations (2.3) and (2.4) naturally suggest the blocked Gibbs sampling procedure described in Algorithm 1.

Algorithm 1: Blocked Gibbs sampling for RBM.

Data: RBM with parameters (W, b, c); T = number of Gibbs steps.
Result: Sample (v, h) from RBM distribution.
Sample v from the uniform distribution on $\{0, 1\}^n$.
for $k = 1$ **to** T **do**
  Sample h given v using Equation (2.4).
  Sample v given h using Equation (2.3).
end
return (v, h).

This procedure is very efficient, since all units in the same layer can be updated simultaneously. This is in fact the principal advantage of RBMs when compared to other energy models with more complex graph topologies.

2.2. **Training.** RBMs can be trained using likelihood maximization via stochastic gradient ascent. The gradient of the log-likelihood function with respect to any model parameter $\theta = (W, b, c)$ satisfies the following general expression:

$$\frac{\partial \mathcal{L}(\theta|v)}{\partial \theta} = \mathbb{E}_{H|V}\left[\frac{\partial E(v,h)}{\partial \theta}\right] - \mathbb{E}_{V,H}\left[\frac{\partial E(v,h)}{\partial \theta}\right] \tag{2.5}$$

The first expectation, corresponding to the conditional distribution of H given V, can be evaluated directly on the dataset (or on a mini-batch of training samples). The second summand in (2.5) involves the distribution (2.1) and therefore it is hard to compute. We have to approximate this second term by a Monte Carlo method such as the blocked Gibbs sampling in Algorithm 1. But unfortunately this routine would have to be used at each paraemeter update, which makes overall training computationally demanding, as full convergence to the model distribution may require in practice as many as 10^3 Gibbs steps [56].

The breakthrough by Hinton [47] was to observe that the model can be efficiently trained by evaluating the second expectation in (2.5) by running Algorithm 1 for only few Gibbs steps at each parameter update and this training method came to be known as Contrastive Divergence (CD). In this paper we employ a popular improvement of the CD algorithm, namely Persistent Contrastive Divergence (PCD) [93, 88]. PCD is reported schematically

below, where for simplicity we present the version in which all the dataset is explored at each step.

Algorithm 2: k-steps Persistent Contrastive Divergence.

Data: Initial parameters $\theta_0 = (W_0, b_0, c_0)$; M datapoints $x_1, \dots, x_M \in \mathbb{R}^n$; T = number of iterations.

Result: Maximal likelihood estimator θ_T

Set $x_1^{[0]} = x_1 \dots x_M^{[0]} = x_M$

for $\ell = 1$ **to** T **do**

  for $j = 1$ **to** M **do**

    For $p(v, h; \theta_{\ell-1})$ run Algorithm 1 for k steps starting at $x_{[j]}^{[\ell-1]}$. Store the final sample $x_{[j]}^{[\ell]}$.

  end

  Compute the approximated log-likelihood gradient

$$\frac{1}{M} \sum_{j=1}^{M} \left(\mathbb{E}_{H|V=x_j} \left[\frac{\partial E(v,h)}{\partial \theta} \right] - \mathbb{E}_{H|V=x_j^{[\ell]}} \left[\frac{\partial E(v,h)}{\partial \theta} \right] \right) .$$

  Update $\theta_\ell \leftarrow \theta_{\ell-1}$.

end

3. The credit RBM model

To model a credit risk portfolio, we work with an RBM with binary units by identifying each visible unit with the default indicator function of an obligor in the portfolio (i.e. $V_i = \mathbb{1}_{\mathcal{D}_i}$ where $\mathcal{D}_i$ is the event that the i-th obligor defaults), so that the total number of defaults in the portfolio is given by $L_n = \sum_{i=1}^n V_i$. The hidden units will instead correspond to binary statistical latent factors, exactly as in a mixture model.

This model amounts in practice to a particular choice of parametrization for the obligors' default probabilities, as is clear from Eq. (2.3). Nevertheless, RBMs are capable of considerable representational power, which crucially depends on the number of hidden units. In particular, RBMs have been shown to be universal approximators for distributions on $\{0,1\}^n$ [59, 30, 71], which implies that the credit RBM model is capable of learning loss distributions with any dependence structure, which is to say any copula.

This flexibility allows us to dispense with specific parametric assumptions and can be shown empirically to yield better finite sample fits for portfolio loss distributions.

To test this claim we train a credit RBM model on an empirical dataset of daily one-year default probabilities for 30 publicly listed US companies from 3 January 2000 to 30 June 2020 (for a total of 5156 datapoints per company). More information on the dataset can be found in Appendix A. The dataset is partitioned into a training dataset (80% of datapoints) and a test dataset (20% of datapoints), reserved for out-of-sample performance evaluation.

In order to train the binary RBM on a dataset of default probabilities, we replace the visible units by their activation probabilities (2.3), as in [50].

The RBM is then trained on the training dataset using Persistent Contrastive Divergence (PCD) for 100 Gibbs steps, 5000 epochs, a linearly decreasing learning rate with initial value $\eta = 2 \cdot 10^{-3}$, and mini-batch size equal to 250.

The number of hidden units is here chosen by $k = 5$-fold cross-validation. Model selection for generative models is more complicated than in the case of supervised learning models, because probability distributions admit many possible and orthogonal performance metrics [87, 91, 15, 11]. In our case, we have chosen 250 hidden units by using the following metrics for model selection: KDE-based log-likelihood estimation [40, 87], Maximum Mean Discrepancy (MMD) [44], L^2 reconstruction error [92], and Number of statistically Different Bins (NDBs) [77].

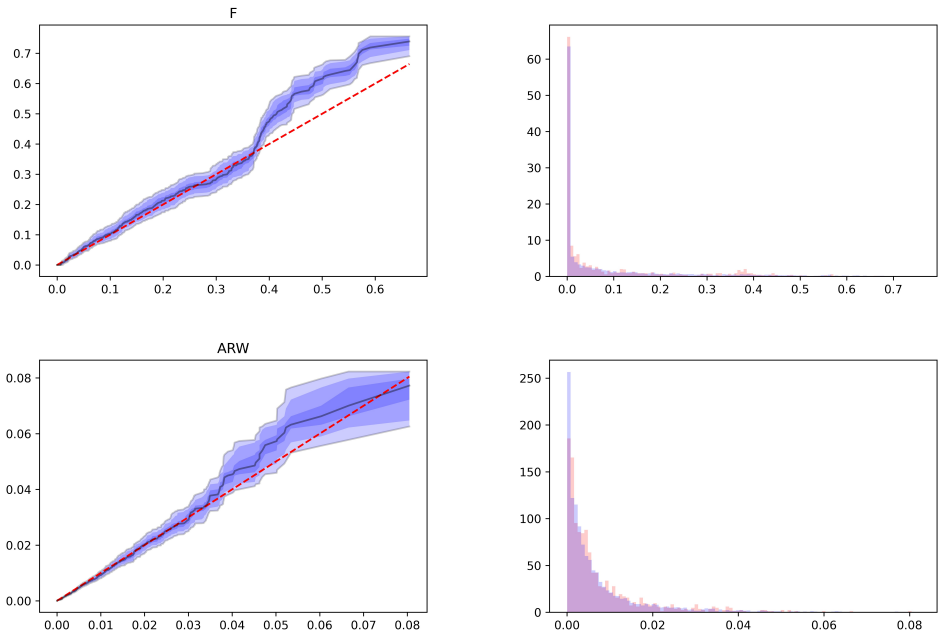


FIGURE 2. Out-of-sample comparison between empirical and RBM-generated default probabilities for Ford Motor Co. (ticker: F) and Arrow Electronics Inc. (ticker: ARW). Left: empirical QQ-plots with 99%, 95%, and 75% bootstrapped confidence intervals. Right: corresponding histograms with empirical data in red and RBM-generated data in blue.

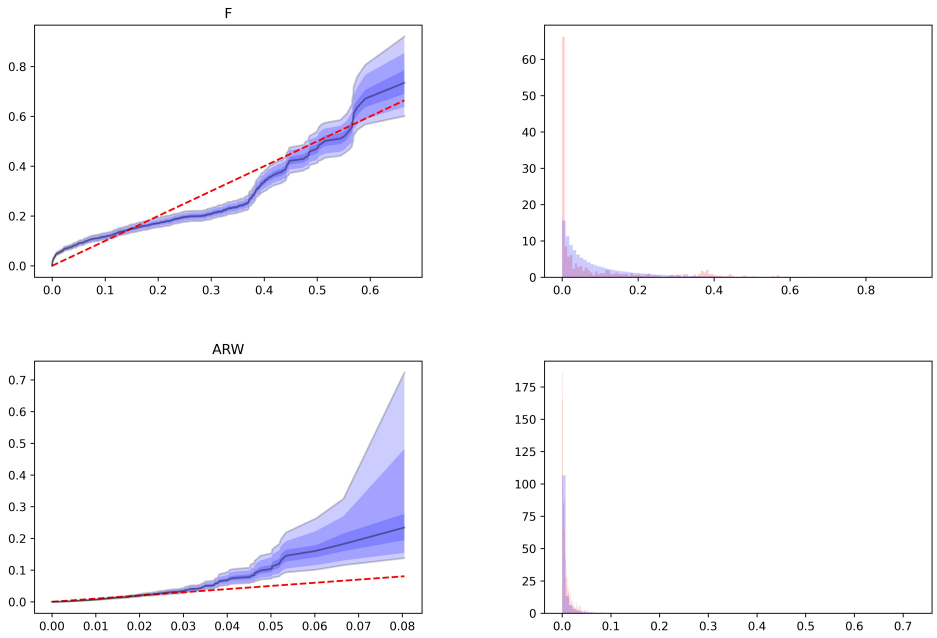


FIGURE 3. Out-of-sample comparison between empirical and one-factor Gaussian copula default probabilities for Ford Motor Co. (ticker: F) and Arrow Electronics Inc. (ticker: ARW). Left: empirical QQ-plots with 99%, 95%, and 75% bootstrapped confidence intervals. Right: corresponding histograms with empirical data in red and Gaussian copula data in blue.

