UNIVERSITY OF AMSTERDAM

UvA-DARE (Digital Academic Repository)

Contagious defaults in a credit portfolio: a Bayesian network approach

Anagnostou, I.; Sanchez Rivero, J.; Sourabh, S.; Kandhai, D.

DOI
10.2139/ssrn.3446615
10.21314/JCR.2020.257

Publication date
2020

Document Version
Submitted manuscript

Published in
Journal of Credit Risk

Link to publication

Citation for published version (APA):
Anagnostou, I., Sanchez Rivero, J., Sourabh, S., & Kandhai, D. (2020). Contagious defaults in a credit portfolio: a Bayesian network approach. *Journal of Credit Risk*, *16*(1), 1-26. https://doi.org/10.2139/ssrn.3446615, https://doi.org/10.21314/JCR.2020.257

General rights
It is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), other than for strictly personal, individual use, unless the work is under an open content license (like Creative Commons).

Disclaimer/Complaints regulations
If you believe that digital publication of certain material infringes any of your rights or (privacy) interests, please let the Library know, stating your reasons. In case of a legitimate complaint, the Library will make the material inaccessible and/or remove it from the website. Please Ask the Library: https://uba.uva.nl/en/contact, or a letter to: Library of the University of Amsterdam, Secretariat, P.O. Box 19185, 1000 GD Amsterdam, The Netherlands. You will be contacted as soon as possible.

UvA-DARE is a service provided by the library of the University of Amsterdam (https://dare.uva.nl)

Download date:12 Oct 2026

Contagious defaults in a credit portfolio: a Bayesian network approach

Ioannis Anagnostou*[1,2], Javier Sánchez Rivero[2], Sumit Sourabh[1,2], and Drona Kandhai[1,2]

[1]Computational Science Lab, University of Amsterdam, Science Park 904, Amsterdam 1098XH, The Netherlands
[2]Quantitative Analytics, ING Bank, Foppingadreef 7, Amsterdam 1102BD, The Netherlands

August 5, 2019

Abstract

Robustness of credit portfolio models is of great interest for financial institutions and regulators, since misspecified models translate to insufficient capital buffers and a crisis-prone financial system. In this paper, we propose a method to enhance credit portfolio models based on the model of Merton by incorporating contagion effects. While in most models the risks related to financial interconnectedness are neglected, we use Bayesian network methods to uncover the direct and indirect relationships between credits, while maintaining the convenient representation of factor models. A range of techniques to learn the structure and parameters of financial networks from real Credit Default Swaps (CDS) data is studied and evaluated. Our approach is demonstrated in detail in a stylized portfolio and the impact on standard risk metrics is estimated.

Keywords – Portfolio Credit Risk; Bayesian Learning; Credit Default Swaps; Default Contagion; Probabilistic Graphical Models; Network Theory

1 Introduction

In recent years, there has been an increasing interest in modelling credit risk by practitioners as well as academics (see e.g., Gregory, 2015, Green, Kenyon, & Dennis, 2014, Sourabh, Hofer, & Kandhai, 2018, De Graaf, Feng, Kandhai, & Oosterlee, 2014, de Graaf, Kandhai, & Reisinger, 2018, Simaitis, de Graaf, Hari, & Kandhai, 2016, Anagnostou & Kandhai, 2019). Portfolio credit risk models are concerned with the occurrence of large losses due to defaults or deteriorations in credit quality. In practice, these models have a wide range of applications, such as regulatory and economic capital measurements, portfolio management, and risk-adjusted pricing. The robustness of such models is of great interest both for financial institutions and regulators, since misspecified models could lead to insufficient capital buffers, which in turn would result in a crisis-prone financial system and the need for regular bail-outs.

The key challenge in portfolio credit risk modelling is the incorporation of default dependence. Joint defaults of many issuers to which a portfolio is exposed to may lead to extreme losses. Therefore, understanding the relationship between default events is crucial. In most portfolio credit risk models existing in the literature, defaults of individual issuers depend on a set of common underlying risk factors, describing the state of a sector, region, or the economy as a whole. Notable examples of this approach are the Asymptotic Single Risk Factor (ASRF) model (Gordy, 2003) in the Basel regulations and industrial adaptations of Merton model (Merton, 1974) such as the CreditMetrics (JP Morgan, 1997) and KMV models (Bohn & Kealhofer, 2001; Crosbie & Bohn, 2002).

*i.anagnostou@uva.nl

Electronic copy available at: https://ssrn.com/abstract=3446615

Many researchers have challenged the claim that default dependence can be fully explained by dependence on common underlying factors, on the grounds that such models often fail to capture default clustering that does occur from time to time. Schönbucher & Schubert, 2001 suggest that in most cases the default correlations that can be achieved with common factors are not as high as the ones in empirical data. Das, Duffie, Kapadia, & Saita, 2007 perform statistical tests and reject the hypothesis that factor correlations can sufficiently explain the empirically observed default correlations. Thus, it becomes evident that an additional channel of default dependence needs to be considered.

Besides dependence on common factors, joint defaults might occur as a result of direct links between issuers, a phenomenon known as contagion. Davis & Lo, 2001 were among the first ones who introduced contagion in credit risk models, by considering that any default might infect to any other issuer in the portfolio. Jarrow & Yu, 2001 tried to generalize already existing models, included particular specifications of the issuers and focused on their effect on bonds and credit derivatives. Egloff, Leippold, & Vanini, 2007 introduced network theory to allow for a variety of infections, however the model required detailed information making its application more difficult than expected.

Following the financial crisis, there has been a significant interest in using network-based methods for financial stability and systemic risk (see e.g., Battiston, Gatti, Gallegati, Greenwald, & Stiglitz, 2012, Cont, Moussa, & Santos, 2013, Squartini, Van Lelyveld, & Garlaschelli, 2013, Battiston et al., 2016, Poledna, Thurner, Farmer, & Geanakoplos, 2014, Musmeci, Nicosia, Aste, Di Matteo, & Latora, 2017). Next to network models, there is a growing literature on particle systems with mean-field interaction, considered, e.g., in Hambly, Ledger, & Sojmark, 2018; Hambly & Sojmark, 2018; Kaushansky & Reisinger, 2018; Kaushansky, Lipton, & Reisinger, 2018; Nadtochiy, Shkolnikov, et al., 2019. Nevertheless, the use of these methods for valuation and measurement of risk charges such as capital is limited. In a recent study, Anagnostou, Sourabh, & Kandhai, 2018 introduced a portfolio credit risk model that can account for both channels of default dependence: common underlying factors and contagion from sovereigns to corporates and sub-sovereigns. The authors augment systematic risk factors with a contagious default mechanism where the default probabilities of issuers in the portfolio are immediately affected by a sovereign default. To estimate the contagion effect they use a network constructed from CDS time series and introduce CountryRank, a network based metric that approximates the probability of default of a node conditional on the infectious default of a sovereign. The article presents a thorough approach of how contagion effects can be introduced to portfolio credit risk modes using complex networks. However, the underlying network in Anagnostou et al., 2018 is based on one-to-one relationships between issuers. While this can capture the direct relationships effectively, it is well-known that the associations between entities might be indirect and often mediated through others. In this article we use Probabilistic Graphical Models (PGM) to learn the network using the data in a holistic manner. This extends the one-to-one approach and provides a more natural and accurate representation of the network. Moreover, we can efficiently approximate the joint default probability distribution of the issuers in a PGM.

PGMs are a powerful framework for representing complex relationships using probability distributions. Their ability to model associations in complex datasets has proven them particularly useful for a wide range of machine learning problems, including natural language processing (Galley, McKeown, Hirschberg, & Shriberg, 2004), medical diagnosis (Beinlich, Suermondt, Chavez, & Cooper, 1989), and genetic linkage analysis (Fishelson & Geiger, 2004). One of the most important classes of PGMs is Bayesian networks (BNs). More recently, there have been attempts to utilize BNs for solving financial problems. In particular, Denev, 2013 presented a method to calculate portfolio losses in the presence of stress events using BNs. Nevertheless, his approach, relies on the ability of the risk manager to identify causal links and subjectively assign probabilities. Chong & Klüppelberg, 2018 developed a structural default model for interconnected financial institutions, but the need for balance sheet data makes its applicability limited. Kitwiwattanachai, 2015 used credit default swaps (CDS) data to learn the structure of interbank networks, which would enable policy makers to make decisions on the too-big-to-fail problem. In order to learn the BN, the author uses the log of CDS spreads under the assumption of normality. However, this strong assumption is barely supported by empirical evidence.

In this paper, we overcome the need for making assumptions about the distribution of CDS spreads by introducing a discretization method based on the notion of *modified ϵ-drawups* (Kaushik & Battiston, 2013; Anagnostou et al., 2018). This transformation enables us to utilize algorithms

Electronic copy available at: https://ssrn.com/abstract=3446615

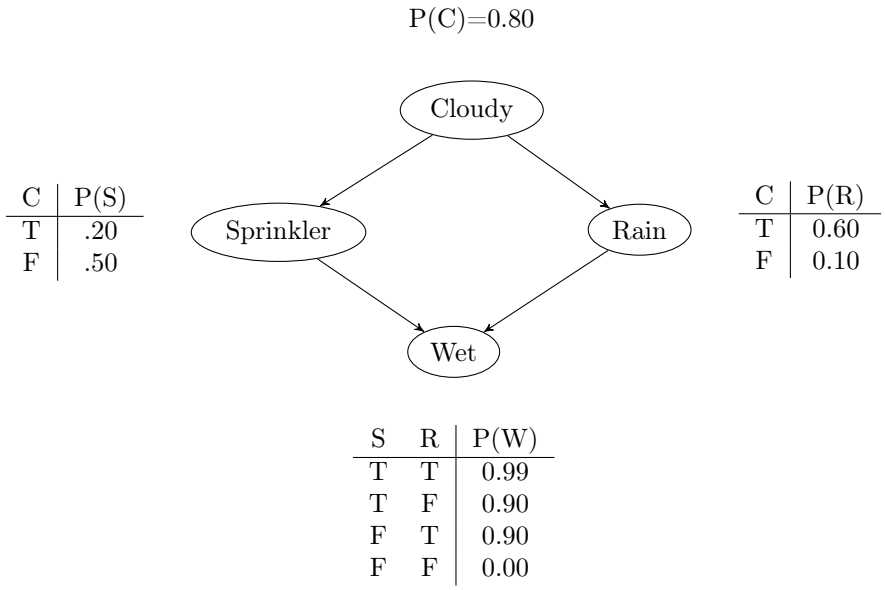


Figure 1: A Bayesian network with Conditional Probability Tables (CPTs).

for structure and parameter learning that assume discrete random variables, without having to sacrifice the interpretability of the resulting models. We use the discretized CDS time series to learn a BN of interactions between issuers, and to estimate the contagion effects following a sovereign default. Different techniques to learn the structure and parameters of financial networks are studied and evaluated, with the results confirming that the structures are robust. In order to investigate the impact of these effects on credit losses, we carry out simulations and calculate the percentiles of the loss distribution in the presence of contagion. Finally, we perform a comparative analysis with the results obtained by Anagnostou et al., 2018.

