Learning Network Structure of Financial Institutions from CDS Data

Chanatip Kitwiwattanachai*

January 14, 2015

*University of Connecticut, School of Business, 2100 Hillside Road, Storrs, CT 06269, phone: 860-486-1275, email: ckitwiwattanachai@business.uconn.edu

Learning Network Structure of Financial Institutions from CDS Data

Abstract

Default contagion is a concern especially after witnessing the financial crisis. How the financial institutions are connected is crucial to how likely they will default together. However, the information on financial network structure is not available. This paper builds a probabilistic graphical model relating the network structure to observable prices. The financial network structure can be learned using CDS data. Conditional default probability and expected conditional loss can be calculated accordingly. As the information is extracted from the CDS price, all quantities are under the risk-neutral measure, and thus account for the market fear that can trigger bank run. This knowledge can be used to gauge systemic risk and inform policy decisions.[1]

[1]This paper is supported by the GARP Risk Management Research Program 2013

Introduction

During the financial crisis, the government had to spend billions of dollars to bail-out large banks. The argument was that the default of these large banks would result in the domino effect which would cause defaults of other banks and eventually the collapse of the entire financial system. However, there is little information about how these banks are connected and which banks are worth saving. This paper intends to study how banks are interconnected using observable data.

The knowledge of how banks are connected is crucial to the understanding of systemic risk.[2] Banks are not required to disclose their positions with other banks, and thus this information is not observable.[3] The network structure is an important input into the network model such as Eisenberg and Noe (2001). Correlation matrix of stock returns is generally used to address the level of interconnectivity. However, it is a rather crude measure which only addresses pairwise interconnectivity of any two banks. Is there a way to learn about the structure of the whole network in a consistent way from the available data?

This paper proposes learning the connection structure from observable data. In the computer science or engineering literature, a method called probabilistic graphical model has been developed to address a general problem of connection structure of random variables. The model has been applied to learn consumer behavior from their shopping histories. Such knowledge can then be applied to recommend other shopping items. Another popular application is in medical diagnosis where the system learns about the connection structure of diseases and symptoms. With a well-trained system, patients can report symptoms and the system can compute the most likely cause. An example of scientific discovery using this method is Sachs et al. (2005) who use machine

[2]This paper refers to banks in a broad sense, including insurance companies, hedge funds and other financial institutions.

[3]The positions in this case can be thought of as cash flows from one bank to another, as a net result of their trading and lending activities. These cash flows relate the value of one bank to another, and thus also relate their credit risk, and subsequently their CDS spreads. However, as the CDS spread also accounts for the market belief (risk-neutral probability), the structure that we want to learn is not exactly the cash flow structure, but the structure of defaults implied by the market.

learning techniques to learn the causal protein signaling networks from multi-parameter single-cell data. Since the model is general in nature, we can apply this model to learn about the connection structure of financial institutions.

When defaults are concerned, credit default swaps (CDS) data will be the most appropriate source of information.[4] If the market is efficient, prices of securities will reflect all available information. During the financial crisis, stock prices and CDS spreads of many financial institutions moved together. During the European debt crisis, the spreads of sovereign credit default swaps also moved together. The market expects one firm's default will increase another firm's default probability. While data in many markets are available, CDS data tie directly to default probabilities, and thus are the most direct information source for default contagion. One potential concern is that the nonlinearity of CDS spreads may not be compatible with the graphical model, which usually requires a linear relationship between variables. Kitwiwattanachai and Pearson (2014) develop a model to link CDS spreads to the distance-to-default, that will linearize the CDS dynamics and make the data appropriate for the graphical model. In practice, one may simply use a log transformation to alleviate concerns about model misspecification, but the calculation for default probability and expected loss will still rely on the model.

How important is the spillover effect to default correlations? Glasserman and Young (2013) show that expected losses from network effects are actually small regardless of network topology. Using data from the European Banking Authority (EBA), they show that banks are more likely to default from common shocks, rather than from the spillover effect from other banks through the network. From the surface, network structure seems to matter less than we thought.

However, they mention that losses from spillover effects are small unless shocks are magnified by some other mechanism, such as bankruptcy costs and mark-to-market revaluations of assets. Borrower's deteriorating credit quality can create mark-to-market losses for a lender well before the point of default. Learning network structure

[4]It is worth clarifying that this paper is not about CDS trading positions of banks, but rather the probability of default of banks and how these probabilities are connected. In other words, we try to learn how default probabilities are connected using the CDS data as a proxy.

from CDS data will find the connection structure according to the market perception (or under the risk-neutral measure). This will account for the mark-to-market values of financial institutions and thus shed light on how the market will value other nodes when one node defaults. Such network structure can address the amplifying mechanism that will magnify the expected losses beyond simple spillover effects.

The input for the model is the CDS data for each financial institution. The data are first transformed into a log scale and then regressed on systematic risk factors, such as market returns and implied volatility (VIX), to filter out the systematic risks that affect all the nodes. The residuals are then passed through the Bayesian network model where algorithms find the network structure with the highest BIC score (likelihood score with penalty for more complex network). Parameters are then estimated from the learned network structure. The learned network structure together with parameters is then used to calculate conditional default probability and expected conditional loss. Firms on top of the network, such as Goldman Sachs, Bank of America and Morgan Stanley, can affect other firms lower down the chain, while firms on the bottom of the network, such as Lehman Brothers, AIG and UBS, have low impact on other firms in the network. The results can help answer the question whether Lehman Brothers should have been saved.

In theory, when all factors or all relevant nodes in the network are observable, and data and computational power are unlimited, the learned network structure will converge to the true network. In practice, this ideal condition is usually not the case. The factors driving CDS spreads are still not fully accounted for (Das et al. (2007) and Collin-Dufresne et al. (2001)). Which financial institutions are important for systemic risk are also debatable. The model implementation in this paper is thus demonstrative rather than comprehensive.

After the calculation of the main network structure, I also extend the model by adding a couple of omitted nodes (Bear Sterns and Merrill Lynch) back to the model. The resulting network structure is largely similar to the previous one, but also contains some significant and nontrivial changes. On the one hand, the model is relatively robust to the choice of financial institutions to include in the calculation. On the other hand, since there are significant changes to the network structure, the model implementation and interpretation still rely on the domain knowledge and judgment of the researcher.

This paper is, to the best of my knowledge, the first to apply the probabilistic graphical model to uncover network structure of financial institutions. The immediate benefit of such network structure is the ability to calculate conditional default probability and expected conditional loss, which can help gauge the systemic risk or whether any bank is too-big-to-fail. Reference books for probabilistic graphical model include Daphne and Friedman (2009) and Murphy (2012). Nagarajan et al. (2013) and Scutari and Denis (2014) include software implementation in R together with examples from biology. Sachs et al. (2005) apply the probabilistic graphical model to scientific discovery in protein signaling. Eisenberg and Noe (2001) show the impact of network connection on default contagion while Glasserman and Young (2013) show that such contagion is unlikely. Kitwiwattanachai and Pearson (2014) explore default correlations using CDS data and structural models but do not attempt to uncover the connection structure.

The recent paper examining the network structure using security prices is Billio et al. (2012). They propose econometric measures of connectedness based on principal component analysis and Granger causality. My paper instead uses a probabilistic graphical model and compares the results with Granger causality. This paper further demonstrates the calculation of conditional default probability and expected conditional loss, which is the relative advantage of using CDS data, and was absent in Billio et al. (2012) who used equity returns. This paper belongs to a strand of literature investigating the impact of default contagion or systemic risk in general.

The paper proceeds as follows. Section 1 provides a background for the probabilistic graphical model. Section 2 provides theoretical dynamics and distribution of CDS spreads. Section 3 provides empirical analysis using CDS data and shows the network structure implied from the CDS data. Section 4 explores implications of the learned network structure, in particular conditional default probabilities and expected conditional loss. Section 5 extends the model by adding a few omitted nodes to the calculation. Section 6 concludes.

1 Probabilistic Graphical Model

This section provides a brief introduction to the probabilistic graphical model. A complete treatment can be found in Koller and Friedman (2009) or Scutari and Denis (2014).

1.1 Bayesian Network Representation

A graphical model is composed of a set of random variables $\mathbf{X} = \{X_1, X_2, ..., X_p\}$ describing the quantities of interest (in this case, the CDS spreads of financial institutions). The multivariate probability distribution of $\mathbf{X}$ is called the global distribution of the data while the univariate distribution associated with each $X_i \in \mathbf{X}$ is called local distribution.

A directed acyclic graph (DAG), denoted $G = (V, A)$, is composed of nodes $v \in V$ and directed arcs $a \in A$. Each node v corresponds to one variable $X_i \in \mathbf{X}$. The arc a represents probabilistic dependency. If there is an arc from X_i to X_j, it means that X_j depends on X_i. X_i is a *parent* of X_j, and X_j is a *child* of X_i. In the context of linear regression, X_j is the dependent variable and X_i is the independent variable. Furthermore, if there is a path leading from X_i to X_j, X_i is called *ancestor* of X_j, and X_j is called *descendant* of X_i. A cycle is not allowed in DAG, i.e., there cannot be a path from any X_i back to X_i.

Figure 1 shows an example of a Bayesian network taken from Koller and Friedman (2009). The *student* example shows the dependency of variables related to student's performance. Difficulty (D) stands for the difficulty of the class. Intelligence (I) stands for student's intelligence. Grade (G) is the grade the student receives in the class. SAT (S) is the SAT score and Letter (L) stands for the quality of the letter of recommendation. The graphical model shows that grade depends on difficulty and intelligence, SAT depends on intelligence and the letter of recommendation depends on the grade.

To represent the global distribution of $\mathbf{X}$ would require a large number of parameters. Even in the simplest case of binary variables, a joint distribution requires $2^p - 1$ numbers, the probabilities of 2^p different assignments of the random variables. The

Bayesian network representation simplifies the calculation of the global distribution by decomposition. The general formulation for the decomposition of the global distribution $Pr(\mathbf{X})$ is:

$$Pr(\mathbf{X}) = \prod_{i=1}^{p} Pr(X_i|Pa_{X_i}) \tag{1}$$

where Pa_{X_i} is the set of parents of X_i. For the *student* example, the global distribution decomposes $Pr(I, D, G, S, L)$ decomposes to:

$$Pr(I, D, G, S, L) = Pr(I)Pr(D)Pr(G|I, D)Pr(S|I)Pr(L|G)$$

1.2 Structure and Parameter Learning

If the network structure and parameters are available, the Bayesian network can be used for inference, such as finding the probability of a good grade given the information about difficulty and intelligence. However, if the network structure and parameters are not available, one must first find the structure and parameters through model selection and estimation, also called *learning* from the machine learning literature.

Bayesian network learning is performed as a two-step process, structure learning and parameter learning. Structure learning finds the most likely network structure given the data, similar to maximum likelihood over all the configurations of network structures. Parameter learning then estimates the parameters of each node by maximum likelihood over the parameter space. This step concerns only with the local distribution, thus avoiding the curse of dimensionality.

To be precise, let the Bayesian network model be represented by a pair $< G, \theta_G >$, where G is the network structure and θ_G is the associated parameters. Bayesian network learning involves maximizing the likelihood score:

$$\begin{aligned} \max_{G,\theta_G} L(< G, \theta_G >: D) &= \max_{G}[\max_{\theta_G} L(< G, \theta_G >: D)] \\ &= \max_{G}[L(< G, \hat{\theta}_G >: D)] \end{aligned} \tag{2}$$

where $L(.)$ is the chosen likelihood score function and D is the given data. In other words, to find the maximum likelihood $< G, \theta_G >$ pair, we should find the graph structure G that achieves the highest likelihood, using the MLE parameters for G ($\hat{\theta}_G$).

