UNIVERSITY OF LIVERPOOL Management School

Working Paper in Economics

202024

July 2020

Applying a Bayesian Network to VaR Calculations

Emma Apps

Applying a Bayesian Network to VaR Calculations

Abstract

This paper focuses considers a methodology for deriving stock returns and VaR through the application of a Bayesian Network (BN). A network map is specified where the returns for three stocks are deemed to be conditionally dependent on two factors. The latter are defined having previously considered literature relating to the financial crisis and risk contagion. Subsequently, two factors are identified as influencing the individual stock returns – one relating to liquidity and the other relating to the market. Following application of the Gaussian Bayesian Network, regressions generate models for the said returns. The latter are then used to simulate time series of stock returns and those outcomes are compared to the original data series. The BN specification is found to be a satisfactory alternative for the modelling of stock returns. Furthermore, the resulting quantiles are shown to be more prudent estimates in relation to VaR calculations at the 5% level and, therefore, can result in increases in regulatory capital.

Applying a Bayesian Network to VaR Calculations

1 Introduction

In their survey of 31 quantitative measures of systemic risk, Bisias et al (2012) identify a research method in relation to Network Analysis in general. Specifically, a small network of factors is defined as being systemically important in relation to their impact on the returns of a set of financial entities. Where each factor is regarded as commonly significant to each entity. Existing research tends to focus on applying such networks in the assessment of how events spread through a financial system and interconnectedness in general. For example, simulating how the failure of one bank can trigger the domino effects across many and whether certain ones are more resilient to the default than others. Indeed, Chan-Lau et al (2009) and the IMF (2009a) use network models to assess the impact of a failing bank on others given respective exposures between them. While the specified networks *can* be used to quantify VaR losses at the bank level following the original default and subsequent domino effect, this is rarely discussed. This paper thereby attempts to contribute to existing literature by applying a Bayesian Network of two factors to determine their impact on the returns of three UK banking stocks and their three-stock portfolio in terms of VaR.

Identification of the factors isn't necessarily intuitive but existing literature in relation to financial linkages and reasons for the spread in financial crises, can be drawn upon. For example, issues around market liquidity are raised and I suggest that the latter is an important factor when assessing impacts on stock returns. There are certain market indicators of the overall health and strength of liquidity among financial institutions, such as the LIBOR-OIS spread in the UK and the TED-spread in the US. Indeed, Hull and White

(2013) suggest that, despite both spreads being stable and largely ignored pre-financial crisis, both are now used as the summary indicators of liquidity following their extreme movements in 2007 and 2008. Subsequently, in order to define a workable network, I begin with just two factors – firstly the aforementioned liquidity factor and secondly, the influence of the wider financials' sector on each stock. In terms of visualising the network, there are a series of nodes connected to each other by edges – where the latter represent the relationship between the nodes. Thereafter, the resulting model is used to simulate returns data for each bank and their three-stock portfolio and quantify their respective 5% and 1% quantiles - where the latter can be used in a VaR calculation and be reasonably applied as an alternative to the RiskMetrics approach. The network itself is specified using Bayesian techniques as presented by Scutari and Denis (2015) and Shonoy and Shonoy (2000).

This paper is divided into several parts. Section 2 highlights the recent literature in relation to Network Analysis and measuring systemic risk but also general applications of Bayesian Networks (BN). Section 3 presents the data, identifying each time series and summary statistics. Application of the BN to this data set in modelling stock returns is presented in section 4 – including specification of the network, the underlying probability distributions and tests of conditional independence and model specifications. The process for simulating stock returns is also discussed. Results are detailed in section 5 – specifically the respective significance of the partial correlations, the parameters of the BN model specifications and the comparisons of the simulated summary statistics and quantiles versus those of the actual historical returns. The paper ends with concluding remarks.

2 Relevant Literature

2.1 Network Analysis

A network rationale has been applied in a diverse range of social and behavioural science contexts. For example, considering how large corporations differ in the extent to which they offer support or assistance to local communities in which they have a presence. Corporate and social responsibility dictates that they should be actively involved in their communities but how much of that is influenced by the activities of other corporations? A network can be used to model how such community involvement is influenced by their interactions and relationships with other corporations. Likewise, in any decision-making process involving several individuals or groups, a network approach can be used to understand how individuals within a group influence each other in the decision-making process. A common underlying theme is how the units within the network interact – they are not viewed in isolation. According to Faust and Wasserman (1994, pp. 7):

> "The network perspective differs in fundamental ways from standard social and behavioural science research methods. Rather than focus on attributes of autonomous individual units, the associations among these attributes, or the usefulness of one or more attributes for predicting the level of another attribute are theorised and modelled through a network."

Such associations and relationships can be witnessed in many other contexts, certainly within science, finance and economics. Indeed, the interlinkages and interconnectedness between financial institutions and within financial systems, leading to spreading in crises are directly relevant (Diao et al 2000). From a scientific perspective, networks are used in a variety of contexts – engineering, biology, ecology, medicine. For example, they are used to analyse ecological systems and specifically how the food chains and ecosystems are connected. In relation to public health, Luke and Harris (2007) present their use in the study of how diseases are transmitted, specifically HIV and AIDS. Applications in medical and microbiology contexts are popular – for instance, Barabasi et al (2011) use network-based methods in genetics to identify molecular linkages and subsequent gene mutations.

Of course, this paper is interested in their application in finance and economics, specifically in relation to systemic risk. Given the focus on liquidity issues, particularly in the interbank markets, the work of Chan-Lau et al (2009), is relevant. In using network models, they highlight the impact of institutional failure when there are exposures within such markets – where the network illustrates the domino effect between connected banks when exposed to a failing institution. Furthermore, Bilio et al (2012) go beyond the inter-bank markets in their application of Granger-causality networks to the study of interconnectedness between a variety of investor sub-groups, specifically hedge funds, banks, brokers and insurance companies. Likewise, the IMF (2012) assess linkages within the global over-the-counter derivative markets and the identification of systemically important financial intermediaries. In each case, as suggested by Battiston et al (2012), there is no widely accepted, single methodology to determine the systemically important nodes or factors within the network – it is very much linked to interpretation and the underlying data set (financial instrument, market, sub-sector, region). The latter indicates the degree of qualitative judgement required in defining the network in the first instance. Nevertheless, Allen and Babus (2009, pp. 367) argue that network analysis can assist our understanding of financial systems and specifically risk contagion, given the interconnections revealed by the 2008 financial crisis. Furthermore, aside from defining the network itself, they suggest that it can then be usefully applied in formulating a regulatory framework for supervising financial institutions, an objective entwined within this paper. Consistent with Allen and Babus (2009), Hu at al (2012) allude to the deficiencies of pre-existing methods in measuring exposures to systemic risk, given the significant widespread losses post 2008. Accordingly, they too suggest a network-based approach as a more appropriate and accurate measurement and monitoring process.

