SYSTEMIC RISK TRANSMISSION IN EUROPEAN SECTORAL CDS USING BAYESIAN NETWORKS

Jesúa López Máñez

Trabajo de investigación 21/003

Master en Banca y Finanzas Cuantitativas

Tutores: Dra. Laura Ballester
Dr. José Manuel Pavia

Universidad Complutense de Madrid

Universidad del País Vasco

Universidad de Valencia

Universidad de Castilla-La Mancha

www.finanzascuantitativas.com

SYSTEMIC RISK TRANSMISSION IN EUROPEAN SECTORAL CDS USING BAYESIAN NETWORKS

Jesúa López Máñez

Supervisors: Laura Ballester Miquel
Jose Manuel Pavía Miralles

MSc in Banking and Quantitative Finance

Complutense University of Madrid

University of the Basque Country

University of Valencia

University of Castilla-la Mancha

www.finanzascuantitativas.com

Abstract

Credit derivative contracts, including Credit Default Swaps (CDS), are very important financial instruments in international markets that have played a key role in allowing the transfer, in whole or in part, of the credit risk inherent in an underlying asset or set of assets without transferring ownership.

The objective of this work is to study the transmission of systemic risk among different European sectoral CDS during the current COVID-19 crisis. This main objective has been subdivided and concretized into three goals: understanding the structure and the dynamics of risk transmission underlying European sectoral CDS, learning what the proportion of new systemic risk resulting from risk transmission among sectors is, and developing a coherent model into which possible prior information on the transmission of systemic risk can be introduced.

The methodology used to analyze the transmission of systemic risk among the different European sectoral CDS is based on the use of Dynamic Bayesian Networks, characterized by Network Structure Learning and Parameter Learning. The structure of the Dynamic Bayesian Networks is built from the Vector Autoregressive Moving Average Models and the whole study is carried out in the context of Bayesian statistics. Network Structure Learning is performed by using independence tests and network scores, combined with exhaustive and heuristic exploration methods. Parameter Learning is undertaken using simulation methods based on Markov Chain Monte Carlo.

The application of this methodology to analyze the problem addressed provides coherent and relevant findings. It follows that only the relationships between the original series and the series delayed with one or two lags are relevant. By carrying out the complete Bayesian Network Learning, the risk transmission structure is ascertained and a complete understanding of the way in which systemic risk is transmitted in the Dynamic Bayesian Network is obtained. The underlying risk transmission structure is robust and stationary over time, thus pointing to a clear transmission of systemic risk. The new systemic risk returns are explained between 5% and 40% by the transmission among the different CDS series. The effects of lagged series on sectoral CDS are either positive or negative, indicating which relationships among the CDS series show a direct transmission of risk and which relationships correct the transmission marked by the system. The proposed modeling allows for a more advanced analysis of the transmission of systemic risk among European sectoral CDS series and longer-term or more complex relationships.

Keywords: CDS; Systemic Risk; Credit Risk; Transmission; Markov Chain Monte Carlo; Vector Autoregressive Moving Average Models; Dynamic Bayesian Networks

Contents

1 Introduction

Credit derivative contracts are fundamental financial instruments in international markets. Credit derivatives, including Credit Default Swaps (CDS), have played a key role in allowing the transfer, in whole or in part, of the credit risk inherent in an underlying asset or set of assets without transferring ownership. The increasing use of this product has led to the creation of CDS for a large number of companies, sectors and countries. The CDS market has become increasingly prominent, their trading volume has grown significantly, and they are now considered important indicators of credit quality and have come to be used routinely as a proxy for credit risk. The growing stature of CDS has led to greater interest in their study, giving rise to a productive branch of literature, highly topical and relevant. There are multiple lines of active research in which CDS are the protagonist and which follow the more general investigations into credit risk. Among these lines, the literature related to the transmission of credit risk stands out. The limitations still present in the study of the transmission of credit risk mean that taking a fresh approach, such as that offered by Bayesian statistics, should be welcome in view of the potential innovations it can provide.

The objective of this research is to study the transmission of systemic risk among different CDS, in particular among European sectoral CDS during the current COVID-19 crisis. The transmission of systemic risk is a complex problem that is difficult to study given the large number of factors involved and the diversity of possible dynamics. Because of the non-triviality of the problem addressed, the goal of analyzing the transmission of systemic risk is of particular interest. This main objective is in turn made up of several smaller ones, which together respond to the overall aim of analyzing the transmission of systemic risk. The first goal is to understand the structure of risk transmission underlying the European sectoral CDS. The second is to determine the dynamics of systemic risk transmission, in general, and learn how risk is transmitted among sectors, in particular. The third aim is to determine the proportion of new systemic risk that results from risk transmission among sectors, as well as from new market innovations. Moreover, we would like to provide an alternative to the classical transmission analyses in which possible expert information on the transmission of systemic risk can be introduced.

This research is the result of combining two areas of academic literature: the financial literature that analyzes the transmission of credit risk and the mathematical literature concerning Bayesian statistics. These two lines of research are current and valid, giving rise to multiple publications and new studies. Within the vast literature on CDS, this study is framed within the line of research related to the analysis of risk transmission, while in the Bayesian statistics literature it falls in the area of investigation dealing with dynamic models. The intersection between these two bodies of literature is found in network modeling. Within the analysis of credit risk transmission, the network approach, which analyzes how credit risk is transmitted among various nodes in a network, has recently become popular [Brownlees et al., 2021, Sun et al., 2020, Kanno, 2020, Chen et al., 2020, Gandy and Veraart, 2019, Agosto et al., 2020]. In addition, within the Bayesian statistics literature, many papers related to the network approach, specifically Bayesian networks, have been produced [Scutari and Denis, 2021, Heckerman, 2021, Fortunato et al., 2019]. The literature related to the application of networks to analyze the transmission of credit risk in the context of Bayesian statistics is still to be explored, and this current study is presented as the beginning of a fresh line of research. Thus, the novel contribution of this work is that it analyzes the transmission of credit risk through networks in the context of Bayesian statistics.

The methodology used to analyze the transmission of systemic risk among the different European sectoral CDS in this study is based on the use of Bayesian networks. In order to analyze the transmission of systemic risk, or credit risk specifically, multiple methodologies are used, which are approached from different points of view depending on the main objective. Given the goals of this work and the temporary nature of the problem, the use of Dynamic Bayesian Networks has been proposed to address the problem studied. These networks are characterized by two fundamental elements: Network Structure Learning and Parameter Learning, which make up the entire statistical learning process, from model construction and selection to model estimation and diagnosis. In our case, the structure of the Dynamic Bayesian Networks is built from the Vector Autoregressive Moving Average Models, which determine the possible dynamic relationships present in the CDS network. The whole study is carried out in the context of Bayesian statistics, which presents a new approach to address this problem and provides the necessary advantages to meet the objectives set. Network Structure Learning is performed using unconditional independence tests, conditional independence tests and network scores, combined with exhaustive and heuristic exploration methods. Parameter Learning is carried out using simulation methods based on Markov Chain Monte Carlo (MCMC), using both methods based on maximum likelihood and penalized methods. In short, the methodology employed in this work can be summarized in the use of Dynamic Bayesian Networks based on Vector Autoregressive Moving Average Models in the context of Bayesian statistics, using various techniques to perform Network Structure Learning and Parameter Learning.

The analysis of the transmission of systemic risk among the different European sectoral CDS during the current COVID-19 crisis has been correct and has led us to important results. The initial approach to the modeling of the problem has been the consideration of a Dynamic Bayesian Network in which the relationships between the original sectoral CDS series and the delayed series with between one and five lags are considered. The first finding is that of all the relationships considered for our Bayesian network, only a few, namely those between the original series and the series delayed with one or two lags, are relevant. From the conditional and unconditional independence tests we have verified that there are no relevant transmission relationships between the original series and the series delayed with three, four or five lags, in the same way that there are no relevant transmission relationships instantaneously when working with daily data. The second finding of the study is the systemic risk transmission structure among the different European CDS. By performing structure learning following the various strategies proposed, we have been able to show how systemic risk is transmitted among the different European CDS, obtaining the risk transmission structure corresponding to the Dynamic Bayesian Network. This structure has been verified by checking that the transmission relationships are robust and consistent with the complete data series. The underlying risk transmission structure is robust and stationary over time, thus pointing to a clear transmission of systemic risk. The third finding of the study involves a complete understanding of the way in which systemic risk is transmitted in the network. By analyzing the posterior distributions of all the parameters, we have come to understand how risk is transmitted along the Dynamic Bayesian Network, as well as learning what proportion of the new risk is a consequence of the transmission of risk from previous time points. We have found that the new systemic risk is explained between 5% and 40% by the transmission among the different CDS series, which is a high but reasonable proportion that tells us much about the importance of risk transmission. The effects of lagged series on sectoral CDS are either positive, as a direct propagation of systemic risk, or negative, as a correction of the propagated risk. The proposed modeling allows for a more advanced analysis of the transmission of systemic risk among European sectoral CDS series.

The format of this work follows a classical structure in which the fundamental elements of a piece of research are presented in an orderly fashion. This brief introduction is followed by a complete review of the literature in Section 2, which summarizes the different lines of research followed and explains where exactly this research fits in. Next, Section 3 presents the European sectoral CDS data we have worked with throughout the study, and makes a brief descriptive analysis of them, indicating how they will be analyzed. This is followed by Section 4 in which the methodology used throughout this work is explained and the different elements that make it up, as well as their interrelationships, are discussed. More specifically, this modeling section begins with an introduction to Bayesian statistics followed by a presentation on Bayesian networks, a brief refresher on Vector Autoregressive Moving Average Models and a thorough introduction to Bayesian Network Learning. Subsequently, the details of the implementation and the results of the study are explained in great detail in Section 5, this being the most important segment of the work. Specifically, in this section devoted to implementation and results we present the findings for Structure Learning followed by those for Parameter Learning. Finally, in Section 6, the conclusions resulting from the study are given, summarizing the most significant results and briefly suggesting some future lines of research.

2 Review of the Literature

This research is the result of combining two areas of literature, one that analyzes the transmission of credit risk and another dealing with Bayesian statistics. The intersection between the two is found in network modeling. Within the analysis of credit risk transmission, the network approach, which analyzes how credit risk is transmitted among various nodes in a network, has become popular. In addition, in terms of the Bayesian statistics literature, many papers related to the network approach, specifically Bayesian networks, have been produced. Thus, the novel contribution of this work is that it analyzes the transmission of credit risk through networks in the context of Bayesian statistics.

Credit derivative contracts, including CDS, have become popular in the last two decades, gaining even more strength due to the recent financial crises [Duffie, 1999, Hull and White, 2000, Hull and White, 2001, Hull and White, 2003]. The literature looking at CDS is very extensive and dynamic, with several lines of research open, one of which is centered on analyzing the transmission of credit risk [Arora et al., 2012], offering multiple approaches and methodologies [Benoit et al., 2016, Hull et al., 2004, Longstaff et al., 2005]. While the literature related to credit risk extends to all those underlying assets for which CDS exist, the works that analyze credit risk transmission mostly examine sovereign CDS [Kartal, 2020, Ismailescu and Kazemi, 2010]. This current study targets European sectoral CDS, an area in which the literature is much less abundant. Studies of systemic risk transmission focus on the dynamics of transmission and, in turn, on the strength of such transmission [Corbet et al., 2021]. In this study we will follow the same strategy to hone in on systemic risk transmission in European sectoral CDS.

The Bayesian approach, based on Bayes theorem, combines the construction of complex models with the inclusion of known prior information about the parameters in the models [Gelman et al., 2014]. Bayesian statistics allows a new approach to be taken with respect to frequentist statistics, both in conceptual and practical terms. In the Bayesian context, the parameters can be considered as random variables, thus providing a more realistic approach in many cases where the parameter is of interest in itself [Lindley and Smith, 1972]. In addition, the Bayesian approach to analysis not only uses the data, but the results are also based on a priori sources of information known prior to data sampling. Bayesian inference is based on simulation methods, which allow one to calculate the a posteriori distributions of the parameters when it is not possible to sample directly [Gilks et al., 1996]. The literature on Bayesian statistics is very extensive due to its infinite number of fields of application and the advantages the Bayesian approach can provide with respect to frequentist statistics [Bauwens et al., 1999]. Within the diverse literature on Bayesian statistics, in this study we focus on the literature associated with dynamic models, due to the nature of the problem studied [West and Harrison, 1989].

Bayesian networks are a class of probabilistic models encompassed within Bayesian statistics. Bayesian belief networks are very useful and powerful models that combine probabilistic reasoning and graphical modelling [Spiegelhalter et al., 1993] and can successfully manage the different elements of uncertainty and causality in complex problems [Cowell et al., 1999, Jensen and Nielsen, 2007]. These models are composed of a combination of random variables and the relationships between them, defined from a directed acyclic graph [Kjaerulff and Madsen, 2008]. The nature of these models is clearly dynamic, due to the possibility of introducing series as the nodes of the network itself. The literature on Bayesian networks is less extensive, though increasingly frequent in recent years

[Scutari and Denis, 2021]. The fields of application of Bayesian networks are also very diverse, as they are compatible with most contexts in which the goal is to analyze the relationship between time series [Nagarajan et al., 2013, Heckerman, 2021]. Bayesian networks are compatible with multiple models that define the structure, such as the Vector Autoregressive Moving Average Models [Nicholls and Hall, 1979, Neapolitan, 2004]. In the case of introducing an autoregressive structure in Bayesian networks there are Dynamic Bayesian Networks, which are of special interest for studying the relationships among financial time series [Neil et al., 2005, Fortunato et al., 2019]. Dynamic Bayesian Networks take into account the temporal nature of the data and consider the possible relationships between the series over time, which is essential for analyzing the transmission of systemic risk.

One of the lines of research in the analysis of the transmission of systemic risk is a network approach in which the analysis of credit risk transmission consists in analyzing the transmission of risk among various nodes of a network [Chen et al., 2020, Gandy and Veraart, 2019, Peltonen et al., 2014]. This approach presents the transmission of credit risk among different time series as relationships within a network, which are directed from one series to another [Agosto et al., 2020, Markose et al., 2012]. From the structures of the network and the relationships determined, it is possible to analyze how shocks are transmitted between series [Kanno, 2020]. This approach is widely used in the literature to study risk contagion and risk spillover [Brownlees et al., 2021, Sun et al., 2020]). The network approach has important advantages over other alternative analyses due to the good results that these models obtain and their interpretability. The network analysis of credit risk transmission has been applied to CDS data.

Combining the network approach to perform the analysis of credit risk transmission on CDS data with the possibility of working with Dynamic Bayesian Networks, we arrive at our proposed line of research, which, despite the limited literature, arises naturally from the two topics raised. This work thus makes important contributions in the different areas that make up the study. On the one hand, our data involves European sectoral CDS, which is a novel contribution since the literature on sectoral CDS is very scarce and the little that does exist does not employ a network approach. On the other hand, the Dynamic Bayesian Networks approach introduces a new methodology, not previously considered and with great potential. Finally, this work analyses data associated with the current COVID-19 crisis, a period for which few risk transmission studies have been conducted. For all these reasons, this study makes a novel and informative contribution to the literature.

3 Data and Preliminary Analysis

In this study we will analyze the transmission of systemic risk among the different European industrial sectors during the COVID-19 crisis. This transmission will be measured from the study of the spread of the different sectoral CDS associated with each of them, by analyzing the temporal evolution of the whole set.

CDS are key financial instruments to enable the transfer of credit risk. We can define a CDS as a financial derivative contract that exchanges the credit risk of a product for a premium. CDS are financial swap agreements in which the seller of the CDS will compensate the buyer in the event of a debt default. In this way, a CDS allows the buyer to insure himself in the event of default of the reference asset. CDS are mainly used for two purposes, to hedge risk and for speculation. CDS are defined by the credit risk being exchanged and by the term over which it is exchanged. The CDS spreads are quoted on the market in basis points. We consider CDS contracts maturing in five years, since they are the most liquid and the most traded credit risk derivative [Norden and Weber, 2009, Ballester et al., 2016].

Specifically, throughout this work we have taken the daily data corresponding to each of the industrial sectors from December 2007, when its use became widespread, until April 2021. However, the study focuses only on the period from December 02, 2019 to April 30, 2021, a period that includes the current COVID-19 financial crisis. Subsequently, the entire period of data will be used to carry out different diagnoses and studies complementary to the analysis of the problem, which will also be explained.

The data used have been obtained from Thomson Reuters Datastream. The data downloaded corresponds to the period starting in December 2007 since, although previous data is available for some sectoral CDS, for most of them no further information is available. Because the focus of the study has been on the current COVID-19 crisis, only data from December 2019 to April 2021 have been taken, a total of 3690 observations, 369 for each of the 10 sectoral CDS.

The data includes several European sectoral CDS series, a total of ten sectoral CDS series. The sectoral CDS, together with their Datastream codes, correspond to the banking sector (DSEBK5E), the consumer goods sector (DSECG5E), the electricity sector (DSEEP5E), the energy sector (DSEEC5E), the manufacturing sector (DSEMF5E), other financial enterprises (DSEOF5E), the services sector (DSESC5E), the sovereign (DSESV5E), the telephony sector (DSETL5E) and the transport sector (DSETR5E). Therefore, nine series of sectoral CDS and one series of sovereign sectoral CDS have been taken, which we will call for simplicity also as sectoral CDS.

Data corresponding to daily frequencies are used in order to analyze how systemic risk spreads is transmitted between the different industrial sectors in Europe. This daily frequency allows the study described above. The study with daily data allows for the elimination of the noise inherent to higher frequency data. In addition, working with daily data allows a better interpretation of the results in the models used.

The data are therefore composed of the spreads of the ten sector CDS described above. By working with the CDS spreads we are working with a measure of credit risk. Thus, we must focus on the movements of the data themselves to understand the underlying risk, since CDS spreads are in themselves a measure of the probability of default.

In our case we are interested in the estimation of the model itself. Thus, we will use the whole data set for model fitting. The fact of selecting the whole set as the training sample has an important risk associated with it, namely overfitting. This potential problem will not be a risk for our model in practice, since the construction of the model penalizes overfitting significantly. In short, it makes sense to take the whole data set for model fitting, so we select all 3690 observations for model construction and estimation.

The models are based solely on the values of the CDS spreads themselves, without considering possible exogenous variables. Not introducing other exogenous variables is due to the interest in knowing how systematic risk is transmitted among the different industrial sectors. Thus, the models considered only take into account the values of the spreads of the different sectoral CDS, both at present and at previous points in time.

In Figure 1 we can see the time evolution of the spread of each of the ten sector CDS analyzed, which indicates the time evolution of the risk. It is therefore essential to analyze the level of the CDS and their time evolution, as well as to compare them with each other, in order to get an initial idea of the problem.

First, in Figure 1 we can see how the different graphs have similar behaviors over time. It is easy to see how, although the level of risk in each of the sectors is notably different, the comovements between the different CDS series are evident. The relationships between the CDS seem to indicate generally positive correlations between the risks of each of the sectors. The apparent correlation between sectoral CDS justifies the use of the models presented in Section 4. This analysis is extended later in this study by performing a formal analysis of the correlations, as well as an analysis of the correlations between lagged time moments, seeing that a multivariate autoregressive structure should be introduced.

On the other hand, throughout the period studied, from December 2019 to April 2021, we can observe several moments in which the CDS have undergone significant alterations in level, rises or falls. The most important movements are easily identifiable and are explained by the current COVID-19 financial crisis. Initially, horizontal movements are presented in the spread of the different sector CDS, from the month of December 2019 until the beginning of March 2020, when COVID-19 expanded globally becoming pandemic. During the month of March, it can be seen how the CDS spread increased considerably, reaching levels much higher than those observed the previous months, reaching the maximums shown in the table 1. Thus, the uncertainty present in the markets significantly altered the levels at which the different sectoral CDS were traded, marking the first major movement in the period under analysis. Subsequently, a pronounced decrease in the value of the spread of all sectoral CDS can be observed, corresponding to the beginning of June 2020, when in Europe there were announcements regarding the possibility of trips abroad in the summer. This news regarding a reactivation of the economy was very well received in the market, thus ensuring that the risk of default in the various sectors would be notably lower. Finally, a last notable decline can be seen during the month of November 2020, as a result of the news concerning the discovery of the first effective vaccines against COVID-19. In short, the three events described above, strongly associated with the COVID-19 crisis, are the ones that have determined the largest movements during the period from December 2019 to April 2021.

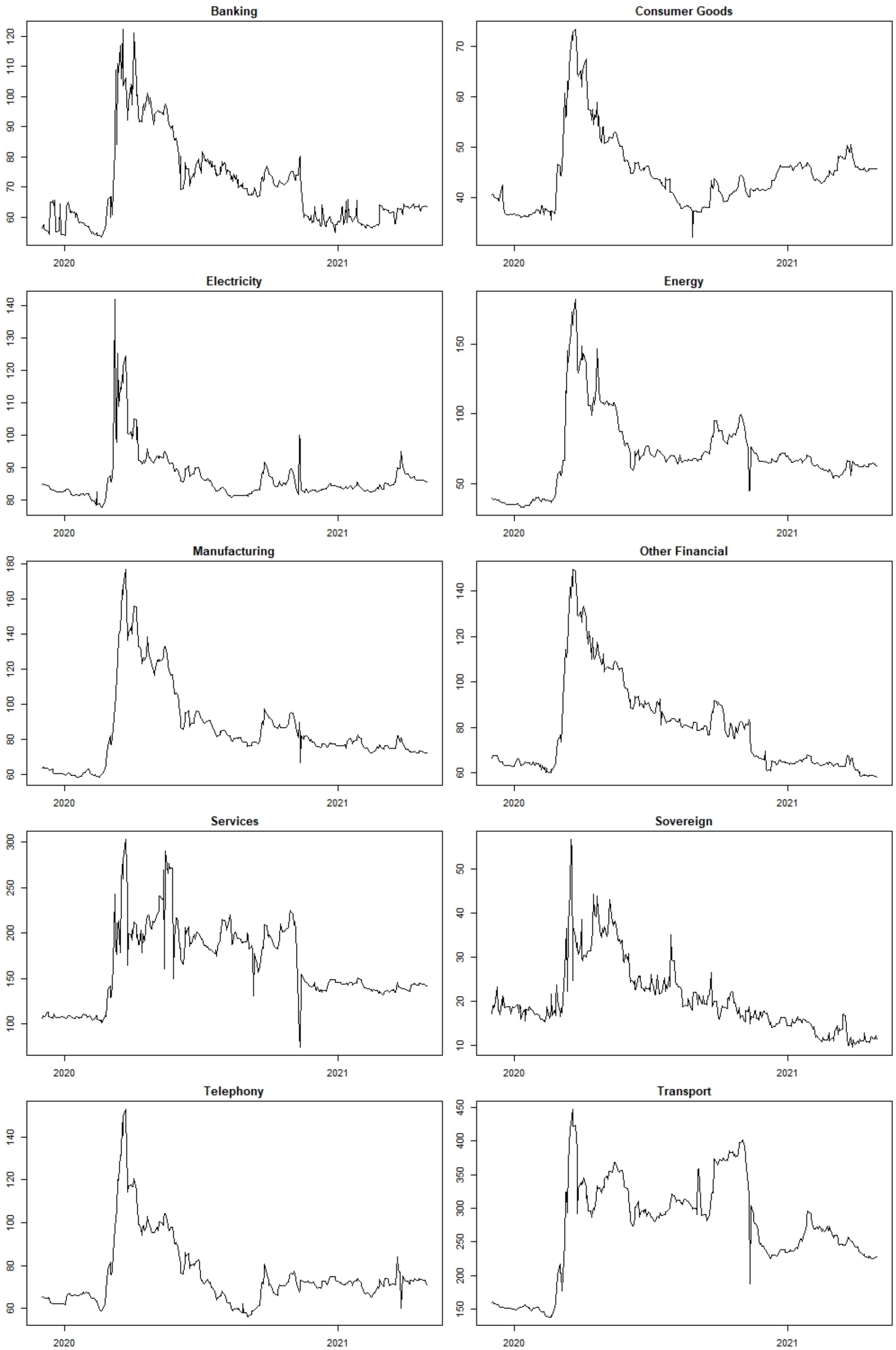


Figure 1: European sectoral CDS spread returns correlation.

	mean	sd	Min.	1st Qu.	Median	3rd Qu.	Max
Banking	70.29	14.25	53.37	58.90	65.75	75.73	122.34
Consumer Goods	44.68	7.27	32.09	39.69	43.71	46.33	73.34
Electricity	86.75	7.42	77.76	82.83	84.47	88.09	141.92
Energy	71.45	27.17	32.96	59.50	66.65	78.75	182.59
Manufacturing	85.86	23.02	58.42	74.39	78.99	90.49	176.76
Other Financial	79.30	19.96	58.19	64.06	71.75	88.69	149.67
Services	164.06	43.01	74.84	136.53	148.59	197.47	303.03
Sovereign	20.63	7.94	9.69	15.81	18.66	23.22	56.67
Telephony	75.67	16.12	55.90	66.21	71.69	76.45	152.92
Transport	269.5	72.43	138.1	230.4	277.1	316.1	447.1

Table 1: Summary of the European sectoral CDS spreads.