In Figure 2 we compare out-of-sample the empirical marginal distributions of the default probabilities with the marginal distributions generated by the RBM model for two representative companies in the dataset, Ford Motor Co. and Arrow Electronics Inc.

As can be seen the marginal distributions are satisfactorily learned. For comparison, the same out-of-sample plots are shown in Figure 3 but for a one-factor Gaussian copula model with heterogeneous correlation coefficients, trained using likelihood maximization (see Appendix B for more details on this fitting procedure).

The one-factor Gaussian copula model appears to fit the tail of the marginal distributions slightly worse than the RBM model. But the difference in performance is much starker if we look at the distribution of the aggregate portfolio losses. This distribution is statistically more difficult to learn, since it requires accurate estimates not only of each individual marginal distribution, but also of their entire default dependence structure among firms.

Figure 4 shows a comparison of the out-of-sample estimation of the tail probability function of relative portfolio losses, using Monte Carlo simulations under the empirical real data distribution, the RBM model, a one-factor Gaussian copula, and a one-factor t copula.

As the figure shows, the Value at Risk of the portfolio under the two copula models is substantially underestimated up to confidence levels of 99.9%, while the finite sample properties of the RBM model match those of the empirical portfolio remarkably well out-of-sample.

This increase in performance is entirely due to the non-parametric nature of the RBM model and its universal approximation property, which makes it resilient to model misspecification.

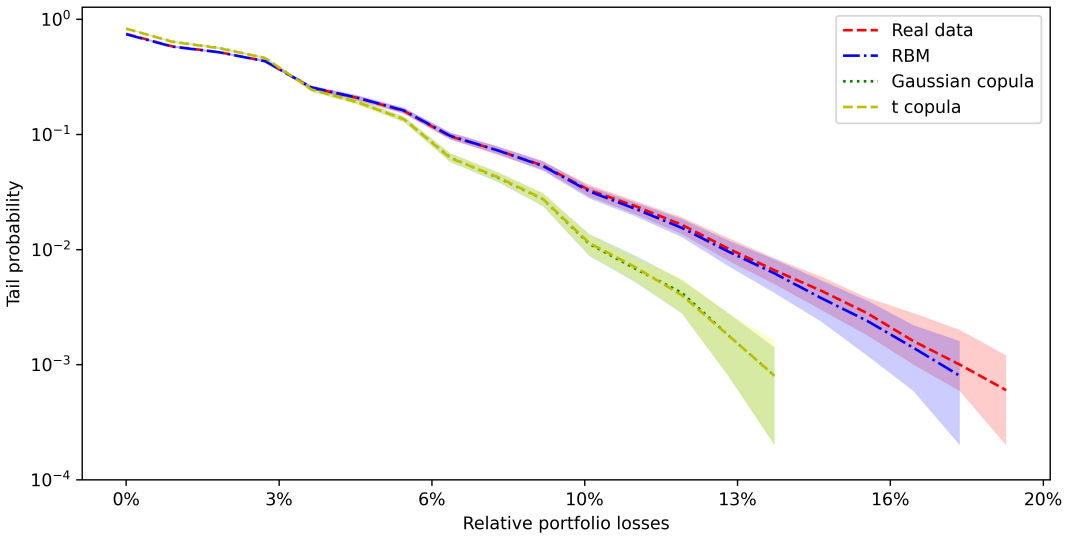


FIGURE 4. Out-of-samples comparison of Monte Carlo estimations of the tail function of relative portfolio losses, i.e. $\mathbb{P}(L_n/n > x)$. Shaded area denote 95% confidence intervals. Recovery rates are iid random variables following a Beta$(1/2, 1/2)$ distribution.

4. Importance Sampling for RBMs

4.1. **Introduction to importance sampling in credit risk.** In this section we will see how a credit RBM model can be used to perform importance sampling (IS) estimation of the full tail function of portfolio losses in a computationally efficient way.

It is well-known that, even in the case of the simplest parametric credit risk model, such as the one-factor Gaussian copula model with equicorrelation structure, it is not possible to derive a non-asymptotic closed-form formula for the portfolio loss distribution. Semi-analytical or advanced simulation-based techniques are the only solution for estimating risk measures or tail functions at very high confidence levels.

Examples of semi-analytical methods are the saddle-point approximation [65, 42], the Panjer recursion [54, 13], the Stein-Chen's method for Gaussian and Poisson approximation [25, 26], the large deviations principle [22] and, more recently, mod-ϕ approximation schemes [67].

As far as simulation-based methods are concerned, the starting point is the observation that a naive Monte Carlo approach is computationally expensive and is likely to underestimate risk measures, especially for high confidence levels. To overcome this problem, many authors have investigated the use of importance sampling techniques for portfolio credit risk models [5, 38, 39, 55, 68, 24]. The approaches developed so far apply exclusively to the Gaussian copula model and even in the Gaussian case they rely on approximate optimization, so that an extension to more general factor models appears difficult.

In the case of credit RBM models, instead, it is possible to perform efficient and accurate importance sampling estimation of the entire tail function of portfolio losses and it is therefore possible to compute risk measures at arbitrarily high confidence intervals.

4.2. **Importance sampling for RBMs.** In order to perform importance sampling for a credit RBM model, we need to find a change of measure from the distribution of the visible units

$$\mathbb{P} = \mathbb{E}_{H\,|\,V}[p(V, H; \theta)]$$

to a new distribution, $\mathbb{Q}$, under which losses are more frequent. It is natural to consider the change of measure suggested by the large deviations upper bound for $L_n = \sum_{i=1}^n V_i$, which leads to the following parametrized family $(\mathbb{Q}_t)_{t\in\mathbb{R}^+}$:

$$\frac{\mathrm{d}\mathbb{Q}_t}{\mathrm{d}\mathbb{P}}(v) = \frac{e^{t\sum_{i=1}^n v_i}}{\Gamma(t)}, \quad t \in \mathbb{R}^+$$

where $\Gamma(t) = \mathbb{E}\left[e^{t\sum_{i=1}^n V_i}\right]$ denotes the moment generating function of L_n. This is readily computed from Eq. (2.1) and (2.2) as:

$$\begin{aligned}
\Gamma(t) &= \mathbb{E}\left[e^{tL_n}\right] \\
&= \sum_{v\in\{0,1\}^n} \sum_{h\in\{0,1\}^m} \frac{1}{Z} \exp\left(t\sum_{i=1}^n v_i\right) \exp\left(-E(v,h)\right) \\
&= \sum_{v\in\{0,1\}^n} \sum_{h\in\{0,1\}^m} \frac{1}{Z} \exp\left(h^T W v + (b+tu)^T v + c^T h\right) \\
&= \frac{\tilde{Z}_t}{Z}, \qquad (4.1)
\end{aligned}$$

where $u \in \mathbb{R}^n$ is such that $u_i = 1$ for all $i = 1 \dots n$ and $\tilde{Z}_t$ denotes the partition function of an RBM with parameters $(W, b + tu, c)$.

To understand better the behavior of the tilted measure, $\mathbb{Q}_t$, we further notice that:

$$\begin{aligned}
\frac{\mathrm{d}\mathbb{Q}_t}{\mathrm{d}\mathbb{P}}(v) &= \frac{Z}{\tilde{Z}_t} \exp\left(t\sum_{i=1}^n v_i\right) \\
&= \frac{Z}{\tilde{Z}_t} \exp\left(t\sum_{i=1}^n v_i\right) \frac{\sum_{h\in\{0,1\}^m} \exp\left(-E(v,h)\right)}{\sum_{h\in\{0,1\}^m} \exp\left(-E(v,h)\right)} \\
&= \left(\frac{1}{\tilde{Z}_t} \sum_{h\in\{0,1\}^m} \exp\left(t\sum_{i=1}^n v_i - E(v,h)\right)\right) \Big/ \left(\frac{1}{Z} \sum_{h\in\{0,1\}^m} \exp\left(-E(v,h)\right)\right)
\end{aligned}$$

Therefore under the measure $\mathbb{Q}_t$ the visible units, $V = (V_1, \dots, V_n)$, follow the distribution of an RBM with parameters $(W, b + tu, c)$ and partition function $\tilde{Z}_t$.

This exponentially tilted RBM is identical to the original RBM in all but the visible bias parameter, which is shifted to the right by t and therefore results in higher portfolio losses, as can be deduced from Eq. (2.3) by recalling that the sigmoid function is strictly increasing.

If we are interested in estimating tail probabilities of the form $\mathbb{P}\left(\sum_{i=1}^n V_i > x\right)$ we can find the optimal tilting parameter, t^*, by minimizing the corresponding large deviations upper bound:

$$t^* = \inf_{t\in\mathbb{R}^+} \{\log\Gamma(t) - tx\} \implies t^* \text{ such that } \mathbb{E}_{\mathbb{Q}_{t^*}}\left[\sum_{i=1}^n V_i\right] = x \tag{4.2}$$

In practice it is not necessary to tilt the RBM for each value of x at each we want to evaluate the tail function. Rather, the entire probability tail function can be computed efficiently from a single tilted RBM (provided the tilting parameter, t^*, is sufficiently high), using the following IS estimator:

$$\mathbb{P}\left(\sum_{i=1}^n V_i > x\right) \approx \frac{1}{M}\sum_{\ell=1}^M \mathbb{1}_{\{\sum_{i=1}^n v_i^{(\ell)} > x\}} \exp\left(-t^*\sum_{i=1}^n v_i^{(\ell)}\right)\frac{\tilde{Z}_{t^*}}{Z} \tag{4.3}$$

where $(v^{(\ell)} = (v_1^{(\ell)}, \ldots, v_n^{(\ell)}))_{\ell=1}^M$ is a sample of visible units obtained from a tilted RBM with parameters $(W, b + t^*u, c)$.