Unsurprisingly, there has been an upsurge in interest in research in this area - several empiricists identify the importance of the use of network analysis. For example, Markose et al (2012) apply a network to investigate the connections between banks in the Credit Default Swap market – the latter market being identified as a key determinant of substantial losses in 2008. In some cases, there is the final realisation that, given their widespread application in science and medicine, surely analogies can be drawn in finance. For instance, Haldane and May (2011) apply the dynamics of food webs in an ecological context to modelling the stability of a given financial system. A leading empiriscist in relation to network theory, Kimmo Soramaki, has several publications focusing on applications in finance. For instance, Soramaki et al (2016) simplify complex network structures in order to filter or highlight the most important determinants of correlations between returns of European stocks. Earlier studies focus on the interbank payment systems and, specifically, the creation of a network representing how payments are transferred between financial institutions (see Soramaki et al (2007)). The latter highlights the key players in such markets and the degrees of connectedness between them but also makes the point that the “minor” players in the market are also connected to the more tightly connected core of major players. Given the financial linkages, the network illustrates the severe impact of any subsequent disruption to it and the issues arising in transferring and accessing capital through the interbank markets. This is further explored by Soramaki and Cook (2013) and Soramaki and Langfield (2016), whereby, following a bank’s failure, the disruption to the payment network is identified along with systemically important institutions and the resulting impacts on individual network participants. A common theme, once again, is the interbank markets. It is clear that, whether referring to literature immediately following the crisis or more recently, that theme remains - the liquidity issues generating from within the inter-bank markets. Similar to Chan-Lau et al (2009), Krause and Giansante (2012) also focus on the exposures within those markets and use a network of connected banks to model how failure of one spreads through the network. Subsequently, a factor encompassed within the BN defined in section 4, relates to liquidity – denoted by a particular spread quoted in the inter-bank markets.

2.2 Network Analysis from a Bayesian Perspective

Bayesian Networks are encompassed within the framework of network analysis and incorporate graphical theories and conditional dependencies between variables in the graphical network. Within the literature there are several instances of the application of BNs to data sets, not necessarily from a finance perspective. Indeed, almost any event conditional on the probability of a prior event can be analysed using this concept. In geographical and environmental studies, for example, a BN is used to evaluate flood plains and the extent of flooding given certain extreme prior events such as changes in sea level and improvements in coastal defenses (see Narayan et al 2018). They are also applied within the context of Social Corporate Responsibility in assessing a corporation's likely compliance with child labour regulations across their supply chain network. The BN is used to determine the likelihood of breaches to such regulations using available data on suppliers, their employee demographics and the frequency of child labour incidents (see Thoni et al 2018). From a medical research perspective, BNs are also readily applied. For instance, in assessing links between patients diagnosed with clinical depression and variability in their heart rates and also in identifying important factors in relation to survival rates from lung cancer (see Anisa and Lin 2017).

From a risk management perspective, there is ample literature relevant to their application, particularly in relation to operational risk. Essentially, various factors are identified and inserted into the BN with estimates made of associated loss distributions resulting from the various risk factors. According to Cowell et al (2007), such techniques can be applied in insurance settings when assessing the financial impact of cases of fraud upon the insuring company and thereby in the setting aside of adequate regulatory capital in relation to such cases. From a banking perspective, Aquaro et al (2010) present their application in relation to losses sustained through cases of failed internal processes, human error, IT failures and certain litigation cases. The factors leading to the losses in each situation become part of the BN and the associated loss distributions are generated. Clearly, there is some degree of subjectivity in identifying the break-downs in the internal processes or human interventions leading to loss making errors but, this is a commonality across all BNs, regardless of the arena in which they are being applied. In all cases, analogies can certainly be drawn with VaR and the need to ascertain the loss quantiles from subsequent returns' distributions. Indeed, Martin et al (2005) apply BNs to specifically model the severest loss inducing events, referred to as the long tails or unexpected losses from an operational loss perspective – similar of course to the 1% and 5% VaR scenarios. Of direct relevance to this paper is the research of Hager and Andersen (2010) who seek to model loss severity across all activities of a financial institution and not just from an operational perspective. This is done through the identification of influencing factors – which I argue can be liquidity and market based.

Other literature identifies the importance of risk contagion through BN modelling of default probabilities resulting from financial linkages – for example Giudici and Spelta (2016) and Chong and Kluppelberg (2017). Furthermore, interconnectedness is also examined through the effect of exposures within the interbank markets. A BN is applied to model individual institutional liabilities within that market and the subsequent impact on other banks in the event that a participant in the network defaults. Gandy and Veraart

(2016) illustrate that the BN can be used to stress test differing assumed levels of inter-bank liabilities and likelihood of default conditional on another bank defaulting. At the very least it indicates the importance of the inter-bank markets once more, if not specifically assessing the impact on bank returns' distributions. Despite all of the literature under review, however, there appears to be a lack of focus on application of BNs specifically in modelling stock returns and losses applied in VaR estimations. This paper seeks to provide a workable alternative approach to modelling both, whilst also considering the importance of the entire financials sector, interlinkages and the impact of reducing liquidity within the inter-bank markets.

3 The Data

In order to produce the appropriate network, and specifically assess the impact of the chosen factors, the data is gathered for the period 14th December 2000 to 29th June 2012 – implying 2,914 daily observations for each variable. This timeframe incorporates the financial crisis but also adequate periods pre and post crisis. The data is sourced from Bloomberg (excluding the portfolio) and the variables are as follows, where the daily return and the daily percentage change ensure stationarity:

- Daily returns for Barclays Bank stock;
- Daily returns for Lloyds Bank stock;
- Daily returns for HSBC stock;
- Daily percentage change in the 3-month Sterling LIBOR vs. 3-month sterling overnight index swap spread (OIS);

- Daily returns for the MSCI Financials' Sector Index:
- Daily returns for the three-stock portfolio.

The data set representing the impact of the market on each stock is the MSCI Financials Sector Index, within which all three banking stocks have a percentage weighting. The chosen liquidity factor is represented by the daily percentage change in the difference between the 3-month sterling LIBOR rate and the 3-month sterling overnight indexed swap rate. Overnight indexed swaps are interest rate swaps whereby a fixed rate of interest is exchanged for floating and the latter is the average of a daily overnight rate. In deriving the floating rate payment, the intention is to replicate the aggregate amount of interest that would be earned from rolling over a sequence of daily loans at an appropriate overnight rate. Given that we are applying a 3-month time-frame, it implies rolling over a sequence of daily loans, for 90 days at an overnight rate – where that rate is determined in the UK by the Bank of England and referred to as SONIA (sterling overnight index average). The 3-month sterling libor rate is the average interest rate at which a selection of banks lend British pounds to one another for a period of 3 months.

In deriving the daily returns for each bank and the nominated market index, the following is applied:

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \qquad (3.1)$$

Where: p_t refers to the closing price of the stock or index at time t.
p_{t-1} refers to the closing price of the stock or index at time t-1.

r_t refers to the daily return of the stock or index at time t.

With regards the three-stock portfolio, we begin with a total initial investment of £30,000,000, split equally between the stocks – representing an equal weighting of 33.33% and £10,000,000 invested in each stock in the portfolio. As the prices of the component stocks change in value, their weights in the portfolio change, as does the notional value of the portfolio. The daily return on the portfolio is derived as follows:

$$r_{port,t} = \frac{Notional\ Value_{port,t} - Notional\ Value_{port,t-1}}{Notional\ Value_{port,t-1}} \qquad (3.2)$$

The notional value of the portfolio each day is derived as follows:

$$NV_{port,t} = \left[(1 + r_{t,B}) \times NV_{t-1,B}\right] + \left[(1 + r_{t,H}) \times NV_{t-1,H}\right] + \left[(1 + r_{t,L}) \times NV_{t-1,L}\right]$$
(3.3)

Where: $NV_{port,t}$ refers to the notional value of the 3-stock portfolio at time t.
$r_{t,B}$ refers to the daily return of Barclays at time t;
$NV_{t-1,B}$ refers to the notional value of investment in Barclays at time t-1;
$r_{t,H}$ refers to the daily return of HSBC at time t;
$NV_{t-1,H}$ refers to the notional value of investment in HSBC at time t-1;
$r_{t,L}$ refers to the daily return of Lloyds at time t;
$NV_{t-1,L}$ refers to the notional value of investment in Lloyds at time t-1.