Comparing the different graphs in Figure 1, we can see that the sectors whose CDS spreads are higher are the services and transport sectors. These two sectors seem to have a higher default risk, so they tend to trade with higher spreads than the rest. Similarly, in proportional terms, these two sectors are the ones that were most affected by the COVID-19 crisis. However, it should be noted that the behavior of sector CDS prices in each sector was different. That said, in this paper we do not aim to study the level of the price of the different CDS as a whole, but rather the transmission of systemic risk between sectors, so we will be more interested in the interaction between these time series of spread values.

Once we have analyzed the general behavior of the different time series of spreads over time, it is interesting to make a brief numerical summary of the information contained. For this purpose we can calculate some basic statistics for each of the series. These statistics include measures of central tendency, the mean and median, and measures of dispersion, the quartiles and standard deviation. These statistics are included in the table 1.

In the table 1 we can see that the different series of the spreads of the various European sectoral CDS are notably different in level and behavior. In this sense it can be seen that some sectoral CDS are traded on average at lower values than others, denoting a lower risk of default in the companies of that industrial sector. On the other hand, some sectors have a higher volatility than others, showing a more uncertain behavior and, probably, more dependent on the situation caused by the COVID-19 financial crisis. On the other hand, we can see how spreads follow a distribution with a clear positive asymmetry in most cases. This fact is clearly related to the negative asymmetry observable in most market price series, since these prices are measuring a risk.

In order to know how the systemic risk propagates among the different CDS, we will consider a model in which the only covariates are the spreads of the rest of the CDS, as well as autoregressive terms of the others or of oneself. However, this paper is not interested in analyzing the level present in the different CDS spreads, but rather in how they are correlated, so we focus on analyzing the comovements. We can begin to see the relationship between the different sectors by calculating the correlations between the different series.

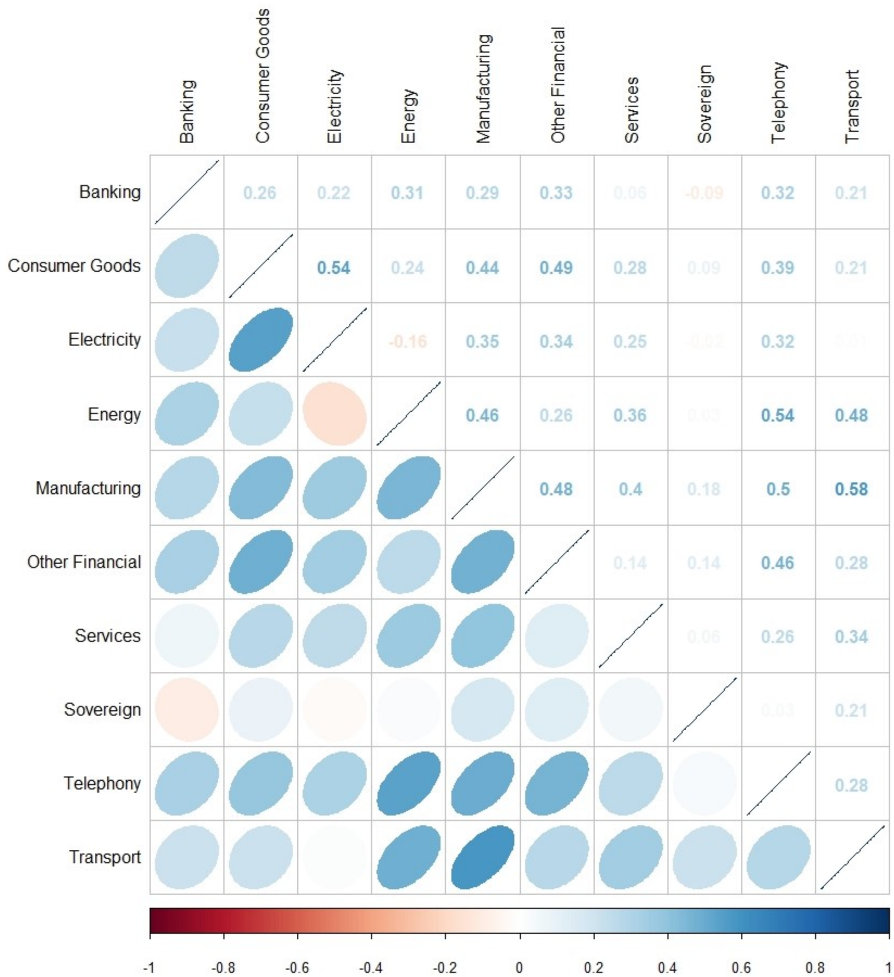


Figure 2: European sectoral CDS spread returns correlation.

In Figure 2 we can see the two to two correlations between the ten time series studied. Looking at the plot the first striking fact is that all correlations between any pair of sectoral CDS are positive, which is truly intuitive, consequence of the fact that all industrial sectors are part of the same economy and spreads are measures of risk. Thus, when one of the industrial sectors has suffered shocks that have caused an increase in the spread of the associated CDS sector, then the rest of the sectors have suffered, to a greater or lesser extent, a transmission of this shock that has caused an increase in the spread of the rest of the CDS sector. If we look in detail we can see how some CDS sector pairs have a higher correlation than others. We also highlight that the sovereign CDS sector is the one with the lowest correlation with the rest of the CDS sector.

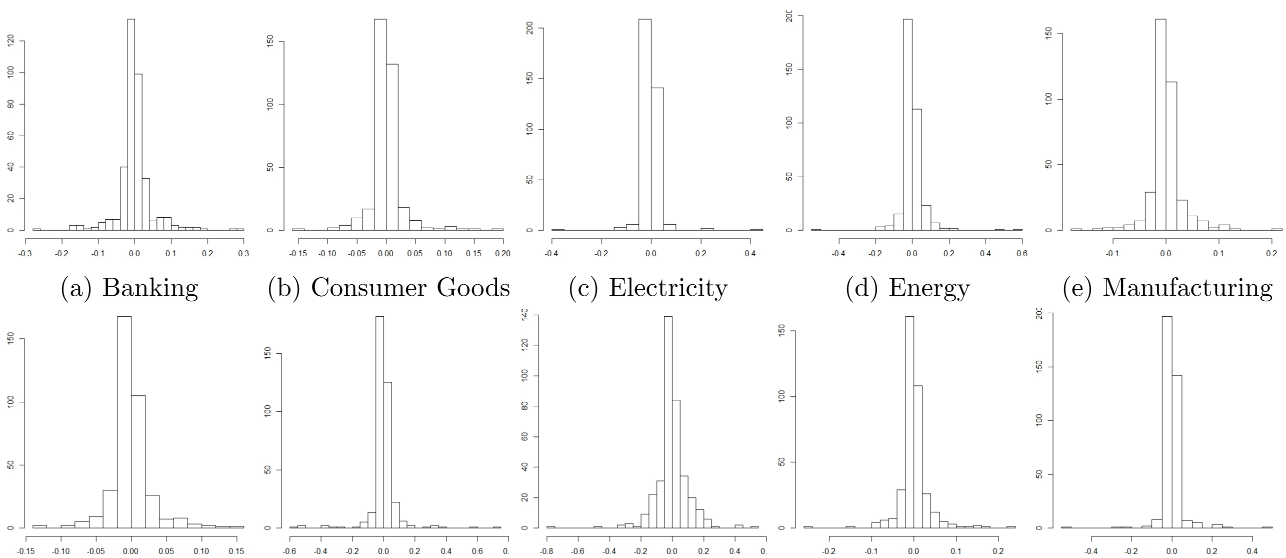
The correlations presented, however, do not show the complete relationship between the different series. Throughout this paper we will not only consider the correlations between the different series of the sectoral CDS, but we will also consider the possible partial correlations that the series have with all of them at lagged time points. Moreover, once the information on the existing lagged relationship is provided, it is important not only to study the correlation between the series, but also to study the conditional correlations. Therefore, although the figure 2 presents a first approach to the existing relationships between the time series, throughout this study we will consider much deeper information, evaluating also the partial correlations and the conditional correlations.

In order to analyze how the movements between the different CDS spreads are related, we have to perform a previous treatment of the data. Specifically, the treatment we will follow consists of working with the logarithmic returns of the different series. Thus, from the time series of the CDS spreads we obtain the logarithmic returns

$$X_i(t) = \log\left(\frac{Z_i(t)}{Z_i(t-1)}\right),$$

where $Z_i(t)$ indicates the sectoral CDS spread i at t and $X_i(t)$ indicates the logarithmic returns of such sectoral CDS.

The graphical representation of sectoral CDS spread returns over time is not a clear aid to interpretation. By plotting the returns of a given series over time, we lose information about the overall performance. Thus, although it is true that the representation of the series of returns can provide information regarding the periods of higher and lower volatility, in our case we do not consider this study to be of great interest. However, it may be interesting to include a graph showing how all the CDS series returns are distributed.



(a) Banking (b) Consumer Goods (c) Electricity (d) Energy (e) Manufacturing

(f) Other Financial (g) Services (h) Sovereign (i) Telephony (j) Transport

Figure 3: European sectoral CDS spread returns histogram.

In Figure 3 we can see how the returns observed in the different sectoral CDS spreads are distributed. We can see how they all have a distribution centered on zero, i.e., the mean of the returns in the ten series is approximately zero. On the other hand, the observed distribution seems to have a high kurtosis, forming distributions much more shaded than a Normal distribution, as we can see in the table 2. Thus, many values close to zero and only a few significantly away from this value are observed.

In the table 2 we present some key statistics to complete our understanding of the distribution of returns. In this table we include information on standard deviation, to summarize the dispersion of the distribution, and skewness and kurtosis, to analyze the shape of the distribution. We do not include information on the mean and median since in most cases they are close to zero.

	sd	Skew.	Kurt.
Banking	0.0496	0.6120	12.3700
Consumer Goods	0.0277	1.6539	17.1442
Electricity	0.0384	2.9620	75.9043
Energy	0.0634	1.6332	39.9708
Manufacturing	0.0317	0.5300	13.4129
Other Financial	0.0288	0.6121	10.1981
Services	0.0937	0.2190	27.8316
Sovereign	0.1042	-0.8695	14.6214
Telephony	0.0348	0.4366	18.0611
Transport	0.0549	-0.3743	44.9938

Table 2: Summary of the returns of the European sectoral CDS spreads.

By analyzing the results of the table 2 we can see the similarities and differences in the behavior of the returns of each of the return series. In particular, it is easy to see how the standard deviation of the returns is notably different between the different sectoral CDS. Thus, the sectoral CDS of sovereign bonds has a high volatility, while the volatility of the sectoral CDS of consumer goods companies and financial companies is very low. On the other hand, it can be seen that in most cases the distribution of yields has a slight positive skewness. Moreover, all the distributions present a significant excess of kurtosis with respect to the Gaussian distribution. In short, the yield series have an asymmetric distribution with a high kurtosis, despite the fact that the volatility present in each of them is different.

Based on the results presented in table 2 we can conclude that the returns follow a non-Gaussian distribution, with a higher kurtosis. This, however, will not be contradictory to the Gaussian distribution assumptions made in Section 4. This is because we will assume a Gaussian distribution for the model residuals, and not for the CDS returns.

All the methodology we will present below is based on the use of the described logarithmic yields. Working with these new yields will allow us to propose models that capture the transmission of risk between the different CDS. Therefore, in summary, the data used throughout this paper consists of the time series of the spread yields of the ten sectoral CDS described above.

4 Methods

In this section we present the methods used to analyze the transmission of systemic risk between industrial sectors in Europe.

Network modeling is particularly interesting when considering each industrial sector as a unit. In this way, we will represent each of the industrial sectors as a node within the network composed of all the industrial sectors. Systemic risk spreads among the different nodes of the network. Therefore, our objective is to analyze the transmission of this systemic risk, i.e., how it spreads from one node to the rest.

We work in the framework of Bayesian statistics. Thus, we propose the modeling of the problem using Bayesian networks. Specifically, to take into account the temporal nature of the data, we will use dynamic Bayesian networks. Within these networks we will introduce relationships marked by autoregressive and moving average models. In order to understand how systemic risk propagates in Bayesian networks, we will have to analyze the structure of the network and understand the relationships between the nodes. We will perform the so-called network Structure Learning and Parameter Learning, by means of which we will come to understand risk transmission. With all this we will be able to address the objective of analyzing systemic risk transmission. In addition, we add the possibility of introducing expert information to improve the analysis.

This section begins by presenting the methodological bases of the Bayesian approach. Then, Bayesian networks, whose theoretical foundations are essential for the modeling of the problem, are introduced. Subsequently, the Bayesian approach of the Vector Autoregressive Moving Average Models is presented. Finally, the way of working with Dynamic Bayesian Networks is presented, both for model selection and model estimation.

4.1 Bayesian Statistics

Bayesian inference is a process of fitting a probabilistic model to a data set and summarizing the results by probability distributions over model parameters and unobserved data, such as predictions for new observations [Gelman et al., 2014].

The process of performing a Bayesian data analysis can be summarized in three steps. The first step is to formulate a complete probabilistic model, including probability distributions for all variables, both observed and unobserved. The model should be consistent, based on the information we have about the problem and the data collection process. The second step is to perform Bayesian inference on the model from the observed data. Calculate and interpret posterior distributions on the parameters. Lastly, the third step is to evaluate the fit of the model and the implications of the posterior distribution. See if the fit of the model to the data is good, if the conclusions are reasonable, and how sensitive the model results are to the assumptions of the first step.

The statistical conclusions of Bayesian inference about the parameters, θ, or about the unobserved data, $\tilde{y}$, are made in terms of probability. These probability statements are conditional on the observed data of the variables y; we can denote this as $p(\theta|y)$ or $p(\tilde{y}|y)$. These probabilities are also conditional on the observed values of the covariates x.

4.1.1 Bayes Rule and Bayesian Inference

To make probability statements about θ once given y data, we have to start from a model that proposes a joint probability distribution of θ and y. The joint density function can be written as a product of two densities, which we can call prior distribution $p(\theta)$ and data distribution $p(y|\theta)$:

$$p(\theta, y) = p(\theta)p(y|\theta).$$

Inference from the known values of the variable y is performed using a property on the conditional probability, known as Bayes Rule, which gives us the formula for the posterior distribution

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)}, \tag{1}$$

with $p(y) = \sum_\theta p(\theta)p(y|\theta)$, the summation is performed for all possible values of θ, if the set is discrete; in the case that θ is continuous we will have that $p(y) = \int p(\theta)p(y|\theta)d\theta$.

From this formula we can interpret $p(y|y)$ as the sampling information and $p(\theta)$ as the prior knowledge of the parameters and $p(\theta|y)$ as the knowledge about the parameters θ updated from the data. Therefore, we could summarize that the a posterior information is obtained with the a prior information and the sample information.

To perform inference on unobserved values, called predictive inference, we can follow a similar logic. Before considering the observed data of y, the distribution of this data y is

$$p(y) = \int p(y, \theta)d\theta = \int p(\theta)p(y|\theta)d\theta.$$

It is usually called marginal distribution of y or predictive prior distribution. It is called prior because it is not conditional on the observed values and predictive because with this distribution we can predict the observed values.

Once we have observed the observed data y we can predict the unknown observations $\tilde{y}$, if they follow the same model. The distribution $\tilde{y}$ is called the posterior predictive distribution. It is called posterior because it is conditional

$$\begin{aligned} p(\tilde{y}|y) &= \int p(\tilde{y}, \theta|y)d\theta = \\ &= \int p(\tilde{y}|\theta, y)p(\theta|y)d\theta = \\ &= \int p(\tilde{y}|\theta)p(\theta|y)d\theta. \end{aligned}$$

The second and third lines show the posterior predictive distribution as an average of the conditional predictions on the posterior distribution of θ. The last step follows from the assumption of independence of y and $\tilde{y}$ given θ.

4.1.2 Likelihood, Subjectivity and Priors

Using Bayes Rule with a selected probabilistic model tells us that the data y affect the posterior information only through $p(y|\theta)$, which is called the likelihood function when considered as a function of θ for the fixed y. In this sense the inference verifies what is called the likelihood principle, which says that for a sample of the data and two probabilistic models $p(y|\theta)$ that have the same likelihood function then they will yield the same inference for θ. The likelihood principle makes sense within the context of a model or family of models that we have taken for a particular analysis.

All methods using probability are subjective in the sense that they assume a mathematical idealization of the world. Bayesian methods, as probabilistic methods, are subjective in that they depend on the choice of a prior distribution. Bayesian statistics is often criticized for subjectivity, however, in most problems one needs to specify both the likelihood and the priors of the model. Therefore, subjectivity can occur to a lesser or greater extent within Bayesian methods, but it is not an intrinsic feature of Bayesian methods. We will see that the priors can be categorized according to how informative they are, so that the model will depend more or less on the selection and will be more or less subjective.

The prior distributions can be considered according to two interpretations. On the one hand, they can be interpreted as representing a population of possible values of the parameter. On the other hand, in the more subjective sense, we can consider them as the place where to express our knowledge and uncertainty about the parameter, considering it as a random variable that follows this prior distribution.

In many problems there is no perfectly relevant population of values containing the actual value of the parameter. In general, the prior distribution should include all plausible values of the parameter but the distribution need not necessarily be concentrated around the true value, because often the information contained in the data will outweigh any reasonable prior probability specification.

What all this tells us is that there are different types of previews depending on the influence they have. On the one hand we have informative priors, which are those priors that use the information on the parameter to make it very specific and narrow it down. On the opposite side are the non-informative priors, which are those priors that play a minimal role in the subsequent distribution.

4.1.3 Posterior and Conjugates

The process of Bayesian inference involves moving from a prior distribution, $p(\theta)$, to a posterior distribution, $p(\theta|y)$, and it is natural to think that some general relationships hold between these two distributions.

One of the relationships is between the means of the posterior and the prior. Specifically, the relationship is that the mean of the prior is the expectation of the mean of the posterior,

$$E(\theta) = E(E(\theta|y)).$$

Another relationship tells us about variance. The posterior distribution, which includes the data information, has less variance than the prior distribution. This statement is a result of the fact that

$$var(\theta) = E(var(var(\theta|y)) + var(E(\theta|y)).$$

This result is interesting because it tells us that the variance of the posterior is smaller in mean than the variance of the prior by an amount that depends on the variance in the mean of the posterior around the distribution of possible data.

The calculation of posterior distributions is performed as explained above. Specifically, if we work with continuous variables the calculation, following the result shown in (1), is

$$p(\theta|y) = \frac{p(\theta, y)}{p(y)} = \frac{p(\theta)p(y|\theta)}{p(y)} = \frac{p(\theta)p(y|\theta)}{\int p(\theta)p(y|\theta)d\theta}.$$

Performing the integral calculation can be a very complicated task. In order to obtain the result, computational techniques can be used. A case in which the computations are made much easier is if we use the so-called conjugate prior distributions, they can be formally defined as shown below. If $\mathcal{F}$ is the class of all density functions $p(y, \theta)$, we say that a class $\mathcal{P}$ of a prior distributions is said to be a conjugate family for $\mathcal{F}$ if

$$p(\theta|y) \in \mathcal{P},\ \forall p(\cdot|y) \in \mathcal{F}\ y\ \forall p(\cdot) \in \mathcal{P}.$$

In the same way we can say that Γ, a family of distributions for θ, $p(\theta)$, is conjugate with respect to the family $\mathcal{F} = \{p(y|\theta)\}$ if

$$p(\theta|y) \in \Gamma,\ \forall p(\cdot|y) \in \mathcal{F}\ y\ \forall p(\cdot) \in \Gamma.$$

Of particular interest are the families of natural conjugate priors, which appear when considering $\mathcal{P}$ as the family with the same functional form as the likelihood.

Some of the advantages of the conjugate prior distributions are that they simplify the calculations in obtaining the posterior, facilitate the description of the results, allow an interpretation of the prior as an equivalent experiment and are very useful in the construction of more complicated models. However, except in very simple cases, they are often too rigid to represent the information. These advantages and disadvantages should be taken into account when using conjugate priors.

4.1.4 Bayesian Hierarchical Models

Many statistical applications involve multiple parameters that are related or connected to each other in some way through the structure of the problem. Thus, the joint probability model must reflect their dependence.

As we have introduced so far, we view the parameters θ_j as a random sample with a given distribution. A key element is to see that the observed data can be used to estimate the distribution of the θ_j parameters, even though the parameter values are themselves unobserved. The idea of positing a hierarchical model arises naturally; with the observed data we will conditionally model some parameters, which in turn are given by a certain probability function in terms of other parameters, known as hyperparameters. This hierarchical way of thinking will help us to understand multi-parameter problems and will also mark the development of computational strategies.