4.3. **Estimation of ratios of partition functions.** One last numerical issue pertains to the computation of the ratio $\tilde{Z}_{t^*}/Z$ in Eq. (4.3), which in general cannot be done exactly (with the exception of very small portfolios) due to the computational intractability of the partition function.

An efficient estimation procedure relying on Annealed Importance Sampling (AIS) was presented in [79] and can be used to obtain fast and accurate estimations of this ratio.

To briefly introduce this procedure, let us start by considering the simple Monte Carlo estimator for this ratio obtained from Equation (4.1), which is as follows:

$$Z_{t^*}/Z \approx \frac{1}{M}\sum_{\ell=1}^M \exp\left(t^*\sum_{i=1}^M v_i^{(\ell)}\right), \tag{4.4}$$

where $(v^{(\ell)} = (v_1^{(\ell)}, \ldots, v_n^{(\ell)}))_{\ell=1}^M$ is a sample of visible units obtained from the RBM with parameters (W, b, c). This estimator may suffer from high variance, especially for high values of t^*, which is unfortunately exactly the regime we are interested in.

The AIS estimation procedure interprets the exponential term in Equation (4.4) as the ratio of two unnormalized RBM distributions:

$$Z_{t^*}/Z \approx \frac{1}{M}\sum_{\ell=1}^M \frac{\exp\left(t^*\sum_{i=1}^M v_i^{(\ell)} - E(v^{(\ell)}, h)\right)}{\exp\left(-E(v^{(\ell)}, h)\right)} \tag{4.5}$$

where the energy at the numerator belongs to the tilted RBM with parameters $(W, b + t^*u, c)$, while the one at the denominator to the RBM with parameters (W, b, c). Indeed if

evaluated on a Monte Carlo sampling from the RBM with parameters (W, b, c), the distributions of these two RBMs cannot in general be expected to be sufficiently close. This is the main reason of the high variance of the estimator.

Algorithm 3: Annealed Importance Sampling (AIS) estimation of Z_{t^*}/Z.

Data: $t^* \geq 0$; RBM with parameters (W, b, c); T temperatures, $(t_k)_{k=1}^T$, uniformly spaced on $[0, t^*]$, M number of AIS runs.
Result: Estimate of Z_{t^*}/Z with standard deviation.
for $j = 1$ **to** M **do**
  $\varepsilon^{(j)} \leftarrow 1$;
  for $k = 1$ **to** T **do**
    if $k = 1$ **then**
      Sample v^k from RBM (W, b, c) with random initialization;
    else
      Sample v^k from one-step Gibbs sampling of RBM $(W, b + t_k u, c)$ initialized with v^{k-1};
    end
    $\varepsilon^{(j)} \leftarrow \varepsilon^{(j)} \cdot \exp\left((t_k - t_{k-1}) \sum_{i=1}^n v_i^k\right)$
  end
end
Compute $\hat{\varepsilon} = \frac{1}{M} \sum_{j=1}^M \varepsilon^{(j)}$;
Compute $\hat{\sigma}^2 = \frac{1}{M-1} \sum_{j=1}^M (\varepsilon^{(j)} - \hat{\varepsilon})^2$;
return $\hat{\varepsilon}$ *with standard deviation* $\hat{\sigma}/\sqrt{M}$.

The solution put forward by the AIS procedure is to define a sequence of intermediate RBMs with parameters $(W, b+t_k u, c)$, where $(t_k)_{k=1}^T$ form a grid on $[0, t^*]$ of suitable thickness. The parameters $(t_k)_{k=1}^T$ are also called *temperatures* and allow us to interpolate smoothly from one distribution to the other.

The ratio in Equation (4.5) can then be approximated by a telescoping product of ratios, each computed for a pair of consecutive RBMs in the temperature sequence, (t_{k-1}, t_k), under the distribution of temperature t_k. In this way, the ratios are always computed on distributions that are very close to each other and the estimator is gradually tilted away from the original RBM to the fully tilted one, thus performing an annealed version of importance sampling. A complete description of the AIS estimation procedure can be found in Algorithm 3. As the number of temperatures and the number of AIS runs increase, the accuracy of the estimator increases. In practice, good results can be obtained for $T \approx 20000$ and even low values of M, such as $M \approx 100$.

4.4. **Performance of the IS estimator.** We test the IS estimator in Equation (4.3) by training the credit RBM model on a synthetic dataset of default probabilities generated from a one-factor Gaussian copula portfolio comprising 250 obligors with heterogeneous average default probabilities (uniformly distributed between 2% and 10%) and with equicorrelation coefficient $\rho = 0.2$.

The optimal tilting, t^*, is chosen in such a way as to shift the mean relative portfolio loss from 6% to approximately 9.5%. This can be accomplished simply by trial and error, tilting the RBM of an arbitrary value t and then evaluating the expected portfolio losses under this tilting via simple Monte Carlo sampling. For our choice of t^* the histogram of the relative portfolio losses is shifted to the right as shown in Figure 5.

We can then compare the IS estimator in Equation (4.3) with a simple Monte Carlo sampling from an RBM that has not been tilted. The results are shown in Figure 6. The two estimators agree on the initial portion of the tail, up to tail probabilities approximately of the order 10^{-3}. Subsequently the variance of the simple Monte Carlo estimator starts diverging and the estimator itself vanishes, while the IS estimator, using the same number of samples, continues to map the tail function up to probabilities of order almost 10^{-8}. Higher values of t^* can be used to map the entire tail function to any degree of accuracy in a similar way, with no need to increase the number of simulations.

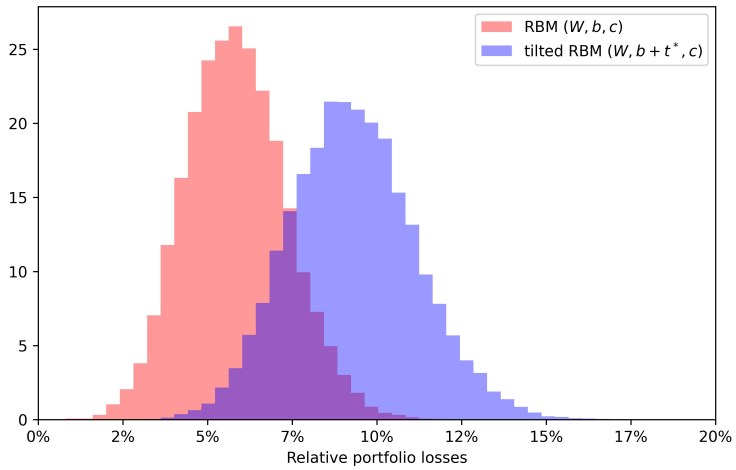


FIGURE 5. Histogram of relative portfolio losses for the original RBM and the tilted RBM.

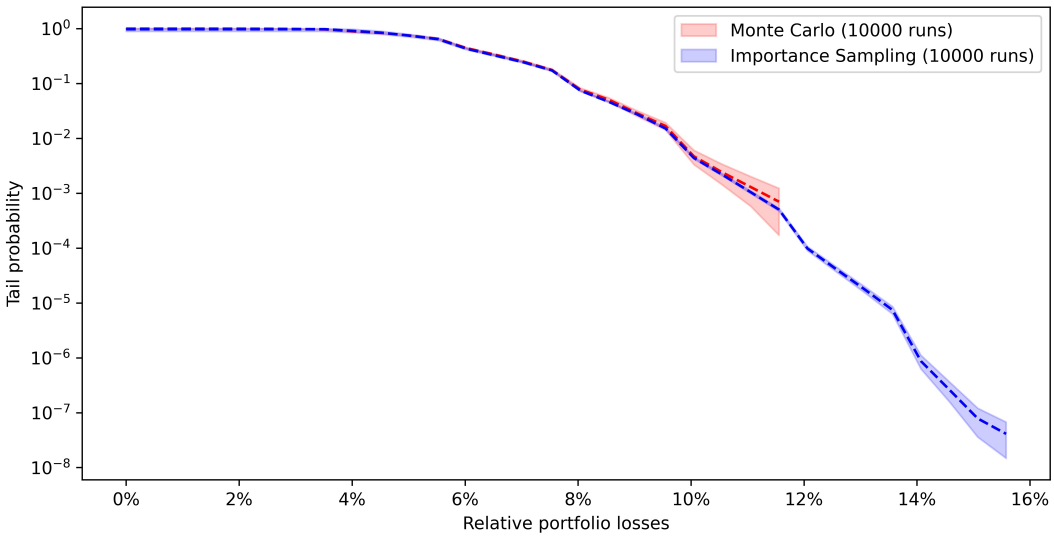


FIGURE 6. Estimated tail functions obtained by sampling from the RBM model using Monte Carlo simulations (in red) and using the importance sampling algorithm. Both estimators use 10000 samples. Colored shaded areas represent 95% confidence intervals.

5. Sector structure via hidden units' receptive fields

Latent factors in mixture models provide a low-dimensional representation of the dependence structure of the covariates. The detection of factors that model fine dependencies between subgroups of covariates is of particular importance in finance, as it plays a role in portfolio diversification (by identifying sources of concentration risk) and in portfolio optimization (simplifying many operations, such as covariance matrix inversion [20]). There is a rich literature in empirical finance concerned with the detection of sector structure from correlation matrices and its use in portfolio optimization [19, 94, 6, 60]. We will show how to design similar procedures for the RBM model, by leveraging the specifics of its architecture.