The graphs are presented for each time series of returns' data in figures 3.1 to 3.5 plus an indication of how the LIBOR-OIS spread moved in the period under review in figure 3.6. The former illustrate stationarity in the time series and significant volatility in the 2007-

2008 time-frame of the financial crisis. In relation to the LIBOR-OIS spread, there are noticeable peaks associated with certain key events. For instance, in September 2007, the spread reached 85 basis points in response to the Bank of England announcing emergency funding to rescue Northern Rock and, three months later as the crisis began to unfold, reached an all-time high of 108 basis points. At its worst, following the insolvency of Lehman Brothers in the autumn of 2008, the spread was around 300 basis points. The Augmented Dickey Fuller tests at various lags in table 3.1, indicate the stationarity in the time series of returns for each variable:

Table 3.1: Augmented Dickey Fuller tests for each variable

	LIBOROIS	Market	Barclays	HSBC	Lloyds
1 Lag	-47.0068*	-38.5601*	-36.5903*	-41.127*	-38.3401*
2 lags	-40.925*	-33.0937*	-30.3896*	-33.8231*	-31.5608*
3 lags	-33.7785*	-27.1927*	-25.4275*	-28.0044*	-27.0613*
4 lags	-29.4624*	-26.0475*	-23.5451*	-25.5346*	-26.8292*
5 lags	-24.4209*	-24.5432*	-21.8710*	-24.7745*	-25.0340*
6 lags	-22.0149*	-22.1618*	-20.8262*	-21.6686*	-23.5661*
7 lags	-20.7559*	-19.8397*	-19.6280*	-19.5955*	-21.9056
8 lags	-18.6174*	-18.4505*	-17.2105*	-18.2352*	-20.2806*
9 lags	-17.3829*	-17.9787*	-17.1425*	-17.7670*	-18.0233*
10 lags	-16.8473*	-17.4447*	-16.3418*	-17.9372*	-16.7871*

Note: Critical values of -3.43, -2.86, -2.57. * denotes test statistic < critical values at all level

Figure 3.1: Time Series of Barclays Daily Returns.

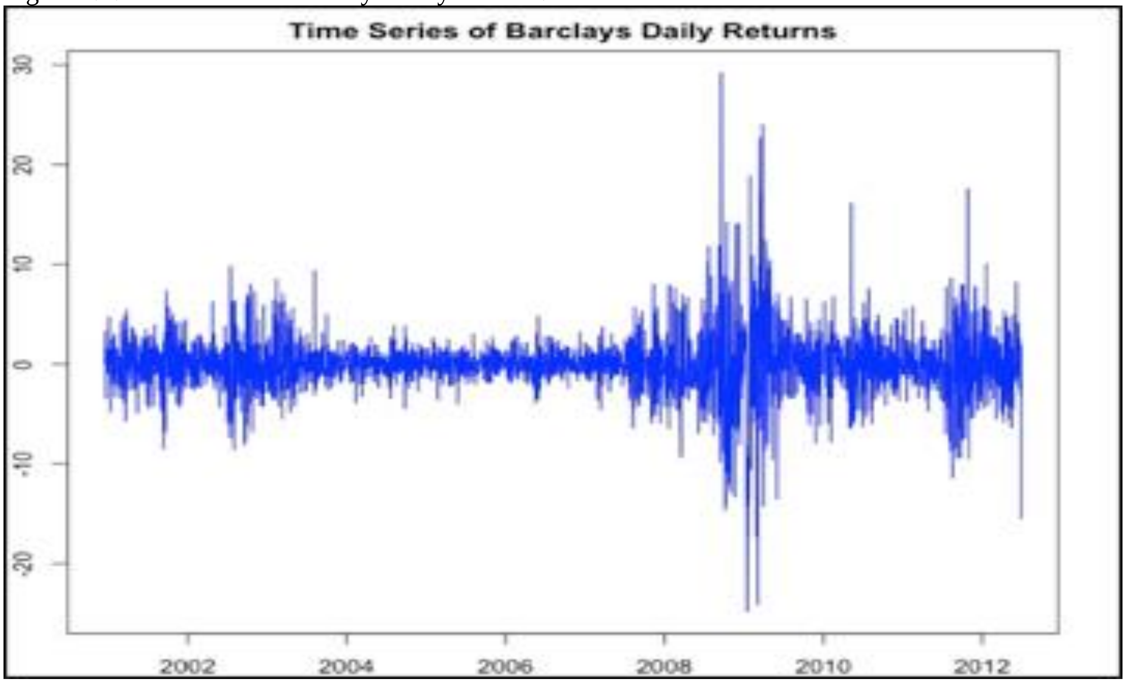


Figure 3.2: Time Series of HSBC Daily Returns.

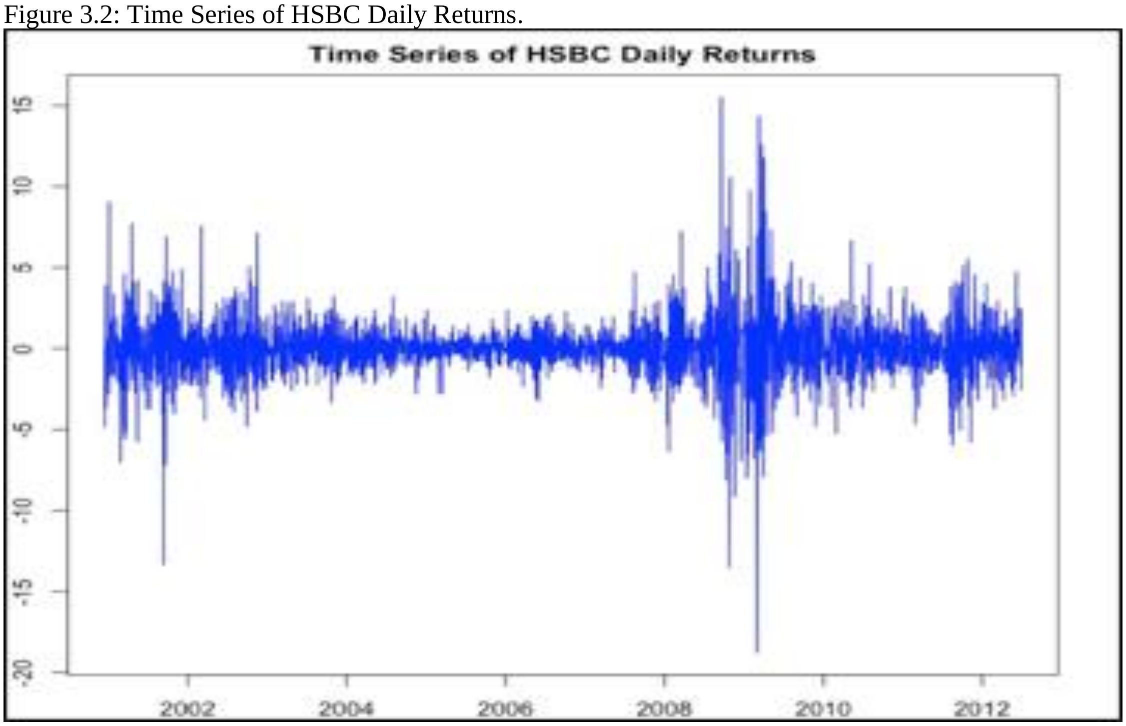


Figure 3.3: Time Series of Lloyds Daily Returns.