Define:

$$score_L(G : D) = l(\hat{\theta}_G, D) \tag{3}$$

where $l(\hat{\theta}_G, D)$ is the logarithm of likelihood score function $(log(L(< G, \hat{\theta}_G >: D)))$. Bayesian network learning thus means finding G that maximizes the $score_L(G : D)$.

Similar to model selection problem, for structure learning, a major limitation of the likelihood score ($score_L(G : D)$) is that it always prefers complex structures (with more connections) over simple ones. The likelihood score always increases with an additional arc between the nodes. Thus, the maximum likelihood network will usually be a fully connected one. To avoid overfitting, Bayesian information criterion (BIC) is introduced as an alternative score function:

$$score_{BIC}(G : D) = l(\hat{\theta}_G, D) - \frac{logN}{2} Dim[G] \tag{4}$$

where N is the number of observations and $Dim[G]$ is the model dimension, or the number of independent parameters in G. The BIC score penalizes more complex models by the second term ($Dim[G]$). The BIC score has been shown to be consistent, i.e., as $N \to \infty$, the true G will have the maximum BIC score. Bayesian network learning thus means finding G that maximizes the $score_{BIC}(G : D)$, using the likelihood computed from the decomposition in (1).

1.3 Equivalence Class and CPDAG

Two DAGs can have the same score and thus belong to the same equivalence class. For example, consider two DAGs, $X_i \to X_j \to X_k$, and $X_i \leftarrow X_j \to X_k$. The likelihood from these two DAGs are the same as shown below:

$$\begin{aligned} Pr(X_i)Pr(X_j|X_i)Pr(X_k|X_j) &= Pr(X_i, X_j)Pr(X_k|X_j) \\ &= Pr(X_i|X_j)Pr(X_j)Pr(X_k|X_j) \end{aligned}$$

Bayesian network learning cannot differentiate between the DAGs in the same equivalence class because they have the same likelihood. An equivalence class can be represented by *completed partially directed graph* (CPDAG). For CPDAG, arcs that are necessary for the score and ensure no cycle will maintain directions. The remaining arcs will still be present but do not preserve directions, becoming undirected arcs. The CPDAG for the example two DAGs is $X_i - X_j - X_k$. An example of structures that must maintain arc directions is the v-structure, $X_i \rightarrow X_j \leftarrow X_k$.

1.4 Algorithms and Model Averaging

The space of network structures is not continuous. Thus, one cannot use simple optimization algorithms to obtain the highest BIC score. Moreover, there are a prohibitively large number of possible configurations of network structures, making it impossible to enumerate all possible BIC scores for a network of reasonable size. One must use a search algorithm to obtain the network structure with the highest BIC score.

There are many search algorithms in the literature but for this paper I will focus only on the hill-climbing algorithm. The algorithm explores the search space starting from a network structure (usually an empty network without any arc) and adding, deleting or reversing one arc at a time until the score can no longer be improved. As is common for any greedy algorithm, the solution obtained may be the local instead of global maximum. One can use random restart to make sure that the global maximum is obtained.

Several networks may have similar scores and one may not be comfortable choosing only the network with the highest score to represent the true structure. One can improve the quality of the structure learned by averaging over multiple CPDAGs. One possible approach is to apply bootstrap resampling of the data and learn a set of network structures, say 1000. The final average structure will contain the arcs that are present in the majority, say 85%, of the set of learned structures and in the direction that appears most frequently. The significance threshold for the arcs to be present in the final average structure can be set by the user or one can use the default threshold from the software package.

2 CDS Dynamics

Most software packages for Bayesian network allow only discrete or Gaussian random variables, and only a linear relationship among them. CDS data are not discrete (although discretization algorithm is also available to transform the continuous data into discrete). This section considers if the CDS data can fit the Gaussian distribution.

Each node in the Bayesian network is a financial institution with the corresponding CDS data. Kitwiwattanachai and Pearson (2014) provide a link between CDS dynamics and asset dynamics. I briefly discuss the setup and results here while full details can be looked up in the paper.

As is standard in structural models, the underlying asset value V follows a geometric Brownian motion. Under the risk-neutral probability,

$$d \ln V(u) = (r - \sigma^2/2)du + \sigma dW(u), \tag{5}$$

where r is the interest rate, σ is the asset return volatility, and W is a Brownian motion. Default occurs when the asset value reaches a default boundary with time u value $B(u)$ that grows at a deterministic rate equal to the expected growth rate of $\ln V$ under the risk-neutral probability, i.e.

$$B(u) = B(0)e^{(r-\sigma^2/2)u}. \tag{6}$$

Consider the distance-to-default $m(V(t), B(t)) = \ln(V(t)/B(t))/\sigma$. The default event that the asset value V reaches the boundary B is identical to the event that the distance-to-default reaches zero. Using (6) and the solution to (5), the distance-to-default

$$\begin{aligned} m(V(t), B(t)) &= \frac{\ln(V(t)/B(t))}{\sigma} \\ &= \frac{\ln V(0) + (r - \sigma^2/2)t + \sigma W(t) - \ln B(0) - (r - \sigma^2/2)t}{\sigma} \\ &= \frac{\ln(V(0)/B(0))}{\sigma} + W(t) \end{aligned} \tag{7}$$

is a Brownian motion with initial value $\frac{\ln(V(0)/B(0))}{\sigma}$. Similarly,

$$dm(t) = dW(t) \tag{8}$$

The dynamic of distance-to-default is Gaussian, which is the assumed distribution for the Bayesian network. While there are other extensions of this basic model in Kitwiwattanachai and Pearson (2014), I focus only on this basic model for the approximate distribution.

The remaining missing link is the map between CDS spreads and distance-to-default. Kitwiwattanachai and Pearson (2014) also shows that CDS spreads (S) can be written as a function of distance-to-default (m) = $S(m)$. Figure 2 shows the plot between CDS spreads and distance-to-default from the model. In principle, one can then invert the plot to find distance-to-default (m) from CDS spreads (S) and obtain the distance-to-default dynamics as an input to the probabilistic graphical model. However, one may be uncomfortable trusting the model in the first place.

To alleviate the concern of model misspecification, I use the log of CDS spreads instead. Figure 3 shows the plot between the log of CDS spreads and the distance-to-default. The plot is almost linear with some nonlinearities when CDS spreads are extremely high (distance-to-default is extremely low) on the left end. The linear relationship motivates the assumption that the dynamic of the log of CDS spreads is Gaussian, because the dynamic of the distance-to-default itself is Gaussian (see equation (8)). Thus, for the purposes of Bayesian network learning, I use log of CDS spreads as an input to the model. We will investigate the normality assumption of the log of CDS spreads again in the empirical analysis in section 3.2.3.

For the purposes of inference, such as computing conditional default probability and expected conditional loss, however, one needs the implied distance-to-default from the model. Thus, after learning the network structure and parameters, I return to the model to calculate distance-to-default and default probabilities.

3 Empirical Analysis and Implied Network Structure

3.1 Data

CDS data are from Markit starting from January 2001. However, not all firms have CDS traded from the first day of 2001. Some firms' CDS start being traded much later. I use only 5-year CDS for senior unsecured debt because they are the most prevalent and the most liquid. The data are available daily. However, the daily quotes tend to be sticky because CDS may not be traded every day and the Markit filtering algorithm among dealers' quotes may smooth out any small daily changes. Monthly data are preferred but the data may not be sufficient for statistical analysis. Thus, I use weekly data for all the analyses.

The firms are selected from the list of G-SIFIs according to Financial Stability Board (FSB) document on 4 November 2011. SIFIs are financial institutions whose distress or disorderly failure would cause significant disruption to the wider financial system and economic activity. Some of the firms in the list do not have CDS until as late as 2008 and some banks' CDS data have duplicates. I drop banks with insufficient data and duplicates and non-US banks that are unlikely to be the source of systemic risk in the last financial crisis. I also include Lehman Brothers back to the list because the objective of the paper is to study the network structure leading up to the financial crisis, and to answer questions such as "should Lehman Brothers have been saved?"

The final data set includes 11 financial institutions with CDS data from June 2001 to August 2008 (because Lehman Brothers declared bankruptcy by September 2008). Table 1 shows the name and market capitalization of the banks in the data set. The list includes mostly big US banks. The market capitalization is calculated as of January 2007 to avoid the impact of financial crisis. Table 2 shows summary statistics of CDS data during the sample period. The statistics are calculated both on raw spreads in basis points and log of spreads. All of the banks have relatively low CDS spreads and thus low default probabilities before the financial crisis. The mean and standard deviation of CDS spreads are also not very different across banks.

Figure 4 shows an example of time series of CDS spreads for 3 banks, Citigroup, Goldman Sachs and Lehman Brothers. The spreads are highly correlated throughout the entire period. The financial crisis significantly affects the banks' default probabilities, as CDS spreads significantly increase at the onset of the crisis around 2007 and 2008.

3.2 Empirical Analysis

Since the main focus is on the connection between financial institutions, I first filter out the systematic factors that will affect all CDS spreads. The risk factors that affect CDS spreads include market returns and market volatility (see, for example, Ericsson et al. (2009)). I run regressions of log of CDS spreads on S&P 500 index returns and VIX to remove the systematic factors and only use the residuals for the Bayesian network analysis. In particular, I run the regression:

$$\Delta log S_{i,t} = \alpha_i + \beta_m R_{m,t} + \beta_v \Delta VIX_t + \epsilon_{i,t} \tag{9}$$

where Δ denotes weekly changes, $R_{m,t}$ is the S&P500 weekly returns at time t, and VIX is the CBOE implied volatility. The residuals $\epsilon_{i,t}$ in theory should be normally distributed and will be used for the Bayesian network learning.

3.2.1 Comparison with Granger Causality

As mentioned before, a simple correlation matrix or linear regression cannot establish the network structure because the direction of causality can go either way. A typical method to establish causality in economics is Granger causality or Vector Autoregression (VAR). By running regression of one variable on the lagged value of other variables, causality can be determined because only the past can determine the future, but not vice versa. In this section I run VAR with the CDS data to see if the network structure can be found this way.

I run VAR(1) regression using the residuals obtained from (9). In this setup, VAR is equivalent to running separate linear regressions. For example, for AIG:

$$AIG_t = \alpha + \beta_1 AIG_{t-1} + \beta_2 BAC_{t-1} + \beta_3 BARCLAYS_{t-1} + ... + \beta_{11} WFC_{t-1} + \epsilon_t$$

The same regression can be run for CDS data for all firms by changing the left-hand side. The significance of the coefficient can establish the direction of causality from the regressor to the dependent variable on the left hand side.

Table 3 shows an example of VAR results for AIG, GS and LEH in columns 1, 2, and 3 accordingly. The first column suggests that CITI, LEH and UBS can cause AIG's spreads to move. However, CITI and UBS are marginally significant at 10% significance level. Moreover, even though LEH is highly significant, the sign is negative, indicating that an increase in LEH's CDS will reduce AIG's CDS. This result goes against the intuition of systemic risk that one firm's default risk will increase another firm's default risk.

The second column suggests that CITI, LEH and WFC can affect GS's default risk. However, LEH and WFC have the negative sign. GS is also negatively autocorrelated, suggesting a mean reversion process, which does not exist in AIG. The third column suggests that AIG can affect LEH's default risk with a positive sign. This result creates a problem because AIG can Granger-cause LEH, while LEH can also Granger-cause AIG. The final direction is undetermined.

VAR results are not convincing because of the negative coefficients instead of positive and undetermined direction of causality. CDS spread changes may not be appropriate for Granger causality test because changes in prices may mostly be contemporaneous, if the market is efficient.