In the credit RBM model the latent factors are extracted from data during training and stored as random binary hidden units. In this section we see how a closer analysis of these hidden units can help us interpret the underlying statistical factors and gain more insight into the dependence structure of the data.

We train a credit RBM model on a synthetic dataset of default probabilities generated from a high-dimensional multi-factor Gaussian copula model with 106 factors and 100 obligors, divided in five groups or *sectors* of 20 obligors each.

More specifically, the default probability of each obligor depends on three factors: a global factor, which is common to all obligors in the portfolio; a sector factor, which is shared among all obligors of the same sector; and finally an idiosyncratic factor, which is unique to each obligor. The sample correlation matrix of the default probabilities is shown in Figure 7, where all obligors have been shuffled, in order to hide the block structure of the matrix. Our goal is to perform correlation clustering using a trained RBM. In other words, we want

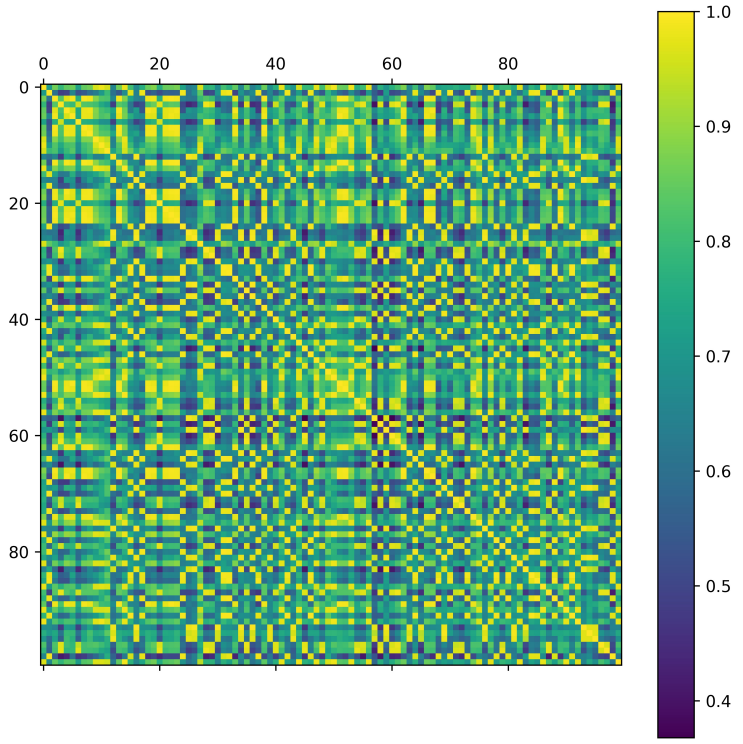


FIGURE 7. Correlation matrix of the artificially generated dataset with sector structure given by a multi-factor Gaussian copula.

to recover the way in which obligors are partitioned in sectors, so that when this partition is used to reorder the obligors in the sequence specified by its blocks, we can recover the correlation matrix in its original block configuration (up to permutations of the blocks).

The starting point is to analyze the so called *receptive fields* of the hidden units, to borrow a term from biology. For each hidden unit we want to identify which visible units tend to activate it more often. In particular we are interested in activation patterns that result in higher defaults, i.e. higher values of the visible units themselves.

From Equation 2.3 we can readily see that the j-th hidden unit leads to higher default probabilities when $H_j = 1$ and $W_{ji} > 0$, as well as when $H_j = 0$ and $W_{ji} < 0$, with stronger impact for higher values of $|W_{ji}|$. For a given hidden unit, say the j-th hidden unit, we can therefore suppose that it encodes information about high default values of obligor i if $|W_{ji}|$ takes a high value.

One way of displaying this quantity for the j-th hidden unit is to plot $|W_{ji}|$ as a function of $i = 1, \ldots, n$, normalized by its maximum value across all visible units, i.e. $\tilde{W}_{ji} := |W_{ji}| / \max_{i=1,\ldots,n} |W_{ji}|$. Figure 8 shows exactly these plots for four hidden units in the trained RBM.

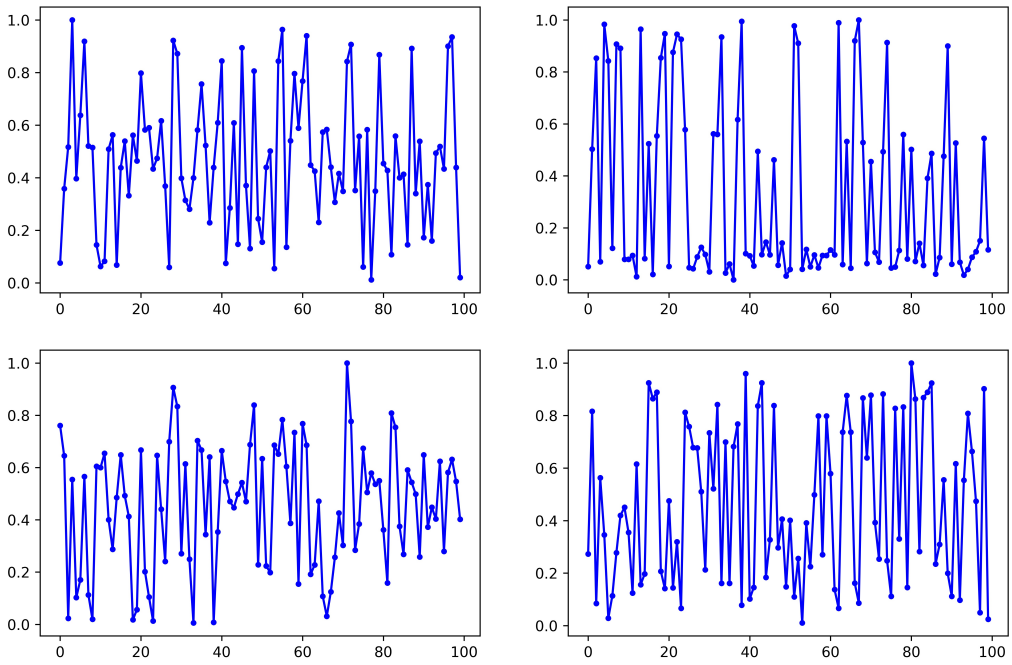


FIGURE 8. Plot of $\tilde{W}_{ji}$ for $i = 1, \ldots, n$ and four randomly selected hidden units.

The hidden units clearly display a broad activation pattern, ranging from zero (i.e. no sensitivity to the corresponding visible unit) to one (maximum sensitivity), with a few selective activation peaks, that correspond to specific subgroups of visible units. In other words, each hidden unit appears to be specialized in modelling a local dependence, corresponding to the default of only few obligors.

These local dependencies can be graphically represented in an undirected, weighted graph, $G = (V, E)$, with vertex set $V = \{1, \ldots, n\}$, corresponding to the set of portfolio obligors. The edges of the graph are added according to the following rule: given a fixed threshold, $\varepsilon \in (0, 1)$, we create an edge, $e = (l, k)$, between two vertices, l and k, if there is at least one hidden unit, $j \in \{1, \ldots, m\}$, such that both $\tilde{W}_{jl}$ and $\tilde{W}_{jk}$ are above $1 - \varepsilon$.

Graphically, for each hidden unit we threshold the values of $(\tilde{W}_{ji})_{i=1}^{n}$, as shown in Figure 9, and create edges between all pairs of visible units above this threshold. In this way each hidden unit induces a clique on the graph G.

Of course two different hidden units can give rise to the same edge. This information gets stored in the edge weight. More specifically, the number of hidden units that give rise to the same edge are counted in the edge weight, i.e. if $e = (l, k) \in E$, then

$$w_e = |\{j \in \{1, \ldots, m\} \ : \ \tilde{W}_{jl}, \tilde{W}_{jk} > 1 - \varepsilon\}|.$$

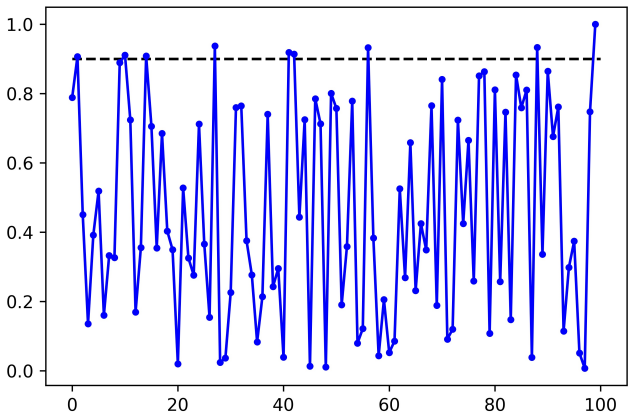


FIGURE 9. Thresholded plot of $\tilde{W}_{ji}$ for $i = 1, \ldots, n$ for $\varepsilon = 0.1$. The hidden unit is randomly chosen among the hidden units of the RBM.

This implies that visible units that belong to the receptive fields of many hidden units are connected by stronger edges.

As the threshold value ε is varied, the graph G will in general display different topologies. Figure 10 shows three possible choices of ε and the corresponding graphs. The nodes have been colored according to their sector under the true data model, which is an information that is not explicitly available to the RBM, of course.

For low values of ε, as in figure (A), we fail to capture the full receptive field of most hidden units, so that the resulting graph is highly disconnected and the many connected components do not form sectors.