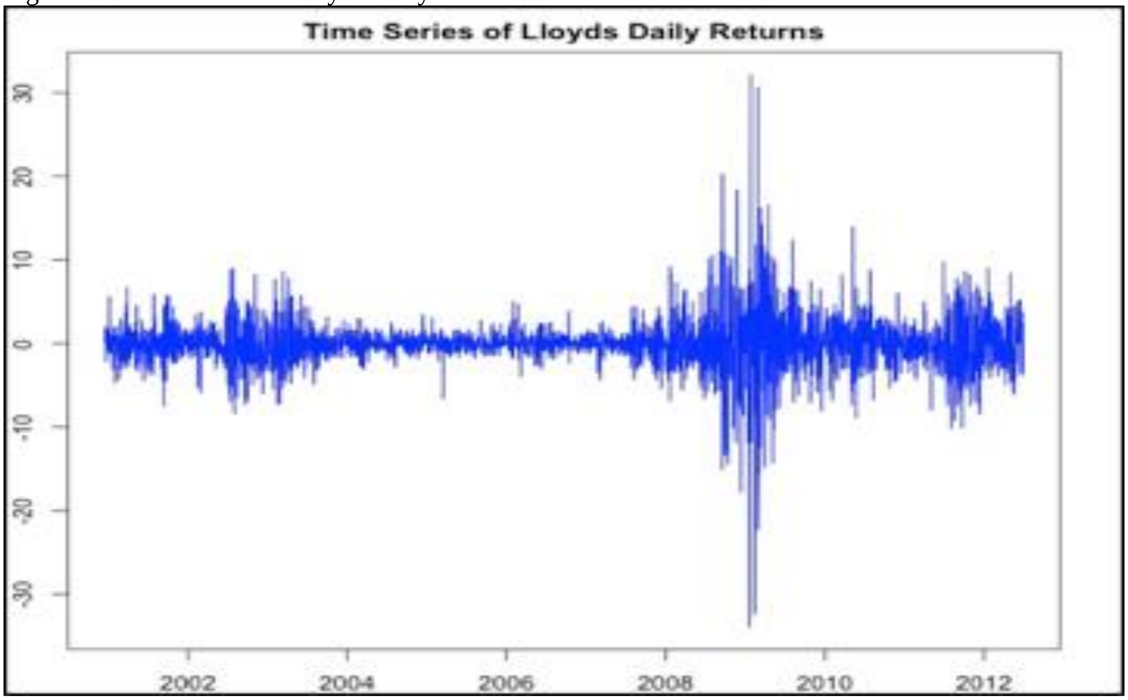


Figure 3.4: Time Series of Market Daily Returns.

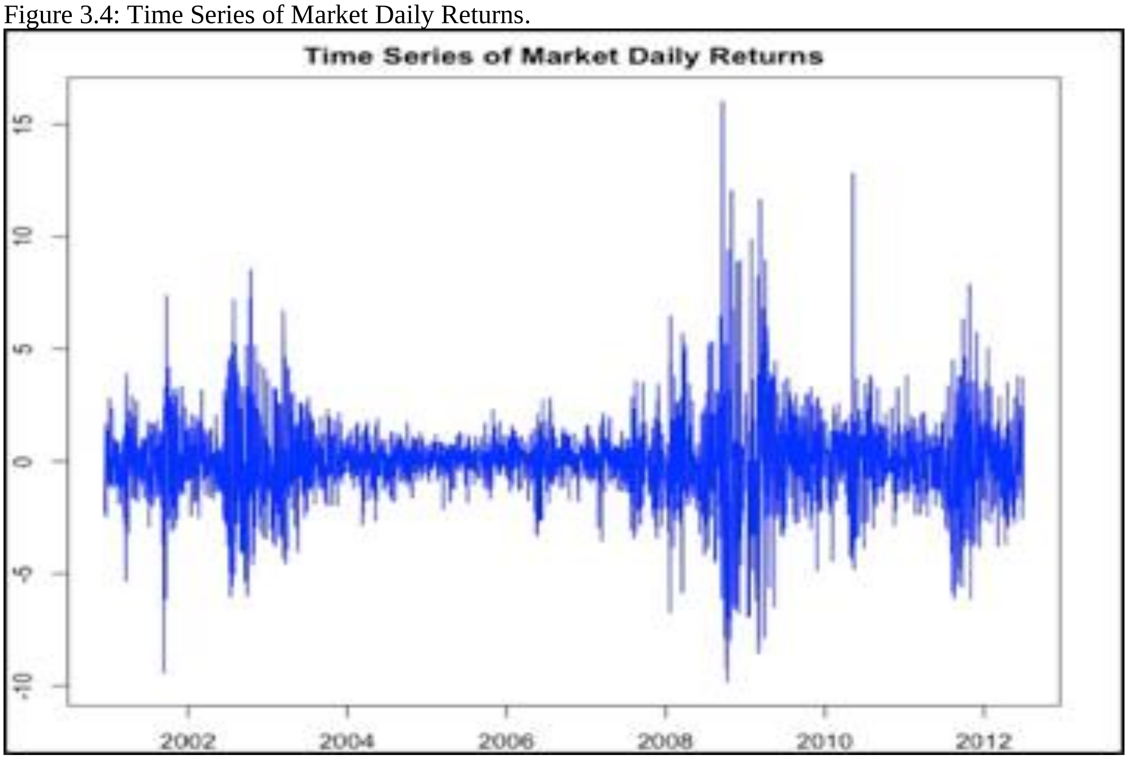


Figure 3.5: Time Series of Portfolio Daily Returns.

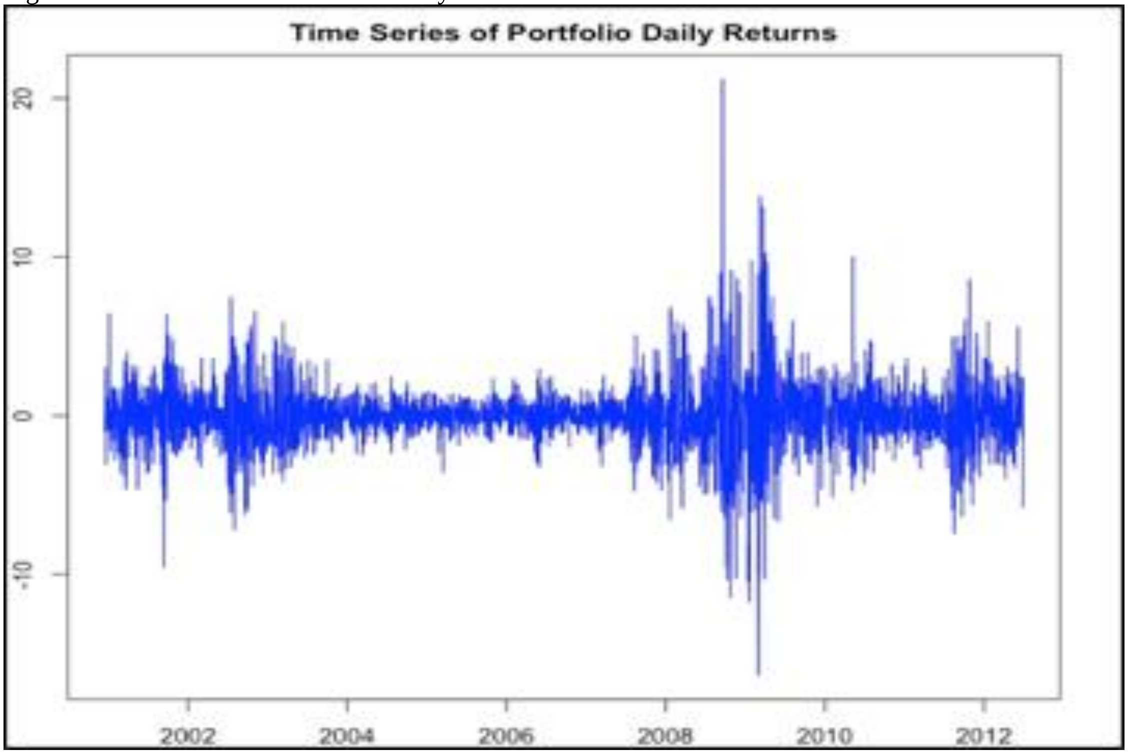


Figure 3.6: Graph of the LIBOR-OIS spread.