3.2.2 Software Packages for Bayesian Network

The theory and algorithms for probabilistic graphical model are well developed. There are many software packages that implement algorithms to construct and manipulate the Bayesian network model with different focuses and functionalities. This paper uses the R package **bnlearn**, which implements discrete and Gaussian Bayesian network together with structure learning, parameter learning and inference algorithms. The package has been used mostly in the biological research area.

To visualize the Bayesian network, I use another R package **Rgraphviz** which can interact with **bnlearn**. The Bayesian network graph generated for this paper is the product of these software packages. See Appendix for examples of specific functions used to generate results.

3.2.3 Node Distribution

Figure 5 shows the distribution of each node in the Bayesian network. The first 11 plots show the distribution of CDS of the banks in the network, while the last plot shows the distribution of S&P 500 index returns for comparison. All distributions are weekly and are the residuals of the systematic risk factors as shown in (9). The distributions are not too far from, but also not exactly, the normal distribution. For example, there are some outliers for LEH, GS and MS with realizations extremely far away from the mean. The extreme outliers appear near the end of the sample period, at the onset of the financial crisis, which explains the sharp increase in the CDS spreads. Most distributions appear to have fat tails.

Table 4 shows summary statistics of the distribution of each node in the network, together with the distribution of S&P 500 for comparison. The mean is always 0 by construction (because of residuals) and thus not reported. The standard deviations for all firms are similar at around 0.1 to 0.12 with JPM and LEH higher than others. Most have slight positive skewness and high kurtosis (fat tails). LEH has the highest positive skewness and kurtosis, as the firm's CDS significantly increases near default. S&P 500 has lower standard deviation, skewness and kurtosis, and thus is closer to a normal distribution.

While the distributions of the nodes are not perfectly Gaussian, the maximum likelihood algorithm (in this case, maximizing the BIC score) should still find the most likely network structure. Such exercise is similar to the argument that Quasi Maximum Likelihood Estimation (QMLE) can be used for nonnormal distribution if only the mean and standard deviation are concerned. The resulting estimates are consistent but not efficient, and thus the resulting network structure will also be consistent. Further development can extend the software library to accept distributions other than Gaussian and define the score for nonnormal distributions to maximize accordingly.

The main concern will be for inference when computing the conditional default probability, because the probability will depend on the assumed distribution. This paper maintains the Gaussian assumption, but Bayesian network simulation with non-normal distribution is also available in R package such as **rjags** or the general package **JAGS** (Just Another Gibbs Sampler).

3.2.4 Learned Bayesian Network

Each node in the Bayesian network is normally distributed. The data for each node are obtained from the residuals in (9). The log transformation and (8) ensure linearity and normal distribution. With the specified input, the software package uses hill-climbing algorithm to find the network structure with the highest BIC score. To ensure robustness, model averaging using bootstrap samples is performed and the final result is the network structure containing arcs that are present in the majority of the learned structures, in the direction that appears most frequently.

The learned Bayesian network is shown in Figure 6. The corresponding CPDAG is shown in Figure 7. The learned parameters are reported in Table 5. The R code to generate these results is in the Appendix.

Figure 6 shows the relationship between financial institutions as implied by the CDS data. For example, GS can affect WFC, WFC can affect LEH, and LEH can affect AIG. GS can also affect DB which then affects AIG. The CPDAG in Figure 7 is similar to DAG in Figure 6 but some arcs, such as $GS \rightarrow MS$ and $GS \rightarrow BAC$, lose direction. This indicates that the direction can go either way, and the BIC score will be the same. For this network, the learning algorithm cannot differentiate between $GS \rightarrow MS$ and $MS \rightarrow GS$, or $GS \rightarrow BAC$ and $BAC \rightarrow GS$.

Table 5 shows the learned parameters for the Bayesian network. The table corresponds exactly to Figure 6 with the coefficients indicating the strength of the relationship from parents to children. For example, AIG has 3 parents: BARC, DB and LEH with coefficients 0.162, 0.218 and 0.228, respectively. Since the Bayesian network model follows Gaussian distribution, these estimated coefficients are equivalent to ones obtained from a linear regression (MLE and linear regressions yield the same estimates in this setting), with familiar interpretation. For example, one unit of increase in (log of) CDS of BARC can increase 0.162 unit of (log of) CDS of AIG. The impact is 0.218 for DB and 0.228 for LEH respectively. These coefficients are important for the computation of conditional default probability, which will be discussed in the next section.

Some interesting patterns emerge from the network structure. At the top of the network are GS, MS, BAC, and CITI, the familiar big and powerful banks that can influence almost all other firms in the system. JPM is not related to other banks, perhaps indicating their superior risk management practice that shields themselves from the volatile environment. In fact, during the crisis, JPM did survive almost intact. At the bottom of the network are AIG and UBS, which are influenced by other firms, either directly or indirectly through the chain. Other firms in the middle receive influences from the top and pass down influences to the bottom.

Should Lehman Brothers have been saved? From the network structure, LEH is near the bottom and thus would not exert influence that could collapse the entire financial system. LEH can still impact AIG, which is indeed what we witnessed during the crisis. After Lehman Brothers collapsed, the next firm in line was AIG, which may have gone bankrupt if the government had not stepped in to help. While defaults are always costly, LEH's default was one of the least systemic defaults possible among the selected firms.

Simply looking at the network structure can inform policy makers for their decision, but more precise answers can be delivered from such a network model. Next section discusses interesting queries that can be answered by the model.

4 Implications

Once the Bayesian network model and parameters have been learned, one can answer probabilistic questions using the model. In the context of systemic risk, we are concerned with conditional default probability and expected conditional loss. For example, if GS defaults, what is the probability that AIG will also default, and what is the expected loss for the entire system?

Such questions can be answered directly by the Bayesian network model. Defaults occur when the asset value is below the default barrier, or when the distance-to-default (m) is below 0. Given the CDS spreads, I can find the corresponding distance-to-default from the model that links CDS spreads to distance-to-default, i.e., by inverting Figure 2. The calculation in this section thus depends on the structural model in Kitwiwattanachai and Pearson (2014) that relates CDS spreads to distance-to-default. One can also use an alternative model if needed while maintaining the same network structure learned previously.

Table 6 shows the distance-to-default of all firms in the network as of 31 July 2008. This date is just before the financial crisis in August and September 2008. The CDS spreads start to widen especially for LEH, displaying signs of trouble. The corresponding distance-to-default (m's) are centered at around 3.5 but can go as low as 2.53 for LEH or as high as 4.20 for DB.

With the distance-to-default one can calculate the corresponding default probabilities as the probability that m falls below 0 within a certain period, typically 1 year. The **bnlearn** package also provides a simulation capability for such computation. To calculate 1-year probability, the variance needs to be adjusted to be the variance for m, not for the log of CDS spreads. With the guidance from (8), I readjust the variance to 1 for all nodes. As mentioned earlier, the default probability will be calculated assuming normal distribution.

The corresponding default probabilities are reported in the last column of Table 6 and the code is available in the Appendix. In general firms with low CDS spreads will also have low default probabilities. Most firms have very low default probabilities below 1 %, except for LEH, AIG and MS, with LEH leading at almost 5%. Network structure also plays a role in default probabilities - firms lower in the network chain can receive uncertainty from firms on top of the network, and thus have higher default probability even with similar distance-to-default (for example, GS and BARCLAYS)

4.1 Conditional Default Probability

The interesting question is "what will happen if firm X defaults?" The model can simulate conditional default probability by setting the distance-to-default of the firm of interest to be below 0, and examine the probability that other firms' distance-to-default also fall below 0. The horizon of interest is 1 year.

Table 7 shows the results for conditional default probabilities. Panel A shows the probability calculated from the network structure. Since we are interested in the additional default probability for other firms if the firm of interest defaults, Panel B shows the difference between the conditional default probability and the original default probability in Table 6. The left column is the firm of interest. The *Influence* column shows firms that are influenced by the firm on the left, the children or descendants. For example, CITI directly influences BARCLAYS, UBS, and WFC. BARCLAYS influences DB, UBS and AIG. WFC influences LEH and AIG. Thus, CITI can influence BARCLAYS, UBS, WFC, DB, AIG and LEH. Panel A shows the default probabilities for these firms, if CITI defaults in the next year.

Default probabilities of children and descendants increase substantially if the parents default. From unconditional probability below 1%, the conditional default probabilities can increase to 10% or 20%, and in some cases over 60% if the link is strong. GS's default can affect almost all other nodes in the network because GS sits on top of the network, with the strongest impact on LEH and MS. LEH's default, on the other hand, only increases default probability of AIG to 8.88%, a substantial amount but relatively low compared to other nodes in the network. On the other hand, LEH is affected tremendously by defaults of its ancestors, BAC, CITI, GS, MS and WFC, each driving up LEH's default probability up to around 30%, and in some cases over 60%.

4.2 Expected Conditional Loss

The Bayesian network model allows us to calculate conditional default probability. The next immediate question is the expected conditional loss from the default, i.e., ”what is the expected loss to other firms if firm X defaults?” With the available conditional default probability, finding expected loss is relatively straightforward. I multiply conditional default probability in Panel B of Table 7 with the corresponding market capitalization from Table 1. I assume that all the equity will be wiped out in case of default. Additional loss to debtholders many incur but is not considered here. The result is the expected *incremental* loss, which is purely from the systemic risk as a result of other firm's default, in addition to the firm's own expected loss from systematic factors and its own activity.

Table 8 reports the results for expected loss. The cumulative loss is the sum of all the expected losses conditional on the node's default. Ranking can also be assigned based on the cumulative loss. As expected, the nodes on top of the network rank high on the expected loss. GS, BAC and MS are systemically more important than UBS, AIG and JPM. While GS seems extremely important with cumulative loss near $200 billion, one must keep in mind that this is partly due to the assumed arc direction $GS \rightarrow MS$ and $GS \rightarrow BAC$, which is the most frequent direction obtained from the learning algorithm. However, as CPDAG in Figure 7 shows, the arc direction can point to the opposite direction without changing the likelihood of the network. The opposite arc direction will also change the expected loss attributed to GS, BAC and MS.

Should Lehman Brothers have been saved? Since the expected loss is only $13.25 billion (excluding its own loss), the answer is probably no.

5 Extensions

So far we have learned the network structure of firms that survived the financial crisis, except for Lehman Brothers which has been included to answer the hypothetical question. The purpose of the exercise is demonstrative rather than comprehensive. Many firms are omitted for the advantage of interpretation.[5] In this section I add a few more firms to the Bayesian network calculation. The firms of interest are Bear Sterns and Merrill Lynch, which were important firms before the financial crisis, but were acquired by other firms rather than facing a default as Lehman Brothers.

We will learn that the network structure with additional firms maintains the basic structure, but also exhibits differences from the network learned in the previous section. Thus, the network structure learning exercise is not a simple mechanical process, but

[5]For example, small firms or international firms are omitted, because intuitively they should not play an important role in systemic risk. This will help the algorithm to learn the network structure more efficiently, and the resulting network structure is easier to interpret.

requires deep domain knowledge of the user to select relevant firms and interpret the results.

5.1 Adding Bear Sterns

Bear Sterns was a prominent investment bank and securities trading and brokerage firm that failed in the 2008 financial crisis, and was sold to JPMorgan Chase. The CDS data for Bear Sterns are available from January 2001 to July 2008. As the network model requires complete data for all nodes, the CDS data used for this section end on 7 July 2008 for all firms, and thus are somewhat shorter than the data used in the analysis so far which extend to August 2008.[6] The symbol for Bear Sterns is BSC.