The rest of the article is organized as follows. Section 2 presents BNs and outlines the methods for learning their structure and parameters. Section 3 demonstrates a method to learn BNs for CDS data. Section 4 gives a brief description of factor models for portfolio credit risk, along with a model for credit contagion. In Section 5, we present empirical analysis on a synthetic test portfolio. Finally, in Section 6 we summarize and draw conclusions. Additional information is included in the Appendices.

2 Bayesian networks

A Bayesian network (BN) is a graphical model that allows us to represent and reason about an uncertain domain. The nodes in a BN represent random variables, and the edges represent the direct dependencies between variables (Korb & Nicholson, 2003).

The graphical structure $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ of a BN is a *directed acyclic graph* (DAG), where $\mathcal{V} = \{X_1, ..., X_n\}$ is a finite vertex set and $\mathcal{E} \subseteq \{(i, j) : X_i, X_j \in \mathcal{V}, i \neq j\}$ is a set of edges without any self-loops. The DAG defines a factorization of the joint probability distribution of $\mathcal{V}$, into a set of local probability distributions, one for each variable. The form of the factorization states that every random variable X_i directly depends only on its parents Pa_{X_i}:

$$P(X_1, \ldots, X_n) = \prod_{i=1}^{n} P(X_i | \text{Pa}_{X_i}) \tag{1}$$

BNs have this useful property of *conditional independence* that allows us to represent a joint distribution in a tractable manner. For example, even if the random variables $X_1, \ldots, X_n$ follow a binomial distribution, we would need $2^n - 1$ probabilities to represent their joint distribution. With a BN, the order of representation of joint distribution is linear in the number of variables.

In order to demonstrate this representation, we provide a commonly used example from (Jensen & Nielsen, 2007), illustrated in Figure 1. We consider that the grass can appear wet in the morning either because the sprinkler was on during the night or because it rained. Note that these events

Electronic copy available at: https://ssrn.com/abstract=3446615

are not mutually exclusive: it is possible that the sprinkler was on and it rained at the same time. Thus we have two binary-valued random variables, *Sprinkler* (S) and *Rain* (R). We also have two more binary-valued variables, *Cloudy* (C) and *Wet* (W), which are correlated to both *Sprinkler* and *Rain*. The strength of these relationships is shown in the Conditional Probability Tables (CPTs). For example, we see that $P(W|S,R) = 0.99$, and thus, $P(\neg W|S,R) = 1 - 0.99 = 0.01$. Since the C node has no parents, its CPT specifies the prior probability that it is cloudy, which in this case is equal to 0.5. Overall, our probability space has $2^4 = 16$ values which correspond to all the possible assignments of these four variables. By the chain rule of probability, the joint probability of all the nodes in the graph is:

$$P(C,S,R,W) = P(C)P(S|C)P(R|C,S)P(W|C,S,R). \tag{2}$$

By using conditional independence relationships, this can be rewritten as:

$$P(C,S,R,W) = P(C)P(S|C)P(R|C)P(W|S,R), \tag{3}$$

where it was possible to simplify the third term because *Rain* is independent of *Sprinkler* given its parent *Cloudy*, and the last term because *Wet* is independent of *Cloudy* given its parents *Season* and *Rain*. Thus, it is clear that the conditional independence relationships allow for a more compact representation of the joint probability distribution.

Although, in theory, there are many possible options for the distributions of the random variables in a BN, the literature has mainly focused on two cases (Nagarajan, Scutari, & Lèbre, 2013):

- *Multinomial variables*: this representation is used for discrete/categorical data and is often referred to as the discrete case. This assumption is the most common in the literature, and the corresponding Bayesian networks are called discrete Bayesian networks.
- *Multivariate normal variables*: used for continuous data and therefore referred to as the continuous case. These Bayesian networks are referred to as Gaussian Bayesian networks.

In this work, the nodes in BNs represent random variables characterizing issuers of debt and the edges how these issuers influence each other. More specifically, the random variables of interest are the probabilities of default. It should be noted that, since BNs are DAGs, the existence of cycles or loops is neglected in our analysis. It can be argued, however, that the magnitude of such second order effects can be, in fact, negligible as far as the contagion process is concerned[1]. To learn the structure and parameters of the BNs we use time series of CDS spreads. The rest of this section describes the process of learning the structure and parameters of a BN from data.

2.1 Learning

In order to estimate the joint probability distribution from a BN, we first need to learn both the structure and the parameters of the network from data. In the first place we will explain the parameter learning and afterwards the structure learning. The reason for this order is the use of some parameter estimation techniques in the latter.

Parameter learning Suppose we have a collection of n random variables $X_1,\ldots,X_n$ such that the number of states of the random variable X_i are $1,2,\ldots,r_i$ and the number of configurations of parents of X_i are $1,2,\ldots,q_i$. The parameters which have to be estimated in this case are:

$$\theta_{ijk} = P(X_i = j \,|\, \mathrm{Pa}_{X_i} = k), \qquad i \in \{1,\ldots,n\}, j \in \{1,\ldots,r_i\} \text{ and } k \in \{1,\ldots,q_i\}.$$

We use $\theta_{\mathcal{G}} = \{\theta_{ijk} \,|\, i \in \{1,\ldots,n\},\ j \in \{1,\ldots,r_i\},\ k \in \{1,\ldots,q_i\}\}$ to denote the parameter vector.

Let us assume that we know the structure of a BN. There are two different methods for learning the parameters: Maximum Likelihood Estimation (MLE) and Bayesian estimation. MLE is based

[1]Several contributions in the contagion literature, for instance (May & Arinaminpathy, 2009), (Gai & Kapadia, 2010), (Gleeson, Hurd, Melnik, & Hackett, 2012), and (Hurd & Gleeson, 2013), take advantage of conditional independence relationships in order to approximate the probability of contagion with a closed-form expression. The authors of the above papers test their approximations to the contagion probability when applied to finite networks that entail cycles by performing numerical simulations, and show that the analytic approximations work surprisingly well. These results can be seen as a test on the actual relevance of second order, cycle effects in contagion processes. Taking this into consideration, we believe that we can neglect the existence of cycles in our financial networks without compromising our contagion analysis.

Electronic copy available at: https://ssrn.com/abstract=3446615

on choosing the parameters which maximize the likelihood of the data. Given a data set $\mathcal{D}$, the MLE method chooses parameters $\hat{\theta}_{\mathcal{G}}$ such that:

$$\hat{\theta}_{\mathcal{G}} = \arg\max_{\theta_{\mathcal{G}} \in \Theta_{\mathcal{G}}} L(\theta_{\mathcal{G}} : \mathcal{D}) = \arg\max_{\theta_{\mathcal{G}} \in \Theta_{\mathcal{G}}} P(\mathcal{D} : \theta_{\mathcal{G}}) = \arg\max_{\theta_{\mathcal{G}} \in \Theta_{\mathcal{G}}} \prod_{m} P(\xi[m] : \theta_{\mathcal{G}}),$$

where $\Theta_{\mathcal{G}}$ is the space of possible values of $\theta_{\mathcal{G}}$, $\xi[m]$ is the m-th instance of $\mathcal{D}$.

The alternative Bayesian estimation method is based on assuming a prior distribution over the parameters $P(\theta_{\mathcal{G}})$, and updating it with each instance of the data to obtain the posterior distribution, $P(\theta_{\mathcal{G}} \,|\, \mathcal{D})$ using the Bayes rule as follows:

$$P(\theta_{\mathcal{G}} \,|\, \xi[1], \ldots, \xi[M]) = \frac{P(\xi[1], \ldots, \xi[M] \,|\, \theta_{\mathcal{G}})\, P(\theta_{\mathcal{G}})}{P(\xi[1], \ldots, \xi[M])}, \tag{4}$$

where the denominator is a normalizing factor and $P(\xi[1], \ldots, \xi[M] \,|\, \theta_{\mathcal{G}})$ is the likelihood.

The choice of the prior distribution is key for the Bayesian estimation procedure. From Equation 4, we can see that the posterior distribution is proportional to the product of the likelihood and the prior. Therefore, we need to choose the prior in such a way that it can be updated easily after each new sample, while maintaining the form of the posterior distribution. It is well-known that the Dirichlet distribution is the conjugate prior for the multinomial distribution (Koller & Friedman, 2009, Section 17.3.2), which means that if the prior distribution of the multinomial parameters is Dirichlet then the posterior distribution is also a Dirichlet distribution. Since we deal with multinomial variables in our case, we choose Dirichlet distribution as the prior for our experiments.

Structure learning The structure learning for a BN is essentially an optimization problem where we minimize a score over the search space of possible configurations of the network. The score measures how likely a particular structure is based on the data, and is divided into two categories: likelihood scores and Bayesian scores.