In relation to the summary statistics presented in table 3.2, the mean daily returns appear close to zero and the minimum returns reflect the substantial losses during the financial crisis.

Table 3.2: Summary Statistics for LIBOROIS % change, Market, Stock and Portfolio Daily Returns.

	LIBOROIS	Market	Barclays	HSBC	Lloyds	Portfolio
Max	141.6667	16.0399	29.2357	15.5148	32.2159	21.2197
Min	-67.6692	-9.8446	-24.8464	-18.7788	-33.9479	-16.3841
Median	0.0000	0.0000	-0.0501	0.0000	-0.0504	-0.0076
Mean	0.7612	-0.0223	-0.0076	0.0024	-0.0429	-0.0183

4 Application of a Gaussian Bayesian Network to Continuous Data

4.1 Proposed Network Structure

In applying Bayesian Networks to modelling data, they are useful in the situation where information is incomplete and uncertainty exists over the key determinants of the dependent variable. According to Shenoy and Shenoy (2000), there is initially a degree of qualitative judgement and subjectivity in specifying the factors to include in the graphical representation of the network. However, in subsequently applying quantitative tests of the model and simulating posterior data distributions, certain inferences can be made about its validity. In this instance the proposed network is being applied to model portfolio returns based on certain inputs or factors added to it. Furthermore, given the simulated posterior return distribution of the portfolio, a cut-off return is derived for use in a VaR calculation, where the cut-offs refer to the 1% and 5% quantiles of the said distribution.

In specifying a Gaussian Bayesian Network, I am modelling continuous data sets with the underlying assumption of multivariate normality. With regards the variables defined in section 3, I denote them with the following abbreviations:

- Daily returns for Barclays Bank stock→B
- Daily returns for Lloyds Bank stock→L
- Daily returns for HSBC stock→H
- Daily percentage change in the 3-month Sterling LIBOR vs. 3-month sterling overnight index swap spread (OIS) →S
- Daily returns for the MSCI Financials' Sector Index→M
- Daily returns for the three-stock portfolio→P

Prior to tests of conditional independence, the suggested relationships between variables are as follows:

B is directly influenced by S and M, L is directly influenced by S and M, H is directly influenced by S and M and P is directly influenced by B, L and H. Consequently, the proposed relationships are defined as follows:

$$\{S, M\} \rightarrow B,\ \{S, M\} \rightarrow L,\quad \{S, M\} \rightarrow H,\ \{B, L, H\} \rightarrow P \tag{4.1}$$

4.2 Proposed Network Graph and Probability Distribution

Based upon the above suggested relationships between the variables a graphical representation can be defined – as presented in figure 4.2.1. It is referred to as a directed acyclic graph (DAG) and contains a series of arcs and nodes. The former reflect the direct dependencies between variables and the latter reflect the variables within the network. Each variable or node has its own distribution – for example, ‘B” has a distribution or time series of daily returns. If an arc exists from one variable to another, the latter variable is dependent upon the former, otherwise known as the parent. The overall distribution, encompassing all variables and suggested dependencies, can be depicted as follows:

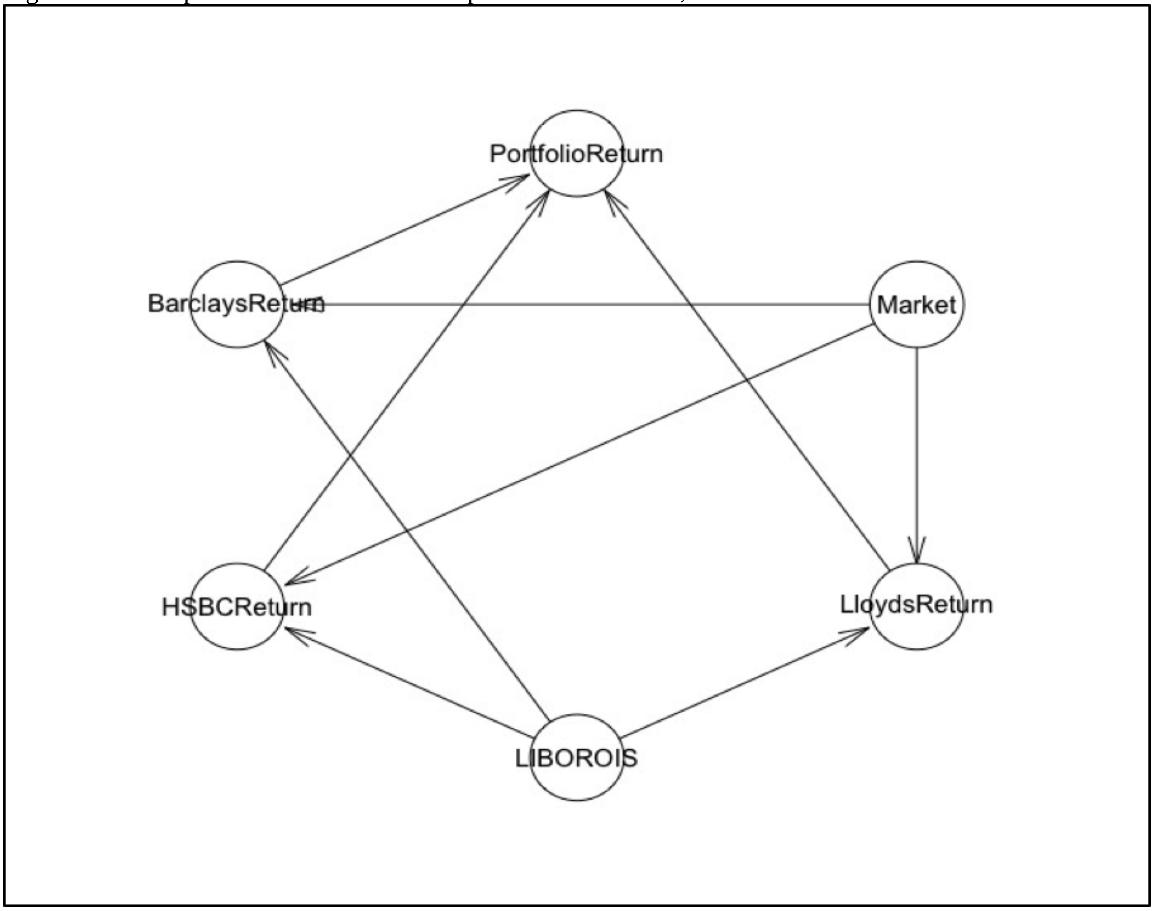
$$\Pr(S, M, B, L, H, P) = \Pr(S)\Pr(M)\Pr(B|S, M)\Pr(H|S, M)\Pr(L|S, M)\Pr(P|B, H, L)$$

Furthermore, the distributions at each node can be expressed as:

$$B|S = s, M = m \qquad H|S = s, M = m \qquad L|S = s, M = m \qquad P|B = b, H = h, L = l$$

where, the distribution at each node is conditional on the values of its parents. Rather than determining the overall joint probability distribution encompassing all variables from the outset, the Bayesian Network (BN) approach breaks the distribution into sub-groups and derives the local distributions at each node. Scutari and Denis (2015) present that specifying a joint probability distribution is rather difficult and complex given the numbers of variables and correlations requiring estimation. Therefore, the BN overcomes this modelling issue through specifying the local distribution at each node conditional on the values of the parents.