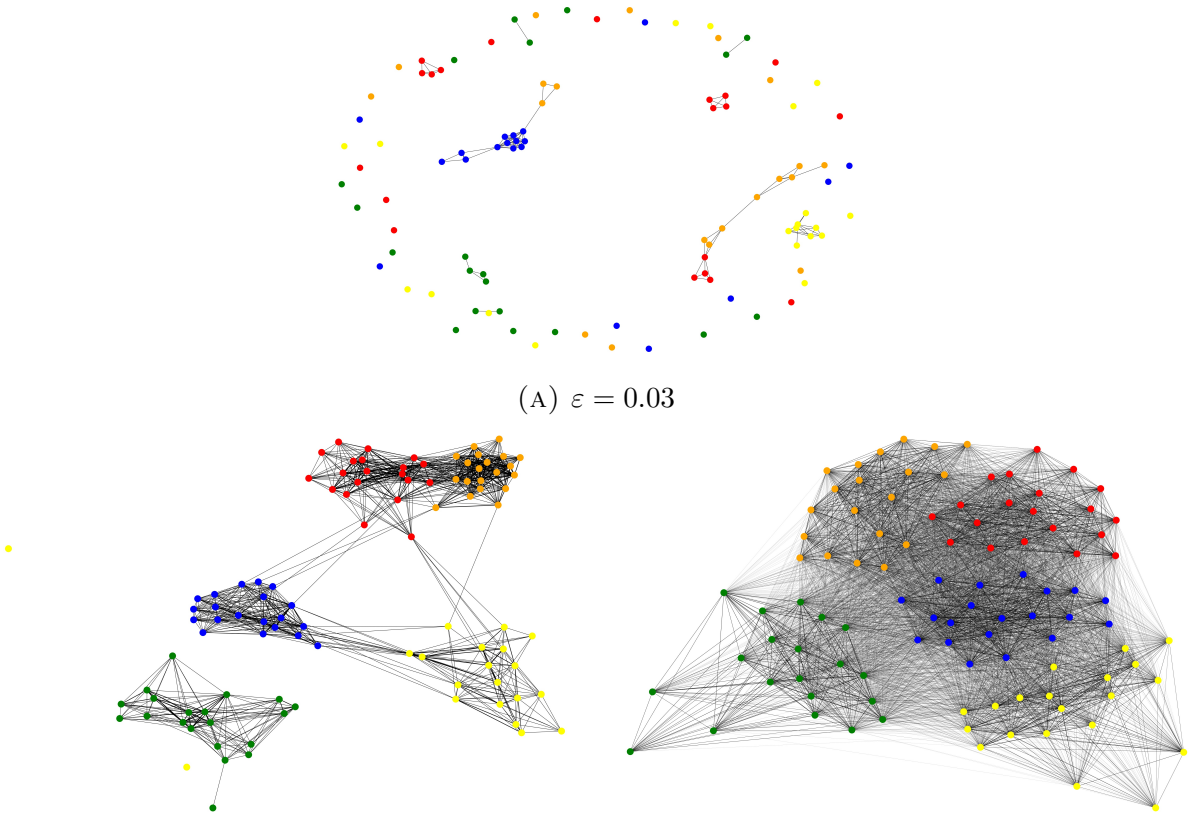
For intermediate values (see (B)) we see that the graph is able to identify the true sector structure of the data, since all nodes of the same color form tightly connected components, with only few and peripheral connection across sectors.

For high values of ε (see (C)), instead, the local nature of the hidden units' receptive field is lost as we include more and more global correlations and the sectors become more intertwined.

The graphical representations shown in Figure 10 are certainly useful, but in order to perform correlation clustering, we must be able to eventually extract a partition of the vertices.

The identification of a candidate partition corresponding to the underlying sector structure in G can be fully automated by using any community detection algorithm for weighted graphs. One natural choice is the well-known Newman's greedy modularity maximization algorithm [74, 73], of which a version, the Clauset-Newman-Moore's algorithm [16], has the added benefit of being very efficient also on extremely large graphs ($\approx 10^7$ vertices). These algorithms detect communities of tightly connected vertices and return as final output a partition of the vertices such that, loosely speaking, edges within the partition blocks are maximized and edges across partition blocks are minimized.

In Figure 11 we show the number of sectors detected via greedy modularity maximization on the graph G as ε is varied. By its very definition, modularity tends to result in many



(A) $\varepsilon = 0.03$

(B) $\varepsilon \approx 0.075$

(C) $\varepsilon = 0.25$

FIGURE 10. Representations of G for three values of the thresholding parameter ε. Nodes are colored according to their sector under the true data model. The graph layout is computed using the Fruchterman-Reingold force-directed algorithm [35].

communities both for highly disconnected graphs and for highly connected ones, so that the number of sectors as a function of ε has a global minimum and thus allows the identification of an optimal partition, corresponding to the minimum number of sectors.

On our dataset this minimum number of sectors is seven and is attained at $\varepsilon \approx 0.075$. The corresponding graph was already shown in Figure 10 (b) and it correctly identifies all five sectors, with the exception of the addition of two spurious sectors corresponding to two isolated yellow vertices (left and lower left of the image).

6. Stress testing via conditional Gibbs sampling

Stress testing is a risk management tool, adopted by financial institutions and regulatory authorities, which is used to assess the impact of a set of adverse macroeconomic scenarios on the performance of an institution's asset portfolio [76].

The first use of stress testing in banking regulation dates back to the Basel I accord [75], in which large banking institutions were required to conduct regular stress tests by simulating losses on their trading book under adverse macroeconomic scenarios, of a nature

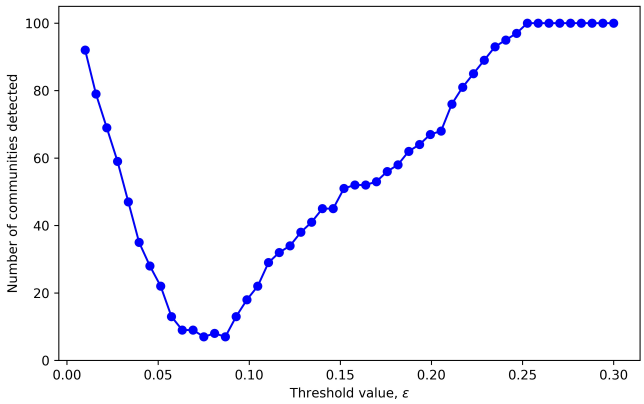


FIGURE 11. Number of communities detected on G using greedy modularity maximization, as a function of the threshold parameter, ε.

and magnitude similar to historically important market shocks. With Basel II [7] stress testing techniques started being applied also to risks other than market risk and played an important role for the determination of capital requirements. Finally, after 2007 several regulatory bodies began mandating regular stress test exercises. In 2012 the Federal Reserve Board (FRB) consolidated old practices into a new framework, which nowadays comprises the annual Dodd-Frank Act stress test and the Comprehensive Capital Analysis and Review (CCAR), the results of which are regularly released to the public and constitute a cornerstone of the supervisory activity of the FRB.

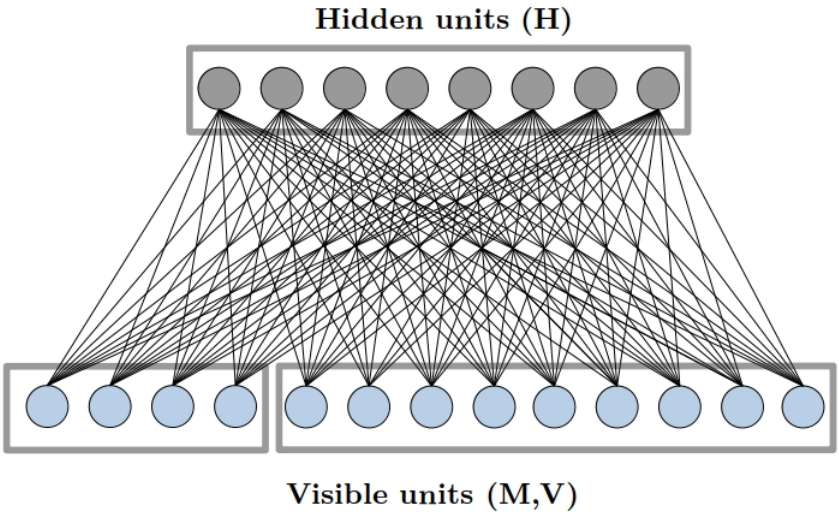


FIGURE 12. Restricted Boltzmann Machine for stress testing. The visible units consist of macroeconomic variables, M, and portofolio losses, V.

In this section we will focus on the FRB's implementation of the Dodd-Frank Act stress testing methodology, which is paradigmatic of virtually all stress testing methodologies implemented worldwide. Under this methodology a set of three macroeconomic scenarios is produced by independent economists: the baseline scenario, the severely adverse scenario, and the alternative severe scenario. Each scenario is specified in terms of 28 macroeconomic variables, divided in 16 domestic variables (e.g. US real and nominal GDP, US unemployment rate and disposable income, etc.) and 12 international variables (e.g. foreign currency exchange rates, real and nominal GDP growth in four different world areas, etc.). The individual financial institutions taking part in the exercise are then required to use their asset-specific, internal risk models (e.g. credit risk, market risk, and liquidity risk models) to assess the impact of each macroeconomic scenario on their overall portfolio profits and losses.

Mathematically speaking, if we denote by $M = (M_1, \ldots, M_k)$ the vector of macroeconomic variables used for scenario specification, a stress test exercise on a credit portfolio amounts to the evaluation of the conditional distribution of the total portfolio losses, L_n, given M. More specifically, one is interested in evaluating the conditional tail function, $\mathbb{P}\left(L_n > x|M\right)$, or the various conditional risk measures, such as, for instance, the conditional Value at Risk, $\mathrm{VaR}_\alpha(L_n|M)$, or even the conditional default probabilities of individual borrowers in a portfolio, $\mathbb{P}\left(V_i = 1|M\right)$.