The likelihood scores rely mainly on the likelihood function, which is the probability of sampling the data given the structure, $L(\mathcal{G} \,|\, \mathcal{D}) = P(\mathcal{D} \,|\, \mathcal{G})$. The notation $\langle \mathcal{G}, \theta_{\mathcal{G}} \rangle$ denotes a BN, where $\mathcal{G}$ represents the structure and $\theta_{\mathcal{G}}$ the parameters of the network. The structure of the network is chosen so as to maximize the likelihood score, using the MLE parameters.

$$\max_{\mathcal{G}, \theta_{\mathcal{G}}} L(\langle \mathcal{G}, \theta_{\mathcal{G}} \rangle : \mathcal{D}) = \max_{\mathcal{G}} \left[\max_{\theta_{\mathcal{G}}} L(\langle \mathcal{G}, \theta_{\mathcal{G}} \rangle : \mathcal{D}) \right] = \max_{\mathcal{G}} L(\langle \mathcal{G}, \hat{\theta}_{\mathcal{G}} \rangle : \mathcal{D}).$$

The Bayesian scores have a similar approach as the Bayesian estimation for the parameters. First, we define a prior distribution over the structure $P(\mathcal{G})$ and a conditional prior over the parameters $P(\theta_{\mathcal{G}} \,|\, \mathcal{G})$. Then, we obtain the posterior distribution $P(\mathcal{G} \,|\, \mathcal{D})$ using the Bayes rule. Similar to the Bayesian estimation for the parameters, the denominator is a normalizing factor and it remains the same for all the structures. Then, the score can be defined by taking the logarithm of the numerator:

$$\text{score}_B(\mathcal{G} : \mathcal{D}) = \log P(\mathcal{D} \,|\, \mathcal{G}) + \log P(\mathcal{G}),$$

where the second term, the prior, makes no significant difference in the score (see, e.g. Koller & Friedman, 2009, Section 18.3.2). The first term, called *marginal likelihood*, is the average over all the possible choices of $\theta_{\mathcal{G}}$ based on the conditional probability provided before:

$$P(\mathcal{D} \,|\, \mathcal{G}) = \int_{\Theta_{\mathcal{G}}} P(\mathcal{D} \,|\, \theta_{\mathcal{G}}, \mathcal{G})\, P(\theta_{\mathcal{G}} \,|\, \mathcal{G})\, d\theta_{\mathcal{G}}. \tag{5}$$

The average over all the possible parameters makes the model more conservative, and hence it tries to avoid over-fitting as we take into account the sensitivity to the values of the parameters. Finally, we briefly describe the search space and the optimization procedure for structure learning. The search space is a network itself where each of the nodes is a candidate structure $\mathcal{G}$. Given a node $\mathcal{G}$ corresponding to a structure in the search space network, the neighbours of $\mathcal{G}$ are structures obtained by either adding an edge, deleting an edge or reversing an edge. The Hill-Climbing algorithm follows the steps below:

1. Set initial BN structure to a network without edges $\mathcal{G}$.

Electronic copy available at: https://ssrn.com/abstract=3446615

2. Compute its score.

3. Consider all the neighbours of $\mathcal{G}$, obtained by either adding, deleting or reversing an edge in $\mathcal{G}$.

4. Choose the neighbour $\mathcal{G}_{\text{best}}$ which leads to the best improvement in the score.

5. Set $\mathcal{G} \leftarrow \mathcal{G}_{\text{best}}$ and repeat until no further improvement to the score is possible.

Some improvements can be made to this algorithm: (Koller & Friedman, 2009, Section 18.4.3, Glover, 1995).

3 Learning Bayesian networks from CDS data

As mentioned in Section 2, the BN literature has mostly focused on multinomial and multivariate normal data. In Kitwiwattanachai, 2015, the authors make the assumption that the residuals of the regressions on the log returns of the CDS spreads are normally distributed which is not supported by empirical data. In this work, we choose to work with discrete Bayesian networks. Therefore, we transform the continuous CDS time series data into a discrete distribution. For details on CDS contracts, the reader is referred to Section A in the Appendix.

3.1 CDS dataset

The data used for the construction of the network are credit default swaps (CDS) spreads for different maturities obtained from Markit. These consist of daily CDS liquid spreads of Russian issuers from September 14$^{\text{th}}$ 2010 until August 15$^{\text{th}}$ 2015. This period is of particular interest because of the financial crisis in Russia in 2014-2015, which followed a sharp depreciation of the Russian ruble. Apart from the spreads, the dataset also includes information about the recovery rates, region, sector and average of the ratings from Standard & Poor's, Moody's, and Fitch Group of each issuer. We use the CDS spreads of issuers for the 5 year tenor for our analysis, since they are the most liquid quotes. Since recovery rates are not the same for all issuers, we have to normalize CDS spreads to do a consistent analysis. The normalization of CDS spreads for the recovery rate RR is done as follows:

$$\widehat{S} = S\frac{0.6}{(1-RR)}, \tag{6}$$

where $\widehat{S}$ denotes the normalized CDS spread, which corresponds to a recovery rate of 40%. The choice of normalizing the CDS spreads using a recovery rate of 40% is based on the literature (see e.g., Das & Hanouna, 2009) where the recovery rate is often assumed to be a constant. We use the normalized CDS spread of the five year tenor in our analysis. Figure 2 shows the normalized spreads for the Russian issuers in the portfolio.

3.2 Discretization of CDS data

We use the notion of modified ϵ-drawups to transform the continuous CDS time series data into a discrete distribution. Modified ϵ-drawups build up on the notion of ϵ-drawups (Kaushik & Battiston, 2013, Sornette & Zhou, 2006) and can detect instances in the time series where it shows significant upward jumps. Note that, in the context of CDS spreads, upward jumps translate to rapid deteriorations in credit quality.

Definition 1. A *modified ϵ-drawup* is defined as an upward movement in the time series at a local minimum, in which the amplitude of the movement, that is, the difference between the subsequent local maximum and the local minimum, is greater than a threshold ϵ. We record such a local minimum in the time series as a modified ϵ-drawup.

An illustration of the above definition is shown in Figure 3. The ϵ parameter at time t is set to be the standard deviation in the time series between days $t-n$ and t, where n is chosen to be 10 days consistent with the choice in Kaushik & Battiston, 2013. The CDS data can be transformed into a discrete distribution for learning the BN as follows. Firstly, we compute the modified ϵ-drawups for each of the time series. Thus, each issuer i will either have a modified ϵ-drawup at time t or not. We define a binary random variable X_i^t corresponding to the issuer

Electronic copy available at: https://ssrn.com/abstract=3446615

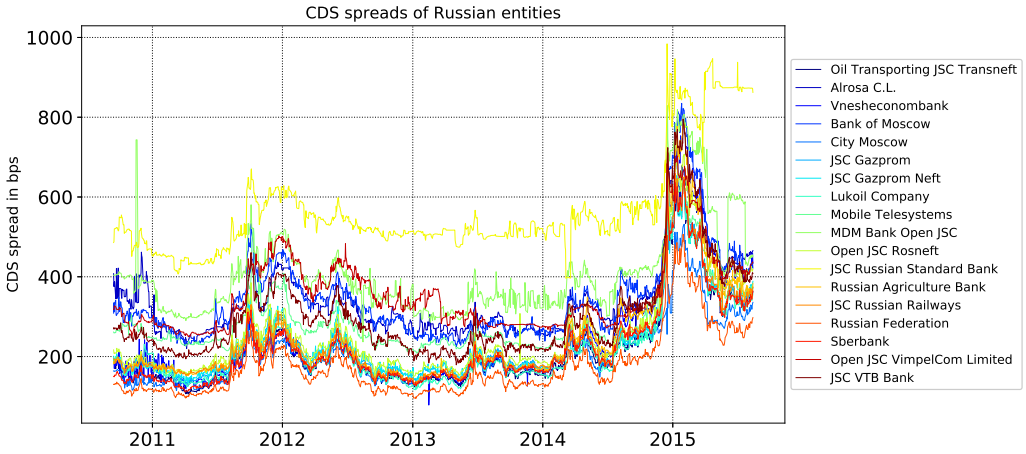


Figure 2: Normalized five year CDS spread of Russian Issuers.

i such that $X_i^t = 1$, if issuer i has a modified ϵ-drawup on day t, and 0, otherwise. Note that a company cannot have two modified ϵ-drawups on consecutive days by definition.

An additional step which is needed to prepare the CDS data for learning the co-dependence of defaults is to introduce the concept of time-lag. This allows us to capture the fact that issuers can impact each other with a slight delay. For time-lag, we introduce a new categorical value, 0.5, in the following way. Let a company i have a modified ϵ-drawup on day t, so $X_i^t = 1$. If a different company j has a modified ϵ-drawup on at least one of the following three days, $t+1, t+2$ or $t+3$, and not on day t, then we set $X_j^t = 0.5$. Hence, the number of modified ϵ-drawups in the time series remains unchanged which ensures that the marginal probability of having a modified ϵ-drawup remains unchanged.

3.3 Bayesian network learning

For learning the structure of the network we use the Hill-Climbing algorithm based on two different scores: Bayesian Information Criterion (BIC) (Schwarz, 1978) which is a likelihood score, and Bayesian Dirichlet Sparse (BDs) (Scutari, 2016), which is a Bayesian score. We refer to Section C in the Appendix for details on the two scores.

For learning the network, we also applied a bootstrapping technique for ensuring the robustness of results (Friedman, Goldszmidt, & Wyner, 1999). For the structure learning, we obtain the structure 1000 times and then we compute the average structure by including the edges which appear in at least 50% of the networks. The data set $\mathcal{D}_k$ used at iteration k is obtained by sampling uniformly $|\mathcal{D}|$ instances from the original training data $\mathcal{D}$.

Once the BN is learnt, we can evaluate the queries for conditional probabilities $P(Q|E)$, of events Q given evidence[2] E. To perform these queries, we use the logic sampling algorithm (Henrion, 1988) which has the following steps. First, it orders the variables in the topological order implied by the structure $\mathcal{G}$. This means that the variables with no parents appear first followed by their children. Next, we set the counters $n_E = 0$ and $n_{E,Q} = 0$. Afterwards we generate a sufficiently large number of samples M where each sample consists of a vector of instances of all the random variables in the network. Note that generating the instance for X_i only requires the values of Pa_{X_i}. Then, for each sample if it includes E, set $n_E = n_E + 1$, and, if it includes both Q and E, set $n_{Q,E} = n_{Q,E} + 1$. Finally, we can estimate $P(Q \mid E)$ by $n_{Q,E}/n_E$. This method is based on a Monte Carlo simulation, therefore a sufficiently large number of simulations is needed to assure a reliable result.

[2] An *event* in the BN terminology refers to a (some) random variable(s) taking a particular value(s). An *evidence* is mathematically the same as an event with the difference that it is known.