Figure 8 shows the learned Bayesian network with Bear Sterns, and Figure 9 the corresponding CPDAG. The network structure is similar to the previous one with GS, MS and BAC at the top, AIG, DB and UBS at the bottom, and JPM isolated from other nodes.[7] BSC stands relatively high in the network, just below GS and MS, and can influence CITI, LEH, WFC and BARCLAYS. LEH is now connected to BARCLAYS, and through BARCLAYS can influence DB, AIG and UBS. Thus, under this network structure, LEH's default will have more impact to the financial system. Interestingly, BSC is connected directly to LEH, indicating that when BSC is in trouble, the next troubling firm will be LEH, which is what we observed in 2008.

Overall, the main network and connection structure remain similar to before, but there are also a few significant changes.

5.2 Adding Merrill Lynch

Merill Lynch was another prominent financial institution that did not make it through the financial crisis and was acquired by the Bank of America in September 2008. The CDS data for Merrill Lynch are available from January 2001 to September 2013.[8]

[6] The shorter data may partially affect the results, as the CDS movement in August 2008 was much more volatile than other periods. Excluding the data in this period may affect the likelihood of the network.

[7] Note that this is before Bear Sterns is acquired by JP Morgan Chase.

[8] Although Merrill Lynch agreed to be acquired by the Bank of America in September 2008, the firm is not completely merged with the Bank of America until October 2013.

However, as the dataset will include Bear Sterns, only the data up until 7 July 2008 are used. The symbol for Merrill Lynch is MER.

Figure 10 shows the learned Bayesian network with Bear Sterns, and Figure 11 the corresponding CPDAG. The network structure is similar to before. MER stands relatively high in the network, just below GS and MS but above BAC. Curiously, this connection between MER and BAC may also hint at the reason why MER was acquired by BAC, and not other firms. LEH can still influence BARCLAYS, which then influences DB, AIG and UBS. Curiously, this connection between LEH and BARCLAYS may also hint at the reason why BARCLAYS purchased LEH's North American investment-banking and trading divisions (along with the headquarter building), after bankruptcy, and not other firms.[9]

Interestingly, both BSC and MER can influence LEH. In fact, LEH is the only firm directly impacted by both BSC and MER. Thus, when BSC and MER are in trouble, the next troubling firm will be LEH. This is also what we observed during the financial crisis. BSC and MER, however, were acquired by JPM and BAC, while LEH could not find an acquirer and eventually had to declare bankruptcy.

Overall, the learned network structure is relatively robust to additional nodes. However, there are a few significant and nontrivial changes. Nevertheless, the network structure reflects the relationship among firms, similar to what we observed during the financial crisis.

The Bayesian network is useful to learn about the connection structure of firms, and gauge the impact of a firm's default to the rest of the network. In a perfect case, when all the factors (or relevant nodes) are known and included in the network, with infinite amount of data and computational power, the learned network structure will be the true structure. In the real world, however, perhaps not all factors that drive CDS spreads are known or observed (Das et al. (2007) and Collin-Dufresne et al. (2001)), and it is not always clear which nodes are relevant. How to select the relevant factors and nodes to include in the Bayesian network learning exercise, and how to interpret

[9] On the contrary, BSC was not acquired by its children (BARCLAYS, CITI, LEH or WFC), but by JPM, which does not seem connected in the network structure. Thus, while the connection structure can "curiously" predict the next event, it is not the rule.

the results, still depend on the knowledge and judgment of the researcher.

6 Conclusion

This paper applies the probabilistic graphical model to learn the network structure of financial institutions from CDS data. The learned structure and parameters can be used to compute conditional default probability and expected conditional loss. The impact of default of one firm can be gauged, allowing policy makers to make decisions on the too-big-to-fail problem. Since the information is extracted from the CDS data, all quantities are under the risk-neutral measure, and thus account for the market fear that can trigger bank run.

The log transformation of the CDS data is sufficient to linearize the CDS dynamics, yielding normal distribution for each node in the network. The algorithm to learn the network structure searches for the network with the highest BIC score. Model averaging is performed for robustness. The resulting network structure shows the highly influential firms such as GS, BAC and MS on top of the network. At the bottom of the network are less influential firms such as LEH, AIG and UBS. The corresponding conditional default probability and expected conditional loss show that the decision to let Lehman Brothers default was probably the right one, because it was not too-big-to-fail.

The learned network structure is relatively robust to the choice of financial institutions. The network structure largely complies with the event observed during the financial crisis. However, as the factors driving CDS spreads included in the model are not comprehensive, researchers must use their own judgment on how to implement the model and interpret the results.

Future research may include refining the distributions of each node in the network and extend the library to include nonnormal distributions. This will improve the estimates for conditional default probability and expected conditional loss. Future research on systemic risk can be built on such platform using the learned network structure as a starting point. Finally, the probabilistic graphical model is a generic mathematical and computational tool which has been successfully applied in the domain of computer science and biology. Finance can also benefit from this well-developed tool, because nowadays everything is connected and network is becoming increasingly important.

Appendix: R Code

```
library(bnlearn)
library(Rgraphviz)

load(Data)   #CDS data (log residauls)
bootCDS =boot.strength(data=Data, R = 1000, algorithm="hc", algorithm.args=list(score="bic-g"))
  #learn network structure, hc algorithm, BIC score
avg.bootCDS = averaged.network(bootCDS)     #model averaging

graphviz.plot(avg.bootCDS)                  #plot the learned structure
graphviz.plot(cpdag(avg.bootResidEVix))     #plot cpdag

fitbn = bn.fit(avg.bootCDS, data=Data)      #estimate parameters

DDList = list()          #set distance-to-default for each node
DDList$AIG = 2.87
DDList$BAC = 3.75
DDList$BARCLAYS = 3.60
DDList$CITI = 3.47
DDList$DB = 4.20
DDList$GS = 3.57
DDList$JPM = 3.85
DDList$LEH = 2.53
DDList$MS = 2.91
DDList$UBS = 3.77
DDList$WFC = 3.78
#-------------------------------------------------------
#modify standard deviation to 1

modfitbn = fitbn
disAIG = list(coef = fitbn$AIG$coefficients, sd = 1)
disBAC = list(coef = fitbn$BAC$coefficients, sd = 1)
disBARCLAYS = list(coef = fitbn$BARCLAYS$coefficients, sd = 1)
disCITI = list(coef = fitbn$CITI$coefficients, sd = 1)
disDB = list(coef = fitbn$DB$coefficients, sd = 1)
disGS = list(coef = fitbn$GS$coefficients, sd = 1)
disJPM = list(coef = fitbn$JPM$coefficients, sd = 1)
disLEH = list(coef = fitbn$LEH$coefficients, sd = 1)
disMS = list(coef = fitbn$MS$coefficients, sd = 1)
disUBS = list(coef = fitbn$UBS$coefficients, sd = 1)
disWFC = list(coef = fitbn$WFC$coefficients, sd = 1)

modfitbn$AIG = disAIG
modfitbn$BAC = disBAC
modfitbn$BARCLAYS = disBARCLAYS
modfitbn$CITI = disCITI
modfitbn$DB = disDB
modfitbn$GS = disGS
modfitbn$JPM = disJPM
modfitbn$LEH = disLEH
modfitbn$MS = disMS
modfitbn$UBS = disUBS
modfitbn$WFC = disWFC
#------------------------------------------------
#calculating unconditional default probability
cpquery(modfitbn , event = (AIG < -(DDList$AIG)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (BAC < -(DDList$BAC)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (BARCLAYS < -(DDList$BARCLAYS)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (CITI < -(DDList$CITI)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (DB < -(DDList$DB)) , evidence = TRUE, n = 10^8)
cpquery(modfitbn , event = (GS < -(DDList$GS)) , evidence = TRUE, n = 10^8)
cpquery(modfitbn , event = (JPM < -(DDList$JPM)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (LEH < -(DDList$LEH)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (MS < -(DDList$MS)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (UBS < -(DDList$UBS)) , evidence = TRUE, n = 10^7)
cpquery(modfitbn , event = (WFC < -(DDList$WFC)) , evidence = TRUE, n = 10^7)

#Conditional default probability (example)
  #AIG default probability , if LEH defaults
cpquery(modfitbn , event = (AIG < -(DDList$AIG)) , evidence = (LEH < -(DDList$LEH)), n = 4*10^7)
  #LEH default probability , if GS defaults
cpquery(modfitbn , event = (LEH < -(DDList$LEH)) , evidence = (GS < -(DDList$GS)), n = 4*10^7)
```

References

[1] bnlearn R package. http://www.bnlearn.com/. Accessed: 2014-09-21.

[2] Monica Billio, Mila Getmansky, Andrew W Lo, and Loriana Pelizzon. Econometric measures of connectedness and systemic risk in the finance and insurance sectors. *Journal of Financial Economics*, 104(3):535–559, 2012.

[3] Financial Stability Board. Policy measures to address systemically important financial institutions. *On-line paper: http://www. financialstabilityboard. org/publications/r_111104bb. pdf*, 2011.

[4] Tim Bollerslev and Jeffrey M Wooldridge. Quasi-maximum likelihood estimation and inference in dynamic models with time-varying covariances. *Econometric reviews*, 11(2):143–172, 1992.

[5] Pierre Collin-Dufresne, Robert S Goldstein, and J Spencer Martin. The determinants of credit spread changes. *The Journal of Finance*, 56(6):2177–2207, 2001.

[6] Sanjiv R Das, Darrell Duffie, Nikunj Kapadia, and Leandro Saita. Common failings: How corporate defaults are correlated. *The Journal of Finance*, 62(1):93–117, 2007.

[7] Larry Eisenberg and Thomas H Noe. Systemic risk in financial systems. *Management Science*, 47(2):236–249, 2001.

[8] Jan Ericsson, Kris Jacobs, and Rodolfo Oviedo. The determinants of credit default swap premia. *Journal of Financial and Quantitative Analysis*, 44(01):109–132, 2009.

[9] Jeff Gentry, Li Long, Robert Gentleman, Seth Falcon, Florian Hahne, Deepayan Sarkar, Kasper Daniel Hansen, PDF R Script A New Interface, PDF R Script How To Plot, A Graph Using Rgraphviz, et al. Provides plotting capabilities for r graph objects. *R package version*, 1(1), 2009.

[10] Paul Glasserman and H Peyton Young. How likely is contagion in financial networks? *Journal of Banking & Finance*, 2014.

[11] Chanatip Kitwiwattanachai and Neil D. Pearson. Inferring asset correlations from cds spreads: A structural model approach. *Review of Asset Pricing Studies, forthcoming*, 2014. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2138448.

[12] Daphne Koller and Nir Friedman. *Probabilistic graphical models: principles and techniques*. MIT press, 2009.

[13] Kevin P Murphy. *Machine learning: a probabilistic perspective*. MIT press, 2012.

[14] Radhakrishnan Nagarajan, Marco Scutari, and Sophie Lèbre. *Bayesian Networks in R*. Springer, 2013.

[15] Karen Sachs, Omar Perez, Dana Pe'er, Douglas A Lauffenburger, and Garry P Nolan. Causal protein-signaling networks derived from multiparameter single-cell data. *Science*, 308(5721):523–529, 2005.

[16] M Scutari and JB Denis. Bayesian networks: with examples in R. 2014.

Table 1: List of financial institutions considered in the Bayesian network model. The market capitalization is calculated as of January 2007 (before the financial crisis).