Figure 4.2.1: Proposed DAG of Relationship Between 2 factors, Stock Returns and Portfolio Returns



4.3 Algebraic Representation of the DAG

The conditional relationships for each of the nodes of the three stocks may be specified as an equation, consistent with the assumptions that 1) every node follows a normal distribution and 2) the equations represent a Gaussian linear model incorporating an intercept, with the node's parents as the explanatory variables. The specifications in this case, for each factor and stock are as follows:

$$S \sim N(\mu_S, \sigma_S^2) \qquad M \sim N(\mu_M, \sigma_M^2) \tag{4.2}$$

$$B \big| S = s, M = m \sim N(\alpha_B + \beta_{1,B} s + \beta_{2,B} m,\ \varepsilon_B^2) \tag{4.3}$$

$$H\big|S = s, M = m \sim N(\alpha_H + \beta_{1,H}s + \beta_{2,H}m,\ \varepsilon_H^2) \quad (4.4)$$

$$L\big|S = s, M = m \sim N(\alpha_L + \beta_{1,L}s + \beta_{2,L}m,\ \varepsilon_L^2) \quad (4.5)$$

Where: α refers to the intercepts, β refers to the regression coefficients for the parents, S and M and ε represents the standard deviation of the residuals.

There is no specification for the three-stock portfolio because its subsequent simulated returns are derived using equations (3.2) and (3.3).

4.4 Testing for Conditional Independence

As each arc in the DAG encompasses a probabilistic dependence, conditional independence tests can be used to assess whether the data actually supports it. In terms of hypotheses, for each variable, the following conditional dependencies are being tested:

$H_0: B\ is\ independent\ from\ M|S$ versus $H_1: B\ is\ not\ independent\ from\ M|S$

$H_0: B\ is\ independent\ from\ S|M$ versus $H_1: B\ is\ not\ independent\ from\ S|M$

$H_0: H\ is\ independent\ from\ M|S$ versus $H_1: H\ is\ not\ independent\ from\ M|S$

$H_0: H\ is\ independent\ from\ S|M$ versus $H_1: H\ is\ not\ independent\ from\ S|M$

$H_0: L\ is\ independent\ from\ M|S$ versus $H_1: L\ is\ not\ independent\ from\ M|S$

$H_0: L\ is\ independent\ from\ S|M$ versus $H_1: L\ is\ not\ independent\ from\ S|M$

The null hypothesis depicts that B, H or L may be independent from M given S or S given M. If the null is proven, the Beta coefficients in equations 4.3 to 4.5 are equal to zero. Using "B" as an example, through the hypotheses, the partial correlation between B and M given

S or S given M, is being tested – denoted by $\rho_{B,M|S}$ or $\rho_{B,S|M}$. The null holds if $\rho_{B,M|S}$ or $\rho_{B,S|M}$ is not statistically different from zero. In the test, the appropriate distribution is a student's t distribution with n – 3 degrees of freedom (where n refers to the total number of observations in each time series of B, H and L and 3 refers to the number of variables in the test e.g. B, S and M).

$$t(\rho_{B,M|S}) = \rho_{B,M|S}\sqrt{\frac{2911}{1-\rho_{B,M|S}^2}} \tag{4.6}$$

The null hypothesis of independence is rejected if the corresponding p-value is less than the 10%, 5% and 1% degrees of significance.

4.5 Simulating the Returns' Distributions

Following the independence tests in section 4.4, the parameters of equations 4.3 to 4.5 are estimated using the maximum likelihood estimator. Each time-series of bank returns, as the response variables, are regressed on the time-series of the daily percentage change in the LIBOROIS spread and the daily returns in the market index. Having determined the parameters of the models proposed by the DAG in section 4.3, they are then used to simulate sets of random variables for each node, B, H and L. Simulation is performed from the BN by generating a sample of random values from the joint distribution of the specified nodes. It is performed following the order implied by the arcs in the DAG – from the parents first, followed by the children (LIBOROIS and the Market being the parents, the 3 banks being the children). For each node, 2,914 random values are generated – depicting estimates of the daily returns for each stock. In each case, the simulation is performed on the basis of both a normal distribution and a student's t-distribution, using the "rnorm" and "rt" functions in R-studio.

5 Results

5.1 Tests of Conditional Independence

For each of the banks, inverse correlation matrices are produced, which are required for the significance tests and generation of p-values. The resulting correlations are presented in tables 5.1.1 to 5.1.3.

Table 5.1.1: Correlation Matrix for Barclays versus 2 parent nodes

	Barclays Returns	LIBOROIS	Market
Barclays Returns	1.0000	0.0332	0.7818
LIBOROIS	0.0332	1.0000	-0.0827
Market	0.78718	-0.0827	1.0000

Table 5.1.2: Correlation Matrix for HSBC versus 2 parent nodes

	HSBC Returns	LIBOROIS	Market
HSBC Returns	1.0000	0.0306	0.7828
LIBOROIS	0.0306	1.0000	-0.0805
Market	0.7828	-0.0805	1.0000

Table 5.1.3: Correlation Matrix for Lloyds versus 2 parent nodes

	Lloyds Returns	LIBOROIS	Market
Lloyds Returns	1.0000	0.0192	0.6953
LIBOROIS	0.0192	1.0000	-0.0787
Market	0.6953	-0.0787	1.0000

The respective significance tests for the partial correlations are presented in table 5.1.4 In all cases, the bank returns have a significant positive correlation with the market (M) given the daily percentage change in the LIBOROIS spread (S) and we can thereby reject the null hypothesis of independence given the extremely small p-values at all levels of significance. In relation to the conditional dependence between the bank returns and the LIBOROIS variable, given the market returns, there is positive correlation but at a low level. Furthermore, the p-values only indicate significance at the 10% level. However, at that level of significance the null hypothesis of independence is rejected and we can surmise that there is a degree of conditional dependence between daily bank returns and the chosen indicator of liquidity in the financial markets. Thereby, both factors, deemed to be the parents in the DAG, can subsequently be applied in the modelling of the bank returns.

Table 5.1.4: Significance Tests of Partial Correlations

	$B{\sim}S\|M$	$B{\sim}M\|S$	$H{\sim}S\|M$	$H{\sim}M\|S$	$L{\sim}S\|M$	$L{\sim}M\|S$
Pearson's Correlation	0.0332	0.7818	0.0306	0.7828	0.0192	0.6953
Degrees of Freedom	2911	2911	2911	2911	2911	2911
P-Value	0.0729*	0.0000***	0.0991*	0.0000***	0.0990*	0.0000***

Note: * denotes significance at 10%, ** significance at 5%, *** significance at 1%.

5.2 Parameters of the BN Model Specification

Following the Gaussian linear regression for each bank, the respective maximum likelihood estimators are produced and presented in table 5.2.1. Values for the intercepts ($\alpha_B, \alpha_H and\ \alpha_L$) and contributions of the parents, as depicted by the Beta coefficients, are provided.

Table 5.2.1: Parameters of the BN Model for the returns of each bank

α_B	α_H	α_L	$\beta_{1,B}$	$\beta_{2,B}$	$\beta_{1,H}$	$\beta_{2,H}$	$\beta_{1,L}$	$\beta_{2,L}$	ε_B^2	ε_H^2	ε_L^2
0.0175	0.0174	0.019	0.0053	1.304	0.0028	0.769	0.0036	1.197	1.95^2	1.15^2	2.32^2

The contributions from the LIBOROIS variable are small but the value of the spread itself is also and percentage changes in the daily returns of any stock are not frequently sizeable. Referring back to equations 4.3 to 4.5, the BN model specifications, following the linear regression, are as follows:

$$B|S = s, M = m \sim N(0.0175 + 0.0053s + 1.304m,\ \ 1.95^2) \quad (5.1)$$

$$H|S = s, M = m \sim N(0.0174 + 0.0028s + 0.769m,\ \ 1.15^2) \quad (5.2)$$

$$L|S = s, M = m \sim N(-0.019 + 0.0036s + 1.197m,\ \ 2.32^2\ ) \quad (5.3)$$

$$S \sim N(0.76, 12.39^2) \qquad M \sim N(-0.0223, 1.88^2)$$

Equations 5.1, 5.2 and 5.3 are then applied in simulating sets of returns for the three bank stocks.