Electronic copy available at: https://ssrn.com/abstract=3446615

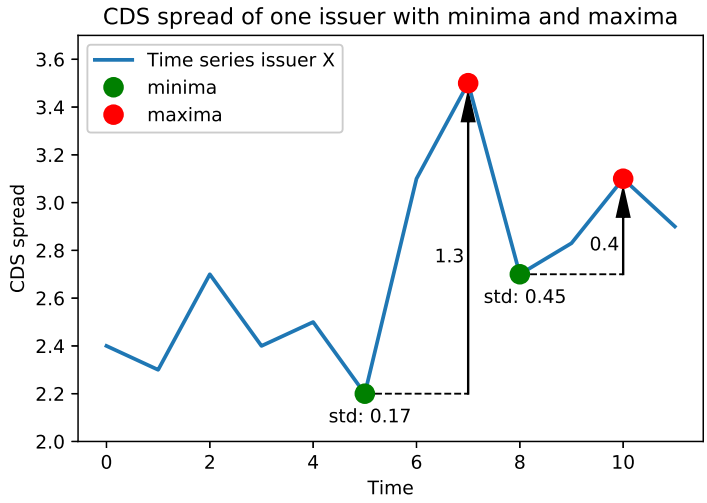


Figure 3: A *modified ϵ-drawup* is defined as an upward movement in the time series at a local minima, in which the amplitude of the movement is greater than a threshold ϵ. The ϵ parameter at time t is set to be the standard deviation in the time series between days $t-n$ and t, where n is chosen to be 10 days. In the above example, the first green minimum is recorded as a *modified ϵ-drawup*, while the second is not.

4 Portfolio credit risk modelling

The most widely used portfolio credit risk models assume that issuers depend on some common underlying factors. Factor models based on the Merton model are particularly popular for portfolio credit risk. The model presented in Anagnostou et al., 2018 extends the muti-factor Merton model to allow for credit contagion. In this section we provide a brief description of factor models for portfolio credit risk, along with an overview of the contagion model from Anagnostou et al., 2018.

4.1 Factor models

Factor models for portfolio credit risk can be motivated by a multivariate firm-value model based on Merton, 1974. This category includes widely used industry models such as CreditMetrics and KMV. Default occurs for an issuer i if a critical variable X_i, representing the standardized asset return, falls below a critical threshold d_i. For a portfolio of m issuers, $(X_1, ..., X_m)' \sim N_m(\mathbf{0}, \Sigma)$ and thus, the probability of default for issuer i is satisfying $p_i = \Phi(d_i)$, where $\Phi(\cdot)$ denotes the cumulative distribution function of the standard normal distribution. The default probabilities are usually estimated by historical default experience using external ratings by agencies or model-based approaches.

In the factor model approach, the critical variables $X_i, i = 1, ..., m$, are linearly dependent on a vector $\mathbf{F}$ of $p < m$ common underlying factors satisfying $\mathbf{F} \sim N_p(0, \Omega)$. Issuer i's standardized asset return is assumed to be driven by an issuer-specific combination $\tilde{F}_i = \boldsymbol{\alpha}_i'\mathbf{F}$ of the systematic factors

$$X_i = \sqrt{\beta_i}\tilde{F}_i + \sqrt{1-\beta_i}\epsilon_i, \tag{7}$$

where $\tilde{F}_i$ and $\epsilon_1, ..., \epsilon_m$ are independent standard normal variables, and ϵ_i represents the idiosyncratic risk. Consequently, β_i can be seen as a measure of sensitivity of X_i to systematic risk, as it represents the proportion of the X_i variation that is explained by the systematic factors. The assumption that $\text{var}(\tilde{F}_i) = 1$ implies that $\boldsymbol{\alpha_i}'\Omega\boldsymbol{\alpha_i} = 1$ for all i. The correlations between asset returns are given by

$$\rho(X_i, X_j) = cov(X_i, X_j) = \sqrt{\beta_i\beta_j}cov(\tilde{F}_i, \tilde{F}_j) = \sqrt{\beta_i\beta_j}\boldsymbol{\alpha}_i'\Omega\boldsymbol{\alpha}_j, \tag{8}$$

since $\tilde{F}_i$ and $\epsilon_1, ..., \epsilon_m$ are independent and standard normal and $\text{var}(X_i) = 1$.

Electronic copy available at: https://ssrn.com/abstract=3446615

4.2 A model for credit contagion

Consider a corporate issuer C_i, and its country of operation S. Denote by p_{C_i} the probability of default of C_i, and by Y_{C_i} the corresponding default indicator variable; $Y_{C_i} = 1 \iff C_i$ defaults and $Y_{C_i} = 0 \iff C_i$ does not default. Under the standard Merton model, default occurs if C_i's standardized asset return X_{C_i} falls below its default threshold d_{C_i}. The critical threshold d_{C_i} is assumed to be equal to $\Phi^{-1}(p_{C_i})$ and is independent of the state of the country of operation S. In the proposed model, a corporate is subject to shocks from its country of operation; its corresponding state is described by a binary state variable. The state is considered to be stressed in the event of sovereign default. In this case, the issuer's default threshold increases, causing it more likely to default, as the contagion effect suggests. In case the corresponding sovereign does not default, the corporates liquidity state is considered stable. We replace the default threshold d_{C_i} with $d^*_{C_i}$, where

$$d^*_{C_i} = \begin{cases} d^{sd}_{C_i} & \text{if the corresponding sovereign defaults,} \\ d^{nsd}_{C_i} & \text{otherwise,} \end{cases} \tag{9}$$

or equivalently

$$d^*_{C_i} = \mathbb{1}_{\{Y_S=1\}} d^{sd}_{C_i} + \mathbb{1}_{\{Y_S=0\}} d^{nsd}_{C_i}, \tag{10}$$

where Y_S is the default indicator of the sovereign. We denote by p_S the probability of default of the sovereign, and by γ_{C_i} the parameter which indicates the increased probability of default of C_i given the default of S. Our objective is to calibrate $d^{sd}_{C_i}$ and $d^{nsd}_{C_i}$ in such way that the overall default rate remains unchanged and $P(Y_{C_i} = 1|Y_S = 1) = \gamma_{C_i}$. Denote by

$$\phi_2(x, y; \rho) := \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left(-\frac{x^2 + y^2 - 2\rho xy}{2(1-\rho^2)}\right) \tag{11}$$

$$\Phi_2(h, k; \rho) := \int_{-\infty}^{h}\int_{-\infty}^{k} \phi_2(x, y; \rho)dydx \tag{12}$$

the density and distribution function of the bivariate standard normal distribution with correlation parameter $\rho \in (-1, 1)$. Note that $d^*_{C_i}(\omega) = d^{sd}_{C_i}$ for $\omega \in \{Y_{C_i} = 1, Y_S = 1\} \subset \{Y_S = 1\}$, and $d^*_{C_i}(\omega) = d^{nsd}_{C_i}$ for $\omega \in \{Y_{C_i} = 1, Y_s = 0\} \subset \{Y_S = 0\}$. We rewrite $P(Y_{C_i} = 1|Y_S = 1)$ in the following way

$$\begin{aligned} P(Y_{C_i} = 1|Y_S = 1) &= \frac{1}{P(Y_S = 1)} P(Y_{C_i} = 1, Y_S = 1) && (13) \\ &= \frac{1}{p_S} P\left[X_{C_i} < d^{sd}_{C_i}, X_S < d_S\right] && (14) \\ &= \frac{1}{p_S} \Phi_2(d^{sd}_{C_i}, d_S; \rho_{SC_i}). && (15) \end{aligned}$$

Using the above representation and given $d_S = \Phi^{-1}(p_S)$ and ρ_{SC_i} one can solve the equation

$$P(Y_{C_i} = 1|Y_S = 1) = \gamma_{C_i}, \tag{16}$$

over $d^{sd}_{C_i}$.

We proceed to the derivation of $d^{nsd}_{C_i}$ in such way that the overall default probability remains equal to p_{C_i}. This constraint is important, since contagion is assumed to have no impact on the average loss. Clearly,

$$\begin{aligned} p_{C_i} &= P(Y_{C_i} = 1) && (17) \\ &= P(Y_{C_i} = 1, Y_S = 1) + P(Y_{C_i} = 1, Y_S = 0) && (18) \\ &= P(Y_{C_i} = 1|Y_S = 1)P(Y_S = 1) + P(Y_{C_i} = 1, Y_S = 0), && (19) \end{aligned}$$

and thus

$$P(Y_{C_i} = 1, Y_S = 0) = p_{C_i} - \gamma_{C_i} \cdot p_S. \tag{20}$$

Electronic copy available at: https://ssrn.com/abstract=3446615

The left-hand side of the above equation can be represented as follows

$$\begin{aligned} P(Y_{C_i}=1, Y_S=0) &= P\left[X_{C_i} < d_{C_i}^{nsd}, X_S > d_S\right] && (21)\\ &= P\left[X_{C_i} < d_{C_i}^{nsd}\right] - P\left[X_{C_i} < d_{C_i}^{nsd}, X_S < d_S\right] && (22)\\ &= \Phi(d_{C_i}^{nsd}) - \Phi_2(d_{C_i}^{nsd}, d_S; \rho_{SC_i}). && (23) \end{aligned}$$

By use of the above and given $d_S = \Phi^{-1}(p_S)$ and ρ_{SC_i} one can solve the previous equation over $d_{C_i}^{nsd}$.

5 Numerical study

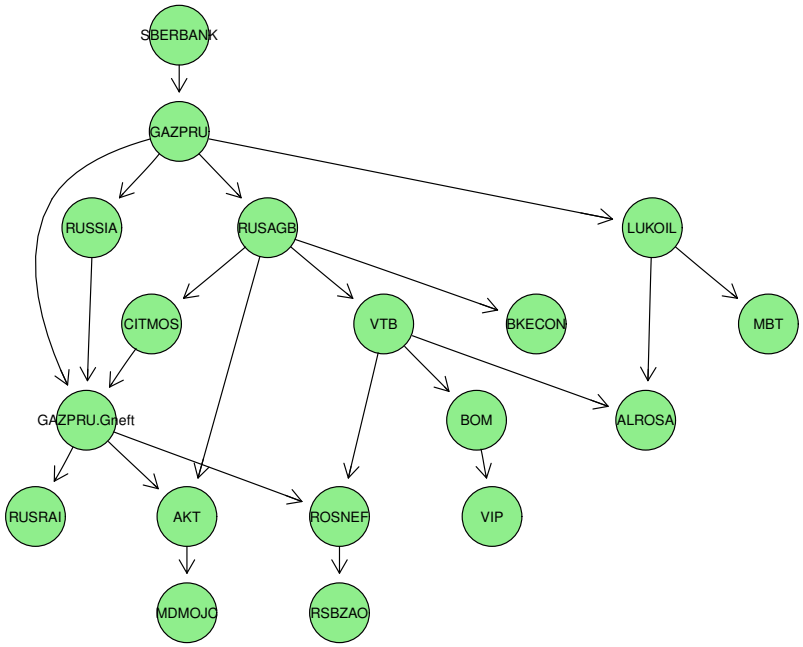
In this section we describe the setup of the problem and the results obtained in the calibrations and simulations.