The joint probability distribution of macroeconomic factors and portfolio losses can be learned by training an RBM with visible units (M, V) on a joint dataset of historical time series (see Figure 12). Once the training of the RBM is successfully completed, any functional of the distribution of V given a particular macroeconomic scenario $M = (m_1, \ldots, m_k)$ can be estimated by conditional blocked Gibbs sampling, which corresponds exactly to a standard blocked Gibbs sampling procedure, as in Algorithm 1, with the exception that the visible units corresponding to the macroeconomic variables, M, are clamped to the values corresponding to the relevant scenario, $(m_1, \ldots, m_k)$, at each iteration. As this Gibbs sampling is performed for sufficiently many steps, the corresponding Markov chain will converge to the conditional distribution of V given the scenario $M = (m_1, \ldots, m_k)$.

To illustrate the kind of analyses made possible by this procedure, we implement the FRB's Dodd-Frank stress test of June 2020 on the same empirical credit portfolio we used in Section 3 (see Appendix A for more information), augmented by the historical time series of quarterly values of the reference macroeconomic variables used by the FRB.

Figure 13 shows the empirical histograms of conditional portfolio losses under the three scenarios (baseline, severely adverse, and alternative severe) predicted by the FRB for 2020Q4. The severely adverse and alternative severe scenarios have a very similar impact on the portfolio loss distribution and both determine a clear shift to the right towards higher losses, with respect to the baseline scenario.

Conditional sampling from the RBM allows also the computation of stressed measures of risk, such as the Value at Risk, as depicted in Figure 14. The two riskier scenarios determine a statistically significant increase of the VaR at almost all confidence levels and the credit RBM model can be used to provide a quantitative estimate (with associated statistical uncertainty) of the additional capital required to cover the losses under each macroeconomic scenario.

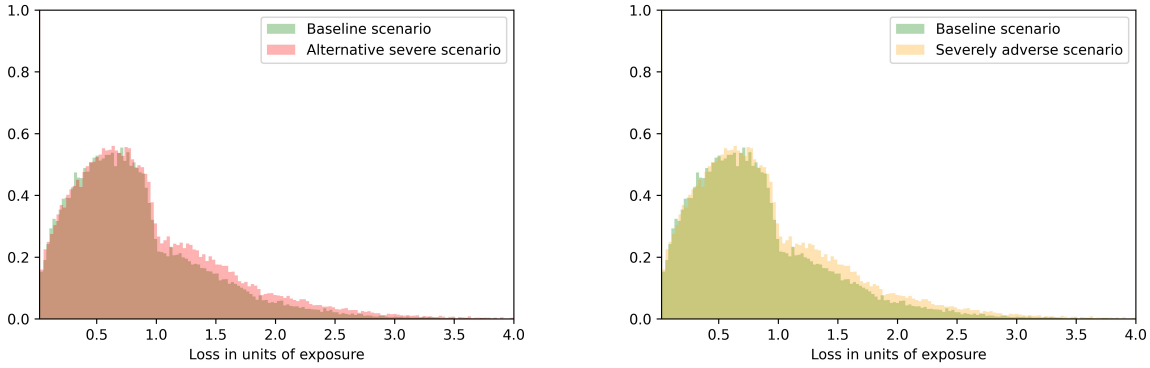


FIGURE 13. Histograms of conditional portfolio losses sampled under the three FRB's 2020Q4 scenarios. Comparison of the baseline scenario with the alternative severe scenario (left) and with the severely adverse scenario (right).

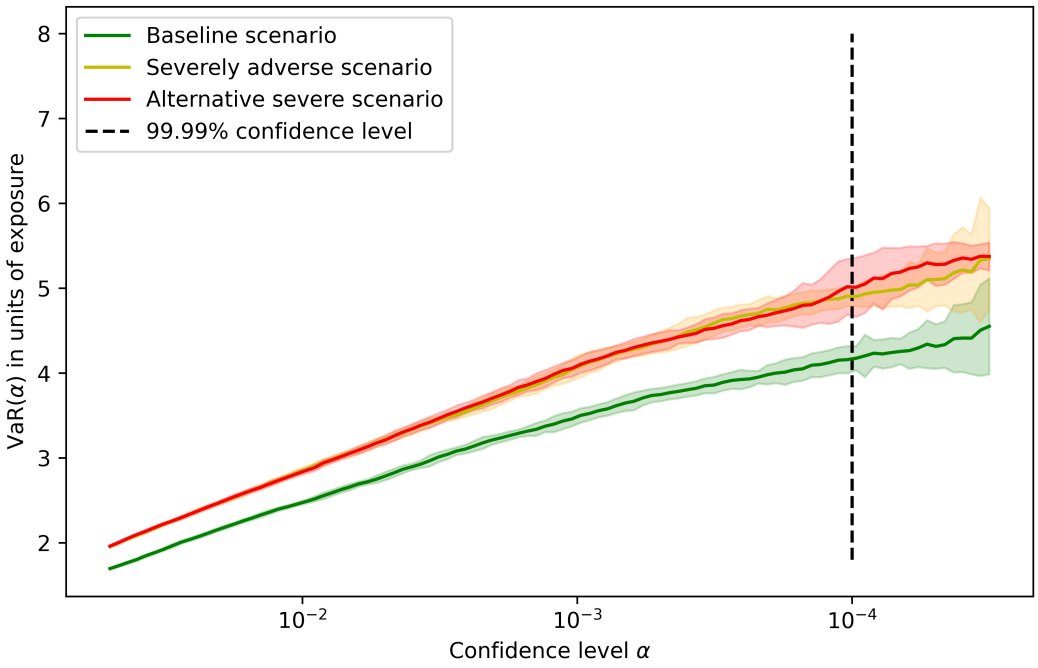


FIGURE 14. Estimation of stressed Value at Risk for different confidence levels and FRB's 2020Q4 stress test scenarios. For ease of reference, the vertical black line marks the 99.99% confidence level. All estimations are based on Monte Carlo simulations with 100000 runs. The colored areas represent 95% confidence intervals.

7. Conclusions

In this paper we have introduced a credit risk model based on Restricted Boltzmann Machines (RBMs) and investigated its performance across several financial tasks. RBMs are universal approximators for loss distributions and can thus provide a non-parametric alternative to the commonly used copula models, which are known to generate substantial model risk due to their over-reliance on parametric assumptions.

At the same time the model is shown to retain mathematically appealing properties, such as a conditional independence structure and an explicit formula for the model distribution, which make it amenable to many of the same mathematical techniques known for classical credit risk mixture models.

We have shown on an empirical dataset that the credit RBM model provides better out-of-sample fits of the empirical loss distribution than a Gaussian or t copula model and that, in particular, it provides better estimates for risk measures.

We have also introduced a novel importance sampling estimator for the tail probabilities of total portfolio losses under the model, thus showing that risk measures can be computed efficiently and accurately to any confidence level. This is a substantial advantage with respect to most other credit risk models, which typically must rely on computationally expensive Monte Carlo simulations.

It has been shown that the model can also be used for automated correlation clustering and sector detection, so that the latent factors extracted by the RBM model can be interpreted in terms of local dependencies among obligors. This could provide practitioners with an important tool for portfolio management and identification of concentration risk.

Finally, we have seen how the model naturally lends itself to the implementation of portfolio stress tests, by simulating three macroeconomic scenarios from the FRB's June 2020 Dodd-Frank Act stress test on an empirical portfolio.

Appendix A. Dataset description

The empirical dataset consists of daily one-year default probabilities for 30 investment-grade publicly listed US companies from 3 January 2000 to 30 June 2020 (for a total of 5156 datapoints per company).

The default probabilities have been estimated using the Merton's firm-value model [69]. According to this model the firm's market capitalization can be viewed as a call option on the firm's asset with strike equal to the firm's total liabilities. It follows that the (unobservable) market value of the firm's assets can then be computed at each point in time by inverting the corresponding Black-Scholes formula for the implied value of the underlying asset. This is a well-known estimation procedure, a complete description of which can be found in [28, Section 10.3].

Figure 15 shows the time series of the average default probability in the dataset. The peaks in default probability correspond to big economic and financial shocks, such as the September 11 attacks, the Venezuelan oil strike (end 2002, beginning 2003), the financial crisis of 2007-2008 and the onset of the Covid pandemic in 2020. The average default probability in the dataset across time and all firms is approximately 5,5%.

Data sources for all model inputs are listed below:

- Daily market capitalization data from Compustat (North America Daily - Security daily):

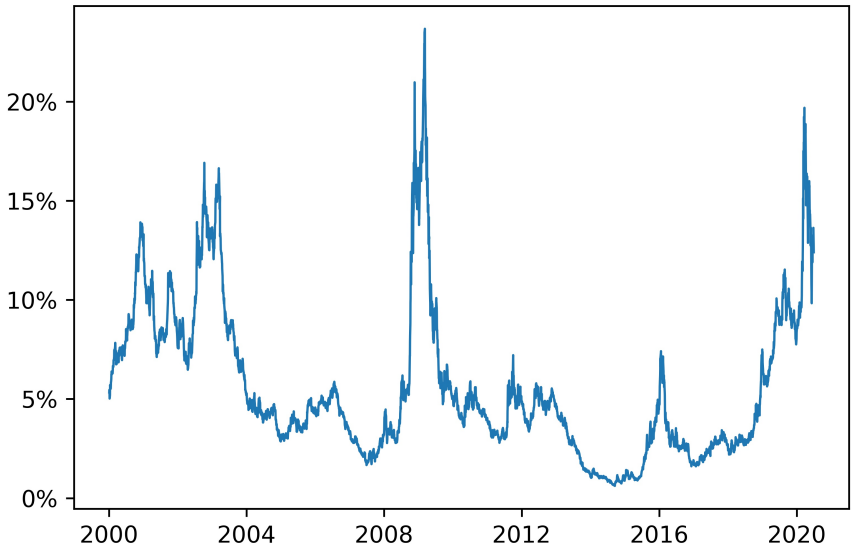


FIGURE 15. Time series of average default probability in the dataset.