In practice, non-hierarchical models are often inappropriate for hierarchical data; those with few parameters cannot fit large data sets accurately, while those with many parameters often tend to over-fit the data. In contrast, hierarchical models can have enough parameters to fit the data well, while using a family of distributions to structure the dependencies between parameters.

4.1.5 Simulation

Working with complex Bayesian hierarchical models, it can be complicated to obtain the posterior distribution analytically. To obtain the posterior distribution, one of the most commonly used possibilities is simulation. We will perform simulation using methods based on Markov Chain Monte Carlo (MCMC). Different simulation techniques have been developed to carry out this process, being Gibbs Sampling one of the most used. In this study we have used the implementation of this method through the program "WinBUGS" [Lunn et al., 2009], called directly from R [R Core Team, 2020].

4.2 Bayesian Networks

4.2.1 Bayesian Networks Fundamentals

Bayesian Network (BNs) are a class of probabilistic graphical models. Graphs can be used to easily represent the probabilistic structure of multivariate data using BNs. BNs are composed of a combination of random variables and the relationship between these variables given by a directed acyclic graph. The set of random variables is denoted by $\mathbf{X} = \{X_1, X_2, \ldots, X_k\}$, describing the quantities of the variables of interest. The distribution of $\mathbf{X}$ is called the global distribution of the data, that is a multivariate probability distribution. The distributions of each $X_i \in \mathbf{X}$ are called local distribution, that are the univariate probability distributions associated to the global distribution of the data.

The Directed Acyclic Graph (DAG) is denoted by G = ($\mathbf{V}$, A). $\mathbf{V}$ denotes the set of nodes, with each node $v \in \mathbf{V}$ associated to a variable $X_i \in \mathbf{X}$. A denotes the directed arcs that connect the nodes, each $a \in A$ is a directed arc. If no arc connects two nodes, the related variables are either independent or conditionally independent depending on the rest of the variables. There is an extensive theory on how it is possible to map the nodes of a BNs and their connections [Scutari and Denis, 2021]. Similarly, a theoretical mathematical basis underlies the conditional independence between nodes. We will not go into details on how to represent the existence of dependency or independence relationships between nodes.

Given a probability distribution P over a set of variables $\mathbf{X}$, a BN is a DAG $G = (\mathbf{X}, A)$ such that it is a minimal independency map (I-map) of P, such that none of the arcs can be eliminated without destroying the independence structure. The BN is denoted $\mathbf{B} = (G, \mathbf{X})$.

4.2.2 Bayesian Networks Properties

Assuming that the DAG is an I-map leads us to the general formulation of the joint probability decomposition, associated with the so-called global distribution,

$$P(\mathbf{X}) = \prod_{i=1}^{k} P(X_i|\Pi_{X_i}) \tag{2}$$

where Π_{X_i} is the set of parent nodes of X_i.

From the equation (2) we can deduce that the so-called local Markov property is verified. This property is equivalent to the result of the probability decomposition taking into account the chain rule. The local Markov property tells us that each node X_i is conditionally independent given its parents to every node that is not one of its descendants. In other words, each node X_i is conditionally independent of nodes X_j such that there is no path from X_i to X_j. The result of this property being verified is that several BNs with different sets of arcs may be encoding the same conditional independence relation and represent the same global distribution. As a consequence, the construction of equivalence classes between the different DAGs is immediate. A large number of graph theory results have been applied to BNs. We will not go into further details on the equivalence between DAGs nor on the definitions of the different types of structures, as this is of no interest in our study.

4.2.3 Dynamic Bayesian Networks and Assumptions

A Dynamic Bayesian Network is one BN in which the temporal nature of the variables studied is taken into account. Thus, a Dynamic BN is represented by different nodes over time and is obtained by expanding the interaction network defined in a BN over time. By determining the directions of the arcs over time we then ensure that the necessary conditions are verified in the BNs. In particular, we will not have problems with possible loops, ensuring the acyclicity of the graph.

In this case we will have that the nodes that can be parents of a node at time t are the nodes corresponding to the past, that is, the past values of the variables. The conditional independence between nodes will be given by the past values. In this way, the independence coincides with the very common way of working with multivariate stochastic processes.

A dynamic BN with a directed acyclic graph G describes a discrete stochastic process $\mathbf{X} = \{X_i(t) : i = 1, \ldots, k, \;\; t = 1, \ldots, n\}$ which takes values in R^k with k variables at n points in time.

Under the implementation of a dynamic BN there are certain underlying assumptions. The first assumption is that the stochastic process following $\mathbf{X}$ is m-order Markovian. This assumption is fundamental for it to make sense to work with a dynamic BN. The second assumption we make is that for all time points $t > 0$ the random variables $\mathbf{X}(t) = (X_1(t), X_2(t), \ldots, X_k(t))$ observed at time t are conditionally independent given the random variables at the previous time points $\mathbf{X}(t-1), \mathbf{X}(t-2), \ldots, \mathbf{X}(t-m)$. This assumption allows for simpler modeling of conditional data. It should be tested carefully as it has important implications. The third assumption made is that the variables are temporally independent of each other. Thus, in no case the time profile of a variable X_i, $(X_i(1), X_i(2), \ldots, X_i(n))$, cannot be written as a linear combination of the rest of the profiles

$(X_j(1), X_j(2), \dots, X_j(n))$, $j \neq i$. This assumption is made so that the proposed modeling does not have problems in inference by taking a higher dimensionality in the parameter set than the data allow. When the k variables are linearly independent, i.e., none of the profiles can be written as a linear combination of the others, the uniqueness of G is guaranteed. As a result, the first and second assumptions permit the existence of a dynamic Bayesian network with graph G that contains only arcs pointing out from a variable observed at previous times $\{t-1, t-2, \dots, t-m\}$ toward a variable observed at time t, with no arcs between concurrently observed variables.

We assume a constant time delay for all interactions, known as time point sampling, which is characterized by the interval between successive time points, in order to limit the number of parameters in the network. Allowing the presence of arcs between variables observed either at the same time or at different times can definitely be used to incorporate simultaneous interactions. However, we must be careful when adding new time moments as the number of model parameters increases exponentially with the number of time moments. Under the assumptions made the probability distribution of $\mathbf{X}$ can be represented as a dynamic BN with a DAG G whose arcs describe exactly the conditional dependence of the current values with respect to the variables on past values.

The Dynamic BN models depend on the number of time delays selected. Extending the number of time delays can lead to spurious conclusions about the network structure. To carry out the model estimation we will make one last assumption. The fourth assumption made is that the process is homogeneous over time, i.e. all network arcs and their directions are invariant over time. In other words, we assume that the phenomenon we are modeling is governed by the same rules throughout the experiment. The network considered is stationary, maintaining relationships over time.

The result of making this assumption is that we now have a total of $(n - m)$ repeated measures observed for each of the variables. Thus, it will be possible to make a representation of the model parameters. Each of the past time moments is represented by a matrix of coefficients of size $k \times k$.

The homogeneity assumption allows us to make the estimation without any problem. However, it is a strong assumption that must be verified. In case this property is not true, other strategies can be used.

4.2.4 Dynamic Bayesian Networks Models

We can work with dynamic BN models in different ways, depending on whether we consider more or fewer assumptions. From the construction of the basic dynamic BN model designed above, the introduction of autoregressive moving average models is intuitive. By introducing moving average autoregressive models within a dynamic BN we can have quite diverse models that can provide good results in the context of the problem. Under the assumptions proposed above, the compatible models are autoregressive models where, in addition, a diagonal variance-covariance matrix is considered.

4.3 Vector Autoregressive Moving Average Models

In this brief section we present a reminder of the Vector Autoregressive Moving Average Models together with the exposition of how these models are treated in the context of Bayesian statistics as a Bayesian hierarchical model.

4.3.1 ARMA Models and Gaussian Distribution

ARMA models consist of two parts, an autoregressive (AR) and a moving average (MA) part [Box et al., 2016]. The notation ARMA(p, q) refers to the model with p autoregressive terms and q moving average terms. We can observe that this model contains AR(p) and MA(q). The ARMA(p, q) model is written as

$$x_t = c + \sum_{i=1}^{p} \varphi_i x_{t-i} + \sum_{i=1}^{q} \theta_i \epsilon_{t-i} + \epsilon_t \tag{3}$$

with $\varphi_1, \ldots, \varphi_p$ and $\theta_1, \ldots, \theta_q$ the parameters, c a constant and $\epsilon_t, \epsilon_{t-1}, \ldots, \epsilon_{t-q}$ the error terms.

The generalization of ARMA models to the multivariate case is given by the Vector ARMA models, VARMA models.

Suppose $\mathbf{x}_t = (x_{1,t}, \ldots, x_{K,t})^T$ is a time series that is composed of a K-dimensional vector of components. Let $\mathcal{F}_t = \{x_t, x_{t-1}, \ldots\}$ be the set of all information up to time t. In this situation we will assume that the conditional distribution of x_t follows a Normal distribution

$$x_t \mid \mathcal{F}_{t-1} \sim N(\mu_t, \Sigma_t)$$

with $\mu_t = (\mu_{1,t}, \ldots, \mu_{K,t})^t$ the vector of means and Σ_t the variances and covariances matrix. The next step is to determine the relationship that the parameter vector μ_t has with the information passed from the $\mathcal{F}_{t-1}$ component.

4.3.2 VARMA Models

In this context the autoregressive and moving average structures are introduced by assuming that the vector of means, μ_t, follows an autoregressive and/or moving average multivariate process. In other words, we will consider that the data x_t follows an autoregressive and/or moving average multivariate process [Nicholls and Hall, 1979]. The result of assuming that the vector of means μ_t follows an ARMA(p, q) type structure will give us a model which we will call VARMA(p, q). Thus, we assume that the parameter vector μ_t follows a multivariate ARMA-type process of orders p and q

$$\mu_t = A_0 + \sum_{j=1}^{p} A_j x_{t-j} + \sum_{j=1}^{q} B_j \mu_{t-j}$$

with p and q positive integers, A_0 a constant vector, $A_1, \ldots, A_p$ and $B_1, \ldots, B_q$ coefficient matrices associated with the variables in previous times.

It might not look like a moving average autoregressive structure but if we consider the artificial random vector $\epsilon_t = x_t - \mu_t$ then we have that

$$x_t = A_0 + \sum_{j=1}^{p} A_j x_{t-j} + \sum_{j=1}^{q} B_j (x_{t-j} - \epsilon_{t-j}) + \epsilon_t =$$
$$= A_0 + \sum_{j=1}^{m} (A_j + B_j) x_{t-j} - \sum_{j=1}^{q} B_j \epsilon_{t-j} + \epsilon_t \tag{4}$$

with m = $\max(p, q)$.

Since $\mu_t = E[x_t \mid \mathcal{F}_{t-1}]$ then the expectation of this random vector ϵ_t is $E[\epsilon_t \mid \mathcal{F}_{t-1}] = 0$. Therefore, there is no doubt that the expression (4), similar to the definition of the equation (3), is a multivariate ARMA(p, q) type process.

4.3.3 Bayesian VARMA Models

VARMA models can be considered in a Bayesian context, the complete modeling of VARMA models is based on the treatment of VARMA models in the context of Bayesian statistics. The main difference in the modeling is that the various parameters present in the model are no longer considered as constants but as random variables. Therefore, Bayesian modeling of VARMA models must include a distribution for each and every parameter of the VARMA model in question. Specifically, a distribution must be proposed for each of the parameter matrices of the ARMA structure, as well as the other parameters of the proposed distribution for the data. Thus, VARMA models can be expressed as the following Bayesian hierarchical model:

$$
\boxed{
\begin{array}{ll}
x_t \mid \mathcal{F}_{t-1} \sim N(\mu_t, \Sigma) & \\
\quad \mu_t = A_0 + \sum_{j=1}^{p} A_j x_{t-j} + \sum_{j=1}^{q} B_j \mu_{t-j} & , \forall t > \max(p, q) \\
\quad A_0 \sim \mathcal{P}(A_0) & \\
\quad A_j \sim \mathcal{P}(A_j) & , \forall j \in \{1, \dots, p\} \\
\quad B_j \sim \mathcal{P}(B_j) & , \forall j \in \{1, \dots, q\} \\
\quad \Sigma \sim \mathcal{P}(\Sigma) & \\
\quad \mu_t \sim \mathcal{P}(\mu_t) & , \forall t \leq \max(p, q)
\end{array}
}
\tag{5}
$$

with x_t the value of the components of the series at time t, $\mathcal{F}_t = \{y_t, y_{t-1}, \dots\}$ the information we have on the series up to time t, that is, the set of values of the components of the series up to that time,

μ_t the vector of means of the Normal distribution associated with time point t, Σ the variances and covariances matrix of the Normal distributions, A_0, A_j and B_j the parameters of the ARMA structure, $\mathcal{P}(A_0)$, $\mathcal{P}(A_j)$ and $\mathcal{P}(B_j)$ the prior distributions of the parameters of the ARMA structure, $\mathcal{P}(\Sigma)$ the prior distributions of Σ and $\mathcal{P}(\mu_t)$ the distributions or the structures we give to the parameters μ_t for the first temporal moments, those for which we cannot give the ARMA structure due to lack of previous information.

4.3.4 Bayesian VARMA Models Implementation

The particular case we consider is that of introducing vague prior distributions. On the other hand, after performing a pretreatment to contrast the importance of the different elements of the modeling using VARMA models, we have simplified the model. Thus, the model finally implemented is an autoregressive model without moving averages, VAR, whose variances and covariances matrix is diagonal. The data used for the returns are stationary in mean and variance, so these VARMA models are reasonable.

4.4 Bayesian Network Learning

Model selection and estimation are known in the field of BNs as learning, as is done in the field of machine learning and artificial intelligence in general. In the case of a BN, these elements are consolidated in a two-step process. First, the model selection is done by Structure Learning. This first step is based on learning the structure of the DAG. In this way, the arcs and their directions must be contrasted. The second step, once the DAG structure has been defined, the model estimation consists of Parameter Learning. It is based on learning about the local distributions implied by the DAG structure. This estimation process can be performed either by unsupervised learning, using only the information provided by a data set, or by supervised learning, using expert information that informs us about the structure of the parameter values at a specific time.

The way of working is based on Bayesian principles. Consider a data set D and a BN $B = (G, \mathbf{X})$. We denote the parameters that determine the global distribution of $\mathbf{X}$ by Θ. We can assume that the parameter set Θ uniquely identifies $\mathbf{X}$. Thus, we can define the BN as $B = (G, \Theta)$.

Under this notation we can describe the BN learning process as

$$P(B|D) = P(G, \Theta|D) = P(G|D)P(\Theta|G, D).$$

The decomposition of $P(G, \Theta|D)$ reflects the two steps described above, where $P(G|D)$ is the probability associated with Structure Learning and $P(\Theta|G, D)$ is the probability associated with Parameter Learning. In this way, the underlying logic behind the learning process is explained.

Structure Learning can be realized in practice by finding the DAG G that maximizes

$$P(G|D) \propto P(G)P(D|G) = P(G) \int P(D|G, \Theta)P(\Theta|G)d\Theta.$$

Using Bayes theorem, the posterior distribution of the DAG, $P(G|D)$, is decomposed into the product of the prior distribution over the possible DAGs, $P(G)$, and the probability of the data given each DAG,

$P(D|G)$. It is obvious that for the calculation of the probability of the data, $P(D|G)$ the parameters Θ must also be estimated. The prior distribution of the DAG $P(G)$ allows to introduce the available information about the conditional independence relations between variables of $\mathbf{X}$. Similarly, once the DAG has been selected, a prior distribution can be entered for the parameters of the model Θ. In this case, the introduction of this previous information is made from $P(\Theta|G)$. In case you do not have any information about the DAG structure or parameters then it may be interesting to select a non-informative prior. In our case we will use vague prior distributions.

4.4.1 Network Structure Learning

The first step in modeling a problem using a BN is what we have called Structure Learning. This Structure Learning process is the equivalent of model selection.

Understanding the DAG of a BN is a very complex task. On the one hand, the space of possible DAGs is very large, the number of DAGs increases exponentially as the number of nodes grows. As a result, only a small fraction of the DAGs can be considered in a reasonable time. On the other hand, this space is complex due to the dichotomous nature of the connections.

The selection of the structure can be done following different algorithm based on statistical criteria. These algorithms are based on two alternative ideas: the use of hypothesis testing or the use of information criteria. The different algorithms for Structure Learning must operate under a number of common assumptions, underlying the definition of the BN itself.

The first of the relationships that must be fulfilled is that the nodes in the DAG and the random variables in X must have a one-to-one correspondence. This implies that multiple nodes that are deterministic functions of a single variable are not allowed. In addition, all dependence and independence relationships between variables in $\mathbf{X}$ must be given in conditional terms. This is a consequence of the nature of the modeling described, the rest of the relationships are not relevant. Moreover, any combination among the possible values of the variables in $\mathbf{X}$ must represent a valid observable event. This property must be verified so that it makes sense to consider the problem in probabilistic terms. Finally, observations are treated as independent realizations of the set of nodes. In this case, we will consider that a certain temporal structure is followed by means of VARMA models, thus forming a dynamic BN.

The algorithms used to perform Structure Learning are very diverse and, in many cases, specific to the problem being addressed. Thus, we will not go into the details of the application of any of these structure construction algorithms. However, it is worthwhile to address what are the fundamentals on which these algorithms are based on. In particular, we will see which are the contrasts performed to analyze the possible conditional independence between variables in X and which are the information criteria to calculate the performance of a network.

The two sets of criteria used to learn about the structure of the DAG are conditional independence tests and network scores. Both criteria are based on statistics in the context of the selected distribution. In our case the statistics on which these criteria are based correspond to the framework of multivariate normality that we are considering.

Conditional independence tests focus on the presence of individual arcs. Each arc indicates probabilistic dependence, conditional independence tests can be used to test whether such probabilistic dependence is compatible with the data.

As far as the tests are concerned, the most common is to work with the exact test for partial correlations. This test can only express the marginal linear dependencies between two variables. The correlation coefficient is

$$\rho_{i,j} = cor(X_i, X_j) = \frac{\frac{1}{n}\sum_{l=1}^{n}(X_i(l) - \overline{X}_i)(X_j(l) - \overline{X}_j)}{\sqrt{\frac{1}{n}\sum_{l=1}^{n}(X_i(l) - \overline{X}_i)^2}\sqrt{\frac{1}{n}\sum_{l=1}^{n}(X_j(l) - \overline{X}_j)^2}}$$

If the null hypothesis of conditional independence is rejected then the inclusion of the arch in the DAG can be considered. Thus, the null hypothesis of these contrasts is that two of the variables are probabilistically independent conditional on past values. On the other hand, the alternative hypothesis is that the two variables are not probabilistically independent conditional on past values.

$$\begin{cases} H_0: & X_i \perp\!\!\!\perp_P X_j | \mathbf{X}(t-1), \ldots, \mathbf{X}(t-m) \\ \\ H_1: & X_i \not\perp\!\!\!\perp_P X_j | \mathbf{X}(t-1), \ldots, \mathbf{X}(t-m) \end{cases}$$

Thus, the contrast in which we are interested is the one based on conditional partial correlation. Specifically, we will have $X_i \perp\!\!\!\perp_P X_j|\mathbf{X}(t-1), \ldots, \mathbf{X}(t-m)$ if and only if $\rho_{i,j|\mathbf{X}(t-1),\ldots,\mathbf{X}(t-m)}$ is not significantly different from zero. Unfortunately, there is no closed expression for conditional partial correlations; however, they can be estimated numerically. To perform this estimation we have to calculate the inverse matrix of the variances and covariances matrix between the variables we are analyzing if they are independent and the set of variables in the past. Working as described above, we obtain an unbiased and efficient numerical estimate of the partial correlation conditional on past moments of the two variables between which we are testing for the existence of an arc.

The transformation of the conditional partial correlation given by

$$t_{i,j|\mathbf{X}(t-1),\ldots,\mathbf{X}(t-m)} = \rho_{i,j|\mathbf{X}(t-1),\ldots,\mathbf{X}(t-m)} \sqrt{\frac{n - n_0}{1 - \rho_{i,j|\mathbf{X}(t-1),\ldots,\mathbf{X}(t-m)}}}$$

is distributed under the null hypothesis as a Student's t distribution with n minus the number of variable involved (number of rows or columns of the correlation matrix) degrees of freedom $(n - n_0)$. Thus, the number of degrees of freedom is calculated by subtracting the number of variables involved in the test from the sample size.

Using a simpler notation we have that if we want to perform a conditional independence test on the partial correlation coefficients $\rho_{XY|\mathbf{Z}}$ of X and Y given $\mathbf{Z}$ then we have that

$$t_{XY|\mathbf{Z}} = \rho_{XY|\mathbf{Z}} \sqrt{\frac{n - |\mathbf{Z}| - 2}{1 - \rho_{XY|\mathbf{Z}}}} \tag{6}$$

is distributed as a Student's t with $n - |\mathbf{Z}| - 2$ degrees of freedom.

If the contrasted variables are conditionally independent then t will be close to zero. Large values, both positive and negative, indicate the presence of conditional dependence.

Alternatively, a Fisher's Z test can be used. The transformation of $\rho_{XY|\mathbf{Z}}$ defined by

$$Z(X,Y|\mathbf{Z}) = log\left(\frac{1+\rho_{XY|\mathbf{Z}}}{1-\rho_{XY|\mathbf{Z}}}\right)\frac{\sqrt{n-|\mathbf{Z}|-3}}{2}$$

where n denotes the number of observations and $|\mathbf{Z}|$ the number of nodes in $\mathbf{Z}$ is asymptotically distributed as a Normal distribution.

In addition to these two tests, we can perform other tests, such as those based on mutual information or mutual information shrinkage, both based on a Chi-square distribution of one degree of freedom. By considering the different tests that can be performed and applying them on each of the possible arcs of interest following a certain algorithm, we can achieve a complete Structure Learning.

We can use **network scores** as an alternative to conditional independence tests. Network scores approach the construction of the DAG structure as a whole, trying to measure the performance of the DAG on the observed data. Thus, network scores are goodness-of-fit statistics that measure how well the DAG captures the dependence structure of the data.

There are many network scores that are commonly used to measure the performance of a DAG. One of them is the Bayesian Information Criterion (BIC). In general terms the BIC can be expressed as

$$BIC(G,\mathcal{D}) = \sum_{i=1}^{p}\left[logP(X_i|\Pi_{X_i}) - \frac{|\Theta_{X_i}|}{2}log(n)\right] \tag{7}$$

Each of the values depends on the conditional likelihood associated with each of the nodes. To calculate the BIC, the cumulative probability function of the assumed local distribution must be used. In our case, each of the local distributions is considered a Normal distribution, so the probability will be that associated with a Normal.

Another widely used network score is the Bayesian Dirichlet equivalent uniform (BDeu or BDe). This other network score calculates the posterior probability of each DAG considering a uniform prior distribution over the space of DAGs and parameters. Both BIC and BDe assign a higher score to those DAGs that best fit the data. Using this criterion, the network scores are useful in order to complete the Structure Learning.