5.1 Bayesian network learning and robustness

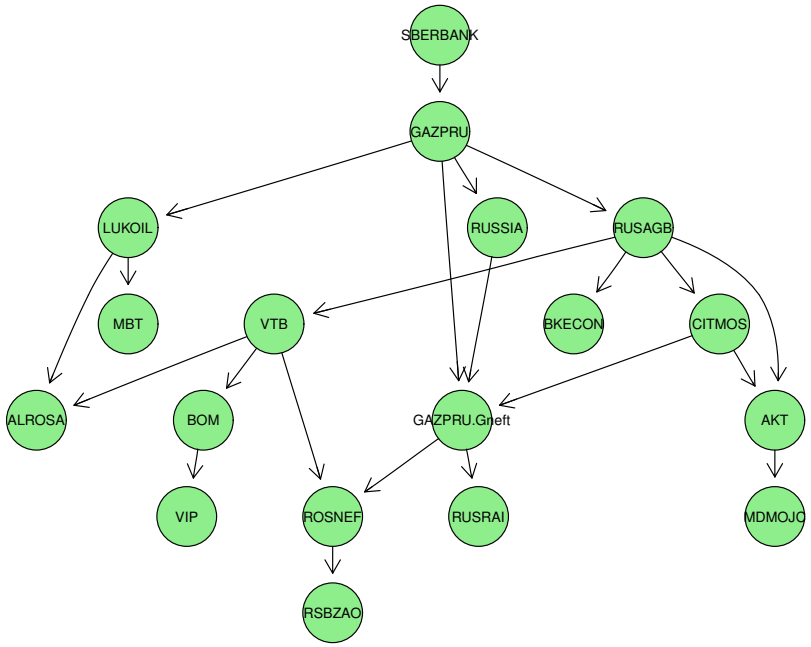
We use CDS data as described in Section 3 to learn the structure and the parameters of the BN. The *bnlearn* (Scutari, 2010) library in R is used for the Hill-Climbing procedure. The numerical experiments were performed using two scores: BDs and BIC. Both scores resulted in similar structure of the network as shown in Figure 5, except for one edge, GAZPRU.Gneft → AKT which is present in the network learnt with BIC, whereas with BDs it is substituted by CITMOS → AKT, as we can see in Figure 5. The table in Appendix A, lists the tickers of the issuers and the complete names.

After learning the network, the next step is to estimate the conditional probability of default of an issuer, conditional on the default of the sovereign. In order to estimate these probabilities we run a Monte Carlo simulation of 100 iterations. For each of these iterations, we calculate the conditional probability using 4×10^5 samples. Finally, we take the mean of the Monte Carlo simulations as an estimate for conditional probabilities. We compare the probabilities for both BIC and BDs scores including mean, standard deviation and absolute difference in Table 3. Figure 4 shows the structure with the nodes colored according to their probability of default given sovereign default to have a more visual explanation. The darker the color of the node, the higher is the probability of default of the node conditional on sovereign default. We see that Gazprom and Gazprom Neft, which are the two nodes connected to the sovereign, are the ones more affected and the issuers which are further from the sovereign have relatively lower conditional default probabilities.

Electronic copy available at: https://ssrn.com/abstract=3446615



(a) Structure with BIC score



(b) Structure with BDs score

Figure 5: Structure obtained with the different scores. Note that the visualization system is mirroring the plot but close inspection reveals that the structure is not that different.

Electronic copy available at: https://ssrn.com/abstract=3446615

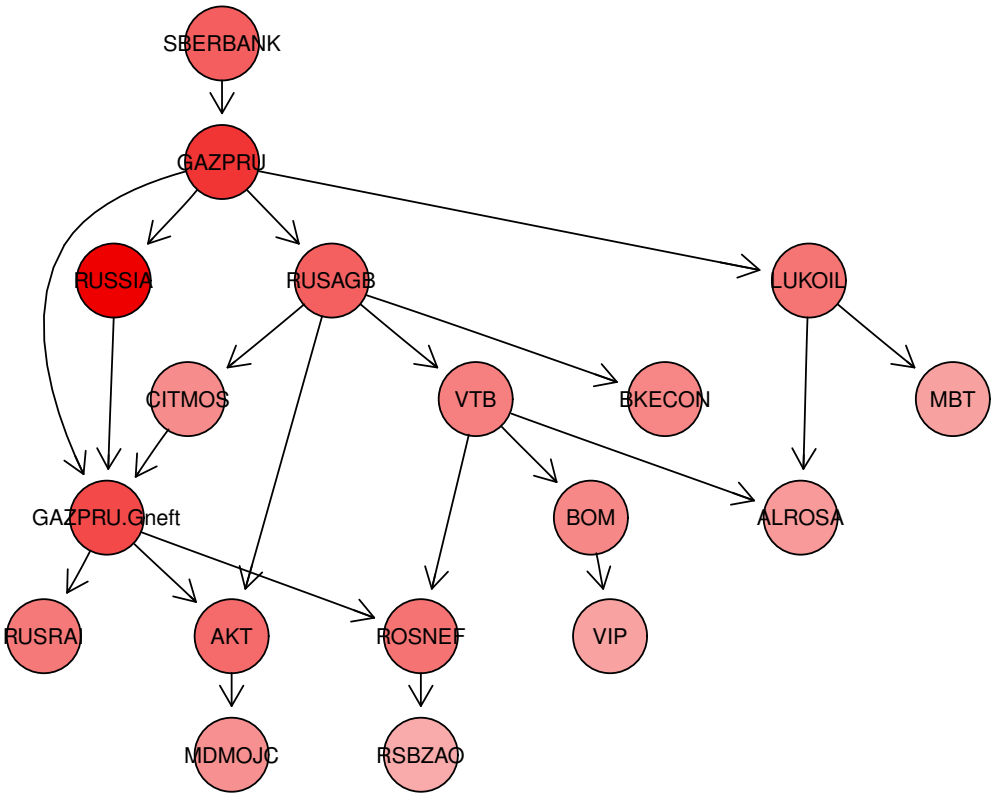


Figure 4: Bayesian network learnt with BIC with colored nodes according to its probability of default given sovereign default. Note that having a darker color means that the node has a higher conditional probability of default.

We observe that the standard deviation of the conditional probabilities estimates is quite small for all the issuers, ranging between 0.002 and 0.0035. This implies that the estimates are quite robust, and gives us a strong confidence on the reliability of the results. Moreover, we notice that the absolute difference of the probabilities from the two scores is smaller than the standard deviation except in two cases.

The main difference we note in the table is Transneft, whose probability of default conditional on sovereign default decreases by more than 0.07 when changing the score from BIC to BDs. This is caused by the change in the structure, which directly affects this issuer. With BIC score, Gazprom Neft is a parent of Transneft whereas City Moscow is not, and using BDs score it is the other way around. We see that with both scores Gazprom Neft is more affected by the sovereign than City Moscow. This is also a confirmation of the fact that the stress spreads faster from one issuer to another if they are directly connected. The second difference in conditional probability is for MDM Bank. If we look at the structures, we see that MDM has only one parent, which is Transneft. As the parent (Transneft) is less affected in the structure learnt by BDs, its child being less affected is in line with intuition. This causes a small but still noticeable difference in the conditional probabilities.

5.2 Comparative analysis

We set up a multi-factor Merton model, as it was described in Section 4. We define a set of systematic factors that will represent region and sector effects. We choose 6 region and 6 sector factors, for which we select appropriate indices, as shown in Table 6. We then use 10 years of index time series to derive the region and sector returns $F_{R(j)}$, $j = 1, ..., 6$ and $F_{S(k)}$, $k = 1, ..., 6$ respectively, and obtain an estimate of the correlation matrix Ω. Subsequently, we map all issuers to one region and one sector factor, $F_{R(i)}$ and $F_{S(i)}$ respectively. For instance, a Dutch bank will

Electronic copy available at: https://ssrn.com/abstract=3446615

be associated with Europe and Financial factors. As a proxy of individual asset returns we use 10 years of equity or CDS time series, depending on the data availability for each issuer. Finally, we standardize the individual returns time series $(X_{i,t})$ and perform the following Ordinary Least Squares regression against the systematic factor returns

$$X_{i,t} = \alpha_{R(i)} F_{R(i),t} + \alpha_{S(k)} F_{S(k),t} + \epsilon_{i,t} \tag{24}$$

to obtain $\hat{\alpha}_{R(i)}$, $\hat{\alpha}_{S(i)}$, and $\hat{\beta}_i = R^2$, where R^2 is the coefficient of determination, and it is higher for issuers whose returns are largely affected by the performance of the systematic factors.

To investigate the properties of the contagion model, we set up a test portfolio. The resulting risk measures for this portfolio are compared to those of the standard latent variable model with no contagion. The portfolio consists of 1 Russian government bond and 17 bonds issued by corporations registered and operating in the Russian Federation. As it is illustrated in Table 1, the issuers are of medium and low credit quality. The sectors represented are shown in Appendix B, Table 2. The portfolio is assumed to be equally weighted with a total notional of €10 million.

Rating	Issuers	%
BBB	1	5.56%
BB	15	83.33%
B	2	11.11%

Table 1: Rating classification for the test portfolio.

Sector	Issuers	%
Materials	5	27.78%
Services	3	16.67%
Financial	7	38.89%
Government	3	16.67%

Table 2: Sector classification for the test portfolio.

In order to generate portfolio loss distributions and derive the associated risk measures we perform Monte Carlo simulations. This process entails generating joint realizations of the systematic and idiosyncratic risk factors, and comparing the resulting critical variables with the corresponding default thresholds. By this comparison we obtain the default indicator Y_i for each issuer and this enables us to calculate the overall portfolio loss for this trial. The only difference between the standard and the contagion model is that in the contagion model we first obtain the default indicators for the sovereigns, and their values determine which default thresholds are going to be used for the corporate issuers. A liquidity horizon of 1 year is assumed throughout and the figures are based on a simulation with 10^6 samples. Moreover, for the results shown we used the probabilities of default computed with the structure learnt with the score BIC. However, this choice does not make any notable difference in the percentiles of the loss distribution because the probabilities were almost the same and such tiny difference would not cause a large disturbance.