  - PRCCD: close price
  - CSHOC: number of outstanding shares
- Balance sheet firms' liabilities data from Compustat (Capital IQ from Standard & Poor's - North America Daily - Fundamentals quarterly)
  - LLTQ: total non-current (i.e. maturity higher than one year) liabilities
  - LCTQ: total current (i.e. maturity up to one year) liabilities
- Risk-free interest rates estimated using the yield to maturity from end-of-day bid-ask average prices on the secondary market for one-month[1] US Treasuries. These data has been aggregated from two different sources:
  - 3 January 2000 to 31 December 2019 from CRSP Treasuries (Risk-free Series (4-week) - Daily): daily series of promised Daily Yield based on nominal price
  - 1 January 2020 to 17 August 2020 from the Federal Reserve's "H.15 - Selected Interest Rates" series, available on federalreserve.gov

APPENDIX B. STATISTICAL ESTIMATION OF FACTOR COPULAS

The most common estimation technique for copula factor models relies on the large homogeneous portfolio (LHP) approximation, which consists in assuming that the portfolio is sufficiently large to justify a Gaussian approximation for the losses distribution and that the obligors are homogeneous, in the sense that correlations can be assumed to be constant and identical across all pairs of obligors.

Under these assumptions the price of many credit instruments admits a closed-form expression (see [89] and [81] for computations in the one-factor Gaussian and t copula cases

[1]More precisely, we have used Treasuries with maturity within 22 and 28 days.

respectively) so that the model can be calibrated on market prices by choosing the equicorrelation coefficient that best fits the data.

This calibration technique typically leads to very poor fits, because a single equicorrelation coefficient is unable to reproduce the observed market prices [14].

In this paper we use instead a likelihood maximization procedure, which has the advantage of allowing heterogeneous correlation coefficients, thus removing the assumptions of the LHP approximation.

In a one-factor Gaussian copula the standardized asset value at maturity is given by:

$$X_i = a_i Z + \sqrt{1 - a_i^2}\, \varepsilon_i, \quad i = 1, \ldots, n$$

where Z is the systematic factor, ε is an firm-dependent noise term, and both both are assumed to be independent standard Gaussian random variables.

It follows therefore that the random vector $(X_1, \ldots, X_n)$ follows a multivariate Gaussian distribution with mean zero and variance Σ, where

$$\Sigma_{ij} = \mathbb{E}\left[X_i X_j\right] = \begin{cases} a_i a_j & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$$

which can be expressed in vector form as

$$\Sigma = aa^T + (I - \operatorname{diag}(aa^T)), \tag{B.1}$$

where $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$ and I denotes the $n \times n$ identity matrix.

Let $x = \{x^{[l]}, l = 1, \ldots, m\}$ be a sample of m realizations of the random vector $(X_1, \ldots, X_n)$, then the log-likelihood fuction of the model evaluated on this sample is given by:

$$\mathcal{L}(a|x) = -\frac{1}{2}\sum_{l=1}^{m})(x^{[l]})^T \Sigma^{-1} x^{[l]} - \frac{1}{2} m \log(|\Sigma|) \tag{B.2}$$

The matrix Σ^{-1} can be computed using the Sherman-Morrison formula [84] applied to Equation (B.1) yielding:

$$(\Sigma^{-1})_{ij} = \frac{\delta_{ij}}{1 - a_i^2} - \left(1 + \sum_{i=1}^{n} \frac{a_i^2}{1 - a_i^2}\right)^{-1} \left(\frac{a_i a_j}{(1 - a_i^2)(1 - a_j^2)}\right).$$

The computation of $|\Sigma|$ follows instead from applying the matrix determinant lemma:

$$\begin{aligned} |\Sigma| &= |I - \operatorname{diag}(aa^T)| + a^T \operatorname{Adj}(I - \operatorname{diag}(aa^T))\, a \\ &= \prod_{i=1}^{n}(1 - a_i^2) + \sum_{i=1}^{n} a_i^2 \prod_{\substack{j=1 \\ j \neq i}}^{n}(1 - a_j^2), \end{aligned}$$

where $\operatorname{Adj}(A)$ is the adjugate of matrix A, i.e. the transpose of its cofactor matrix.

Substituting these two expressions into Equation (B.2) yields a closed-form expression for the model's log-likelihood, which can be then maximized to obtain the maximum likelihood estimator of the factor correlations.

References

1. Kjersti Aas and Giovanni Puccetti, *Bounds on total economic capital: the dnb case study*, Extremes **17** (2014), no. 4, 693–715.
2. David H Ackley, Geoffrey E Hinton, and Terrence J Sejnowski, *A learning algorithm for boltzmann machines*, Cognitive science **9** (1985), no. 1, 147–169.
3. Leif Andersen and Jakob Sidenius, *Extensions to the gaussian copula: Random recovery and random factor loadings*, Journal of Credit Risk Volume **1** (2004), no. 1, 05.
4. Leif Andersen, Jakob Sidenius, and Susanta Basu, *All your hedges in one basket*, RISK-LONDON-RISK MAGAZINE LIMITED- **16** (2003), no. 11, 67–72.
5. Angelo Arvanitis and Jon Gregory, *Credit: The complete guide to pricing, hedging and risk management*, Risk books, 2004.
6. Nikhil Bansal, Avrim Blum, and Shuchi Chawla, *Correlation clustering*, Machine learning **56** (2004), no. 1, 89–113.
7. Basel Committee, *Basel ii revised international capital framework*, 2004.
8. Tim Bedford, Roger Cooke, et al., *Probabilistic risk analysis: foundations and methods*, Cambridge University Press, 2001.
9. Yoshua Bengio, Aaron Courville, and Pascal Vincent, *Representation learning: A review and new perspectives*, IEEE transactions on pattern analysis and machine intelligence **35** (2013), no. 8, 1798–1828.
10. Christian Bluhm, Ludger Overbeck, and Christoph Wagner, *Introduction to credit risk modeling*, Crc Press, 2016.
11. Ali Borji, *Pros and cons of gan evaluation measures*, Computer Vision and Image Understanding **179** (2019), 41–65.
12. Credit Suisse First Boston, *Creditrisk+: A credit risk management framework*, Tech. report, Technical report, Credit Suisse First Boston, 1997.
13. Hans-Juergen Brasch, *A note on efficient pricing and risk calculation of credit basket products*, Tech. report, Working paper, TD Securities. http://www. defaultrisk. com/pp _ crdrv . . . , 2004.
14. Xavier Burtschell, Jon Gregory, and Jean-Paul Laurent, *A comparative analysis of cdo pricing models*, preprint (2005).
15. Nutan Chen, Alexej Klushyn, Richard Kurle, Xueyan Jiang, Justin Bayer, and Patrick Smagt, *Metrics for deep generative models*, International Conference on Artificial Intelligence and Statistics, PMLR, 2018, pp. 1540–1550.
16. Aaron Clauset, Mark EJ Newman, and Cristopher Moore, *Finding community structure in very large networks*, Physical review E **70** (2004), no. 6, 066111.
17. Areski Cousin and Jean-Paul Laurent, *Comparison results for exchangeable credit risk portfolios*, Insurance: Mathematics and Economics **42** (2008), no. 3, 1118–1127.
18. Claudia Czado, *Pair-copula constructions of multivariate copulas*, Copula theory and its applications, Springer, 2010, pp. 93–109.
19. Marcos Lopez De Prado, *Building diversified portfolios that outperform out of sample*, The Journal of Portfolio Management **42** (2016), no. 4, 59–69.
20. ———, *Advances in financial machine learning*, John Wiley & Sons, 2018.
21. Stefano Demarta and Alexander J McNeil, *The t copula and related copulas*, International statistical review **73** (2005), no. 1, 111–129.
22. Amir Dembo, Jean-Dominique Deuschel, and Darrell Duffie, *Large portfolio losses*, Finance and Stochastics **8** (2004), no. 1, 3–16.
23. Darrel Duffie and Kenneth Singleton, *Simulating correlated defaults*, Bank of England London, 1998.
24. Daniel Egloff, Markus Leippold, Stephan Jöhri, and Curdin Dalbert, *Optimal importance sampling for credit portfolios with stochastic approximation*, Available at SSRN 693441 (2005).
25. Nicole El Karoui and Ying Jiao, *Stein's method and zero bias transformation for cdo tranche pricing*, Finance and Stochastics **13** (2009), no. 2, 151–180.
26. Nicole El Karoui, Ying Jiao, and David Kurtz, *Gauss and poisson approximation: applications to cdo tranches pricing*, (2008).