5.3 Simulated Data

Following the simulation of time series of returns for each of the three bank stocks and subsequent three-stock portfolio, a comparison is made between the summary statistics of the original, actual data sets and the simulations, applying both a normal and student's t-distribution. Both are presented in tables 5.3.1 to 5.3.4.

Table 5.3.1: Comparison of Summary Statistics – Actual versus Simulated Returns - Barclays

	Barclays Actual Returns	Simulated Returns (Normal Dist'n)	Simulated Returns (t-distribution)
Mean	-0.00759%	-0.02102%	0.06029%
Max	29.2357%	10.45150%	11.02510%
Min	-24.8464%	-9.75948%	-10.64131%
Median	-0.05013%	0.01165%	0.06097%
Stdev	3.13367%	3.14959%	3.25881%

Table 5.3.2: Comparison of Summary Statistics – Actual versus Simulated Returns - HSBC

	HSBC Actual Returns	Simulated Returns (Normal Dist'n)	Simulated Returns (t-distribution)
Mean	0.002432%	-0.008251%	0.03320%
Max	15.51481%	6.407118%	7.63505%
Min	-18.77880%	-6.580565%	-7.42987%
Median	-0.00000%	0.038593%	0.07663%
Stdev	1.84544%	1.82093%	2.13113%

Table 5.3.3: Comparison of Summary Statistics – Actual versus Simulated Returns - Lloyds

	Lloyds Actual Returns	Simulated Returns (Normal Dist'n)	Simulated Returns (t-distribution)
Mean	-0.04285%	-0.09143%	-0.07486%
Max	32.21586%	10.26661%	10.93075%
Min	-33.94790%	-11.76306%	-11.21700%
Median	-0.007604%	-0.02116%	0.000267%
Stdev	2.13653%	1.61464%	1.802019%

Table 5.3.4: Comparison of Summary Statistics – Actual versus Simulated Returns - Portfolio

	Portfolio Actual Returns	Simulated Returns (Normal Dist'n)	Simulated Returns (t-distribution)
Mean	-0.018333%	-0.04074%	0.011784%
Max	21.21970%	6.19230%	6.413994%
Min	-16.38410%	-5.72795%	-5.939262%
Median	-0.007604%	-0.021160%	0.000267%
Stdev	2.13653%	1.61464%	1.802019%

Given that the mean return is expected to be close to zero, in all cases the simulated values are consistent. Furthermore, the simulated standard deviations match the actuals with reasonable accuracy. Of greater relevance is the model's ability to derive meaningful estimates of minimum returns, given the implications for VaR. For each bank and the portfolio, the estimated minimum returns are significantly different from the actual values. However, given the underlying assumption of a Gaussian BN and normality, it will not necessarily correctly evaluate the tails of the distribution. It is encouraging that, when applying a t-distribution, the resulting minima are larger than in the normal case for three of the four data series – consistent with its ability to model tails more effectively. Despite the clear differences in summary statistics, it is important to consider the relative accuracy of the model in relation to quantiles. After all, they are used as the cut-off points in relation to VaR calculations. Given that the most severe maxima or minima values for daily returns occur so infrequently, they are not necessarily an accurate indicator of the most likely maximum daily loss. Thereby, a comparison of the quantiles from the actual returns and the simulated cases are presented in tables 5.3.5 to 5.3.8.

Table 5.3.5: Comparison of Quantiles – Actual versus Simulated - Barclays

Quantile	Barclays Actual	Simulated (Normal Dist'n)	Over / Under Estimates	Simulated (t-distribution)	Over / Under Estimates1
1%	-8.786%	-7.782%	Under	-7.479%	Under
5%	-4.372%	-5.174%	Over	-5.288%	Over
10%	-3.069%	-3.976%	Over	-4.123%	Over
90%	2.939%	3.854%	Over	4.287%	Over
95%	4.785%	5.035%	Over	5.482%	Over
99%	8.641%	7.322%	Under	7.708%	Under

Table 5.3.6: Comparison of Quantiles – Actual versus Simulated - HSBC

Quantile	HSBC Actual	Simulated (Normal Dist'n)	Over / Under Estimates	Simulated (t-distribution)	Over / Under Estimates1
1%	-5.279%	-4.215%	Under	-4.718%	Under
5%	-2.546%	-3.065%	Over	-3.529%	Over
10%	-1.835%	-2.382%	Over	-2.774%	Over
90%	1.849%	2.289%	Over	2.752%	Over
95%	2.689%	3.008%	Over	3.615%	Over
99%	5.107%	4.154%	Under	5.035%	Under

Table 5.3.7: Comparison of Quantiles – Actual versus Simulated - Lloyds

Quantile	Lloyds Actual	Simulated (Normal Dist'n)	Over / Under Estimates	Simulated (t-distribution)	Over / Under Estimates1
1%	-8.893%	-7.859%	Under	-7.809%	Under
5%	-4.408%	-5.390%	Over	-5.523%	Over
10%	-3.051%	-4.204%	Over	-4.394%	Over
90%	2.920%	3.927%	Over	4.210%	Over
95%	4.468%	5.268%	Over	5.318%	Over
99%	9.079%	7.487%	Under	8.264%	Under

Table 5.3.8: Comparison of Quantiles – Actual versus Simulated - Portfolio

Quantile	Portfolio Actual	Simulated (Normal Dist'n)	Over / Under Estimates	Simulated (t-distribution)	Over / Under Estimates1
1%	-5.874%	-3.881%	Under	-4.236%	Under
5%	-3.099%	-2.766%	Under	-2.997%	Under
10%	-2.187%	-2.106%	Under	-2.312%	Over
90%	2.079%	2.027%	Under	2.300%	Over
95%	3.163%	2.641%	Under	2.938%	Under
99%	6.423%	3.582%	Under	4.318%	Under

In all cases, the simulated 1% quantiles from the simulated returns, are less than those based upon the actual time series of returns. Perhaps not surprising given the underlying normal distribution assumption. However, the related simulated 5% and 10% quantiles are larger than those on an actual basis. This implies greater prudence in subsequent VaR estimates due to the left tail being larger in the simulated cases if the quantiles are used as the appropriate cut-off. Despite the under-estimations at the 1% level, the simulated results are still of use in a practical context due to the industry convention of reporting VaRs at the 5% level for individual stocks. The simulated portfolio quantiles are slightly misleading given that they are impacted by the respective weights of the component stocks. They are, nevertheless, comparable to the portfolio actual quantiles at both the 5% and 10% levels.

Application of the t-distribution, allows for a more realistic modelling of the tails. Consequently, in the simulations, the resulting 5% and 10% quantiles for all stocks are even more prudent than in the normal case and also at the 1% level for HSBC.

Table 5.3.9 illustrates the absolute percentage increases in the 5% and 10% quantiles offered by the simulated results. At the 5% level, the increases in the quantile range from 0.5% to just over 1%. From a regulatory perspective, and the setting aside of regulatory capital based on VaR assessments, an additional 1% would be significant – if we consider the notional values of stocks and equity portfolios.