Alternatively, network scores offer a comparative criterion to see in relative terms which DAG has the best performance among different DAGs. Network scores generally have a high computational cost. The calculation of a network score for each node involves the calculation of a likelihood function and its optimization, which is not immediate. As the number of possible DAGs increases exponentially with the number of nodes, it is impossible to obtain the network score for all of them.

In order to solve this drawback and use network scores, they are combined with heuristic methods. There are many heuristic algorithms used to perform Structure Learning with network scores. The greedy search proposes an initial network structure randomly and adds and removes arcs until it does not find a possible improvement. The genetic algorithms are based on exploring the space of DAGs by means of crossover (combination between networks) and mutation (random alterations). Finally, simulated annealing is an algorithm that allows changes both to improve the network score and to worsen it, but associating different probabilities to these changes depending on the resulting score. By using one of these optimization algorithms, network scores can be used to perform Structure Learning.

4.4.2 Parameter Learning

Once the structure of the BN has been learned, i.e., once Structure Learning has been performed, we can move on to learning about the parameters, to perform Parameter Learning. This new step is based on estimating and updating the parameters of the global distribution.

To perform this estimation task we can go from working with the global distribution to working with the different local distributions, in other words, we can go from working with the joint distribution to working with the individual distribution of each of the nodes. This decomposition can be performed due to the assumptions made about the conditional independence relationships.

The two most common approaches are the use of an estimate in the context of frequentist statistics and the use of an estimate in the context of Bayesian statistics, both an estimation based on maximizing the likelihood and a penalized estimation can be considered.

It should be clear that the strategy we have used to perform Structure Learning does not in any way determine the approach to be used in Parameter Learning. However, it is true that there are some methods to perform a joint analysis, in which Structure Learning is performed at the same time as Parameter Learning.

The simplest possibility is to perform a **maximum likelihood** estimation. This strategy is based on maximizing the global likelihood function or, equivalently, maximizing each and every one of the local likelihood functions. Under the assumptions made, each of the local distributions can be expressed as a classical linear regression model, in which a node is explained by the past nodes. The contribution of the past nodes is additive, and no interaction term is considered. Thus, the maximum likelihood maximization strategy is trivial.

Another possibility for parameter estimation is to employ strategies based on **penalized regression methods**. These methods focus on minimizing the quadratic errors by controlling the parameter values. In this way, large quadratic errors and high parameter values are penalized.

The first of these methods is called ridge regression. This method estimates the parameters as

$$\beta_i = \arg\min_{\beta_i} \sum_{l=1}^{n} (X_i(l) - \beta_i \mathbf{X})^2 + \lambda_2 ||\beta_i||_2^2.$$

Another very common method is called lasso regression. This method estimates the parameters as

$$\beta_i = \arg\min_{\beta_i} \sum_{l=1}^{n} (X_i(l) - \beta_i \mathbf{X})^2 + \lambda_1 ||\beta_i||_1.$$

Finally, the third method used is the so-called elastic net, a combination of the two previous methods. This method estimates the parameters as

$$\beta_i = \arg\min_{\beta_i} \sum_{l=1}^{n} (X_i(l) - \beta_i \mathbf{X})^2 + \lambda_1 ||\beta_i||_1 + \lambda_2 ||\beta_i||_2^2. \tag{8}$$

The use of these strategies to estimate model parameters can in some cases improve the resulting model. However, at the same time these methods require endowing an appropriate value for the model parameters, which can be a bit tricky.

Penalized Methods for Structure Learning

Penalized regression methods can also help to finalize the Structure Learning. Since these regression methods, in particular lasso regression and elastic net, can consider some parameters to be zero, then we can use these penalized methods to remove some of the arcs from the DAG. This way of working can be interesting in the case of not having performed Structure Learning and working with a previously indicated DAG or, also, in the case of considering that the DAG still has an excess of arcs that can lead to overfitting.

5 Results

In this section we present the results of performing a complete analysis of the transmission of systemic risk between industrial sectors in Europe are shown, obtained from using the methods described in Section 4 on the data presented in Section 3. In this work we will perform all the learning associated with the construction of a Bayesian Network model.

We will start by learning the structure of the Bayesian Network and then move on to work on learning the parameters of the resulting model 5.1. Learning the structure of a Bayesian Network consists of multiple statistical tests and the application of different Network scores. In our case, we will present in this order the results of applying the different criteria for the construction of the structure on each of the ten sectoral CDS studied. In this case, we will work on each of the sectoral CDS and, subsequently, we will combine the results obtained in order to form the complete Bayesian Network that has been constructed.

Once the complete structure of the Bayesian Network is known, it is necessary to estimate all the parameters that make up the model 5.2. This estimation is performed by Bayesian inference, that is, working in the context of Bayesian statistics. The estimation in this case will be performed by simulation working under the distributions described above. In addition, these results will be complemented by applying an estimation with penalized methods. This complementary step will allow us to analyze if the estimation of the parameters is correct and to study if the structure learned in the previous section is robust.

Finally, the economic interpretation of the results is included. The Bayesian Network learned is providing us with information on how risk is transmitted between the different industrial sectors in Europe. We will present the relationship between the different sectoral CDS and how systemic risk is transmitted between them.

5.1 Network Structure Learning

The first step to work with a Bayesian network is to learn the structure. The structure is determined by a Directed Acyclic Graph (DAG) between the different nodes considered. In the case of Dynamic Bayesian Networks these DAGs are drawn between the whole set of available nodes, i.e. the different time series and all those delays that we want to consider.

In the case studied we have a total of ten time series of CDS spreads and we have considered lags from the first to the fifth, considering the usual possible daily interactions between the series, which usually occur between the last five lags. Therefore, a DAG should be traced with a total of 60 nodes, 10 nodes from the original series and another 10 for each of the lags. These nodes, in turn, have the restriction that they must be directional towards the 10 nodes of the original series, since the rest of the relationships are marked by previous moments.

In this context, we have to learn the structure of the Dynamic Bayesian Network over a total of more than $10 \cdot 2^{50}$ possibilities for the DAG. This learning is composed of two complementary procedures, one based on the implementation of statistical tests to test each of the two-by-two relationships between nodes, and the other based on the use of network scores.

The Structure Learning process is really computationally expensive. This is due to the large number of DAGs that can be built over the set of nodes arranged. By simply considering the five delays and the relationships to the original series, a total of $10 \cdot 2^{50}$ different DAGs can be constructed. In addition, the possible instantaneous relationships between the series must also be considered. As a consequence, the Structure Learning process cannot be performed by an exhaustive search of all possible models. The two existing possibilities consist either in the use of heuristic model search algorithms or in the simplification of the possible models under consideration.

We have decided to follow a procedure based on performing different statistical tests, from simpler to more complex, in order to simplify the considered modeling and, finally, to perform a search among the possible models. In addition, we support the results obtained with this procedure by using heuristic model search algorithms.

We begin by performing unconditional independence tests for each of the original series with all the series and all the lags, i.e., for each series a total of 59 unconditional independence tests are performed, one for each series and lag, with a number of lags between zero and five. From this first test we try to rule out some of the relationships proposed in the original model. Once the unconditional independence tests have been carried out, conditional independence tests are performed. These tests are the ones that measure the true dependence relationships present in the models and will be the ones that will describe which is the indicated structure for the DAG. Subsequently, heuristic model search algorithms are used to check whether the resulting model matches the model obtained. All these steps are based on the use of the independence tests described in Section 4.4.1. Finally, we employ a Structure Learning based on the use of network scores complemented with a heuristic model selection algorithm. This second possibility provides us with more information about the model selected in the previous step. With all this we arrive to build the final structure of our Bayesian Network.

5.1.1 Unconditional Independence Tests

Given the necessity to build a Dynamic Bayesian Network among the ten sectoral CDS, we start by performing unconditional independence tests. We have a total of ten time series of the sectoral CDS and we will consider structures for our DAG composed of relationships between these ten series and with up to five time lags.

We performed unconditional independence tests on all possible relationships between the original time series and the lagged time series. Specifically, having ten time series and five lags for each of them, a total of 59 unconditional independence tests have been performed for each of the time series. In total, 590 unconditional independence tests, 545 different tests, were performed. These tests are based on the distribution of the statistic described in the equation (6). In particular, these tests are unconditional independence tests so we do not consider any set of variables. In short, in this first step we have performed all the unconditional independence tests described. We do not include in this document the results of performing all these tests explicitly because of their length; moreover, the results will serve as a simple guide for the following steps. In any case, we explain the key results obtained in detail below.

First, it is easy to observe that the different series have, in most cases, unconditional dependence relationships with the rest of the series without lag. Simultaneous correlations seem to be relevant in most cases, a result that agrees with what is observed in Figure 2. Moreover, almost all of these relationships are given by positive correlations. However, it should be noted that not all relationships are as strong as others, even though both may be relevant. In any case, according to the unconditional independence tests the instantaneous relationships, between non-delayed series, should be included in the model.

Secondly, by performing these tests we have been able to see that, likewise, the different series have unconditional dependence relationships with the rest of the series with one or two lags. Thus, the relationships between the series and the series delayed by one or two lags seem to be relevant in many cases. These relationships are more diverse, with some of the correlations being positive, showing direct risk transmission, and others negative, showing market corrections. Similarly as before, not all relationships are as strong as others, with quite a few of them being non-relevant. Therefore, according to the unconditional independence tests the relationships between the series and the series with one or two lags should be included in the model.

Finally, it is easy to observe how the different series have unconditional independence relationships with most series with three, four or five lags. Therefore, the unconditional relationships between the series and the series lagged by three, four or five lags seem not to be relevant in most cases. While it is true that some of the relationships appear to be relevant, it is possible that this is due to spurious relationships. Generally speaking the vast majority of these unconditional relationships are directly irrelevant. That the unconditional relationships between the original series and the series with three, four or five lags are irrelevant does not imply that the conditional relationships are irrelevant. However, by looking at these tests we can conclude that the lack of relationship between the original series and the lagged series seems so clear that it is not worth further investigation of possible conditional relationships. Therefore, following the results obtained in the unconditional independence tests we have decided to stop considering the nodes formed by the series with three, four or five lags.

Among the ten sectoral CDS series we find important differences in the results of the unconditional independence tests. In general terms, we see that for all series the different relationships highlighted above are verified, with relevant relationships with the original series and with the series with one and two lags, and irrelevant relationships with the series with more lags. However, it is observed how some series, such as the manufacturing and transport industrial sectors, have stronger unconditional relationships than the rest. On the other hand, it is observed that the sovereign bond series has less relevant unconditional relationships than the rest.

The unconditional independence tests used in this first step do not actually measure the independence relationship between each original series and the other series with or without delay. These unconditional independence tests measure, in reality, only the independence between each original series and the rest of the series with or without lag without taking into account any other type of information. Therefore, when considering models with more than one covariate, such as the Dynamic Bayesian Network we are constructing, the relationships we are interested in are not unconditional, but conditional on the information provided by the rest of the variables. As a consequence, the conclusions that we have obtained with these first unconditional independence tests should only be a guide as to how to continue the Structure Learning of our Dynamic Bayesian Network, but should not be binding.

Given the results of the implementation of the unconditional independence tests we can redefine the terms in which we are going to model the problem. In particular, from the results of the unconditional independence tests we have decided to consider the possible Dynamic Bayesian Network that is composed from the original series of the sectoral CDS and from these same series with one or two lags. Therefore, when performing the unconditional independence tests we have managed to simplify the modeling by eliminating the series with three, four or five lags. Although it is true that the results of the implementation of the unconditional independence tests are not binding, since they are so clear, we have decided to follow the conclusions reached in order to simplify the models.

5.1.2 Conditional Independence Initial Tests

Once we have performed the successive unconditional independence tests we can move on to the conditional independence tests, necessary to know the true relationships given in the context of the model between the studied time series. We must remember that to learn the structure we must determine the DAG among the whole set of available nodes. Moreover, since we are building a Dynamic Bayesian Network then the DAGs are plotted among the whole set of available nodes, both the original time series and the time series with all the lags we want to consider.

Under the guidelines given by the unconditional independence tests we have reduced the number of nodes on which the DAG is built, since we have eliminated the nodes corresponding to the time series with three, four or five lags. Therefore, in the case we are considering we have only the ten time series of the CDS spreads and we have considered one and two lags for each of these, in order to be able to consider the possible daily interactions between the series. Therefore, the DAG must be plotted with a total of 30 nodes, 10 nodes for the original series and 20 for the lagged series, 10 for each of the lags. These nodes maintain the restriction that they must have relationships directed towards the 10 nodes of the original series, since the rest of the relationships are marked by previous temporal moments.

In this new context, we have to perform the Structure Learning of the Dynamic Bayesian Network on a total of more than $10 \cdot 2^{20}$ possibilities for the DAG. It may already be realistic to perform an exhaustive search among all possible DAGs from the use of the conditional independence tests. However, since this search is based on performing successive conditional independence tests the computational cost remains high.

Before performing an exhaustive search of all the possible DAGs, we are going to start the Structure Learning by calculating the conditional independence tests on the total of the series. We are going to perform the conditional independence tests conditioning for the total set of all the series, the original series and the series with one and two delays, eliminating from this set only the two series between which the possible conditional independence relationship is analyzed. This calculation is really simple, since it only consists of the calculation of 29 conditional independence tests for each of the original series, that is, a total of 290 conditional independence tests, 245 different conditional independence tests.

The result of performing the conditional independence tests, conditioning on the total set of all the other series, provides us with valuable information to continue with the construction of the model. These tests are based on the distribution of the statistic described in the equation (6). In these tests the set $\mathbf{Z}$ on which we are conditioning is not null, unlike in the unconditional independence tests, but consists of all the series except the two series involved in the contrast. The results of these first tests will also not be binding for the final construction of the structure of our Bayesian Dynamics Network. As a consequence, we do not include in this paper the results of these conditional independence tests, due to the size of the number of statistics of interest. In any case, we explain below the most relevant results of these first conditional independence tests below.

First, by performing the successive conditional tests we can observe both results that point to some of the dependence relationships being relevant and others that point to some of the dependence relationships not being relevant. In this sense, it is easy to see how the results obtained when performing the independence tests have been much more restricted, many relationships between series that seemed not to be unconditionally independent now seem to be conditionally independent. By providing the information given by the rest of the variables, many of these relationships that were previously relevant disappear, becoming irrelevant. In the same way, the opposite could happen, that some unconditionally independent relationships become conditionally dependent, however, in our case this is not frequent.

Secondly, the conditional independence tests show very different relationships between the different original series studied, with all other series with or without lags. Some original series appear to have a clear conditional dependence with other series with which no other series does, while other original series have a clear conditional dependence with the same series. In these tests we begin to see the underlying network structure in the propagation of systemic risk that can be observed in the sectoral CDS.

Third, the conditional tests show very weak conditional dependence relationships between the original series. If we analyze the possible conditional dependence relationships between each of the original series with another series without lagging, it can be seen that in many cases conditional independence relationships seem to exist. On the other hand, when the few conditional dependence relationships that may exist between the original series seem to be very weak, there is no clear statistical evidence to think that the series cannot be conditionally independent.

From the results of the conditional independence tests we can better specify how we should proceed with the Structure Learning of our Dynamic Bayesian Network. From the first observation made we can conclude that we should prune our structure, not allowing all possible interconnections, as expected, but limiting them to the true relationships underlying the data. On the other hand, from the second of the observations we can see that we should not choose to mark the same structure for each of the nodes of our network, forcing only the most frequent relationships, but we should analyze for each node which is the related structure. Finally, from the last of the results we deduce that we must make a special treatment with the instantaneous relationships, since they do not have the same nature as the relationships with the delayed series.

The relationships that the original series have with each other, that is, the two-to-two relationships that the different series of sectoral CDS returns have, have a different nature than the relationships that can be traced with the lagged series. When determining a relationship between an original series and a lagged series then the direction of the relationship is marked by time, considering that the transmission relationships propagate over time. However, when tracing a relationship between two original series, the direction of this relationship must also be analyzed, i.e., in which direction the risk is being transmitted. This fact tells us that the strength of the relationship between two original series is reduced by considering only the directional relationship from one to the other. On the other hand, to work correctly with these possible immediate relationships it is more effective to take into account intraday data of the series, which have a different problem and require a different study, as already presented in Section 5. For all these reasons, and taking into account that when performing the conditional independence tests on all the series we have seen that few of the relationships between series were of conditional dependence, we have also decided to discard from our structure those relationships that can occur immediately between the original series.

By eliminating the dependence relationships between the original non-delayed series we are, in turn, simplifying the problem under study in a remarkable way. In the case of considering the possible relationships between the original series, we would have to work with the structure of the entire Dynamic Bayesian Network that we are building in order to carry out Structure Learning. However, by restricting that the relationships between the original series cannot be given then we are gaining the possibility of dividing the Structure Learning in chunks. Specifically, since we do not have to analyze the possible relationships between the original series, we can perform the Structure Learning for each of the ten original series separately, combining the conclusions obtained for each of the series. For each of the sectoral CDS we will analyze how the risk is transmitted between the different sectoral CDS with one or two lags and then define the structure of the DAG as the union of all these individual DAGs. This new approach can help to decrease the complexity of the problem, but also to speed up the computation using parallelization techniques.

Once the decision to eliminate the possible dependence relationships given between the original series has been taken, we are ready to perform an advanced Structure Learning. Definitely, we are going to consider plotting a Dynamic Bayesian Network in which the nodes are the ten series of the original CDS sectoral returns and these same series with one or two lags; furthermore, the relationships that go from the series with one or two lags to the original series are posed. In this final context, we have to perform the Bayesian Dynamics Network structurre learning on exactly $10 \cdot 2^{20}$ possibilities for the DAG. Under these conditions it is already realistic to perform an exhaustive search among all possible DAGs by using the conditional independence tests. However, the computational cost is still high, so we have decided to perform it but, at the same time, to consider possible alternatives that offer a good result for Structure Learning.

Finally we have completely determined the different options that we are going to consider as feasible for our Dynamic Bayesian Network. At this point we can try to perform the Structure Learning using the different techniques that have been previously discussed, thus performing the final Structure Learning. For this final Structure Learning process we will use three different strategies and then compare the results offered by each of them. Specifically, we will first perform an exhaustive search strategy based on conditional independence tests, then we will perform a heuristic search also based on conditional independence tests and, finally, we will perform a search based on network scores.

5.1.3 Conditional Independence Tests

The first of the strategies followed is based on the implementation of an exhaustive search among all possible DAGs by means of conditional independence tests. This strategy ensures that we find the structure that best matches the series of sectoral CDS returns among all the possible structures. On the other hand, this search method is the one that has the highest computational cost associated with it. We will start by presenting the process followed to use the conditional independence tests to find the structure that best fits the data. We will then explain the computational cost of this procedure. Subsequently, we will present a viable alternative with a lower computational cost. Finally, we will present the network structure we have arrived at from this procedure.

In order to perform an exhaustive search for the structure that best fits our data, what we must do is to perform successive conditional independence tests. For each of the original series there are a total of 2^{20} possible DAGs, associated to the combinations that can be made with 20 binary variables that denote the connection or lack of connection between the original series and the 20 series resulting from taking one or two lags. To perform the Structure Learning we must consider the 2^{20} possible structures and perform the 20 conditional independence tests associated with the lagged series. In other words, we must consider the 2^{20} possible sets of covariates and conditionally perform a conditional independence test for each of the lagged series. Once the tests have been performed, we only have to analyze the different results to arrive at the best structures for a Dynamic Bayesian Network for each of the nodes. Specifically, the identification of these structures consists of marking a significance level and looking for the structure for which all the relationships are significant under the set of associated variables and for which none of the other possible relationships is significant. Therefore, the maximum structure that has all the relevant relationships is sought. It should be noted that under this methodology it would be possible to arrive at several structures that verify the conditions, to which we should then apply some other criterion in order to choose one if the case arises. Finally, once we have the structures associated to each of the original series then we compose the complete structure of our Dynamic Bayesian Network as the union of all the structures, that is, tracing all the relationships included in each of the individual structures.

As we have seen, we are performing $20 \cdot 2^{20}$ conditional independence tests for each of the original series, i.e., a total of $200 \cdot 2^{20}$ conditional independence tests. The realization of all these tests involves a high computational cost, however, it is a feasible cost to carry out this exhaustive search. In order to face this problem we work by means of the parallelization of the calculations, possible thanks to the simplification of the problem by eliminating the relations between the series without delays. The procedure followed consists of dividing the original series into two groups of five and launching the calculations of the $20 \cdot 2^{20}$ conditional independence tests for five of the series at the same time. This treatment significantly reduced the computation time, which in total was more than 12 hours before performing the parallelization. In this way, the exhaustive structure search was accomplished in just under 3 hours.

The results of carrying out this procedure for each of the original series have determined the relevant relationships between the lagged series and each of the original series. We have obtained the complete structure of our Dynamic Bayesian Network by composing all the relationships. In our case, the process has led us to a single structure, so that we have not had to supplement the procedure with any other criteria. The only result that seems interesting from this whole procedure is the structure itself that we have arrived at. This structure is presented in Section 5.1.6. In this structure all conditional independence tests confirm that all relevant relationships have been plotted. We will omit the details about the $200 \cdot 2^{20}$ tests of conditional independence due to their length.

The process has a high computational cost that would make it unfeasible in case of considering a higher number of series or, alternatively, a higher number of delays. This problem means that this way of performing Structure Learning is not a good candidate in general. For this reason, it is usual to consider the possibility of working in a simplified way by performing the conditional independence tests on the total set of all series. Therefore, we have considered the possibility of performing for each of the original series a conditional independence test for each of the lagged series, conditioning on the set of all the lagged series except the series involved in this contrast. This way of working is equivalent to the one we have used in Section 5.1.2 with the difference that in this case we have eliminated the instantaneous relationships. This proposal represents a significant simplification with respect to the exhaustive search, since we go from performing $20 \cdot 2^{20}$ conditional independence tests to only 20 for each of the series. In this way, we will only have to perform a total of 200 conditional independence tests. Once these tests have been performed, the next step is to plot only the relationships that are determined to be relevant. Thus, working in this way we arrive in a simple way to a good proposal for the network structure.

Delayed — Original	DSEBK5E	DSECG5E	DSEEP5E	DSEEC5E	DSEMF5E
DSEBK5E	1st		2nd	2nd	
DSECG5E		1st		1st	1st
DSEEP5E	1st	1st	1st / 2nd	1st	
DSEEC5E	2nd		1st	2nd	1st / 2nd
DSEMF5E		1st / 2nd	1st / 2nd	1st	1st
DSEOF5E	1st		1st / 2nd		1st
DSESC5E				1st	1st
DSESV5E	1st	1st			
DSETL5E	1st	1st		1st	1st
DSETR5E	1st	2nd	2nd	2nd	
Total	7	7	9	8	7

Table 3: Delays that have an associated relevant relationship between the delayed series and the original series DSEBK5E, DSECG5E, DSEEP5E, DSEEC5E and DSEMF5E.