We compared the BN model with the CountryRank model of Anagnostou et al., 2018. In Table 4 we can observe the difference between the probabilities obtained with both methods, using the same data, and same parameters, 10 days to compute the standard deviation and 3 days as time lag. Following the sensitivity analysis by Anagnostou et al., 2018 we can expect that the 15% increase of the mean will not have a substantial impact on the percentiles of the loss distribution. This hypothesis can be confirmed by the results shown in Table 5 and the graph depicted in Figure 6.

6 Concluding remarks

In this article, we presented a novel method of estimating contagion effects from CDS data using BNs. Rather than assuming a certain distribution for CDS spreads, we introduced a method

Electronic copy available at: https://ssrn.com/abstract=3446615

		BIC		BDs		
Order	Issuer	γ_C	s.d.	γ_C	s.d.	Abs diff
1	JSC Gazprom	0.7824	0.0030	0.7809	0.0028	0.0015
2	JSC Gazprom Neft	0.7024	0.0026	0.7018	0.0028	0.0006
3	Sberbank	0.6260	0.0034	0.6261	0.0032	0.0001
4	Russian Agriculture Bank	0.6165	0.0032	0.6167	0.0030	0.0002
5	Oil Transporting JSC Transneft	0.5754	0.0025	0.4983	0.0026	0.0771
6	Lukoil Company	0.5417	0.0024	0.5410	0.0023	0.0007
7	Open JSC Rosneft	0.5394	0.0030	0.5407	0.0025	0.0013
8	JSC Russian Railways	0.5186	0.0029	0.5189	0.0025	0.0003
9	JSC VTB Bank	0.4913	0.0027	0.4911	0.0030	0.0002
10	Vnesheconombank	0.4583	0.0028	0.4583	0.0024	$< 10^{-4}$
11	Bank of Moscow	0.4576	0.0027	0.4572	0.0025	0.0004
12	City Moscow	0.4377	0.0031	0.4375	0.0025	0.0002
13	MDM Bank Open JSC	0.4251	0.0026	0.4028	0.0032	0.0223
14	Alrosa C.L.	0.3890	0.0025	0.3885	0.0028	0.0005
15	Mobile Telesystems	0.3542	0.0025	0.3540	0.0023	0.0002
16	Open JSC VimpelCom Limited	0.3523	0.0022	0.3524	0.0023	0.0001
17	JSC Russian Standard Bank	0.3290	0.0024	0.3296	0.0021	0.0006

Table 3: Probabilities of default given sovereign default with BIC and BDs score.

Order	Issuer	BN	CountryRank	Diff
1	JSC Gazprom	0.7824	0.6220	0.1585
2	JSC Gazprom Neft	0.7024	0.5610	0.1415
3	Sberbank	0.6260	0.5854	0.0406
4	Russian Agric Bank	0.6165	0.5854	0.0312
5	Oil Transporting JSC Transneft	0.5754	0.5732	0.0022
6	Lukoil Company	0.5417	0.3381	0.2036
7	Open JSC Rosneft	0.5394	0.5244	0.0150
8	JSC Russian Railways	0.5186	0.5427	-0.0241
9	JSC VTB Bk	0.4913	0.6098	-0.1184
10	Vnesheconombank	0.4583	0.3339	0.1244
11	Bank of Moscow	0.4576	0.5305	-0.0729
12	City Moscow	0.4377	0.5122	-0.0745
13	MDM Bk Open JSC	0.4251	0.3131	0.1120
14	Alrosa C.L.	0.3890	0.2293	0.1596
15	Mobile Telesystems	0.3542	0.2446	0.1096
16	Open JSC VimpelCom Limited	0.3523	0.2964	0.0559
17	JSC Russian Standard Bank	0.3290	0.0869	0.2420
	Mean	0.5056	0.4405	0.0651

Table 4: Comparison of γ_C using BN and CountryRank model.

Electronic copy available at: https://ssrn.com/abstract=3446615

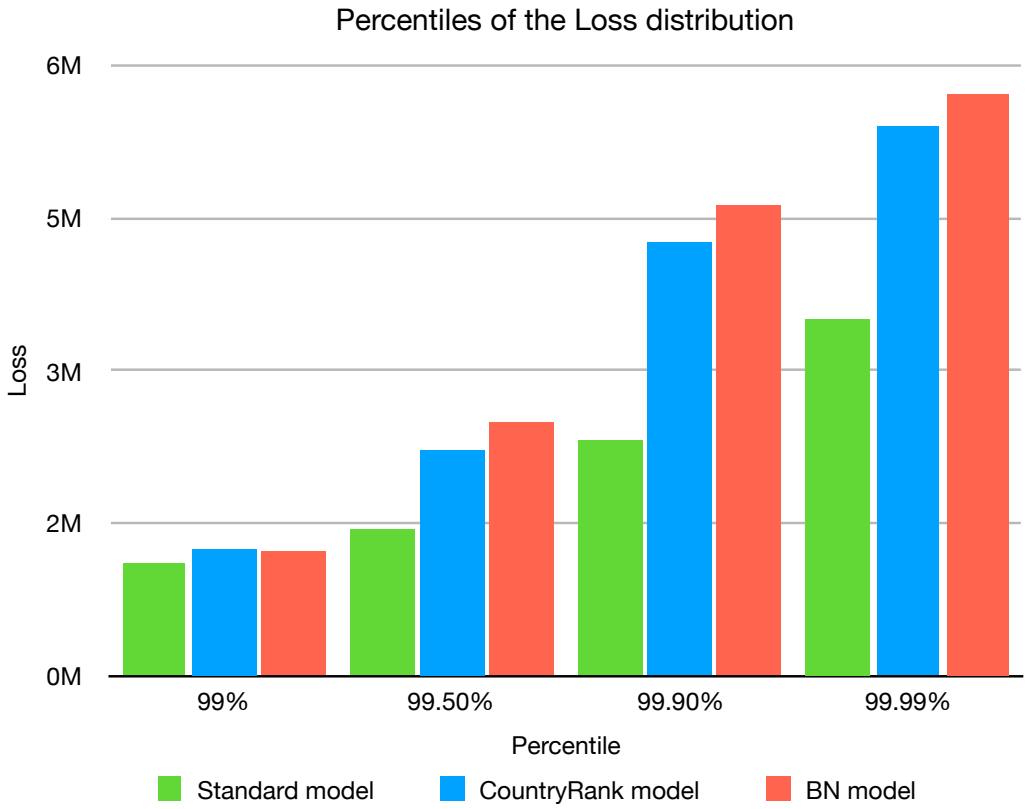


Figure 6: Percentiles of the Loss distribution without and with contagion with the BN and the CountryRank model.

		BN model		CountryRank model	
Percentile	Loss - Standard model	Contagion impact		Contagion impact	
99%	1,115,153	117,341	11%	124,903	11%
99.5%	1,443,579	1,054,844	73%	769,694	53%
99.9%	2,324,088	2,302,667	99%	1,930,166	83%
99.99%	3,514,968	2,198,517	63%	1,883,240	54%

Table 5: Comparison of the percentiles using Bayesian networks and CountryRank model.

for learning BNs using ϵ-drawups. Different techniques to learn the structure and parameters of financial networks were studied and evaluated. We used CDS spreads of issuers in a stylized portfolio and incorporated the conditional probabilities in the credit portfolio model presented by Anagnostou et al., 2018. Simulations were carried out for a stylized portfolio and the impact on standard risk metrics was estimated. Contagion was shown to have a significant impact in the tails of the credit loss distribution, with the results being in line with results obtained by using the CountryRank metric.

The results presented are a first step in the application of BNs on portfolio credit risk models. However, the BN framework we developed is flexible enough to allow for wider applications. For instance, one can extend the contagion so that stress originates from any issuer and not only at the sovereign. Moreover, one can test scenarios where multiple issuers default. Two examples of such scenarios can be found in the Appendix D. These applications can be particularly useful for risk managers, who are often interested in building scenarios for catastrophic risks and testing the resilience of their portfolios to such scenarios.

In order to extend our analysis, we plan to further investigate applications of the developed probabilistic framework in problems beyond credit portfolio modelling. A promising direction is to use our framework in order to identify systemically important nodes in the financial system and measure systemic risk. This could be done by considering, in a recursive manner, the fact that a node is more systemically important if it impacts many systemically important nodes. Another

Electronic copy available at: https://ssrn.com/abstract=3446615

interesting direction is the modelling of wrong-way risk (WWR) arising in the case of a sovereign default in the pricing of Credit Valuation Adjustment (CVA) and Funding Valuation Adjustment (FVA) for interest-rate and foreign exchange derivatives.

Acknowledgement

The authors are grateful to Marcel Boersma and the two anonymous reviewers for the valuable comments.

Disclosure statement

The opinions expressed in this article are solely those of the authors and do not represent in any way those of their current and past employers. The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie Grant Agreement no. 675044 (http://bigdatafinance.eu/), Training for Big Data in Financial Research and Risk Management.

Electronic copy available at: https://ssrn.com/abstract=3446615

References

Anagnostou, I., & Kandhai, D. (2019). Risk factor evolution for counterparty credit risk under a hidden markov model. *Risks*, *7*(2), 66.

Anagnostou, I., Sourabh, S., & Kandhai, D. (2018). Incorporating contagion in portfolio credit risk models using network theory. *Complexity*, *2018*.

Battiston, S., Farmer, J. D., Flache, A., Garlaschelli, D., Haldane, A. G., Heesterbeek, H., ... Scheffer, M. (2016). Complexity theory and financial regulation. *Science*, *351*(6275), 818–819.

Battiston, S., Gatti, D. D., Gallegati, M., Greenwald, B., & Stiglitz, J. E. (2012). Liaisons dangereuses: Increasing connectivity, risk sharing, and systemic risk. *Journal of Economic Dynamics and Control*, *36*(8), 1121 - 1141. Retrieved from http://www.sciencedirect.com/science/article/pii/S0165188912000899 (Quantifying and Understanding Dysfunctions in Financial Markets) doi: http://dx.doi.org/10.1016/j.jedc.2012.04.001

Beinlich, I. A., Suermondt, H. J., Chavez, R. M., & Cooper, G. F. (1989). The alarm monitoring system: A case study with two probabilistic inference techniques for belief networks. In *Aime 89* (pp. 247–256). Springer.