Name	Symbol	Market Capitalization (Billions USD)
American International Group	AIG	177.95
Bank of America	BAC	234.41
Barclays	BARCLAYS	3.35
Citigroup	CITI	270.89
Deutsche Bank	DB	70.40
Goldman Sachs	GS	90.34
JPMorgan Chase	JPM	176.30
Lehman Brothers	LEH	43.59
Morgan Stanley	MS	86.84
UBS	UBS	122.29
Wells Fargo	WFC	121.26

Table 2: Summary statistics for CDS spreads, from January 2001 (or the first available date of CDS data) to August 2008. The statistics are calculated both on raw spreads in basis points and log of spreads.

	raw spreads			log spreads		
Firm	mean	median	std	mean	median	std
AIG	45.89	28.25	53.29	3.49	3.34	0.72
BAC	37.77	29.50	25.52	3.45	3.38	0.58
BARCLAYS	24.22	13.60	28.61	2.81	2.61	0.76
CITI	37.10	27.02	35.96	3.32	3.30	0.72
DB	26.93	19.25	20.74	3.10	2.96	0.56
GS	51.33	41.38	31.36	3.80	3.72	0.51
JPM	43.90	35.55	26.85	3.62	3.57	0.56
LEH	72.06	44.80	68.33	4.00	3.80	0.68
MS	56.09	40.00	45.88	3.82	3.69	0.58
UBS	22.34	12.80	29.19	2.69	2.55	0.79
WFC	32.12	28.37	24.83	3.25	3.35	0.66

Table 3: VAR(1) example for AIG, GS and LEH. Changes in (log) CDS spreads are regressed on lag of changes in (log) CDS spreads. Standard errors are ported in parentheses. *, ** and *** correspond to significance at 10%, 5 % and 1 % level

Lagged Value	AIG	GS	LEH
AIG	-0.059	0.056	0.132*
	(0.069)	(0.058)	(0.069)
BAC	0.115	0.059	0.047
	(0.091)	(0.077)	(0.092)
BARCLAYS	0.023	-0.068	-0.102
	(0.091)	(0.077)	(0.091)
CITI	0.196*	0.162*	0.049
	(0.105)	(0.089)	(0.106)
DB	-0.006	-0.038	0.046
	(0.088)	(0.074)	(0.089)
GS	-0.105	−0.250**	0.102
	(0.129)	(0.109)	(0.130)
JPM	0.035	-0.011	0.018
	(0.040)	(0.034)	(0.040)
LEH	−0.253***	−0.162**	−0.607***
	(0.092)	(0.077)	(0.092)
MS	0.162	0.153	0.127
	(0.142)	(0.119)	(0.142)
UBS	0.149*	0.146**	0.102
	(0.082)	(0.069)	(0.083)
WFC	0.018	−0.144*	0.013
	(0.102)	(0.086)	(0.103)
N	379	379	379
Adj. R^2	0.051	0.047	0.122

Table 4: Summary statistics for CDS spread residuals from systematic risks as shown in (9). The statistics show standard deviation (std), skewness and excess kurtosis. S&P 500 is also shown for comparison.

Firm	std	skewness	excess kurtosis
AIG	0.126	1.566	9.846
BAC	0.107	1.076	6.689
BARCLAYS	0.109	0.109	6.282
CITI	0.102	1.016	5.575
DB	0.114	1.145	10.219
GS	0.106	1.006	7.260
JPM	0.173	-0.283	11.730
LEH	0.132	2.598	24.138
MS	0.103	0.979	4.829
UBS	0.111	0.464	5.692
WFC	0.097	0.448	3.126
S&P 500	0.025	-0.199	1.286

Table 5: Learned parameters for the Bayesian network. Parents for each node are shown with corresponding coefficients for arc strengths.

	Parents										
Firm	AIG	BAC	BARC	CITI	DB	GS	JPM	LEH	MS	UBS	WFC
AIG			0.162		0.218			0.228			0.274
BAC						0.629					
BARCLAYS				0.589							
CITI		0.437				0.076			0.315		
DB			0.521			0.344					
GS											
JPM											
LEH									0.801		0.210
MS						0.839					
UBS			0.311	0.203	0.329						
WFC		0.254		0.318		0.212					

Table 6: Distance-to-default (m) implied from CDS spreads and the corresponding probability of default (PD) as of 31 July 2008.

Firm	CDS Spread	m	PD (%)
AIG	257.71	2.87	1.433
BAC	113.78	3.75	0.746
BARCLAYS	130.82	3.60	0.209
CITI	149.22	3.47	0.362
DB	72.04	4.20	0.059
GS	135.30	3.57	0.018
JPM	102.76	3.85	0.006
LEH	346.52	2.53	4.959
MS	247.96	2.91	1.286
UBS	111.01	3.77	0.257
WFC	110.43	3.78	0.134

Table 7: Conditional default probability. Panel A shows conditional default probability calculated from the learned network structure. Panel B shows the increase in default probability from the original default probability in Table 6.

Panel A: Conditional Probability of Default

	Influence (% PD)										
Firm	AIG	BAC	BARC	CITI	DB	GS	JPM	LEH	MS	UBS	WFC
AIG											
BAC	13.23		3.54	16.68	1.15			28.48		3.35	7.38
BARCLAYS	18.56				4.60					13.14	
CITI	17.19		9.88		1.58			34.21		7.18	5.31
DB	25.20									17.03	
GS	22.50	9.30	3.27	16.39	3.41			64.25	61.52	3.91	7.49
JPM											
LEH	8.88										
MS	10.81		1.45	5.47	0.55			64.25		1.63	1.34
UBS											
WFC	25.17							36.87			

Panel B: Increase in Probability of Default

	Influence (% PD)										
Firm	AIG	BAC	BARC	CITI	DB	GS	JPM	LEH	MS	UBS	WFC
AIG											
BAC	11.80		3.33	16.32	1.09			23.52		3.09	7.25
BARCLAYS	17.13				4.54					12.88	
CITI	15.76		9.67		1.52			29.25		6.92	5.18
DB	23.77									16.77	
GS	21.07	8.55	3.06	16.03	3.35			59.29	60.23	3.65	7.36
JPM											
LEH	7.45										
MS	9.38		1.24	5.11	0.49			59.29		1.37	1.21
UBS											
WFC	23.74							31.91			

Table 8: Expected conditional loss. The table shows expected loss conditional on the default of the firm in the left column, calculated from the product of increase in default probability in Panel B of Table 7 and market capitalization in Table 1. *Cumulative* is the sum of all expected losses. *Ranking* shows ranking by *Cumulative*.

	Expected Loss (Billions USD)												
Firm	AIG	BAC	BARC	CITI	DB	GS	JPM	LEH	MS	UBS	WFC	Cumulative	Ranking
AIG												0	9
BAC	20.99		0.11	44.20	0.77			10.25		3.78	8.79	88.89	2
BARCLAYS	30.48				3.20					15.74		49.42	7
CITI	28.04		0.32		1.07			12.75		8.46	6.28	56.92	5
DB	42.29									20.49		62.79	3
GS	37.49	20.05	0.10	43.42	2.36			25.84	52.31	4.46	8.92	194.96	1
JPM												0	9
LEH	13.25											13.25	8
MS	16.69		0.04	13.84	0.35			25.84		1.68	1.46	59.90	4
UBS												0	9
WFC	42.24							13.91				56.15	6