In short, we have alternatively performed the conditional independence tests by conditioning the set formed by all the uninvolved lagged series to perform a simplified Structure Learning. In this case we cannot guarantee that the network structure we arrive at is the one that best fits our data but, at least, we arrive at a structure that fits correctly. The difference is usually that the structure arrived at may be somewhat more simplified than the best of the structures, i.e., it may discard some of the relationships that would be seen as relevant when performing the exhaustive search. As already mentioned, in this case the structure is drawn by selecting those relationships that are relevant according to the conditional independence tests. The results of these tests are again cumbersome to present, so they are summarized in the Tables 3 and 4, where the relevant relationships according to the conditional independence tests are presented conditionally for the total set, thus presenting the network structure obtained.

Tables 3 and 4 present the relevant relationships that five of the original series, five in each of them, have with the different lagged series. In the rows the lagged series are presented, while in the columns

Delayed — Original	DSEOF5E	DSESC5E	DSESV5E	DSETL5E	DSETR5E
DSEBK5E					
DSECG5E		1st		1st	1st
DSEEP5E	2nd	1st / 2nd	1st / 2nd		
DSEEC5E	2nd				1st / 2nd
DSEMF5E	1st	2nd	1st	1st	1st
DSEOF5E	1st			1st	
DSESC5E		1st			1st / 2nd
DSESV5E			1st / 2nd		
DSETL5E		1st		1st	
DSETR5E		2nd	1st	1st	1st / 2nd
Total	4	7	6	5	8

Table 4: Delays that have an associated relevant relationship between the delayed series and the original series DSEOF5E, DSESC5E, DSESV5E, DSETL5E and DSETR5E.

the original series are presented. In each of the boxes of the table, the number of lags that have a relevant relationship between the lagged series, in the row, and the original series, in the column, is indicated. In this way, the individual structures of each of the original series can be observed by visualizing the relationships indicated by columns.

By analyzing the tables we can see the diversity of structures present for the different original series. In general, we can see that there are more relationships between the original series and the series with a single delay than with two delays. Moreover, it is usual to observe how the series have a first relationship with themselves, in particular, in all cases there is at least one relevant relationship between each original series and this same delayed original series. On the other hand, we note that some of the relationships between the series are particularly complex, requiring the relationship given by both one and two lags. Finally, we highlight the fact that the individual structure proposed for each of the original series is notably ruling out overfitting. As the relevant relationships between the original series and the lagged series are so few, we arrive at a model with a low dimensionality for the parameter space, so that we will not fall into problems associated with overfitting, as it is verified in Section 5.1.4.

It is relevant to note that the results obtained when performing these conditional independence tests with respect to the total set fully coincide with the structure found when performing the exhaustive search based on the successive conditional independence tests that we have presented previously in this Section 5.1.3. As we have already mentioned, it would have been possible for the structure obtained by the exhaustive search and the structure obtained by the conditional independence tests with respect to the total set to be different. We anticipated that it would have been possible for the structure presented by the exhaustive search to include more relationships that from the point of view of the other strategy were irrelevant. However, in this case the coincidence is absolute, which validates the idea of using the conditional independence tests with respect to the total set to perform Structure Learning.

The results presented in the Tables 3 and 4 univocally determine the structure of our Dynamic Bayesian Network. The underlying structure in these tables is the one we will definitely take to perform our study. In Section 5.1.6 we will present the structure underlying these results. Subsequently, we will perform Parameter Learning and risk transmission analysis on this structure, which is fundamental for the rest of the study.

5.1.4 Conditional Independence Heuristic Algorithm

The second of the strategies considered is based on the implementation of a heuristic search of the DAG structure by means of conditional independence tests. This strategy offers an alternative to the exhaustive search presented in the previous Section 5.1.3. The heuristic search has much lower computational costs associated with it than the exhaustive search, since there are areas of the DAG space that are not explored as they do not seem to fit the underlying structure in the data. Exploring only those areas of the DAG space that seem to offer a good structure leads to the presentation of candidate structures for the Dynamic Bayesian Network.

The heuristic structure search works in a similar way to the exhaustive structure search, with the difference that in this case not all possible structures are considered. The idea is to start from a possible structure, which is usually the one that includes all possible relations, and to perform conditional independence tests with respect to successive sets in order to add and remove relations, until a structure is reached in which all relations are relevant and in which no other relation can be added.

In this case the idea is to start from the structure with all the possible relationships and perform a conditional independence test on all of them in order to eliminate some of the non-relevant relationships. Once this first study has been carried out, one of the relationships is eliminated and the process is repeated, performing the conditional tests on the new set of covariates. In the following steps, the possibility of adding a relationship, and not just eliminating it, is considered, so that if a new relevant relationship is found then it is introduced into the structure. The algorithm has small modifications in order to eliminate the possible entry in loops and, in addition, it presents some randomization in the order of elimination and introduction of relationships in order to explore a larger area of the DAGs space that may be feasible. The strategy we have presented works by performing the successive conditional independence tests individually for each of the series, like the exhaustive search. The procedure is also separable for the reasons explained above, so that it is possible to parallelize the work in order to reduce the processing time.

Heuristic search entails a much lower computational cost than exhaustive exploration of all possible structures. Many of the structures are easily discarded, so they are not explored. In this case we have not performed any parallelization and the processing time required was less than half an hour in total. In case of considering a more complex problem, in which we perform Structure Learning with a larger number of nodes, then a parallelization of the computations could be advisable. The number of possible structures that have been analyzed for each of the original series is less than 5%, considering that the rest were of no interest because they did not fit the data. Although it is true that the proportion of the space studied is relatively low, in our case it is not a problem, due to the non-existence of alternative structures with a better behavior, that is, to the very regular shape of the space of the possible DAGs in terms of the conditional independence tests.

The use of this alternative approach has as its main advantage, over the exhaustive structure search, the reduction of the computational cost. However, this reduction in the number of operations entails an important disadvantage, in this case it is possible that the resulting structure is not the best of all possible ones, i.e., we may not take into account all the structures that verify the results sought in the conditional independence tests. Therefore, we can consider the heuristic search as an intermediate proposal between the use of the exhaustive search and the use of the conditional independence tests with respect to the set of all the lagged series, i.e., an intermediate proposal to those we have made in 5.1.3.

We have implemented this new strategy as described above up to the candidate structures. In this case we have arrived at a single structure that verifies the conditions described for the conditional independence tests, just as has happened in the exhaustive search. Evidently, as the heuristic search involves the exploration of only a part of the space of possible DAGs then all the candidate structures offered by the heuristic search will be among the candidate structures of the exhaustive search. As a consequence, the structure obtained in this case coincides completely with the structure presented in Section 5.1.3, identified both by the exhaustive search and by the use of the conditional independence tests with respect to the set of all the lagged series.

In short, the heuristic search has proposed in this case the same structure that we had already identified in Section 5.1.3. This fact tells us, on the one hand, about the underlying evidence in the data of the presented structure and, on the other hand, about the goodness of this heuristic search strategy in order to identify a good structure for the Dynamic Bayesian Network. The identified structure is presented in Section 5.1.6.

5.1.5 Network Scores

Finally, the third strategy we have considered offers a different view of Structure Learning since it consists of identifying the best structure from the network scores. In this case, instead of working with conditional independence tests, we work with performance metrics. This strategy offers a completely different alternative to those presented in Sections 5.1.3 and 5.1.4, which were also based on conditional independence tests.

Performing Structure Learning from network scores consists of evaluating the different possible structures, analyzing how they fit with the available data. These network scores act as a performance metric for the structures and allow us to see which one makes the most sense in the context of the observed data. In our case we are going to use the BIC, presented in (7), in Section 4.4.1. Just as we have used the BIC as a criterion to evaluate the performance of the different structures, other alternative or complementary information criteria could have been used.

The use of network scores has advantages and disadvantages with respect to the use of conditional independence tests. On the one hand, network scores allow the model to be analyzed on the basis of the selected structure once it has been fitted, which facilitates comparison in terms of both estimation and prediction; however, conditional independence tests do not evaluate model fit. In addition, network scores penalize model overfitting by decreasing the score as the number of parameters increases;

conditional independence tests do not penalize possible model overfitting. On the other hand, network scores provide a subjective score of the different structures, as there are different network scores, it is possible that each of them offers slightly different structures, so that they cannot guarantee that the structure is the best fit to the data with respect to other criteria.

When working with network scores, the steps to follow consist, in a very summarized way, of selecting a candidate structure, fitting the model and evaluating it. In this case, we are spared having to perform the 20 conditional independence tests for each of the structures but, in exchange, we have to adjust the model, so the computational cost is similar. Depending on the complexity of the model and the number of nodes, these methods can be more or less expensive.

In this context we have to decide, as with the conditional independence tests, in which way we want to perform the search, exhaustive or heuristic. In our case we have initially performed the exhaustive search, with a very high computational cost, requiring more than 10 hours of work. On the other hand, we have tried to perform the process by means of a heuristic search. This search was performed using the conditional independence tests, in the same way as in Section 5.1.4, significantly reducing the computational cost to one tenth. In the same way as in the previous sections, we could have proposed a parallelization of the process.

Instead of saving a single structure, as was the case with the unconditional independence tests, this strategy allows us to perform a ranking of structures, saving several good candidates. In our case, we have saved the five structures that had the highest network score associated with them. It is interesting to note that the five structures saved when performing the heuristic search coincide completely with the five saved when performing the exhaustive search. In both cases, as is evident, the network scores of the structures are the same, since they are evaluated after the adjustment, which is deterministic. Thus, we have one more proof that the heuristic search we have performed in Section 5.1.4 has been very good, since it has gone through all the best structures according to the network scores.

The results obtained in this search are very interesting. In this case, it could happen that the structures proposed by the network scores are notably different from those obtained by the conditional independence tests, due to the different criteria used. However, the proposed structures are indeed similar. For six of the ten original series the structure with the best network score coincides with the structure proposed by the conditional independence tests. For the other four series the structure proposed by the conditional independence tests obtained the second or third best network score in their respective ranking. Moreover, the five structures proposed by the network scores for each of the original series were in all cases really similar to the selected structures, including only one more or one less relationship with respect to the structure obtained. Once again we will not present the results obtained by these strategies due to their length.

For all these reasons we consider once again that the structure selected in Section 5.1.3 and shown in Section 5.1.6 fits really well with the data. Moreover, the results of this new strategy allow us to verify that the structure of our Dynamic Bayesian Network verifies more desired properties. In particular, we can deduce that the selected structure has an associated model with a good fit and, moreover, that it does not have a significant overfit associated with it.

In short, we have seen how network scores offer a different approach to Structure Learning, but which can be equally interesting for the use of conditional independence tests. In this particular case it can be seen that the structure obtained coincides remarkably with the structure obtained by the conditional independence tests; however, in other cases it may be possible to obtain greater differences. For all these reasons, we consider network scores as alternative candidates to conditional independence tests for Structure Learning.

5.1.6 Network Structure

In the previous sections of this Section 5.1 we have performed the Structure Learning of our Dynamic Bayesian Network. In this last section we are going to make a brief reflection on the whole process of the Structure Learning performed, comparing the results presented by each of the methods, and we will present the DAG obtained in Section 5.1.3 and seconded in Sections 5.1.4 and 5.1.5.

As we have seen throughout this section, performing the Structure Learning of a Bayesian network can be a complex procedure, not trivial at all. Initially, the context in which we are going to consider the possible DAGs for our Bayesian network must be considered, in our case it has been the original series and these same series with between one and five delays. Subsequently, we have carried out a candidate debugging process, by using the unconditional independence tests and by using the conditional independence tests with respect to the total set of all the series not involved; this step, without being necessary, has allowed us to simplify the models, significantly reducing the computational cost and eliminating possible overfitting problems. Once the context was fully established, Structure Learning was carried out.

The Structure Learning has been based on three alternative methods: an exhaustive search using conditional independence tests, a heuristic search also using conditional independence tests, and a search using network scores. Alternatively, we have found that the use of conditional independence tests with respect to the total set of all uninvolved series can provide a good approximation to the underlying structure in the data.

The results offered by the different methods coincide almost completely, presenting as the selected structure the one corresponding to the DAG of Figure 4. As we have previously analyzed, it seems that this structure has a very good behavior, being a magnificent base to implement our models. We can see that this structure is quite complex in appearance, since all the nodes present many unidirectional relationships towards the rest or, even, towards themselves. However, this appearance is remarkably far from reality, since in this graph a total of 30 nodes are implicitly presented, and not 10, since we also have the nodes corresponding to the delayed series, so that the final models have few parameters in relation to the complexity that other possible models associated to other DAGs have.

In the Figure 4 we can see the complete structure of our Dynamic Bayesian Network, underlying the data and presented in the Tables 3 and 4. In this plot the whole set of relevant conditional dependency relationships are represented. On the one hand, yellow represents the relationships between the original series and the series with a single lag. On the other hand, blue represents the relationships between the original series and the series with two lags. Finally, the green color represents the relationships between the original series and the series with both one and two delays. The different edges of the graph represent all the relevant relationships between what would be the 20 nodes corresponding to the delayed series and the 10 nodes of the original series. Therefore, the Figure 4 actually represents the DAG that links the 30 nodes described and composes the structure of our Dynamic Bayesian

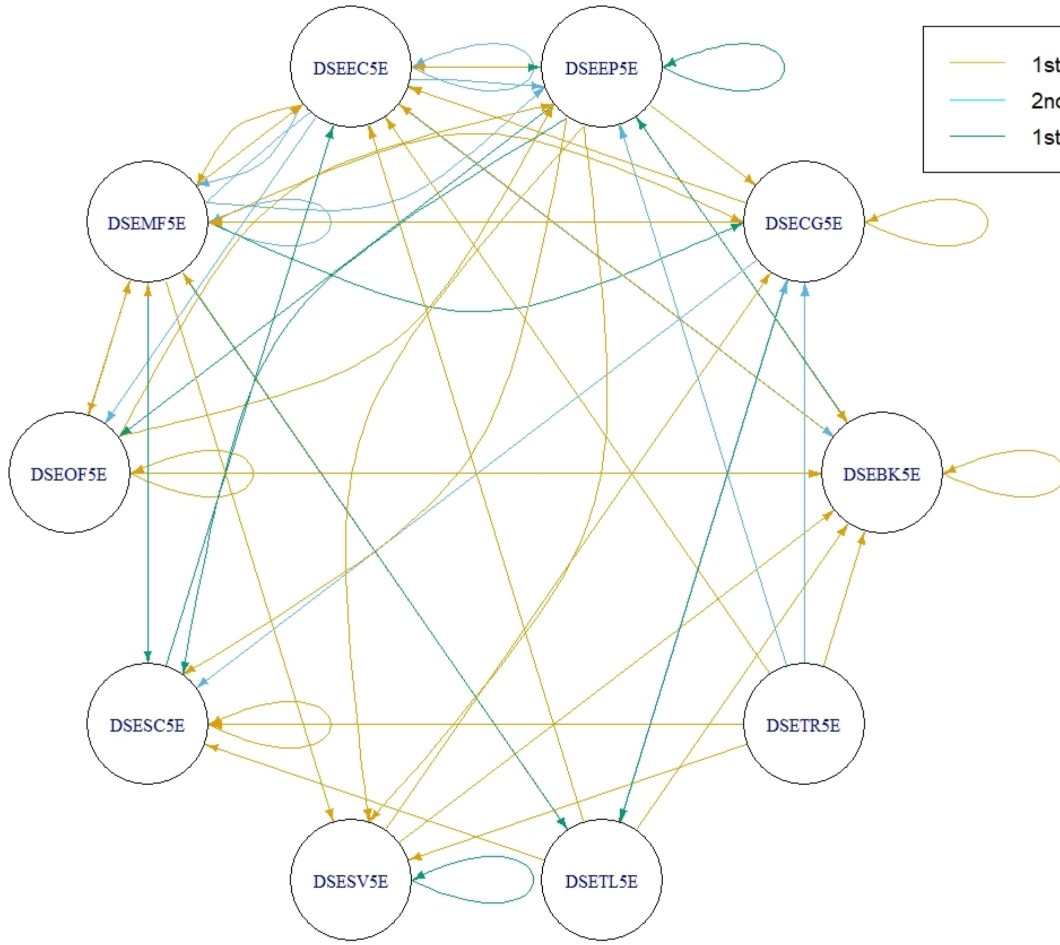


Figure 4: DAG selected by the Structure Learning.

Network. Finally, we must emphasize that the relationships marked in the DAG are directed from the delayed series to the original series. The structure represented in the Figure 4 is composed from the union of the 10 individual structures corresponding to each of the original series.

To finish the Structure Learning we have performed a last validation of the structure. Specifically, we have repeated the Structure Learning process but now with the complete data series, not only the series corresponding to the COVID-19 crisis. In this case we have first used part of the procedure explained in Section 5.1.3, performing a first study using the conditional independence tests with respect to the complete set of time series. We then performed a heuristic search identical to the one presented in Section 5.1.4. The result of these two procedures again coincided with each other and, for the most part, with the 4 structure. Specifically, only for two of the original series has the structure changed, in one of them by adding a new relation and in another by eliminating a relation already present. In this way, we have once again validated the DAG obtained by Structure Learning throughout this Section 5.1.

The Structure Learning of our Dynamic Bayesian Network is now complete. The DAG represented in Figure 4 forms the model with which we work in Section 5.2, from which we will deduce the risk transmission relationships between the different CDS series.

5.2 Parameter Learning

As presented in Section 4.4, the estimation of a Bayesian network consists of two distinct steps, model selection, by means of Structure Learning, and model estimation, by means of Parameter Learning. Once the Structure Learning presented in Section 5.1, by which we have defined the model we are going to work with, has been completed, we are ready to move on to the second step that makes up the implementation of a Bayesian network, the Parameter Learning.

In this new section we are going to perform the estimation of the previously selected model. As described in the methods Section 4, the model implemented throughout this work consists of a Dynamic Bayesian Network, introduced in Section 4.2, on which an autoregressive model is introduced, presented in Section 4.3. The model as a whole is the one presented in equation (5), with the particular characteristic that the work with the model is performed under the terms expressed above, Structure Learning and Parameter Learning, presented in Section 4.4. As a consequence, so far Structure Learning has been performed, which has provided us with information about the autoregressive structure of the model, i.e. it has finally defined the implemented model of the equation (5).

Introducing the DAG presented in Figure 4 as the autoregressive structure of the model presented in equation (5), that is, as the structure presented in the first line of the model, we are ready to work with the distributions. In other words, the model is fully defined and we can proceed to the estimation of the model.

Throughout the Section 5.1 we have been working for simplicity in the context of frequentist statistics. We have taken the liberty because of the clear differentiation of the two components that form the implementation of a Bayesian network. However, the Parameter Learning process presented in this Section 5.2 is performed entirely in the context of Bayesian statistics. For this purpose, we rely on the methodology presented in Section 4.1.

To perform Parameter Learning we work in the context of Bayesian statistics. The treatment consists of estimation by simulation, as described in 4.1.5, whose concrete implementation will be presented in more detail later. Alternatively, the possibility of working with penalized methods is considered. These penalized methods are a modification of the model described above but can provide valuable information on both the Structure Learning and the Parameter Learning performed. The implementation of these penalized methods is also carried out in a Bayesian context. In a complementary way, part of this procedure for implementing the penalized methods has also been carried out in a frequentist context, with the results always coinciding with the difference in approach.

5.2.1 Bayesian Inference

Given the model presented in equation (5) and introduced the autoregressive structure captured by the DAG presented in Figure 4, its estimation is pending. The estimation process is based on Bayesian inference and will be presented below. This estimation process is based on the theory presented in Section 4.1.

As we have explained, we are going to use Bayesian inference to perform the estimation of the model, that is, to perform the so-called Parameter Learning. The first step to perform the Bayesian estimation of the parameters is to write the specific hierarchical model we are going to be working with. In our case the Bayesian hierarchical model in general terms has been presented explicitly in the equation (5). Starting from this model, the first of the particularities considered is the introduction of the autoregressive structure associated with the DAG presented in Figure (4). Once the DAG resulting from Structure Learning has been introduced, what remains to be done to fully determine the Bayesian hierarchical model are the prior distributions.

To perform the Bayesian estimation of the parameters we must specify the prior distributions presented corresponding to the different parameters of the model. In the particular case we are studying, we must introduce the priors corresponding to the parameters belonging to the autoregressive structure described by the DAG and, in addition, the priors corresponding to the distribution of the variances and covariances matrix associated with the distribution assumed for the set of original series, i.e. the multivariate Normal.

Assuming that the structure imposed by the DAG is followed, we will have that the matrices associated to the autoregressive structure are formed by some non-zero parameters, one for each of the relationships presented in the Figure 4, and the rest are parameters equal to zero. Thus, to introduce the prior distributions of the parameters associated with the autoregressive structure, what is done is to present a univariate prior distribution for each of the parameters of the matrices, the non-zero parameters presented in Figure 4, one for each of the relationships. On the other hand, by assuming independence between the original series, i.e., the conditional independence relation of the non-delayed series, then the variances and covariances matrix is a diagonal matrix. As a consequence, we can define the priors for the variances and covariances matrix by working individually with each of the variances of the series, i.e. by defining a prior distribution for each of the variances of the original series.

For the selection of prior distributions we can consider different reasoning. Firstly, if we had expert information on the relationship between the different original series and the lagged series, then we could try to draw prior distributions according to this expert information. However, in the study carried out there is no expert information available from which to extrapolate any conclusions. Having ruled out the possibility of working with expert information, we then have two other reasonable options, namely, these possibilities consist of a more general approach, either based on other data or based on not providing information in these priors. The first of these two options is to fit the model to the available data prior to the COVID-19 financial crisis and use the posterior distribution of the parameters as a prior. The second of the two options is to take uninformative prior distributions by introducing vague prior distributions or uninformative priors distributions.

In this study we have decided to select uninformative prior distributions. The decision to take uninformative prior distributions offers certain advantages and disadvantages. On the one hand, taking non-informative priors offers the opportunity to learn only from what the data provide. In addition, non-informative priors are really simple to implement, since they do not require any prior information. In addition, non-informative priors cannot be criticized for being subjective, since they do not modify the behavior of the posterior distributions as they do not provide information. On the other hand, the fact of taking non-informative priors has the clear and evident disadvantage that they do not provide information, it could be said that their full potential is not exploited. Therefore, the fact of not introducing new information becomes both an advantage and a disadvantage.

We are going to present the specific prior distributions that we have introduced in our model for each of the parameters. Thus, we will present the non-informative prior distributions for both the parameters associated with the autoregressive structure and the prior distributions for the parameters of the variances and covariances matrix of the assumed Normal distribution. Actually we will see that the prior distributions are not completely uninformative, since they are constructed according to the little that is known about the problem, but for simplicity we will call them so.