Bohn, J. R., & Kealhofer, S. (2001). Portfolio management of default risk. *KMV working paper*.

Chong, C., & Klüppelberg, C. (2018). Contagion in financial systems: A bayesian network approach. *SIAM Journal on Financial Mathematics*, *9*(1), 28–53.

Cont, R., Moussa, A., & Santos, E. B. (2013, 005 23). Network structure and systemic risk in banking systems. In J.-P. Fouque & J. A. Langsam (Eds.), *Handbook on systemic risk:* (p. 327-368). Cambridge: Cambridge University Press. Retrieved from https://www.cambridge.org/core/books/handbook-on-systemic-risk/network-structure-and-systemic-risk-in-banking-systems/9BB92BC1B1373738AE32258E09A649B0 doi: 10.1017/CBO9781139151184.018

Crosbie, P. J., & Bohn, J. R. (2002). Modeling default risk. *KMV working paper*.

Das, S. R., Duffie, D., Kapadia, N., & Saita, L. (2007). Common failings: How corporate defaults are correlated. *The Journal of Finance*, *62*(1), 93–117.

Das, S. R., & Hanouna, P. (2009). Implied recovery. *Journal of Economic Dynamics and Control*, *33*(11), 1837 - 1857. Retrieved from //www.sciencedirect.com/science/article/pii/S016518890900102X doi: http://dx.doi.org/10.1016/j.jedc.2009.04.005

Davis, M., & Lo, V. (2001). Infectious defaults. *Quantitative Finance*, *1*(4), 382-387. Retrieved from http://dx.doi.org/10.1080/713665832 doi: 10.1080/713665832

De Graaf, C. S., Feng, Q., Kandhai, D., & Oosterlee, C. W. (2014). Efficient computation of exposure profiles for counterparty credit risk. *International Journal of Theoretical and Applied Finance*, *17*(04), 1450024.

de Graaf, C. S. L., Kandhai, D., & Reisinger, C. (2018). Efficient exposure computation by risk factor decomposition. *Quantitative Finance*, *0*(0), 1-22. Retrieved from https://doi.org/10.1080/14697688.2018.1435902 doi: 10.1080/14697688.2018.1435902

Denev, A. (2013). Credit portfolio models in the presence of forward-looking stress events. *The Journal of Risk Model Validation*, *7*(1), 83.

Egloff, D., Leippold, M., & Vanini, P. (2007). A simple model of credit contagion. *Journal of Banking & Finance*, *31*(8), 2475–2492.

Fishelson, M., & Geiger, D. (2004). Optimizing exact genetic linkage computations. *Journal of Computational Biology*, *11*(2-3), 263–275.

Friedman, N., Goldszmidt, M., & Wyner, A. (1999). Data analysis with bayesian networks: a bootstrap approach. In *Fifteenth conference on uncertainty in artificial intelligence* (pp. 196–205). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.

Gai, P., & Kapadia, S. (2010). Contagion in financial networks. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, *466*(2120), 2401–2423.

Galley, M., McKeown, K., Hirschberg, J., & Shriberg, E. (2004). Identifying agreement and disagreement in conversational speech: Use of bayesian networks to model pragmatic dependencies. In *the 42nd annual meeting on association for computational linguistics* (p. 669).

Gleeson, J. P., Hurd, T., Melnik, S., & Hackett, A. (2012). Systemic risk in banking networks without monte carlo simulation. In *Advances in network analysis and its applications* (pp. 27–56). Springer.

Electronic copy available at: https://ssrn.com/abstract=3446615

Glover, F. (1995). Tabu search fundamentals and uses. *Vasa*. Retrieved from http://medcontent.metapress.com/index/A65RM03P4874243N.pdf{%}5Cnhttp://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.95.3702{&}rep=rep1{&}type=pdf

Gordy, M. (2003). A risk-factor model foundation for ratings-based bank capital rules. *Journal of Financial Intermediation*, *12*(3), 199–232.

Green, A., Kenyon, C., & Dennis, C. (2014). KVA: capital valuation adjustment by replication. *Risk*, *27*(12).

Gregory, J. (2015). *The xVA Challenge: Counterparty Credit Risk, Funding, Collateral, and Capital*. Wiley.

Hambly, B., Ledger, S., & Sojmark, A. (2018). A mckean–vlasov equation with positive feedback and blow-ups. *arXiv preprint arXiv:1801.07703*.

Hambly, B., & Sojmark, A. (2018). An spde model for systemic risk with endogenous contagion. *arXiv preprint arXiv:1801.10088*.

Henrion, M. (1988). Propagating uncertainty in bayesian networks by probabilistic logic sampling. In *Machine intelligence and pattern recognition* (Vol. 5, pp. 149–163). Elsevier.

Hurd, T. R., & Gleeson, J. P. (2013). On watts' cascade model with random link weights. *Journal of Complex Networks*, *1*(1), 25–43.

Jarrow, R. A., & Yu, F. (2001). Counterparty risk and the pricing of defaultable securities. *the Journal of Finance*, *56*(5), 1765–1799.

Jensen, F. V., & Nielsen, T. D. (2007). *Bayesian networks and decision graphs*. Springer.

JP Morgan. (1997). Creditmetrics^{TM}–technical document. *JP Morgan, New York*.

Kaushansky, V., Lipton, A., & Reisinger, C. (2018). Semi-analytical solution of a mckean-vlasov equation with feedback through hitting a boundary. *arXiv preprint arXiv:1808.05311*.

Kaushansky, V., & Reisinger, C. (2018). Simulation of particle systems interacting through hitting times. *arXiv preprint arXiv:1805.11678*.

Kaushik, R., & Battiston, S. (2013). Credit default swaps drawup networks: Too interconnected to be stable. *PloS one*, *8*(7), e61815.

Kitwiwattanachai, C. (2015). Learning Network Structure of Financial Institutions from CDS Data. *Working Paper*.

Koller, D., & Friedman, N. (2009). *Probabilistic graphical models: principles and techniques*. MIT press.

Korb, K. B., & Nicholson, A. E. (2003). *Bayesian artificial intelligence*. CRC press.

May, R. M., & Arinaminpathy, N. (2009). Systemic risk: the dynamics of model banking systems. *Journal of the Royal Society Interface*, *7*(46), 823–838.

Merton, R. C. (1974). On the pricing of corporate debt: The risk structure of interest rates. *The Journal of finance*, *29*(2), 449–470.

Musmeci, N., Nicosia, V., Aste, T., Di Matteo, T., & Latora, V. (2017). The multiplex dependency structure of financial markets. *Complexity*, *2017*.

Nadtochiy, S., Shkolnikov, M., et al. (2019). Particle systems with singular interaction through hitting times: application in systemic risk modeling. *The Annals of Applied Probability*, *29*(1), 89–129.

Nagarajan, R., Scutari, M., & Lèbre, S. (2013). Bayesian networks in r. *Springer*, *122*, 125–127.

O'Kane, D. (2011). *Modelling single-name and multi-name credit derivatives*. Wiley. Retrieved from https://books.google.nl/books?id=tWZ3EgNAZHEC

Poledna, S., Thurner, S., Farmer, J. D., & Geanakoplos, J. (2014). Leverage-induced systemic risk under basle ii and other credit risk policies. *Journal of Banking & Finance*, *42*, 199–212.

Schönbucher, P., & Schubert, D. (2001). Copula-dependent default risk in intensity models. *no. TIK-Report*, *103*.

Schwarz, G. (1978, 03). Estimating the dimension of a model. *Ann. Statist.*, *6*(2), 461–464. Retrieved from https://doi.org/10.1214/aos/1176344136 doi: 10.1214/aos/1176344136

Scutari, M. (2010). Learning bayesian networks with the bnlearn R package. *Journal of Statistical Software*, *35*(3), 1–22. doi: 10.18637/jss.v035.i03

Scutari, M. (2016). An empirical-bayes score for discrete bayesian networks. In *Conference on probabilistic graphical models* (pp. 438–448).

Simaitis, S., de Graaf, C. S. L., Hari, N., & Kandhai, D. (2016). Smile and default: the role of stochastic volatility and interest rates in counterparty credit risk. *Quantitative Finance*, *16*(11), 1725-1740. Retrieved from https://doi.org/10.1080/14697688.2016.1176240 doi: 10.1080/14697688.2016.1176240

Electronic copy available at: https://ssrn.com/abstract=3446615

Sornette, D., & Zhou, W.-X. (2006). Predictability of large future changes in major financial indices. *International Journal of Forecasting*, *22*(1), 153–168.

Sourabh, S., Hofer, M., & Kandhai, D. (2018). Liquidity risk in derivatives valuation: an improved credit proxy method. *Quantitative Finance*, *18*(3), 467-481. Retrieved from https://doi.org/10.1080/14697688.2017.1315166 doi: 10.1080/14697688.2017.1315166

Squartini, T., Van Lelyveld, I., & Garlaschelli, D. (2013). Early-warning signals of topological collapse in interbank networks. *Scientific reports*, *3*, 3357.

Electronic copy available at: https://ssrn.com/abstract=3446615

A CDS and Ticker list

In this section we briefly illustrate what a CDS contract is and after that provide the list with the CDS tickers with a mapping to the names of the issuers in the synthetic portfolio.

A credit default swap is a financial contract in which a buyer B gets insurance from a protection seller A against the default of a third party C. More precisely, given a contractual notional N, regular coupon payments with respect to N and a fixed rate s, the CDS spread, are swapped with a payment of $N(1 - RR)$ in the event of the default of C, where RR, the so-called recovery rate, is the contractual parameter which represents the part of the investment supposed to be recovered in the event of default of C. An extensive description of these contracts including various modifications can be found in O'Kane, 2011.

The following list contains the issuers in the synthetic portfolio and the corresponding ticker which represent them in the networks depicted.

Ticker	Issuer name
AKT	Oil Transporting JSC Transneft
ALROSA	Alrosa C.L.
BKECON	Vnesheconombank
BOM	Bank of Moscow
CITMOS	City Moscow
GAZPRU	JSC Gazprom
GAZPRU.Gneft	JSC Gazprom Neft
LUKOIL	Lukoil Company
MBT	Mobile Telesystems
MDMOJC	MDM Bank Open JSC
ROSNEF	Open JSC Rosneft
RSBZAO	JSC Russian Standard Bank
RUSAGB	Russian Agriculture Bank
RUSRAI	JSC Russian Railways
RUSSIA	Russian Federation
SBERBANK	Sberbank
VIP	Open JSC VimpelCom Limited
VTB	JSC VTB Bank

Note that JSC is the acronym for Joint Stock Company.