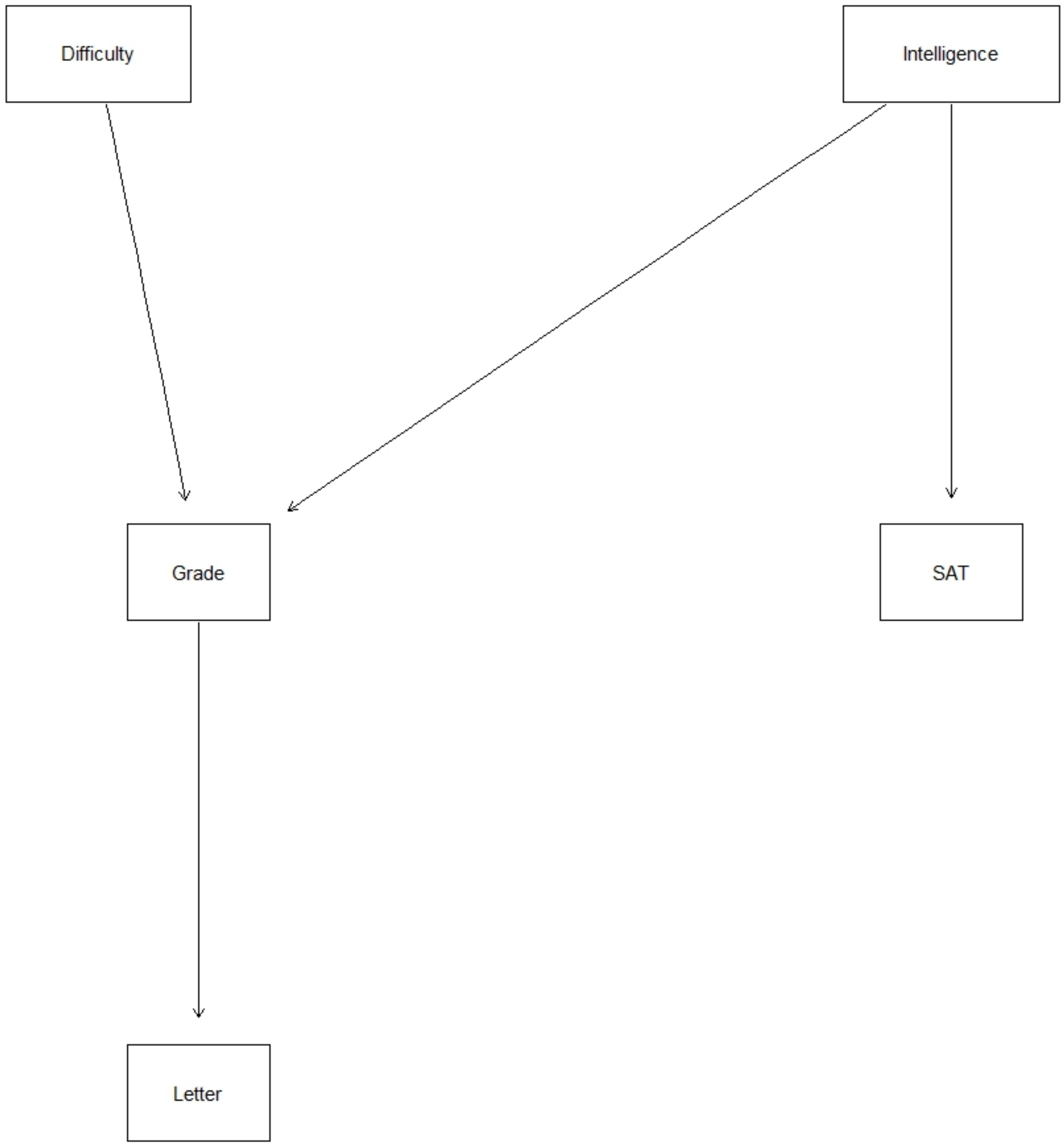

Figure 1: Example of a Bayesian network representing a *student*

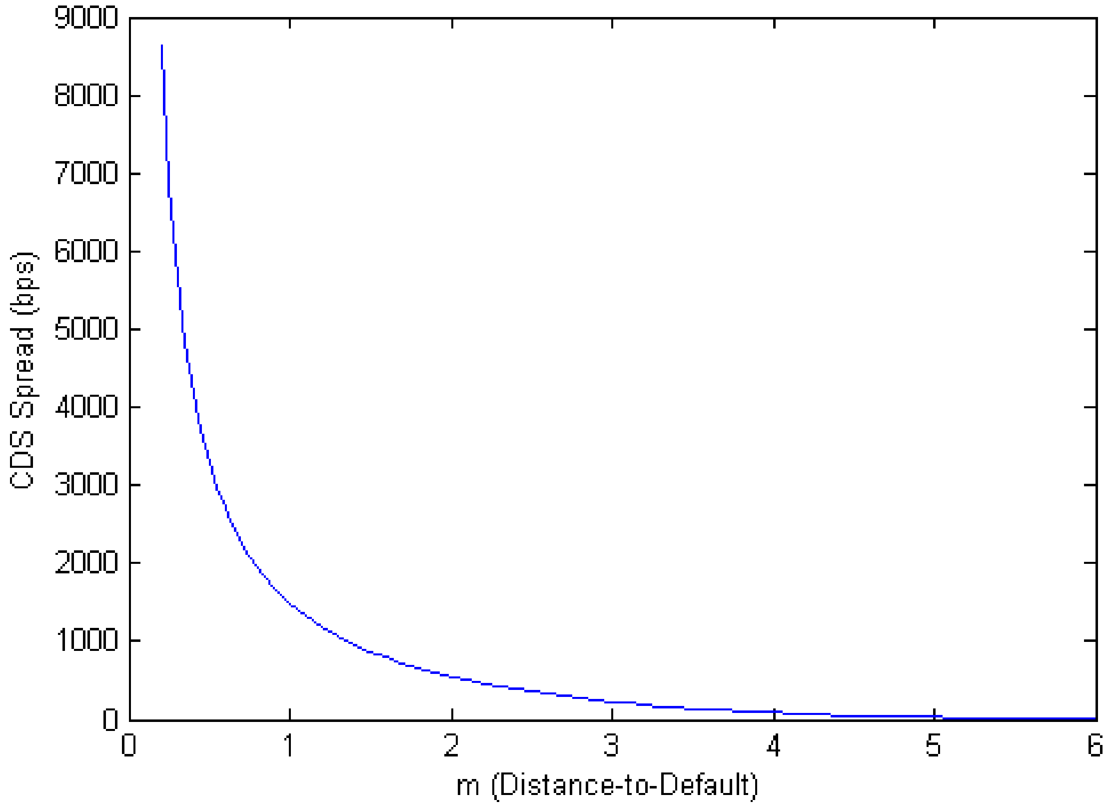

Figure 2: CDS spreads as a function of distance-to-default. The figure displays the CDS spread computed with time to maturity $T = 5$, risk-free rate $r = 0.025$, and expected recovery rate $(\hat{R}) = 0.4$.

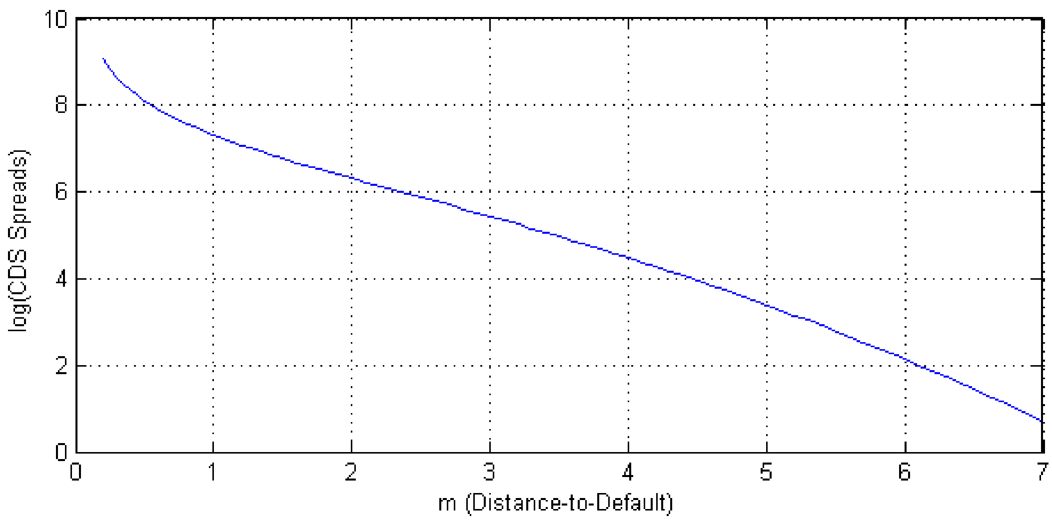

Figure 3: CDS spreads as a function of distance-to-default. The figure displays the log of CDS spread computed with time to maturity $T = 5$, risk-free rate $r = 0.025$, and expected recovery rate $(\hat{R}) = 0.4$.

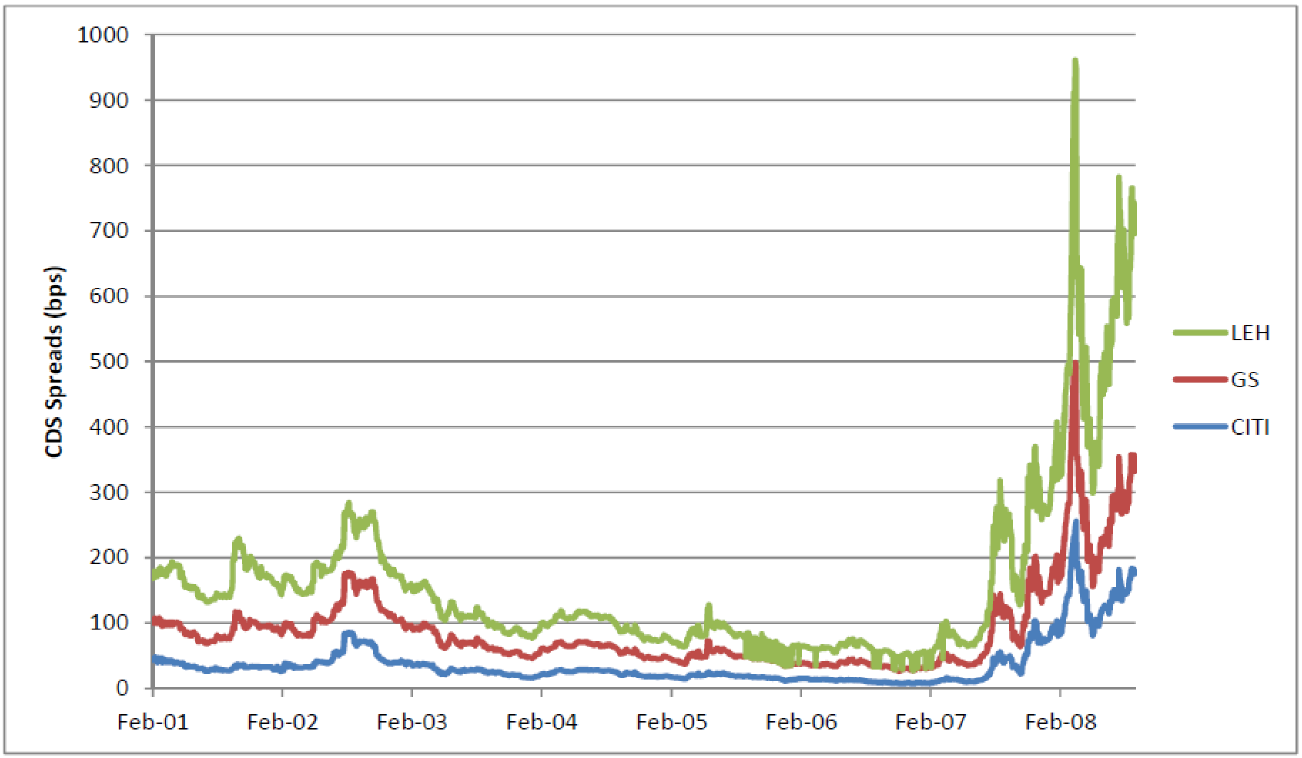

Figure 4: Time series of CDS spreads for Lehman Brothers (LEH), Goldman Sachs (GS) and Citigroup (CITI).

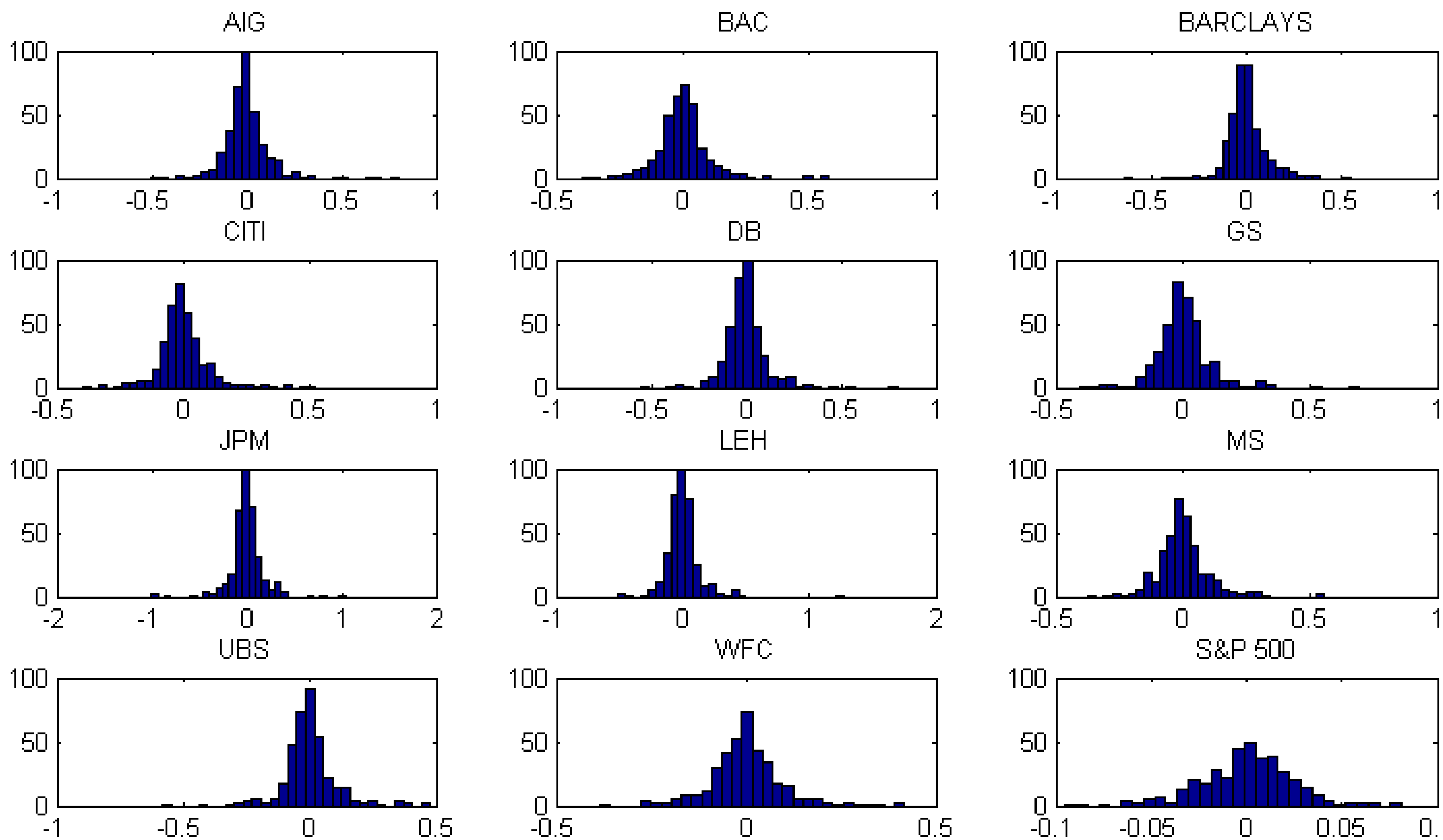

Figure 5: Distributions of nodes in the Bayesian network. The first 11 plots show the distribution of the banks in the network, while the last plot shows the distribution of S&P 500 index returns for comparison. All distributions are weekly and are the residuals of the systematic risk factors as shown in (9).