The parameters of the autoregressive structure measure the propagation of risk between different CDS series. These parameters indicate how each movement in one CDS series, in the form of a yield, affects another CDS series one or two periods later. The effect that the yields of one CDS series usually have on another CDS series several time points later is usually very small, close to zero. As a consequence, the uninformative distribution we are going to construct will be centered on zero. On the other hand, the effect that a shock in one CDS series has on another series at some later point in time can be either positive or negative, as it can be due to a direct transmission of risk or a correction of risk transmitted by other series. Thus, the prior distribution given will be symmetric. Finally, the effect that the returns of one CDS series usually have on another several time points later tends to be damped. As a consequence, the parameter associated with this risk transfer is usually less than one in absolute value. This fact is not always true and depends on the nature of the series, however, in the case analyzed this statement is a reality. Therefore, the prior distribution assumed should accumulate most of the probability in the set formed by the values with modulus less than one, that is, the set $[-1, 1]$. Taking all these elements into account the prior distribution that we are going to introduce for each of the parameters of the autoregressive structure associated to the DAG presented in the Figure 4 is a standard Normal. For the different reasons implemented the standard Normal distribution seems to be a good candidate for prior distribution. As we will see later this prior distribution will act entirely as a non-informative prior, being driven by the data to construct the posterior.

The parameters of the variances and covariances matrix different from zero correspond to the variances of the residuals made when adjusting the original series from the lagged series by means of the multivariate autoregressive structure considered. As the original series are the series of CDS yields then the values taken by these series are considerably less than one in absolute value, being daily yields. As a consequence, the prior distribution that we have to consider for each of the variances must present, at most, probability for values equal to one, since the rest of values are impossible under normal conditions, incompatible with the analyzed data. On the other hand, the variance always takes positive values, so this prior has to accumulate all the probability in values equal to or greater than zero. The prior distribution selected for this purpose is the uniform distribution between zero and one. Without being totally uninformative, this distribution will be flexible enough for the posterior to be a faithful reflection of what the data report.

Once the prior distributions have been determined for all the parameters, both those resulting from Structure Learning and those underlying the initial definition of the model itself, the Bayesian hierarchical model that we are going to use throughout this work is completely defined. The defined model is the one we are going to estimate and the posterior distributions of the parameters are the ones that will inform us about how the risk is transmitted between the different CDS series.

The estimation of the parameters of the Bayesian hierarchical model is performed using simulation methods based on Markov Chain Monte Carlo (MCMC). Specifically, we work using Gibbs Sampling simulation techniques. To perform this simulation we are going to use the program *WinBUGS*. Once the models have been implemented in the *WinBUGS* language, we have carried out the estimation of all the model parameters. We have worked with five MCMC chains of 50100 simulations each, eliminating the first 100 as burn-in period. In addition, we made a thining of the simulations, keeping only one out of five values in order to avoid possible autocorrelation of the simulations.

Once the estimation has been carried out, we have verified that the behavior of the model has been correct. For this purpose, we have analyzed the convergence of the MCMC chains. This verification has been composed of several aspects: to observe if the simulation has completely explored the whole set of values that seem plausible according to the posterior distribution, to study the influence of the initial values on the chain, to analyze if the sample size has been correct and to study if the sample values are independent.

We have started by exploring whether all possible values of the posterior distribution are considered. The objective with this first test is to make sure that the sample drawn is representative of the posterior distribution we wish to sample. In order to make sure that the sample drawn is good, the five MCMC chains discussed above have been included. In order to check that no pathology has arisen, we have graphically represented the five chains considered. In this way, we will be able to see if all the chains have followed a sufficiently random distribution and that, in addition, they have been intertwined, showing convergence. Therefore, we have plotted the MCMC chains for each and every one of the parameters, the 68 associated with the autoregressive structure and the 10 associated with the variances. Specifically, we have made 78 plots, each with the five MCMC chains corresponding to each of the parameters. Once again we will not introduce the plots made due to their length. Working in this way we have been able to observe that the chains cross correctly. Moreover, in all cases the different values of the posterior distributions are considered. Thus, we can state that no pathologies are detected in the simulation and that the different values of the posterior distribution are considered, so we have gone on to analyze the influence of the initial values on the chain.

The values at the beginning of the chain may depend very much on the initial values, so we have performed a burn-in of 100 iterations. We must now consider whether this burn-in has been sufficient and whether the considered part of the chains has not been affected by the initial values. We think that by launching several chains and eliminating the first 100 iterations it will have been enough to not have problems of dependence on the initial values; to assess whether this is true, and the burn-in period has been sufficient, the Gelman-Rubin test has been used. The Gelman-Rubin test analyzes the amplitude close to the first considered values of the chain compared to the amplitude of the chain as a whole, speaking of amplitude as the amplitude of the credibility interval of the parameters. It is considered correct if all parameters have a value for the R-statistic less than 1.1. In this case, the values obtained are practically 1, approximately 1.001 for each and every one of the parameters, which indicates that the burn-in performed is correct. Therefore, we can accept that the chains do not depend excessively on the initial values considered.

Finally, we have now analyzed whether the generated sample size of the inference has been correct.

We must make sure that the simulation includes a sufficiently high effective number of iterations for the results to be reliable. To do this we can look at the effective number of simulations, denoted by "n.eff". This is the most common tool for assessing the goodness of the simulated sample size. This statistic takes into account the length of the chains and their autocorrelation, discarding those simulations that depend too much on the previous ones. Thus, this indicator is telling us about the true size of the simulated sample. In the results obtained we can see that all the different effective numbers of simulations are high, taking values in all cases of more than 30,000, being exactly equal to 50,000 for most parameters. Thus, the different effective numbers of simulations are always between 30000 and 50000, which is the maximum value that can be reached once we have selected the iterations of interest with the thining, subtracted the burn-in, and done all the chains. Therefore, the number of effective iterations observed indicates that there is no significant autocorrelation in the MCMC chains obtained. Since the effective number of simulations is sufficiently high, we can consider the simulation as good.

In conclusion, by means of the described analysis we have ensured that the results obtained from the MCMC inference are adequate. Knowing that the simulation has worked correctly and that the model designed, introducing the structure obtained in the Structure Learning and the described non-informative priors, is compatible with the data. Therefore, from the analysis of the simulated MCMC chains for each of the parameters, both graphically and numerically, we have ensured that the results obtained from this estimation are correct.

Having verified that the results are correct, we can proceed to a case-by-case analysis of the estimation for each of the CDS studied. This analysis will provide us with information on the different parameters of the model, which in turn will explain the risk transmission relationships between the different CDS series. The results of the estimation allow us to observe the complete implemented model, taking into account both the information on the structure and prior distributions and the information on the posterior distributions of all the model parameters.

On the one hand, the posterior distribution of the parameter corresponding to volatility provides information on the residuals made by the model considered, specifically, on the size of the residuals made when estimating CDS returns from this model. These residuals refer to the part of the sectoral CDS returns that is not explained by the returns of the different CDS series in previous times. Thus, the standard deviation of the residuals made by the model provide us with information on the amount of CDS returns that is not explained by the lagged CDS series. Given that CDS price returns refer to increases or decreases in risk, then the estimated standard deviation for each CDS tells us the size of the unexplained risk. By comparing the estimated standard deviations with the standard deviations of the original series we can get to know what is the proportion of the risk of each of the sectoral CDS that has been transmitted from other CDS series at previous time points and what is the proportion of the unexplained risk. In short, the different estimated standard deviations provide us with information on the transmission of systemic risk in terms of the weight of transmitted systemic risk with respect to total systemic risk.

On the other hand, the posterior distribution of the parameters corresponding to the autoregressive structure provide information on the specific transmission of systemic risk between each of the lagged series and the original series. These parameters measure the amount of systemic risk transmitted by a given lagged series over one of the original CDS series. The posterior distribution tells us the full range of possible effects that the lagged series have on the original series. Specifically, these parameters measure how shocks in the lagged series affect the sectoral CDS series.

We will analyze the posterior distributions of all the parameters for each of the sectoral CDS series. Performing this analysis for each of the original series allows us to know how risk propagates from one or two points in time prior to the series. We focus the analysis for each of the CDS as it will allow us to better understand the results, since we will be able to analyze from where and how much systemic risk reaches each CDS series. However, it is true that the analysis as a whole allows us to understand the whole propagation of risk in the system, although we lose interpretability for this.

For each of the CDS we will present a summary table of the posterior distributions for the parameters. These tables include the mean, standard deviation and a series of percentiles of the distribution of all the parameters associated with each of the CDS, i.e. the autoregressive parameters associated with the relationships presented in the DAG obtained in Structure Learning and the standard deviation of the residuals. Subsequently, we present graphically the posterior distributions of the parameters, which allow us to understand in a more visual way the relationships present in our Dynamic Bayesian Network. Finally, we will reflect on the results, analyzing the transmission of the risk underlying the results obtained. In order to keep the analysis simple, we will only study the transmission of risk at a given period, i.e. the transmission marked by the autoregressive structure, although we could simply understand what the transmission is at a longer term, taking into account that this transmission is significantly damped and becomes less and less strong. Having said this, we now move on to work with the different sectoral CDS.

5.2.2 Banking sectoral CDS

We first analyze the CDS associated to the banking sector, DSEBK5E. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that the systemic risk is transmitted towards this sector from the CDS associated to this same banking sector with one lag, DSEBK5E, to the power sector with one lag, DSEEP5E, to the energy sector with two lags, DSEEC5E, to the other financial companies sector with one delay, DSEOF5E, to the sovereign sector with one delay, DSESV5E, to the telephony sector with one delay, DSETL5E, and to the transport sector with one delay, DSETR5E.

A summary of the estimation results of the model associated with the sectoral CDS DSEBK5E is presented in Table 5. On the other hand, the posterior distributions of the 8 parameters associated with this sectoral CDS, 7 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 5.

In the Table 5 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the banking sector, DSEBK5E. In this case we can see that the parameters associated to the sectoral CDS DSEEP5E, DSEEC5E, DSEOF5E, DSESV5E and DSETL5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEBK5E and DSETR5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 5 or in the Figure 5 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSEBK5E,1}$	-0.387	0.050	-0.485	-0.421	-0.387	-0.353	-0.289
$\beta_{DSEEP5E,1}$	0.373	0.062	0.252	0.331	0.373	0.415	0.493
$\beta_{DSEEC5E,2}$	0.132	0.038	0.057	0.106	0.132	0.157	0.206
$\beta_{DSEOF5E,1}$	0.287	0.091	0.110	0.226	0.287	0.348	0.464
$\beta_{DSESV5E,1}$	0.065	0.022	0.023	0.051	0.065	0.080	0.108
$\beta_{DSETL5E,1}$	0.204	0.073	0.060	0.155	0.205	0.254	0.348
$\beta_{DSETR5E,1}$	-0.143	0.047	-0.235	-0.174	-0.143	-0.111	-0.051
σ	0.041	0.002	0.039	0.040	0.041	0.042	0.045

Table 5: Summary of the posterior distributions of the parameters for DSEBK5E.

have an effect with the same sign in the series of the sectoral CDS DSEBK5E.

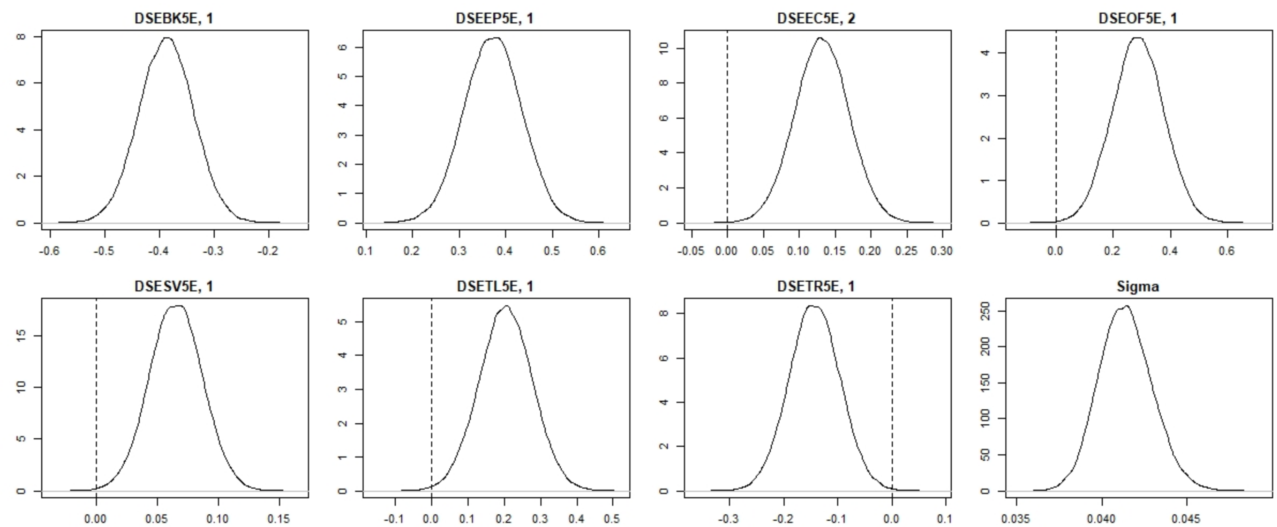


Figure 5: Posterior distributions of the parameters for DSEBK5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 5 corresponds with respect to the standard deviation of the DSEBK5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 32.7%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.3 Consumer goods sectoral CDS

The results for the CDS associated with the consumer goods sector, DSECG5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that systemic risk is transmitted to this sector from the CDS associated to this same consumer goods sector with one lag, DSECG5E, to the electricity sector with one lag, DSEEP5E, to the manufacturing sector with one and two lags, DSEMF5E, to the sovereign sector with one lag, DSESV5E, to the telephony sector with one lag, DSETL5E, and to the transport sector with two lags, DSETR5E.

A summary of the estimation results of the model associated with the sectoral CDS DSECG5E is presented in Table 6. On the other hand, the posterior distributions of the 8 parameters associated with this sectoral CDS, 7 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 6.

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSECG5E,1}$	-0.092	0.018	-0.128	-0.105	-0.092	-0.080	-0.058
$\beta_{DSEEP5E,1}$	-0.006	0.006	-0.017	-0.010	-0.006	-0.003	0.005
$\beta_{DSEMF5E,1}$	0.009	0.005	-0.002	0.005	0.009	0.012	0.020
$\beta_{DSEMF5E,2}$	0.009	0.005	-0.002	0.005	0.009	0.012	0.019
$\beta_{DSESV5E,1}$	0.022	0.007	0.010	0.018	0.022	0.027	0.035
$\beta_{DSETL5E,1}$	0.045	0.009	0.028	0.039	0.045	0.051	0.062
$\beta_{DSETR5E,2}$	0.018	0.011	-0.004	0.010	0.018	0.025	0.039
σ	0.023	0.000	0.022	0.022	0.022	0.023	0.023

Table 6: Summary of the posterior distributions of the parameters for DSECG5E.

In the Table 6 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the consumer goods sector, DSECG5E. In this case we can see that the parameters associated to the sectoral CDS DSEMF5E, DSESV5E, DSETL5E and DSETR5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSECG5E and DSEEP5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 6 or in the Figure 6 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSECG5E. The only exception is the sectoral CDS DSETR5E, which seems to have negative effects in some cases.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 6 corresponds with respect to the standard deviation of the DSECG5E sectoral

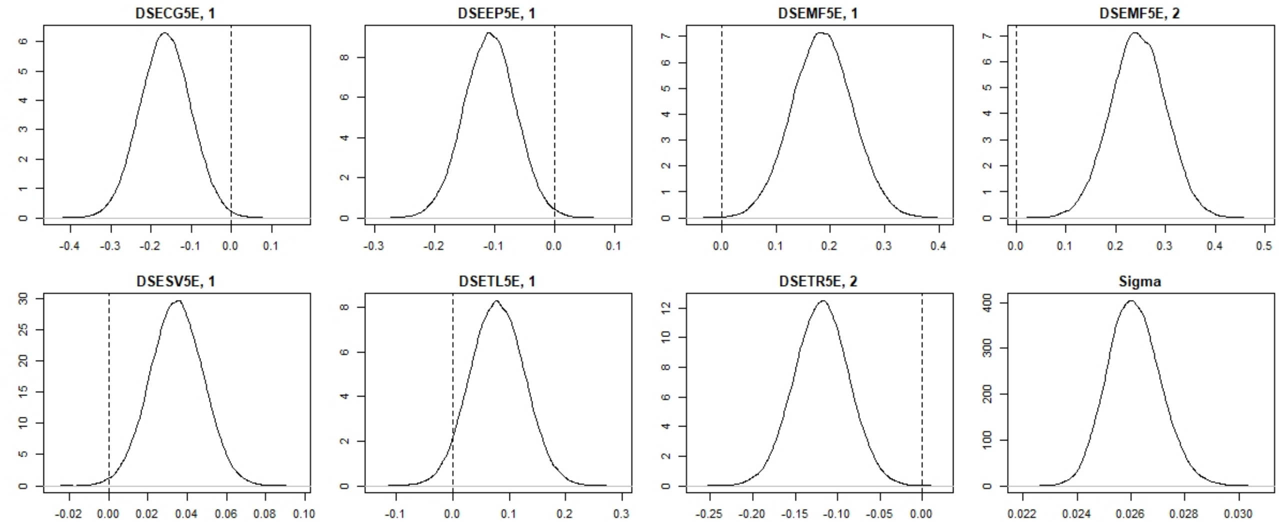


Figure 6: Posterior distributions of the parameters for DSECG5E.

CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 31.1%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.4 Electricity sectoral CDS

The results for the CDS associated with the electricity sector, DSEEP5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that systemic risk is transmitted to this sector from the CDS associated to this same electricity sector with one and two lags, DSEEP5E, to the banking sector with two lags, DSEBK5E, to the energy sector with one lag, DSEEC5E, to the manufacturing sector with one and two lags, DSEMF5E, to the other financial firms sector with one and two lags, DSEOF5E, and to the transport sector with two lags, DSETR5E.

A summary of the estimation results of the model associated with the sectoral CDS DSEEP5E is presented in Table 7. On the other hand, the posterior distributions of the 10 parameters associated with this sectoral CDS, 9 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 7.

In the Table 7 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the electricity sector, DSEEP5E. In this case we can see that the parameters associated to the sectoral CDS DSEMF5E and DSEOF5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEBK5E, DSEEP5E, DSEEC5E and DSETR5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 7 or in the Figure 7 we can see how in all cases the effect of each of the series maintains the same sign, that

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSEBK5E,2}$	-0.065	0.037	-0.137	-0.090	-0.065	-0.041	0.007
$\beta_{DSEEP5E,1}$	-0.776	0.056	-0.886	-0.814	-0.776	-0.738	-0.666
$\beta_{DSEEP5E,2}$	-0.536	0.053	-0.641	-0.572	-0.536	-0.501	-0.432
$\beta_{DSEEC5E,1}$	-0.148	0.032	-0.211	-0.170	-0.148	-0.126	-0.084
$\beta_{DSEMF5E,1}$	0.443	0.071	0.303	0.395	0.443	0.492	0.584
$\beta_{DSEMF5E,2}$	0.351	0.076	0.202	0.300	0.351	0.403	0.500
$\beta_{DSEOF5E,1}$	0.281	0.070	0.144	0.234	0.281	0.327	0.416
$\beta_{DSEOF5E,2}$	0.194	0.068	0.060	0.148	0.194	0.240	0.328
$\beta_{DSETR5E,2}$	-0.154	0.039	-0.230	-0.180	-0.154	-0.128	-0.078
σ	0.030	0.001	0.028	0.030	0.030	0.031	0.033

Table 7: Summary of the posterior distributions of the parameters for DSEEP5E.

is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSEEP5E. The only exception is the sectoral CDS DSEBK5E, which seems to have positive effects in some cases.

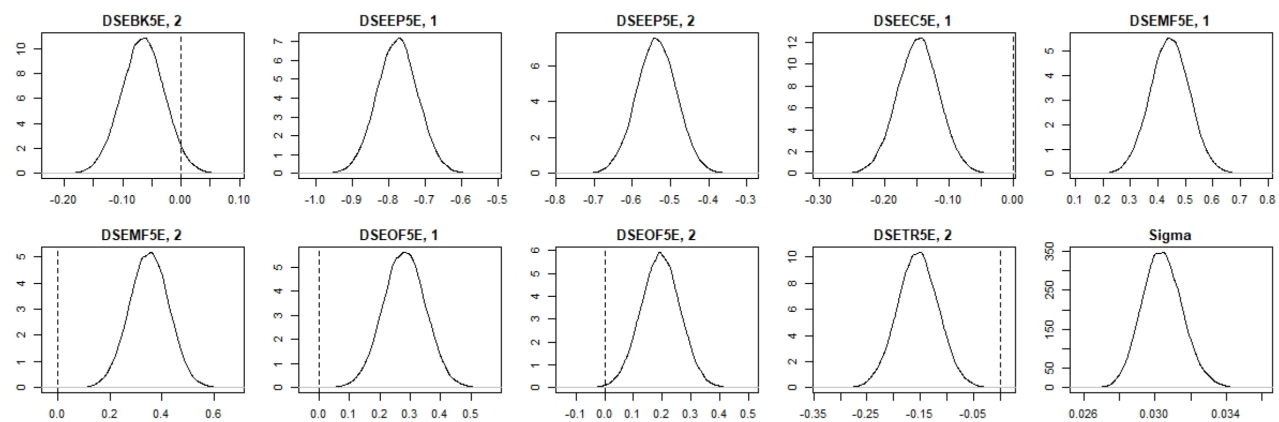


Figure 7: Posterior distributions of the parameters for DSEEP5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 7 corresponds with respect to the standard deviation of the DSEEP5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 39.0%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.5 Energy sectoral CDS

The results for the CDS associated with the energy sector, DSEEC5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that systemic risk is transmitted to this sector from the CDS associated to this same energy sector with two lags, DSEEC5E, to the banking sector with two lags, DSEBK5E, to the consumer goods sector with one lag, DSECG5E, the electricity sector with one delay, DSEEP5E, the manufacturing sector with one delay, DSEMF5E, the services sector with one delay, DSESC5E, the telephony sector with one delay, DSETL5E, and the transport sector with two delays, DSETR5E.

A summary of the estimation results of the model associated with the sectoral CDS DSEEC5E is presented in Table 8. On the other hand, the posterior distributions of the 9 parameters associated with this sectoral CDS, 8 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 8.

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSEBK5E,2}$	0.328	0.063	0.206	0.286	0.328	0.370	0.450
$\beta_{DSECG5E,1}$	0.448	0.134	0.184	0.358	0.449	0.538	0.709
$\beta_{DSEEP5E,1}$	0.253	0.092	0.072	0.191	0.253	0.315	0.434
$\beta_{DSEEC5E,2}$	0.200	0.055	0.092	0.163	0.200	0.237	0.308
$\beta_{DSEMF5E,1}$	-0.238	0.123	-0.479	-0.320	-0.239	-0.156	0.002
$\beta_{DSESC5E,1}$	0.120	0.034	0.054	0.097	0.120	0.143	0.187
$\beta_{DSETL5E,1}$	0.247	0.099	0.054	0.181	0.247	0.314	0.441
$\beta_{DSETR5E,2}$	-0.271	0.061	-0.391	-0.312	-0.271	-0.230	-0.152
σ	0.055	0.002	0.051	0.054	0.055	0.056	0.059

Table 8: Summary of the posterior distributions of the parameters for DSEEC5E.