27. Y Elouerkhaoui, *Credit risk: Correlation with a difference*, Tech. report, working paper, UBS Warburg, 2003.
28. Paul Embrechts, Rdiger Frey, and Alexander McNeil, *Quantitative risk management.*, 2015.
29. Asja Fischer and Christian Igel, *Training restricted boltzmann machines: An introduction*, Pattern Recognition **47** (2014), no. 1, 25–39.
30. Yoav Freund and David Haussler, *Unsupervised learning of distributions of binary vectors using two layer networks*, (1994).
31. Rüdiger Frey and Alexander J McNeil, *Modelling dependent defaults*, Tech. report, ETH Zurich, 2001.
32. ———, *Dependent defaults in models of portfolio credit risk*, Journal of Risk **6** (2003), 59–92.
33. Rüdiger Frey, Alexander J McNeil, and Mark Nyfeler, *Copulas and credit models*, Risk **10** (2001), no. 111114.10.
34. Andrew Friend and Ebbe Rogge, *Correlation at first sight*, Economic Notes **34** (2005), no. 2, 155–183.
35. Thomas MJ Fruchterman and Edward M Reingold, *Graph drawing by force-directed placement*, Software: Practice and experience **21** (1991), no. 11, 1129–1164.
36. Christian Genest and Jock MacKay, *The joy of copulas: Bivariate distributions with uniform marginals*, The American Statistician **40** (1986), no. 4, 280–283.
37. Kay Giesecke, *A simple exponential model for dependent defaults*, The Journal of Fixed Income **13** (2003), no. 3, 74–83.
38. Paul Glasserman, *Monte carlo methods in financial engineering*, vol. 53, Springer, 2004.
39. Paul Glasserman and Jingyi Li, *Importance sampling for portfolio credit risk*, Management science **51** (2005), no. 11, 1643–1656.
40. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio, *Generative adversarial nets*, Advances in neural information processing systems **27** (2014).
41. Michael B Gordy, *A comparative anatomy of credit risk models*, Journal of Banking & Finance **24** (2000), no. 1-2, 119–149.
42. ———, *Saddlepoint approximation of creditrisk+*, Journal of banking & finance **26** (2002), no. 7, 1335–1353.
43. Andrei Greenberg, Roy Mashal, Marco Naldi, and Lutz Schloegl, *Tuning correlation and tail risk to the market prices of liquid tranches*, Lehman Brothers Quantitative Credit Research (2004).
44. Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola, *A kernel two-sample test*, The Journal of Machine Learning Research **13** (2012), no. 1, 723–773.
45. G E Hinton and T J Sejnowski, *Learning and relearning in boltzmann machines,*, Parallel distributed processing (David E Rumelhart and James L McClelland, eds.), MIT press Cambridge, MA, 1986, pp. 282–317.
46. Geoffrey E Hinton, *Analyzing cooperative computation*, Proc. of the 5th Annual Congress of the Cognitive Science Society, 1983, 1983.
47. ———, *Training products of experts by minimizing contrastive divergence*, Neural computation **14** (2002), no. 8, 1771–1800.
48. ———, *Learning multiple layers of representation*, Trends in cognitive sciences **11** (2007), no. 10, 428–434.
49. ———, *A practical guide to training restricted boltzmann machines*, Neural networks: Tricks of the trade, Springer, 2012, pp. 599–619.
50. Geoffrey E Hinton, Simon Osindero, and Yee-Whye Teh, *A fast learning algorithm for deep belief nets*, Neural computation **18** (2006), no. 7, 1527–1554.
51. Geoffrey E Hinton and Ruslan R Salakhutdinov, *Reducing the dimensionality of data with neural networks*, science **313** (2006), no. 5786, 504–507.
52. John C Hull and Alan D White, *Valuing credit default swaps i: No counterparty default risk*, The Journal of Derivatives **8** (2000), no. 1, 29–40.
53. ———, *Valuing credit default swaps ii: Modeling default correlations*, The Journal of derivatives **8** (2001), no. 3, 12–21.
54. ———, *Valuation of a cdo and an n-th to default cds without monte carlo simulation*, The Journal of Derivatives **12** (2004), no. 2, 8–23.

55. Michael Kalkbrener, Hans Lotter, and Ludger Overbeck, *Sensible and efficient capital allocation for credit portfolios*, Risk **17** (2004), no. 1, S19–S24.
56. Alexei Kondratyev and Christian Schwarz, *The market generator*, Available at SSRN 3384948 (2019).
57. H Ugur Koyluoglu and Andrew Hickman, *Reconcilable differences*, Risk **11** (1998), no. 10, 56–62.
58. Dorota Kurowicka and Roger M Cooke, *Uncertainty analysis with high dimensional dependence modelling*, John Wiley & Sons, 2006.
59. Nicolas Le Roux and Yoshua Bengio, *Representational power of restricted boltzmann machines and deep belief networks*, Neural computation **20** (2008), no. 6, 1631–1649.
60. Diego León, Arbey Aragón, Javier Sandoval, Germán Hernández, Andrés Arévalo, and Jaime Niño, *Clustering algorithms for risk-adjusted portfolio construction*, Procedia Computer Science **108** (2017), 1334–1343.
61. David X Li, *On default correlation: A copula function approach*, The Journal of Fixed Income **9** (2000), no. 4, 43–54.
62. Filip Lindskog et al., *Modelling dependence with copulas and applications to risk management*, Swiss Federal Institute of Technology Zurich (2000).
63. Filip Lindskog and Alexander J McNeil, *Common poisson shock models: applications to insurance and credit risk modelling*, ASTIN Bulletin: The Journal of the IAA **33** (2003), no. 2, 209–238.
64. Dilip B Madan, Michael Konikov, and Mircea Marinescu, *Credit and basket default swaps*, Journal of Credit Risk **2** (2004), no. 1, 67–87.
65. R. Martin, K. Thompson, and C. Browne, *Taking to the saddle*, Risk **14** (2001), no. 6, 91–94.
66. Roy Mashal and A Zeevi, *Inferring the dependence structure of financial assets: empirical evidence and implications*, Tech. report, working paper, University of Columbia, 2003.
67. Pierre-Loïc Meliót, Ashkan Nikeghbali, and Gabriele Visentin, *Mod-poisson approximation schemes: applications to credit risk*, (in preparation) (2022).
68. Sandro Merino and Mark A Nyfeler, *Applying importance sampling for estimating coherent credit risk contributions*, Quantitative Finance **4** (2004), no. 2, 199.
69. Robert C Merton, *On the pricing of corporate debt: The risk structure of interest rates*, The Journal of finance **29** (1974), no. 2, 449–470.
70. Guido Montúfar, *Restricted boltzmann machines: Introduction and review*, Information Geometry and Its Applications IV, Springer, 2016, pp. 75–115.
71. Guido Montufar and Nihat Ay, *Refinements of universal approximation results for deep belief networks and restricted boltzmann machines*, Neural computation **23** (2011), no. 5, 1306–1319.
72. Roger B Nelsen, *An introduction to copulas*, Springer Science & Business Media, 2007.
73. Mark EJ Newman, *Analysis of weighted networks*, Physical review E **70** (2004), no. 5, 056131.
74. ———, *Fast algorithm for detecting community structure in networks*, Physical review E **69** (2004), no. 6, 066133.
75. Basel Committee on Banking Supervision, *Amendment to the capital accord to incorporate market risks*, 1996.
76. Adrian Pop, *Stress testing in banking: A critical review*, Preparing for the Next Financial Crisis: Policies, Tools and Models (Esa Jokivuolle and Radu Tunaru, eds.), Cambridge University Press, 2017, pp. 89–107.
77. Eitan Richardson and Yair Weiss, *On gans and gmms*, arXiv preprint arXiv:1805.12462 (2018).
78. Ebbe Rogge and Philipp J Schönbucher, *Modelling dynamic portfolio credit risk*, (2003).
79. Ruslan Salakhutdinov and Iain Murray, *On the quantitative analysis of deep belief networks*, Proceedings of the 25th international conference on Machine learning, 2008, pp. 872–879.
80. L Schloegl, *Modelling correlation skew via mixing copulae and uncertain loss at default*, Presentation at the Credit Workshop, Isaac Newton Institute, 2005.
81. Lutz Schloegl and Dominic O'Kane, *A note on the large homogeneous portfolio approximation with the student-t copula*, Finance and Stochastics **9** (2005), no. 4, 577–584.
82. P Schönbucher, *Taken to the limit: simple and not-so-simple loan loss distributions*, The Best of Wilmott **1** (2002), 143–160.
83. Philipp J Schönbucher and Dirk Schubert, *Copula-dependent default risk in intensity models*, Working paper, Department of Statistics, Bonn University, Citeseer, 2001.

84. Jack Sherman and Winifred J Morrison, *Adjustment of an inverse matrix corresponding to a change in one element of a given matrix*, The Annals of Mathematical Statistics **21** (1950), no. 1, 124–127.
85. Paul Smolensky, *Information processing in dynamical systems: Foundations of harmony theory*, Tech. report, Colorado Univ at Boulder Dept of Computer Science, 1986.
86. Zhao Sun, David Munves, and David Hamilton, *Public firm expected default frequency (edf) credit measures: Methodology, performance, and model extensions*, Moody's Analytics Modeling Methodology **1** (2012), no. 6, 16.
87. Lucas Theis, Aäron van den Oord, and Matthias Bethge, *A note on the evaluation of generative models*, arXiv preprint arXiv:1511.01844 (2015).
88. Tijmen Tieleman, *Training restricted boltzmann machines using approximations to the likelihood gradient*, Proceedings of the 25th international conference on Machine learning, 2008, pp. 1064–1071.
89. Oldrich Vasicek, *The distribution of loan portfolio value*, Risk **15** (2002), no. 12, 160–162.
90. Dennis Wong, *Copula from the limit of a multivariate binary model*, J. Am. Stat. Assoc **4** (2000), 35–67.
91. Yuhuai Wu, Yuri Burda, Ruslan Salakhutdinov, and Roger Grosse, *On the quantitative analysis of decoder-based generative models*, arXiv preprint arXiv:1611.04273 (2016).
92. Sitao Xiang and Hao Li, *On the effects of batch and weight normalization in generative adversarial networks*, arXiv preprint arXiv:1704.03971 (2017).
93. Laurent Younes, *On the convergence of markovian stochastic algorithms with rapidly decreasing ergodicity rates*, Stochastics: An International Journal of Probability and Stochastic Processes **65** (1999), no. 3-4, 177–228.
94. Hannah Cheng Juan Zhan, William Rea, and Alethea Rea, *An application of correlation clustering to portfolio diversification*, arXiv preprint arXiv:1511.07945 (2015).
95. Chunsheng Zhou, *The term structure of credit spreads with jump risk*, Journal of Banking & Finance **25** (2001), no. 11, 2015–2040.

Institute of Mathematics, Universität Zürich, Switzerland

Institute of Mathematics, Universität Zürich, Switzerland

Department of Physics, Universität Zürich, Switzerland

Institute of Mathematics, Universität Zürich, Switzerland