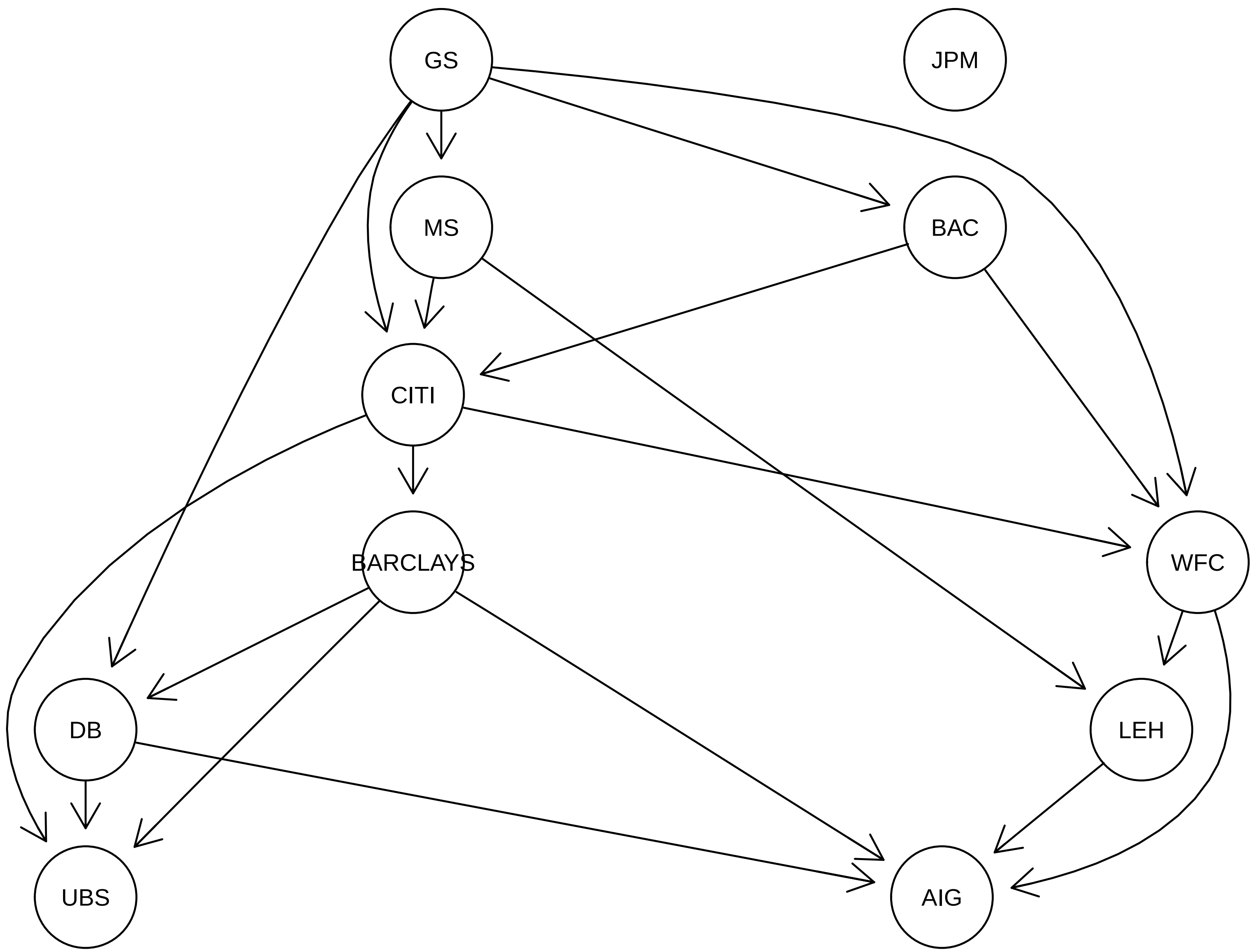

Figure 6: Learned Bayesian network

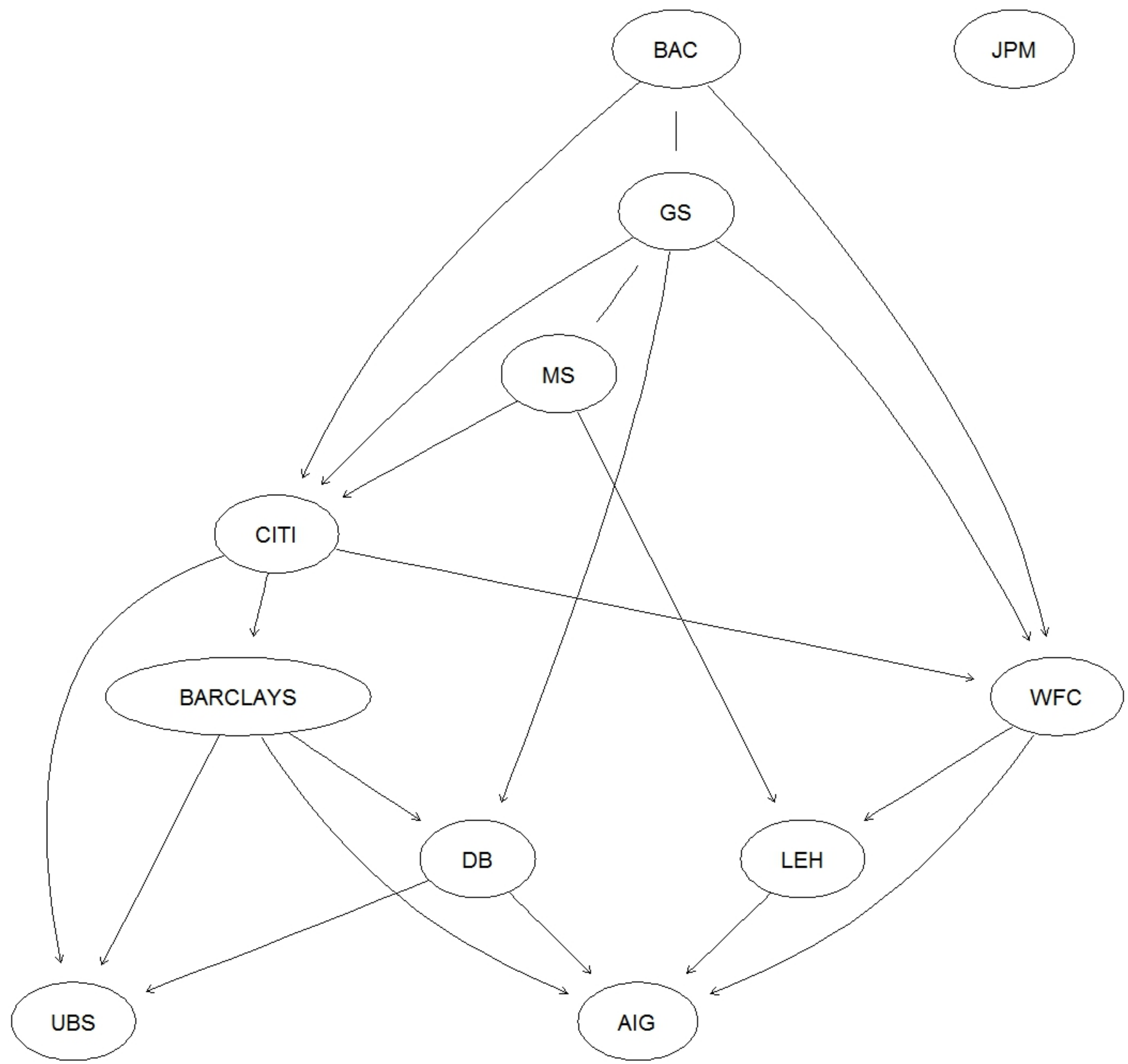

Figure 7: CPDAG of the learned Bayesian network

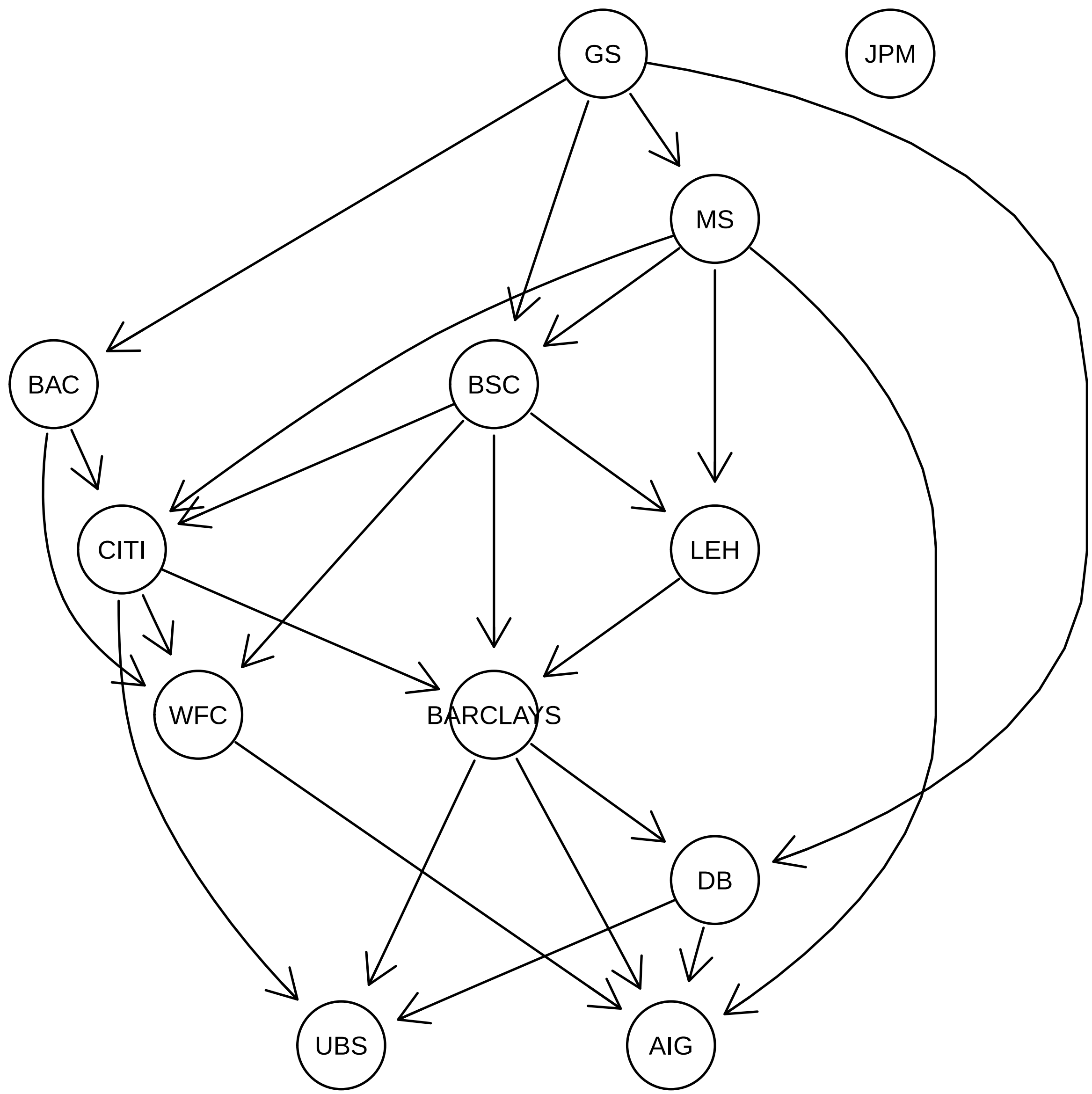

Figure 8: Learned Bayesian network with Bear Sterns (BSC)