In the Table 8 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the energy sector, DSEEC5E. In this case we can see that the parameters associated to the sectoral CDS DSEBK5E, DSECG5E, DSEEP5E, DSEEC5E, DSESC5E and DSETL5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEMF5E and DSETR5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 8 or in the Figure 8 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSEEC5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 8 corresponds with respect to the standard deviation of the DSEEC5E sectoral

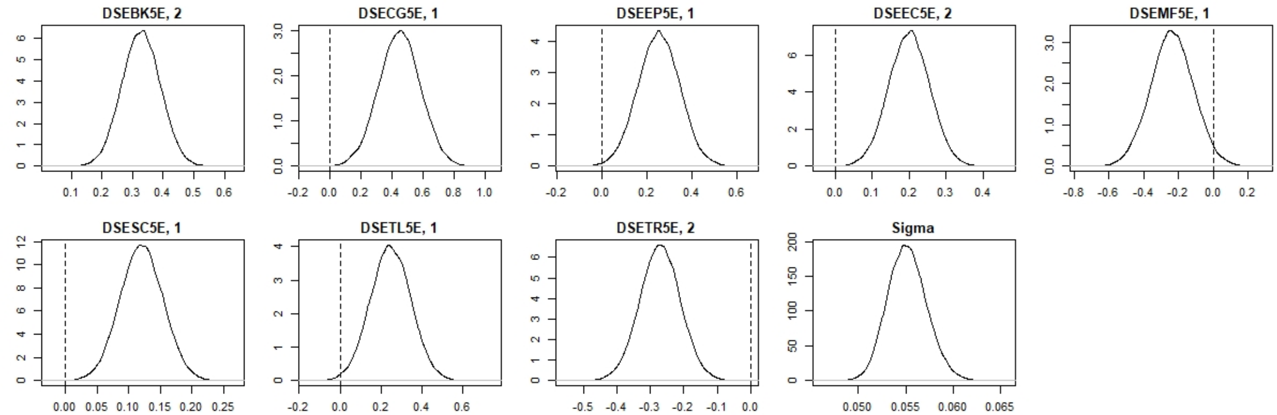


Figure 8: Posterior distributions of the parameters for DSEEC5E.

CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 24.7%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.6 Manufacturing sectoral CDS

The results for the CDS associated with the manufacturing sector, DSEMF5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that systemic risk is transmitted to this sector from the CDS associated to this same manufacturing sector with one lag, DSEMF5E, to the consumer goods sector with one lag, DSECG5E, to the energy sector with one and two lags, DSEEC5E, to the other financial firms sector with one lag, DSEOF5E, to the services sector with one lag, DSESC5E, and to the telephony sector with one lag, DSETL5E.

A summary of the estimation results of the model associated with the sectoral CDS DSEMF5E is presented in Table 9. On the other hand, the posterior distributions of the 8 parameters associated with this sectoral CDS, 7 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 9.

In the Table 9 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the manufacturing sector, DSEMF5E. In this case we can see that the parameters associated to the sectoral CDS DSECG5E, DSEEC5E, DSEOF5E and DSETL5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEMF5E and DSESC5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 9 or in the Figure 9 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSEMF5E. The only exception is the sectoral CDS DSESC5E, which seems to have positive effects in some cases.

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSECG5E,1}$	0.190	0.065	0.063	0.146	0.190	0.233	0.316
$\beta_{DSEEC5E,1}$	0.114	0.029	0.057	0.095	0.114	0.134	0.172
$\beta_{DSEEC5E,2}$	0.145	0.025	0.096	0.128	0.145	0.162	0.194
$\beta_{DSEMF5E,1}$	-0.237	0.065	-0.366	-0.282	-0.237	-0.194	-0.108
$\beta_{DSEOF5E,1}$	0.119	0.065	-0.007	0.075	0.119	0.162	0.247
$\beta_{DSESC5E,1}$	-0.027	0.018	-0.063	-0.040	-0.027	-0.015	0.007
$\beta_{DSETL5E,1}$	0.192	0.057	0.081	0.154	0.192	0.229	0.303
σ	0.028	0.001	0.026	0.028	0.028	0.029	0.030

Table 9: Summary of the posterior distributions of the parameters for DSEMF5E.

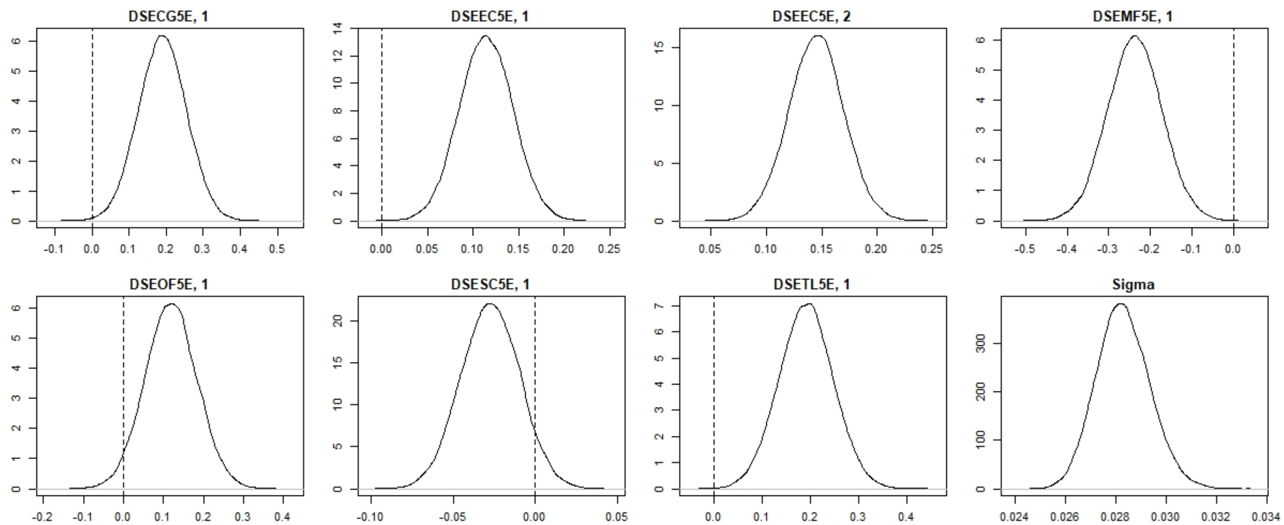


Figure 9: Posterior distributions of the parameters for DSEMF5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 9 corresponds with respect to the standard deviation of the DSEMF5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 22.0%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.7 Other financial enterprises sectoral CDS

The results for the CDS associated with the other financial enterprises sector, DSEOF5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that the systemic risk is transmitted to this sector from the CDS associated to this same sector of other financial firms with one lag, DSEOF5E, to the electricity sector with two lags, DSEEP5E, to the energy sector with two lags, DSEEC5E, and to the manufacturing sector with one lag, DSEMF5E.

A summary of the estimation results of the model associated with the sectoral CDS DSEOF5E is presented in Table 10. On the other hand, the posterior distributions of the 5 parameters associated with this sectoral CDS, 4 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 10.

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSEEP5E,2}$	0.057	0.039	-0.019	0.031	0.057	0.083	0.132
$\beta_{DSEEC5E,2}$	0.076	0.024	0.029	0.060	0.076	0.093	0.123
$\beta_{DSEMF5E,1}$	0.217	0.054	0.111	0.180	0.217	0.254	0.323
$\beta_{DSEOF5E,1}$	-0.042	0.058	-0.156	-0.081	-0.042	-0.002	0.072
σ	0.028	0.001	0.026	0.027	0.028	0.028	0.030

Table 10: Summary of the posterior distributions of the parameters for DSEOF5E.

In the Table 10 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the other financial enterprises sector, DSEOF5E. In this case we can see that the parameters associated to the sectoral CDS DSEEP5E, DSEEC5E and DSEMF5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEOF5E has negative associated parameter, indicating that this perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 10 or in the Figure 10 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSEMF5E. The only exception is the sectoral CDS DSEOF5E, which seems to have positive effects in some cases.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 10 corresponds with respect to the standard deviation of the DSEOF5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 5.5%. This proportion is not high but it is not negligible. This is the case where the model is simpler, has fewer associated parameters, and where the least systemic risk is transmitted.

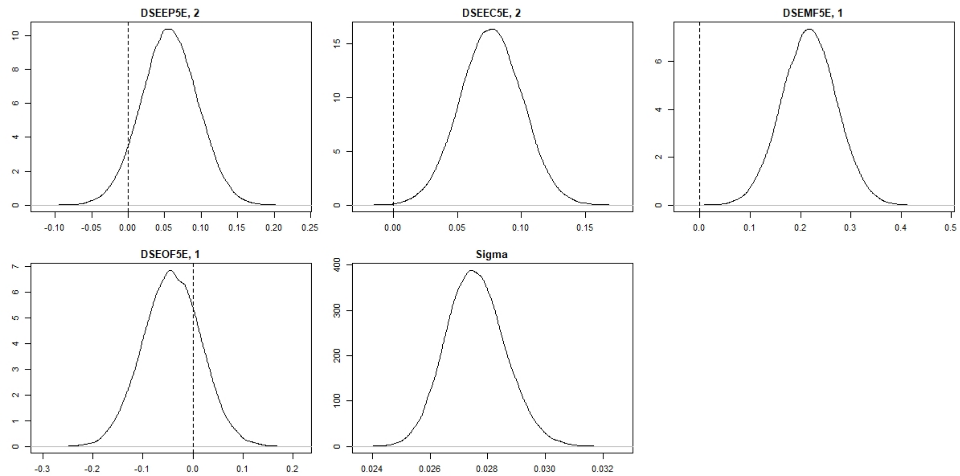


Figure 10: Posterior distributions of the parameters for DSEOF5E.

5.2.8 Services sectoral CDS

The results for the CDS associated with the service sector, DSESC5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that systemic risk is transmitted to this sector from the CDS associated to this same services sector with one lag, DSESC5E, to the consumer goods sector with one lag, DSECG5E, to the electricity sector with one and two lags, DSEEP5E, to the manufacturing sector with two lags, DSEMF5E, to the telephony sector with one lag, DSETL5E, and to the transport sector with two lags, DSETR5E.

A summary of the estimation results of the model associated with the sectoral CDS DSESC5E is presented in Table 11. On the other hand, the posterior distributions of the 8 parameters associated with this sectoral CDS, 7 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 11.

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSECG5E,1}$	0.636	0.199	0.245	0.502	0.637	0.771	1.025
$\beta_{DSEEP5E,1}$	-0.705	0.150	-1.000	-0.805	-0.705	-0.604	-0.411
$\beta_{DSEEP5E,2}$	-0.731	0.139	-1.005	-0.824	-0.732	-0.637	-0.456
$\beta_{DSEMF5E,2}$	0.685	0.200	0.294	0.550	0.685	0.820	1.079
$\beta_{DSESC5E,1}$	-0.242	0.051	-0.342	-0.277	-0.242	-0.208	-0.142
$\beta_{DSETL5E,1}$	0.531	0.143	0.252	0.435	0.532	0.627	0.813
$\beta_{DSETR5E,2}$	-0.490	0.104	-0.694	-0.561	-0.490	-0.420	-0.287
σ	0.085	0.003	0.079	0.083	0.085	0.087	0.091

Table 11: Summary of the posterior distributions of the parameters for DSESC5E.

In the Table 11 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the service sector, DSESC5E. In this case we can see that the parameters associated to the sectoral CDS DSECG5E, DSEMF5E and DSETL5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEEP5E, DSESC5E and DSETR5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 11 or in the Figure 11 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSESC5E.

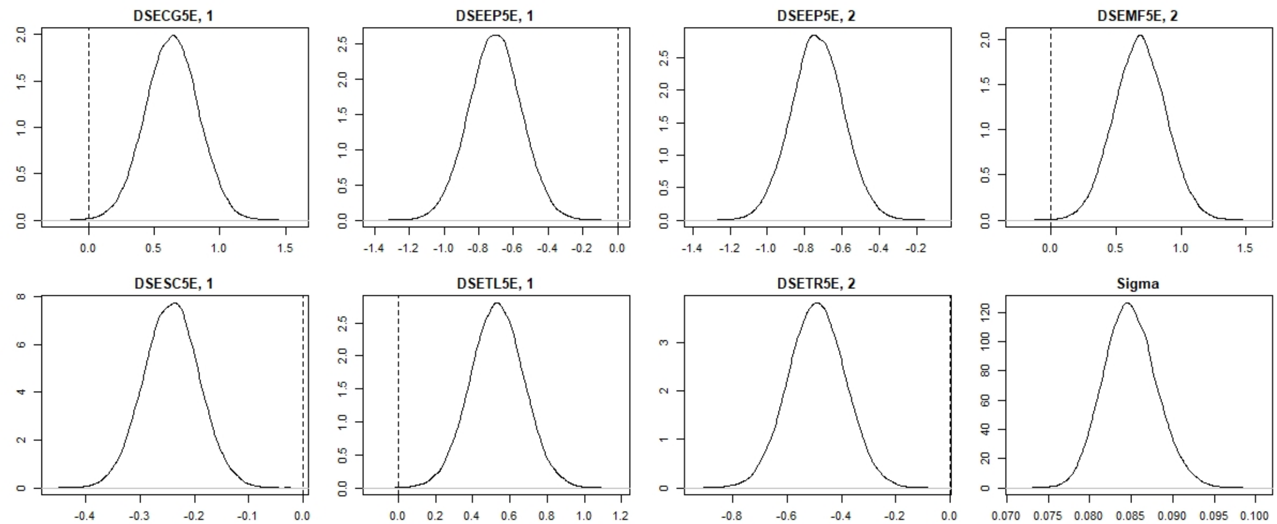


Figure 11: Posterior distributions of the parameters for DSESC5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 11 corresponds with respect to the standard deviation of the DSEMF5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 5.5%. This proportion, as for DSEOF5E, is not high but it is not negligible.

5.2.9 Sovereign sectoral CDS

The results for the CDS associated with the sovereign sector, DSESV5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that the systemic risk is transmitted to this sector from the CDS associated to this same sovereign sector with one and two lags, DSESV5E, to the electricity sector with one and two lags, DSEEP5E, to the manufacturing sector with one lag, DSEMF5E, and to the transport sector with one lag, DSETR5E.

A summary of the estimation results of the model associated with the sectoral CDS DSESV5E is presented in Table 12. On the other hand, the posterior distributions of the 7 parameters associated with this sectoral CDS, 6 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 12.

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSEEP5E,1}$	0.476	0.155	0.171	0.371	0.475	0.580	0.778
$\beta_{DSEEP5E,2}$	0.613	0.141	0.337	0.519	0.613	0.708	0.887
$\beta_{DSEMF5E,1}$	-0.642	0.219	-1.071	-0.789	-0.642	-0.494	-0.209
$\beta_{DSESV5E,1}$	-0.304	0.052	-0.406	-0.339	-0.304	-0.269	-0.202
$\beta_{DSESV5E,2}$	-0.124	0.051	-0.224	-0.158	-0.124	-0.090	-0.026
$\beta_{DSETR5E,1}$	0.402	0.118	0.170	0.322	0.402	0.482	0.633
σ	0.097	0.004	0.090	0.095	0.097	0.099	0.104

Table 12: Summary of the posterior distributions of the parameters for DSESV5E.

In the Table 12 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the sovereign sector, DSESV5E. In this case we can see that the parameters associated to the sectoral CDS DSEEP5E and DSETR5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEMF5E and DSESV5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 12 or in the Figure 12 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSESV5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 12 corresponds with respect to the standard deviation of the DSESV5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 13.3%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.10 Telephony sectoral CDS

The results for the CDS associated with the telephony sector, DSETL5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that systemic risk is transmitted to this sector from the CDS associated to this same telephony sector with a lag, DSETL5E, to the consumer goods sector with a lag, DSECG5E,

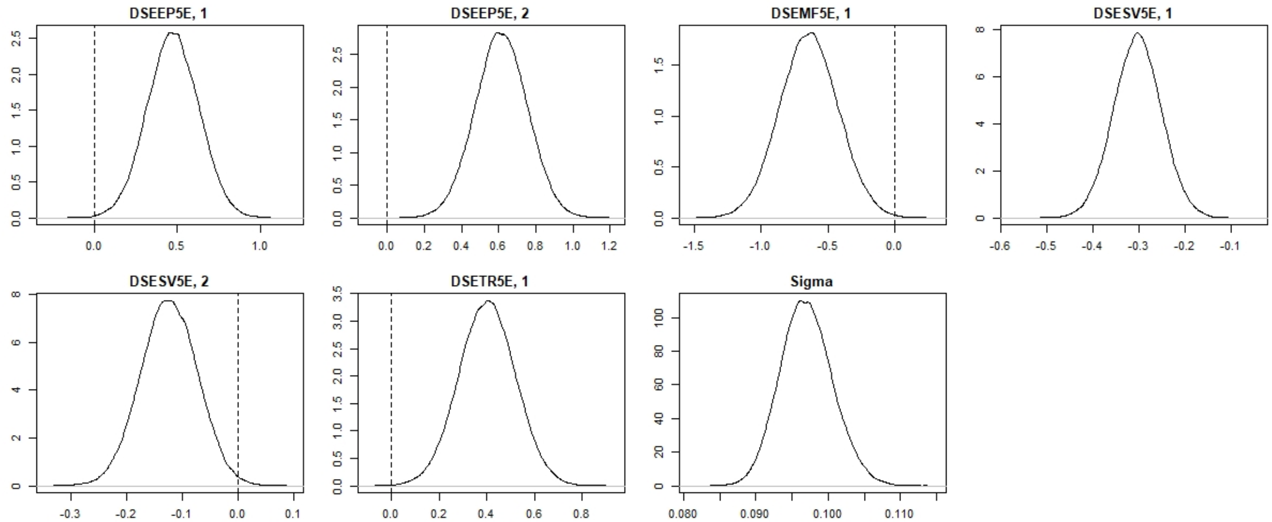


Figure 12: Posterior distributions of the parameters for DSESV5E.

to the manufacturing sector with a lag, DSEMF5E, to the other financial firms sector with a lag, DSEOF5E, and to the transport sector with a lag, DSETR5E.

A summary of the estimation results of the model associated with the sectoral CDS DSETL5E is presented in Table 13. On the other hand, the posterior distributions of the 6 parameters associated with this sectoral CDS, 5 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 13.

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSECG5E,1}$	0.099	0.075	-0.050	0.048	0.099	0.149	0.246
$\beta_{DSEMF5E,1}$	0.270	0.079	0.116	0.216	0.270	0.323	0.425
$\beta_{DSEOF5E,1}$	0.254	0.076	0.107	0.204	0.254	0.305	0.402
$\beta_{DSETL5E,1}$	-0.223	0.061	-0.341	-0.264	-0.223	-0.182	-0.104
$\beta_{DSETR5E,1}$	-0.068	0.039	-0.144	-0.094	-0.068	-0.042	0.008
σ	0.033	0.001	0.031	0.032	0.033	0.034	0.036

Table 13: Summary of the posterior distributions of the parameters for DSETL5E.

In the Table 13 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the telephony sector, DSETL5E. In this case we can see that the parameters associated to the sectoral CDS DSECG5E, DSEMF5E and DSEOF5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSETL5E and DSETR5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand, when observing the distribution of the different parameters in the Table 13 or in the Figure 13 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSETL5E.

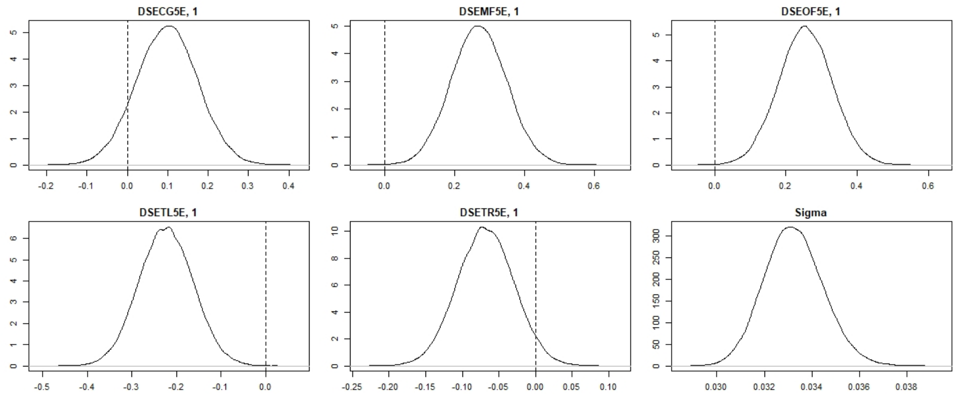


Figure 13: Posterior distributions of the parameters for DSETL5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 13 corresponds with respect to the standard deviation of the DSETL5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 10.1%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.11 Transport sectoral CDS

Lastly, the results for the CDS associated with the transport sector, DSETR5E, are discussed below. Under the structure of our Dynamic Bayesian Network, associated to the DAG presented in the Figure 4, we have seen that systemic risk is transmitted to this sector from the CDS associated to this same transport sector with one and two lags, DSETR5E, to the consumer goods sector with one lag, DSECG5E, to the energy sector with one and two lags, DSEEC5E, to the manufacturing sector with one lag, DSEMF5E, and to the services sector with one lag, DSESC5E.

A summary of the estimation results of the model associated with the sectoral CDS DSETR5E is presented in Table 14. On the other hand, the posterior distributions of the 9 parameters associated with this sectoral CDS, 8 of the relationships with the lagged series and 1 of the standard deviation, are plotted in Figure 14.

In the Table 14 we can see how systemic risk is transmitted from the lagged time series to the CDS associated with the transport sector, DSETR5E. In this case we can see that the parameters associated to the sectoral CDS DSECG5E, DSEEC5E and DSESC5E are positive, indicating that they transmit systemic risk directly, while the sectoral CDS DSEMF5E and DSETR5E have negative associated parameters, indicating that these perform a correction of the transmitted risk. On the other hand,

	mean	sd	2.5%	25%	50%	75%.	97.5%
$\beta_{DSECG5E,1}$	0.564	0.090	0.388	0.503	0.563	0.625	0.739
$\beta_{DSEEC5E,1}$	0.356	0.045	0.267	0.326	0.356	0.386	0.443
$\beta_{DSEEC5E,2}$	0.249	0.045	0.161	0.218	0.249	0.279	0.338
$\beta_{DSEMF5E,1}$	-0.496	0.101	-0.695	-0.565	-0.496	-0.428	-0.297
$\beta_{DSESC5E,1}$	0.127	0.028	0.071	0.107	0.126	0.146	0.182
$\beta_{DSESC5E,2}$	0.144	0.029	0.087	0.124	0.144	0.163	0.201
$\beta_{DSETR5E,1}$	-0.178	0.056	-0.288	-0.215	-0.177	-0.140	-0.067
$\beta_{DSETR5E,2}$	-0.449	0.049	-0.546	-0.482	-0.448	-0.416	-0.353
σ	0.042	0.002	0.039	0.041	0.042	0.043	0.045

Table 14: Summary of the posterior distributions of the parameters for DSETR5E.

when observing the distribution of the different parameters in the Table 14 or in the Figure 14 we can see how in all cases the effect of each of the series maintains the same sign, that is, that under the Dynamic Bayesian Network considered the shocks in the lagged series always have an effect with the same sign in the series of the sectoral CDS DSETR5E.