B Systematic factors

Factor	Index
Europe	MSCI EUROPE
Asia	MSCI AC ASIA
North America	MSCI NORTH AMERICA
Latin America	MSCI EM LATIN AMERICA
Middle East and Africa	MSCI FM AFRICA
Pacific	MSCI PACIFIC
Materials	MSCI WRLD/MATERIALS
Consumer products	MSCI WRLD/CONSUMER DISCR
Services	MSCI WRLD/CONSUMER SVC
Financial	MSCI WRLD/FINANCIALS
Industrial	MSCI WRLD/INDUSTRIALS
Government	ITRAXX SOVX GLOBAL LIQUID INVESTMENT GRADE

Table 6: Systematic factor - Index mapping.

Electronic copy available at: https://ssrn.com/abstract=3446615

C Scores

We briefly describe the two scores namely the BIC and BDs used for BN learning. For details, we refer to Koller & Friedman, 2009, Section 18.3. BIC is a likelihood score defined as:

$$\text{score}_{BIC}(\mathcal{G} : \mathcal{D}) = \log L(\mathcal{G} : \mathcal{D}) - \frac{1}{2}|\mathcal{G}| \log N,$$

where $|\mathcal{G}|$ is the complexity of the network, the number of independent parameters in the network. This score penalizes explicitly the score with the number of independent parameters, hence assigning higher scores to sparser structures.

The following expression is the closed-form derived for the marginal-likelihood of the structure score:

$$P(\mathcal{D} \,|\, \mathcal{G}) = \prod_i \prod_{k=1}^{q_i} \frac{\Gamma(\alpha_{X_i|k})}{\Gamma(\alpha_{X_i|k} + M[k])} \prod_{j=1}^{r_i} \left[\frac{\Gamma(\alpha_{j|k} + M[j,k])}{\Gamma(\alpha_{j|k})} \right], \tag{25}$$

where $\alpha_{X_i|k} = \sum_{j=1}^{r_i} \alpha_{j|k}$.

BDs is derived from a different score, BDeu, which is obtained from 25 by assuming a uniform prior distribution over the parameters (Dirichlet distribution with all the hyperparameters taking the same value α). Let $\alpha_{j|k} = \alpha_i/(r_i q_i)$ and $\alpha_i = \alpha$, where r_i is the number of states of X_i and q_i is the number of configurations of parents of X_i, the number of parents configuration of X_i. Then score BDeu is defined as follows

$$\text{BDeu}(\mathcal{G}, \mathcal{D}; \alpha) = \prod_i \prod_{k=1}^{q_i} \frac{\Gamma(r_i\alpha_i)}{\Gamma(r_i\alpha_i + M[k])} \prod_{j=1}^{r_i} \frac{\Gamma(\alpha_i + M[j,k])}{\Gamma(\alpha_i)}. \tag{26}$$

Scutari, 2016 argues that choosing uniform prior distributions over $\theta_{\mathcal{G}X_i} \,|\, \text{Pa}_{X_i}$ and $\mathcal{G}$ can have a negative effect over the quality of the results obtained with the score BDeu. To avoid this he introduces the score BDs.

In the first place we see that if $P(k) = 0$ for some $k \in \{1, \ldots, q_i\}$ and $i \in [n]$, or if the sample size of $\mathcal{D}$ is very small, it may happen that $M[k] = 0$ for some configurations of Pa_{X_i} which do not appear in $\mathcal{D}$, then we can split

$$\text{BDeu}(\mathcal{G}, \mathcal{D}; \alpha) = \prod_i \left(\left(\prod_{\substack{k=1) \\ M[k]=0}}^{q_i} \frac{\Gamma(r_i\alpha_i)}{\Gamma(r_i\alpha_i)} \prod_{j=1}^{r_i} \frac{\Gamma(\alpha_i)}{\Gamma(\alpha_i)} \right) \left(\prod_{\substack{k=1 \\ M[k]>0}}^{q_i} \frac{\Gamma(r_i\alpha_i)}{\Gamma(r_i\alpha_i + M[k])} \prod_{j=1}^{r_i} \frac{\Gamma(\alpha_i + M[j,k])}{\Gamma(\alpha_i)} \right) \right).$$

We note that as the number of parent configurations which appear in the data $\mathcal{D}$ decreases, the effective imaginary sample size decreases, as

$$\sum_{k:M[k]>0} \sum_j \alpha_i \leq \sum_{k,j} \alpha_i = \alpha. \tag{27}$$

This induces the posterior to converge to the corresponding likelihood estimation and hence leaning towards overfitting and including spurius edges in $\mathcal{G}$. To avoid this problem we define:

$$\tilde{q}_i = |\{k \in \{1, \ldots, q_i\} : M[k] > 0\}| \quad \text{and} \quad \tilde{\alpha}_i = \begin{cases} \alpha/(r_i\tilde{q}_i) & \text{if } \tilde{q}_i > 0, \\ 0 & \text{otherwise.} \end{cases}$$

With this new definition the expression 27 becomes an equality, $\sum_{k:M[k]>0} \sum_j \alpha_i = \alpha$. Moreover, we see that the uniform prior that we just defined is on the conditional distribution which can be estimated from $\mathcal{D}$, so this is a empirical Bayesian score. Finally, we substitute α_i in 26 by $\tilde{\alpha}_i$ and obtain:

$$\text{BDs}(\mathcal{G}, \mathcal{D}; \alpha) = \prod_i \prod_{\substack{k=1 \\ M[k]>0}}^{q_i} \frac{\Gamma(r_i\tilde{\alpha}_i)}{\Gamma(r_i\tilde{\alpha}_i + M[k])} \prod_{j=1}^{r_i} \frac{\Gamma(\tilde{\alpha}_i + M[j,k])}{\Gamma(\tilde{\alpha}_i)}. \tag{28}$$

Electronic copy available at: https://ssrn.com/abstract=3446615

D Stress scenarios

The first scenario is stressing three banks in the network, BKECON, BOM and SBERBANK. Figure 7 depicts the network with the nodes colored according to the probabilities obtained in that scenario, shown next to it. We observe that the values obtained are similar to the ones of the scenario where the sovereign is stressed. This may be due to the strong connections of the sovereign with big companies such as GAZPRU and GAZPRU.Gneft and because the stressed banks are in the periphery of the network.

Issuer	Prob.
BKECON	1
BOM	1
SBERBANK	1
RUSAGB	0.7757
VTB	0.7287
GAZPRU	0.69
AKT	0.5797
GAZPRU.Gneft	0.5593
ROSNEF	0.5514
RUSSIA	0.5351
CITMOS	0.5107
LUKOIL	0.4948
VIP	0.4742
RUSRAI	0.4528
ALROSA	0.4244
MDMOJC	0.4242
MBT	0.3379
RSBZAO	0.3319

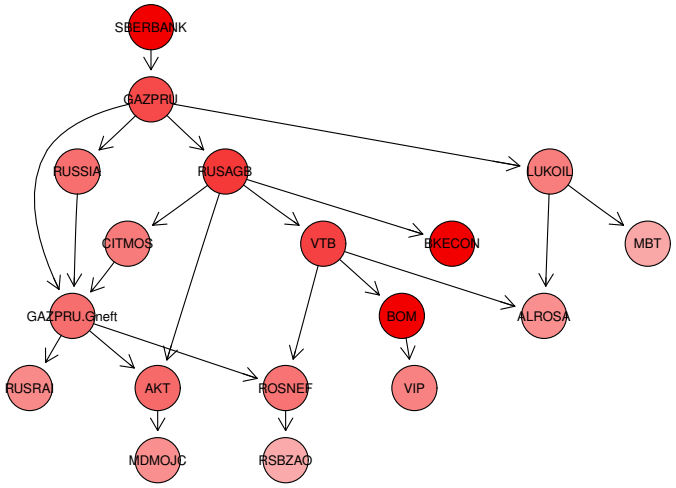


Figure 7: Probabilities and network in the scenario of three banks stressed.

As the Russian crisis of 2014 was strongly related to oil industry, in the second scenario we stress four of the largest oil companies: AKT, GAZPRU, LUKOIL, and ROSNEF. Figure 8 shows the network and the probabilities for this case. It is noticeable that the oil companies have a larger impact on the network. Note also that in this case four nodes are stressed. However, one can see that these nodes are more centered and the rest of the nodes are more stressed.

Issuer	Prob.
AKT	1
GAZPRU	1
LUKOIL	1
ROSNEF	1
GAZPRU.Gneft	0.8382
RUSAGB	0.7867
SBERBANK	0.7563
RUSSIA	0.6469
VTB	0.6028
RUSRAI	0.5814
MDMOJC	0.5321
BKECON	0.5273
CITMOS	0.5133
MBT	0.5078
BOM	0.4972
ALROSA	0.4708
RSBZAO	0.4053
VIP	0.366

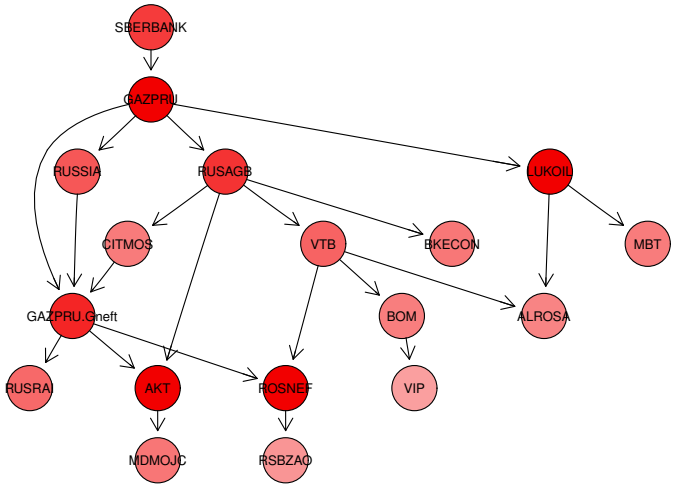


Figure 8: Probabilities and network in the scenario of four oil companies stressed.

Electronic copy available at: https://ssrn.com/abstract=3446615