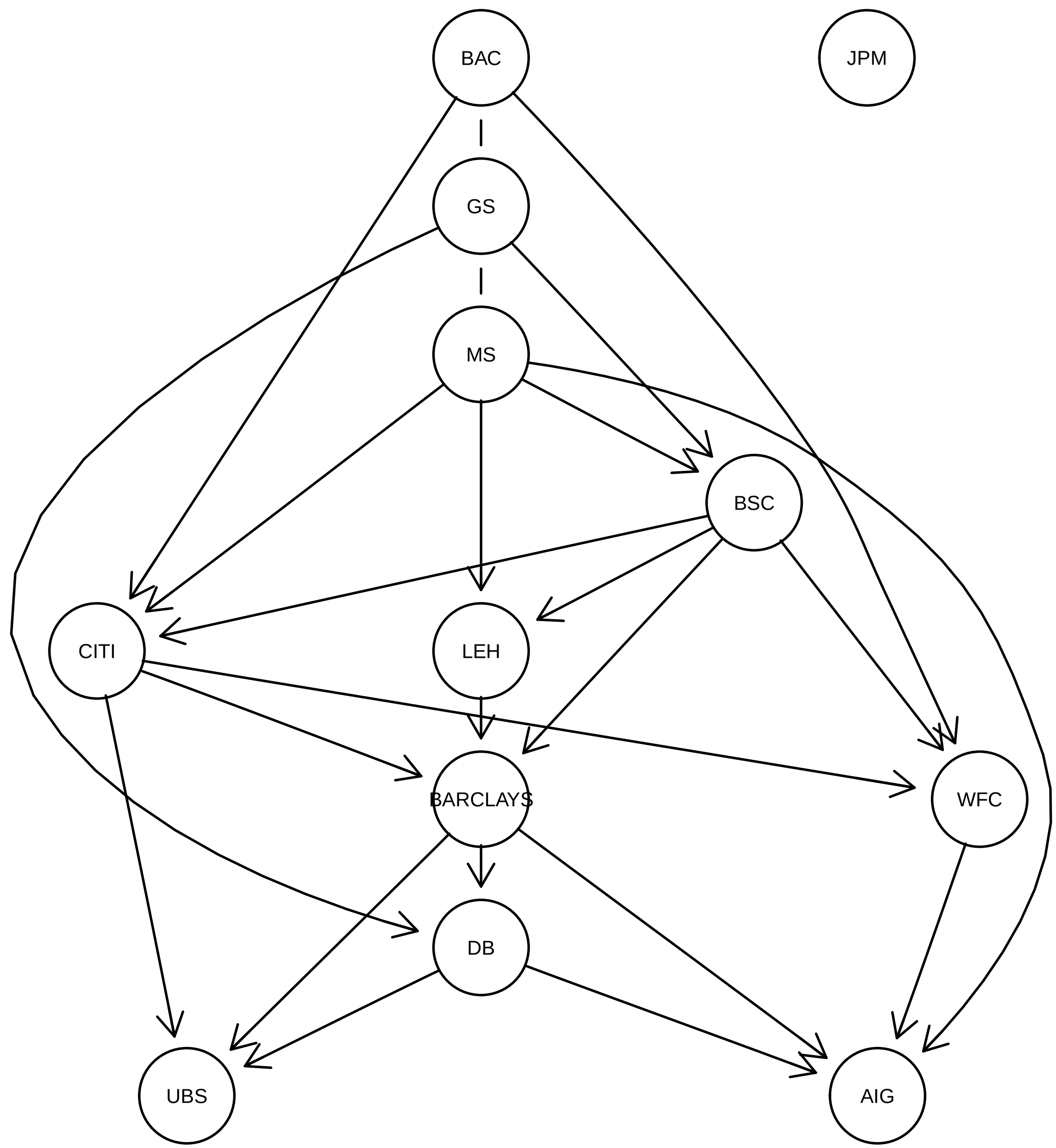

Figure 9: CPDAG of the learned Bayesian network with Bear Sterns (BSC)

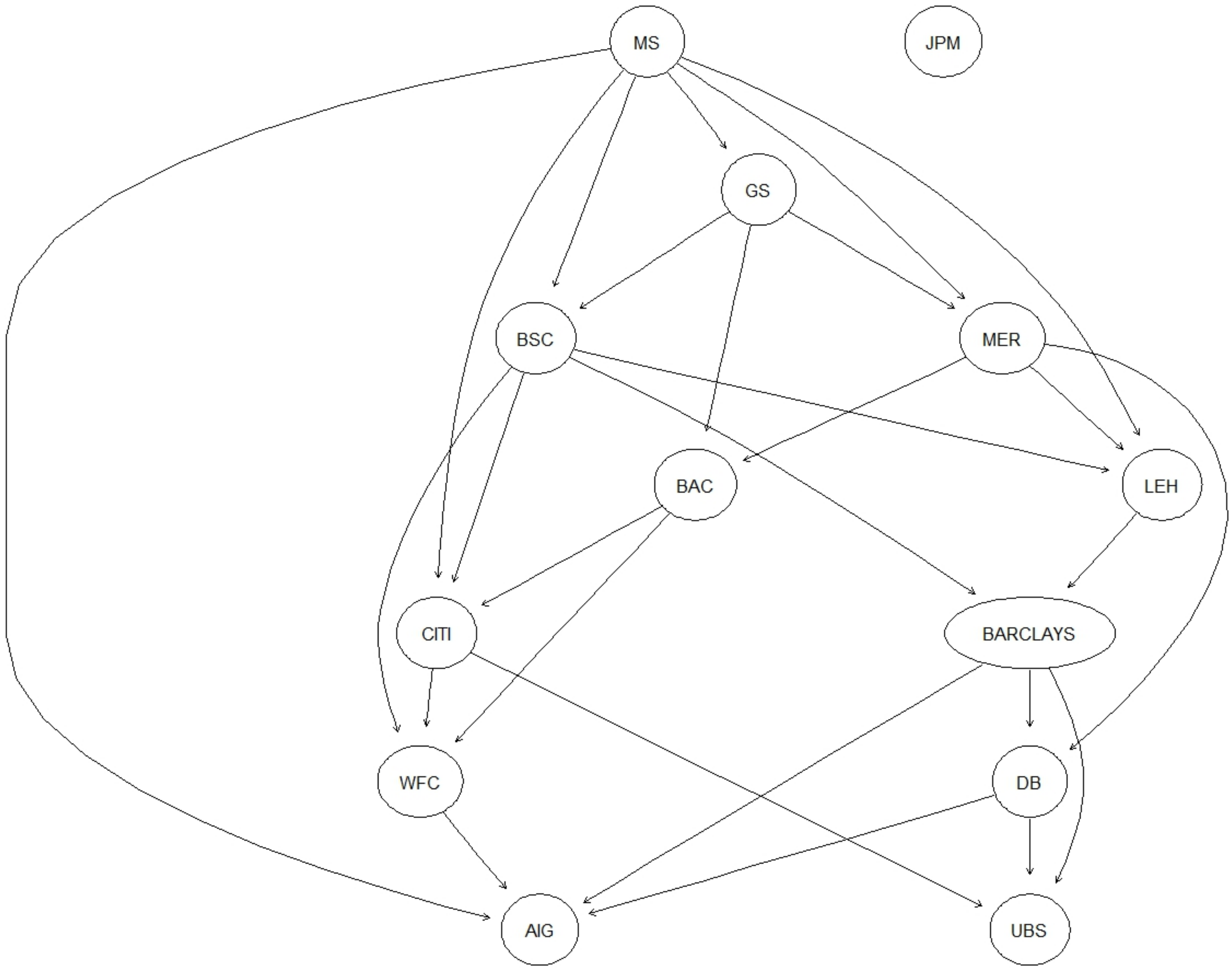

Figure 10: Learned Bayesian network with Bear Sterns (BSC) and Merrill Lynch (MER)

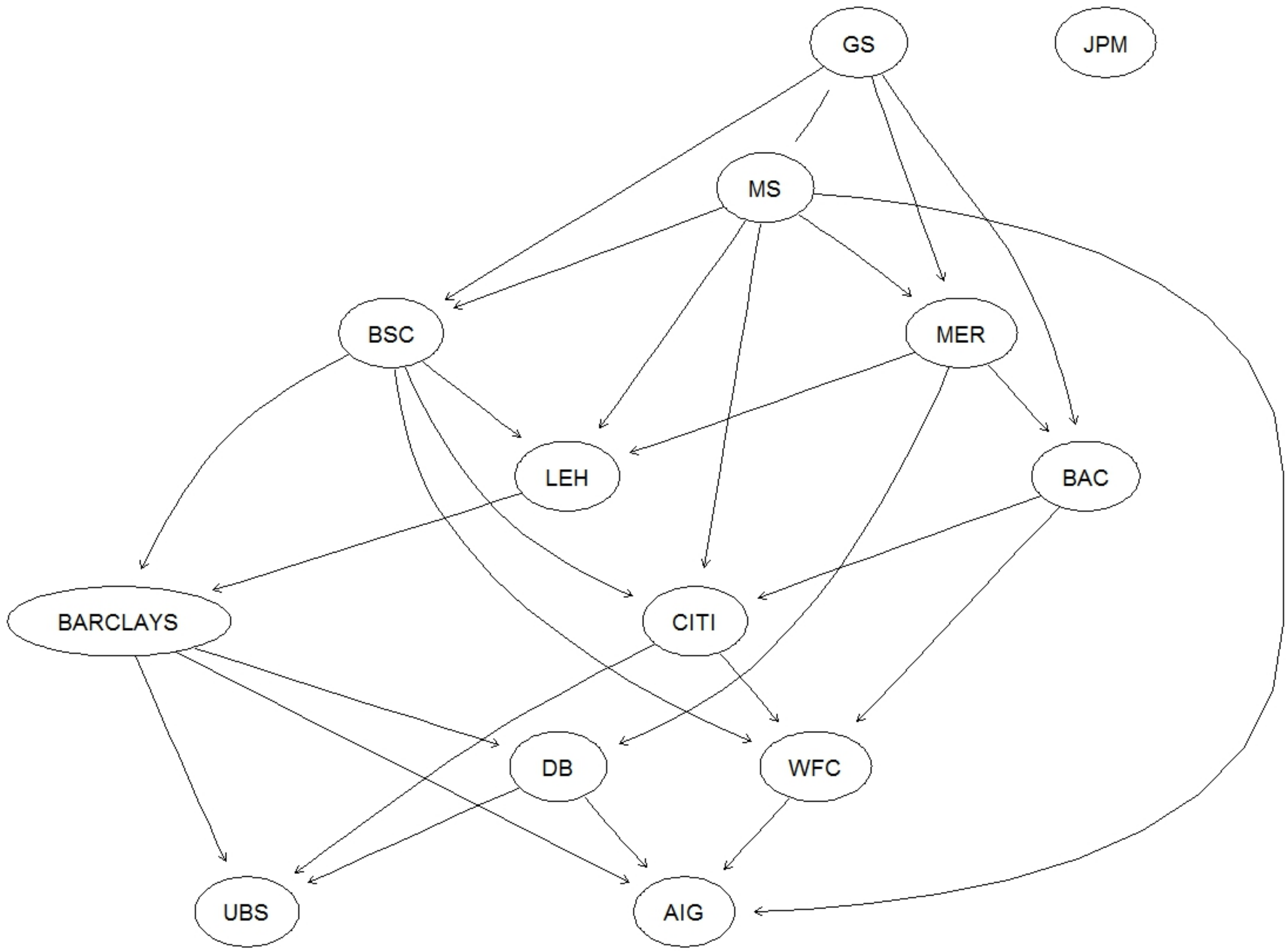

Figure 11: CPDAG of the learned Bayesian network with Bear Sterns (BSC) and Merrill Lynch (MER)