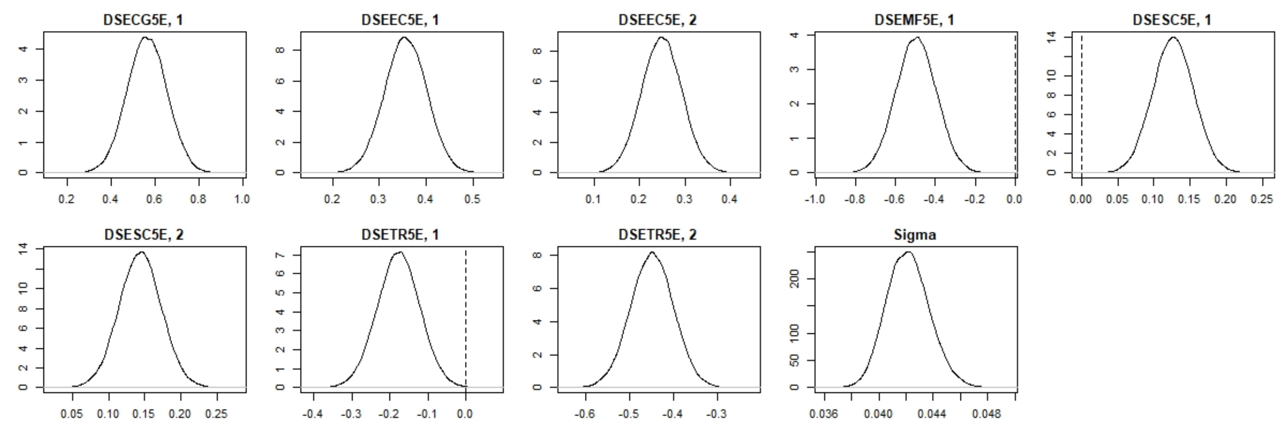


Figure 14: Posterior distributions of the parameters for DSETR5E.

Finally, by analyzing the standard deviation of the residuals and comparing it with the standard deviation of the original series we can see what proportion of the risk is explained by the model and what proportion is still unexplained. As expected, the vast majority of the systemic risk should not be explained by this model since only a part of the risk is transmitted, the rest being a consequence of new innovations in the market. Thus, analyzing the proportion of the variance to which the standard deviation of the Table 14 corresponds with respect to the standard deviation of the DSETR5E sectoral CDS returns indicated in the Table 2 we can see that the proportion of the systemic risk explained by this model is around 41.5%. This proportion is high but reasonable, not negligible but also not suspicious of being the result of overfitting.

5.2.12 Risk Transmission

The estimation of the Dynamic Bayesian Network provides us with information on how systemic risk is transmitted between the different sectors considered. In this brief section we will summarize the information provided by each of the elements of this model on the transmission of systemic risk, draw conclusions on the results obtained in the previous sections, and propose other possible analyses.

As we have already advanced above, the posterior distribution of the standard deviation of the residuals provides us with information on the amount of systemic risk that is not identified as transmitted from the rest of the CDS by the model. Similarly, by comparing this posterior distribution of the standard deviation of the residuals with the standard deviation of the original series we will have information on the proportion of systemic risk that the model identifies as transmitted by the rest of the sectoral CDS. In short, the different estimated standard deviations provide us with information on the transmission of systemic risk in terms of the weight of transmitted systemic risk with respect to total systemic risk. On the other hand, the posterior distribution of the model parameters associated with the autoregressive structure indicates in what specific way the risk propagates from the lagged series to the original series. These posterior distributions provide information on the whole set of values that the effects of shocks in the lagged series on the original series can take.

Looking at all the graphs presented in the previous sections, which contain the posterior distributions of the parameters, we can see what shape they have. Specifically, we can see that the distribution of the different parameters of the Dynamic Bayesian Network is certainly symmetrical, however, a complete normality is not observed due to the high kurtosis. These distributions provide important information on how risk is propagated, since they fully clarify the systemic risk transmitted from each of the lagged series to the sectoral CDS series, including the relevant information on which ranges of values these effects are included and under what probability.

By analyzing the values that the posterior distributions of the parameters corresponding to the autoregressive structure consider as probable, we can get an idea of how the different lagged series affect the original series. In the analysis presented in the previous sections we have that these posterior distributions show how the risk is transmitted one or two periods ahead. The analysis performed is very simple and lends itself to considering alternative analyses.

In particular, it might be interesting to see how a shock in a given CDS series is transmitted over time over the different series from the structure of our Dynamic Bayesian Network. Similarly, it might be interesting to see over time how risk is transmitted from one series to another, fixing the rest of the series. These two approaches, among others, can provide valuable information on the transmission of systemic risk in the analyzed problem.

In short, by analyzing the results of the model estimation associated with each sectoral CDS we have been able to see how systemic risk is propagated to it, understanding from which lagged series and with what effect. In addition, we have been able to appreciate what proportion of the new risk is a consequence of this transmission and what proportion is due to market innovations. The brief results presented in the previous sections can be extended in order to observe more specific transmissions between series or, alternatively, longer-term transmissions of systemic risk. Therefore, it remains open the possibility of focusing on results from the Dynamic Bayesian Network studied to inform us about other desired risk transmission relationships.

5.2.13 Penalized Methods

To finish the work with the model designed, an additional approach based on penalized methods has been carried out. Specifically, we have implemented again the model with which we have worked along the Section 5.2 but in this case performing an estimation with penalized methods. The aim of this new analysis is to observe the robustness of the results.

The estimation performed throughout the rest of this Section 5.2 could have been done from a penalized methods approach. The estimation resulting from employing these penalized methods differs in the results due to their different approach. The penalized methods are based on minimizing the sum of squared errors by penalizing, in turn, the size of the parameters, as represented in the equation (8).

To work with penalized methods in the context of Bayesian statistics we have used the package *bayesreg* from R. This package includes functions to facilitate the implementation of Bayesian penalized regression models. In this way, we implemented the estimation of the Dynamic Bayesian Network whose structure we obtained in the Structure Learning, 5.1, and is presented in the Figure 4. In this case the introduction of the priors was done in the same terms as explained in 5.2.1, also taking non-informative prior distributions.

The result of performing Parameter Learning under the penalized methods is different depending on the penalty parameters used, the β of the equation (8). In the analyzed case we adjusted the model using different values for the β, obtaining initially results practically identical to those we have presented in the rest of the Section 5.2. By greatly increasing the value of the β then the results finally ended up changing markedly. In this case the estimates turned out to be much worse, with very large residuals. In this way, we were able to analyze that the parameter estimates made throughout the Section 5.2 were robust.

On the other hand, the initial stability of the parameter estimates when working with penalized methods indicates that we should not disregard any of the relationships considered by the structure of our Dynamic Bayesian Network. It is not until we consider a sufficiently high value for the β that some of the parameters are estimated to be zero, and in this case the estimation of the model as a whole is clearly erroneous. For all this we have an added reason to consider that the structure learned in Structure Learning is correct.

In short, the estimation of the Dynamic Bayesian Network using the penalized methods in the context of Bayesian statistics confirms that the parameter estimation presented between Sections 5.2.2 and 5.2.11 is correct and robust. Moreover, in turn, it provides us with further evidence to confirm that the structure learned in Structure Learning along Section 5.1 is good and does not have an excess of relations.

6 Conclusions

The aim of this paper is to employ network models to study the transmission of systemic risk among different CDS, in particular European sectoral CDS, focusing on the period of the current COVID-19 financial crisis. In order to analyze the transmission of systemic risk among European sectoral CDS, it was decided to work with network models. Moreover, an approach in the context of Bayesian statistics was chosen in order to consider the transmission parameters as random variables having a certain distribution. Therefore, the work lies at the intersection of two literature topics, those of credit risk and Bayesian statistics, the nexus being network models. Dynamic Bayesian Networks were used to analyze the transmission of systemic risk among sectoral CDS. Next, both Structure Learning and Parameter Learning were employed in order to understand how systemic risk is transmitted among European sectoral CDS.

We modeled the transmission of systemic risk among CDS using networks, specifically Bayesian networks. Since we worked with daily data, the most obvious way to measure the transmission of systemic risk was to consider the relationships that could occur among the CDS series, both immediate and delayed, and thus a Dynamic Bayesian Network was used. Initially, we considered the possibility of introducing into the network instantaneous relationships among the original series and also relationships with series with between one and five delays. In short, we considered a Dynamic Bayesian Network whose structure could have a total of 60 nodes. Having established the network, we then performed Structure Learning.

The first finding of our work has been that of all the relationships considered for our Bayesian network, only a few, those between the original series and the series delayed with one or two lags, are relevant. After considering the Dynamic Bayesian Network in the described terms, the first analysis was to determine which of the possible relationships are actually of interest. Unconditional independence tests were initially implemented in order to rule out some of the proposed relationships. In the same way, conditional independence tests were carried out with respect to the set formed by all the series, in order to continue discarding possible irrelevant relationships. From these tests, we arrived at the first important finding of this work: of all the relationships considered, only a few seem to be of interest. In particular, from the unconditional independence tests we eliminated relationships with series with three, four and five lags, while from the conditional independence tests we excluded instantaneous relationships. Next, we considered the basis on which to perform Structure Learning.

The second result of the study is the systemic risk transmission structure between the different European sectoral CDS. Once the basis for the structure of the Dynamic Bayesian Network was established, we carried out Structure Learning by implementing successive conditional independence tests through an exhaustive search among all possible structures. This onerous procedure was parallelized in order to reduce the computation time. In this way, the structure of the Dynamic Bayesian Network was formed, which summarizes how risk is transmitted among the different European sectoral CDS. Alternatively, the structure was traced from conditional independence tests with respect to the set formed by all the series, arriving at the same structure. In short, based on the conditional independence tests, the structure of the transmission of systemic risk between the different European sectoral CDS has been drawn.

Subsequently, alternative strategies for Structure Learning, based on a heuristic search of the structure or on the use of network scores, were carried out. Specifically, a heuristic search was performed among all possible structures based on conditional independence tests. We also performed a structure search based on network scores, in particular on the BIC. These two alternative ways of working confirmed to us that the structure for the transmission of systemic risk among the different European sectoral CDS is correct.

In short, from the Structure Learning process we obtained the structure of our Dynamic Bayesian Network, associated with the DAG presented in Figure 4. This structure was obtained from the conditional independence tests and validated from both a heuristic search and a network score search. Furthermore, working with a non-coincident data set with the COVID-19 financial crisis, and extending the period, we arrived at an almost identical structure, confirming that the structure holds over time. Therefore, from the Structure Learning we obtained the DAG of the Dynamic Bayesian Network with which we worked in order to analyze the transmission of systemic risk among the different European sectoral CDS during the COVID-19 financial crisis.

Once the Structure Learning was performed, Parameter Learning was carried out entirely in the context of Bayesian statistics. To this end, an estimation was performed using simulation methods based on Markov Chain Monte Carlo. Assuming the DAG learned in the Structure Learning, we finished defining the Bayesian hierarchical model by introducing non-informative prior distributions that were selected according to the intrinsic characteristics of the problem, logically taking advantage of the nature of each of the parameters of the model. Once the model was defined, the estimation was carried out and the correctness of the results was guaranteed by performing a diagnosis of the model and the simulations. In this way, we found that the model is compatible with the data and that the associated results are correct. The results were then analyzed.

The most significant finding of this study is the way in which systemic risk is transmitted in the Dynamic Bayesian Network. The different parameters of the model, both those associated with the autoregressive structure and those associated with the variance and covariance matrix, provide information on the transmission of risk. By analyzing the posterior distributions of all the parameters, we understand how risk is transmitted along the Dynamic Bayesian Network, as well as what proportion of the new risk is a consequence of the transmission of risk from previous time points. On the one hand, by analyzing the posterior distribution of the parameters associated with the standard deviation of the residuals we have been able to understand what proportion of the risk is a consequence of risk transmission among the sectoral CDS series. Thus, we have found that the new systemic risk is explained between 5% and 40% by the transmission among the different CDS series. On the other hand, by analyzing the posterior distribution of the parameters associated with the autoregressive structure of the residuals we have been able to see how exactly it is transmitted along the Dynamic Bayesian Network. Therefore, the posterior distributions have allowed us to understand the transmission of systemic risk among European sectoral CDS during the COVID-19 financial crisis.

By analyzing the posterior distributions we have been able to see the sign and size of each of the parameters of the autoregressive structure, thereby understanding the strength of each of the existing relationships among the sectoral CDS series. In this way, we can determine whether the various relationships indicate a direct propagation of systemic risk or whether, on the contrary, they act as a correction of the propagated risk. By individually studying each of the posterior distributions of the parameters we have verified their symmetry; however, we have also seen that they are distributions that move away from normality due to their high kurtosis.

To conclude the study, we worked with the same Dynamic Bayesian Network but with an estimation based on penalized methods. To this end, we also took the DAG structure learned during Structure Learning and we worked in a Bayesian context with the penalized methods. In this way, we have been able to verify the stability of the estimation obtained through Parameter Learning. Furthermore, we have verified that all the relationships considered by the DAG represented in the Figure 4 are relevant. Therefore, from the penalized methods we have obtained a further proof of the goodness of the model, both of the structure and of the estimation of the parameters.

The results presented are a simple representation of the analysis of the transmission of systemic risk that can be performed from the Dynamic Bayesian Network that we have estimated using Structure Learning and Parameter Learning. Specifically, we have only considered how systemic risk propagates over one or two periods, a single period seen in the model, but this could be extended and longer-term relationships could be studied. Similarly, how risk is transmitted between two particular European sectoral CDS series could be analyzed in more detail. In short, the proposed modeling allows for a more advanced analysis of the transmission of systemic risk among European sectoral CDS series.

Throughout this study, the robustness of the results has been validated by performing multiple alternative processes. In this way, Structure Learning has been carried out using the different strategies proposed, ensuring that the structure of the selected network is determined by the relevant relationships between the different original CDS series and the delayed series. In addition, the accuracy of the estimation results obtained in Parameter Learning has been confirmed.

Given the importance of risk transmission analysis, this study may be relevant from the perspective of policy implementation. Dynamic Bayesian Networks are able to identify the sources of risk and the transmission paths resulting from exogenous shocks. This modeling should be of significant interest to regulators and to supervisory authorities as it provides them with information on market fragility hotspots and allows them to anticipate potential increases in risk in different sectors. In addition, this methodology can be used by financial institutions to carry out internal systemic stress tests of contagion risk. Similarly, this study may be of interest to any financial market participant interested in analyzing and understanding the transmission of systemic risk.

Finally, the modeling of the transmission of systemic risk among CDS by means of Dynamic Bayesian Networks presented in this paper opens various lines of research. The study conducted in this case considers that the variance of the residuals for each of the sectoral CDS series is constant over time; since this assumption does not have to be true, a dynamic model for volatility could be developed. On the other hand, we have rejected the instantaneous relationships given the lack of strength they had under the analysis with daily data; however, a model that took into account the instantaneous relationships could be created, especially if intraday data were used. Finally, the study of risk transmission in a broader network, including other series from other regions, in order to better understand how systemic risk is transmitted at the international level, has yet to be done.

References

[Agosto et al., 2020] Agosto, A., Ahelegbey, D. F., and Giudici, P. (2020). Tree networks to assess financial contagion. *Economic Modelling*, 85:349–366.

[Arora et al., 2012] Arora, N., Gandhi, P., and Longstaff, F. (2012). Counterparty credit risk and the credit default swap market. *Journal of Financial Economics*, 103(2):280–293.

[Ballester et al., 2016] Ballester, L., Casu, B., and Gonzalez-Urteaga, A. (2016). Bank fragility and contagion: Evidence from the bank CDS market. *Journal of Empirical Finance*, 38:394–416.

[Bauwens et al., 1999] Bauwens, L., Lubrano, M., and Richard, J. (1999). *Bayesian inference in dynamic econometric models*. Oxford University Press, Oxford England New York.

[Benoit et al., 2016] Benoit, S., Colliard, J., Hurlin, C., and Pérignon, C. (2016). Where the risks lie: A survey on systemic risk$*. Review of Finance, 21(1) : 109 - -152$.

[Box et al., 2016] Box, G., Jenkins, G., Reinsel, G., and Ljung, G. (2016). *Time series analysis : forecasting and control*. John Wiley & Sons, Inc, Hoboken, New Jersey.

[Brownlees et al., 2021] Brownlees, C., Hans, C., and Nualart, E. (2021). Bank credit risk networks: Evidence from the eurozone. *Journal of Monetary Economics*, 117:585–599.

[Chen et al., 2020] Chen, W., Ho, K., and Yang, L. (2020). Network structures and idiosyncratic contagion in the european sovereign credit default swap market. *International Review of Financial Analysis*, 72:101594.

[Corbet et al., 2021] Corbet, S., Hou, Y. G., Hu, Y., Oxley, L., and Xu, D. (2021). Pandemic-related financial market volatility spillovers: Evidence from the chinese COVID-19 epicentre. *International Review of Economics & Finance*, 71:55–81.

[Cowell et al., 1999] Cowell, R., Dawid, A., Lauritzen, S., and Spiegelhalter, D. (1999). *Probabilistic networks and expert systems*. Springer, New York.

[Duffie, 1999] Duffie, D. (1999). Credit swap valuation. *Financial Analysts Journal*, 55(1):73–87.

[Fortunato et al., 2019] Fortunato, M., Blundell, C., and Vinyals, O. (2019). Bayesian recurrent neural networks.

[Gandy and Veraart, 2019] Gandy, A. and Veraart, L. (2019). Adjustable network reconstruction with applications to CDS exposures. *Journal of Multivariate Analysis*, 172:193–209.

[Gelman et al., 2014] Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*. CRC Press, Boca Raton.

[Gilks et al., 1996] Gilks, W., Richardson, S., and Spiegelhalter, D. (1996). *Markov chain Monte Carlo in practice*. Chapman & Hall, London.

[Heckerman, 2021] Heckerman, D. (2021). A tutorial on learning with bayesian networks.

[Hull et al., 2004] Hull, J., Predescu, M., and White, A. (2004). The relationship between credit default swap spreads, bond yields, and credit rating announcements. *Journal of Banking & Finance*, 28(11):2789–2811.

[Hull and White, 2000] Hull, J. and White, A. (2000). Valuing credit default swaps i. *The Journal of Derivatives*, 8(1):29–40.

[Hull and White, 2001] Hull, J. and White, A. (2001). Valuing credit default swaps II. *The Journal of Derivatives*, 8(3):12–21.

[Hull and White, 2003] Hull, J. C. and White, A. D. (2003). The valuation of credit default swap options. *The Journal of Derivatives*, 10(3):40–50.

[Ismailescu and Kazemi, 2010] Ismailescu, I. and Kazemi, H. (2010). The reaction of emerging market credit default swap spreads to sovereign credit rating changes. *Journal of Banking & Finance*, 34(12):2861–2873.

[Jensen and Nielsen, 2007] Jensen, F. and Nielsen, T. (2007). *Bayesian networks and decision graphs*. Springer, New York.

[Kanno, 2020] Kanno, M. (2020). Interconnectedness and systemic risk in the US CDS market. *The North American Journal of Economics and Finance*, 54:100837.

[Kartal, 2020] Kartal, M. T. (2020). The behavior of sovereign credit default swaps (CDS) spread: evidence from turkey with the effect of covid-19 pandemic. *Quantitative Finance and Economics*, 4(3):489–502.

[Kjaerulff and Madsen, 2008] Kjaerulff, U. and Madsen, A. (2008). *Bayesian networks and influence diagrams : a guide to construction and analysis*. Springer, New York London.

[Lindley and Smith, 1972] Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *Journal of the Royal Statistical Society: Series B (Methodological)*, 34(1):1–18.

[Longstaff et al., 2005] Longstaff, F. A., Mithal, S., and Neis, E. (2005). Corporate yield spreads: Default risk or liquidity? new evidence from the credit default swap market. *The Journal of Finance*, 60(5):2213–2253.

[Lunn et al., 2009] Lunn, D., Spiegelhalter, D., Thomas, A., and Best, N. (2009). The BUGS project: Evolution, critique and future directions. *Statistics in Medicine*, 28(25):3049–3067.

[Markose et al., 2012] Markose, S., Giansante, S., and Shaghaghi, A. (2012). 'too interconnected to fail' financial network of US CDS market: Topological fragility and systemic risk. *Journal of Economic Behavior & Organization*, 83(3):627–646.

[Nagarajan et al., 2013] Nagarajan, R., Scutari, M., and Lèbre, S. (2013). *Bayesian networks in R : with applications in systems biology*. Springer, New York, NY.

[Neapolitan, 2004] Neapolitan, R. (2004). *Learning Bayesian networks*. Pearson Prentice Hall, Upper Saddle River, N.J.

[Neil et al., 2005] Neil, M., Fenton, N., and Tailor, M. (2005). Using bayesian networks to model expected and unexpected operational losses. *Risk Analysis*, 25(4):963–972.

[Nicholls and Hall, 1979] Nicholls, D. F. and Hall, A. D. (1979). The exact likelihood function of multivariate autoregressive-moving average models. *Biometrika*, 66(2):259–264.

[Norden and Weber, 2009] Norden, L. and Weber, M. (2009). The co-movement of credit default swap, bond and stock markets: an empirical analysis. *European Financial Management*, 15(3):529–562.

[Peltonen et al., 2014] Peltonen, T., Scheicher, M., and Vuillemey, G. (2014). The network structure of the CDS market and its determinants. *Journal of Financial Stability*, 13:118–133.

[R Core Team, 2020] R Core Team (2020). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

[Scutari and Denis, 2021] Scutari, M. and Denis, J. (2021). *Bayesian networks : with examples in R*. CRC Press, Boca Raton.

[Spiegelhalter et al., 1993] Spiegelhalter, D. J., Dawid, A. P., Lauritzen, S., and Cowell, R. (1993). Bayesian analysis in expert systems. *Statistical Science*, 8(3).

[Sun et al., 2020] Sun, X., Wang, J., Yao, Y., Li, J., and Li, J. (2020). Spillovers among sovereign CDS, stock and commodity markets: A correlation network perspective. *International Review of Financial Analysis*, 68:101271.

[West and Harrison, 1989] West, M. and Harrison, J. (1989). *Bayesian forecasting and dynamic models*. Springer, New York.

List of Figures

List of Tables