Table 5.3.9: Absolute % increase in 5% and 10% quantiles offered by Simulated Data

	Barclays Normal Dist’n	HSBC Normal Dist’n	Lloyds Normal Dist’n	Barclays t-dist’n	HSBC t-dist’n	Lloyds t-dist’n
5%	0.802%	0.519%	0.982%	0.916%	0.983%	1.115%
10%	0.907%	0.547%	1.153%	1.054%	0.939%	1.343%

Finally, figures 5.3.1 to 5.3.3 reflect the fitted distributions of the simulated stock returns according to the underlying assumption of normality.

Figure 5.3.1: Barclays Fitted Simulated Returns

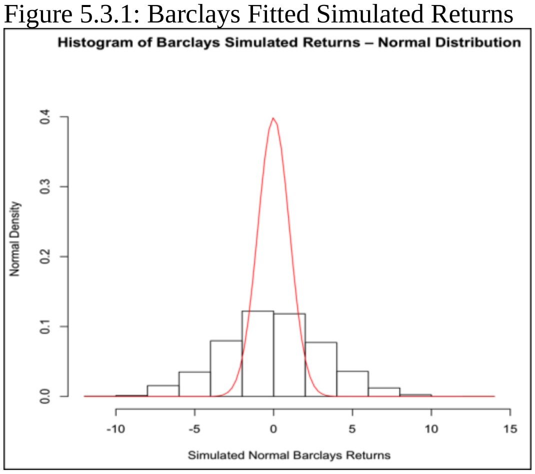


Figure 5.3.3: Lloyds Fitted Simulated Returns

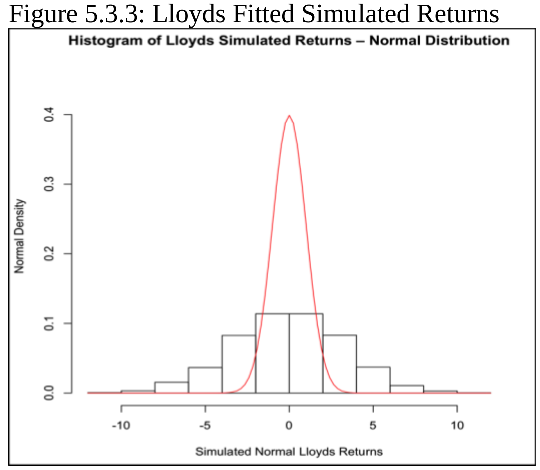
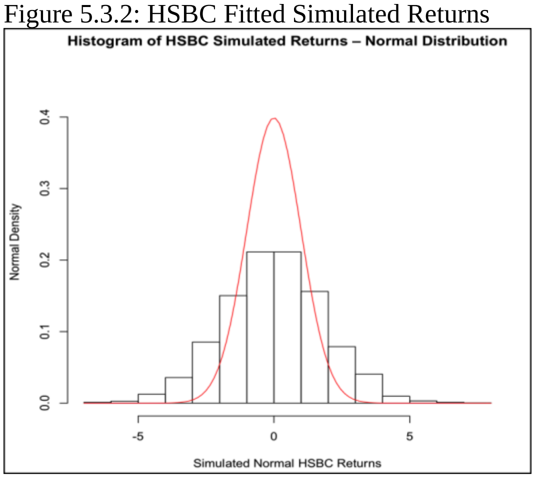


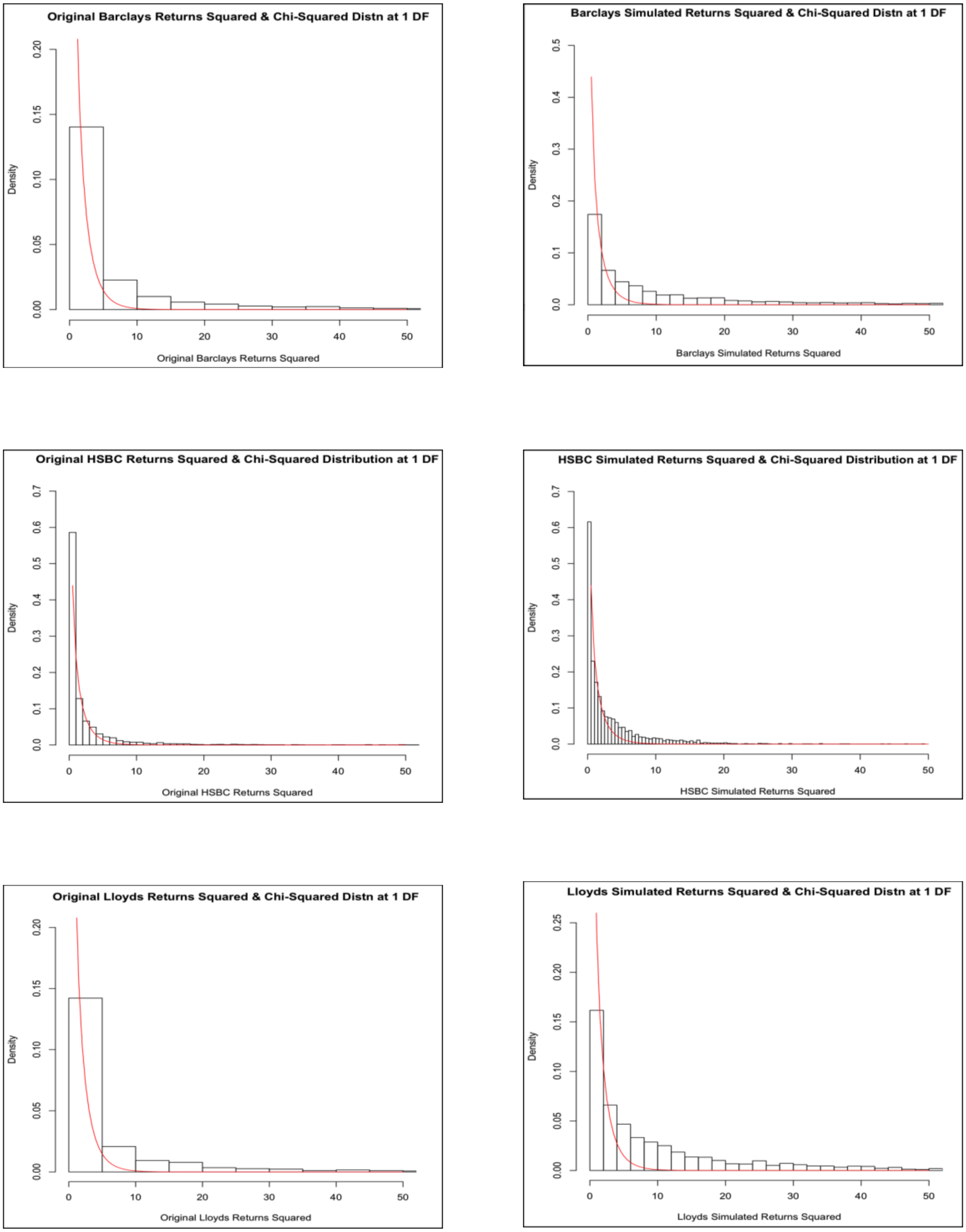
Figure 5.3.2: HSBC Fitted Simulated Returns



Consistent with the quantiles reflected in tables 5.3.5 to 5.3.7, the left tails in the fitted distributions reflect the 10%, 5% and 1% levels. For example, the 1% quantile for Lloyds being -7.859% and reflected in the spread of the left-hand side of figure 5.3.3.

Figure 5.3.4 reflects comparisons between the original and simulated returns for each stock on the basis of an assumed Chi-Squared Distribution. In each case, the outcomes are similar.

Figure 5.3.4: Comparative Graphs of Original versus Simulated Squared Returns and Chi-Squared Distribution:



5.6 Concluding Remarks

This paper provided a BN approach to modelling stock returns. The data was sourced from Bloomberg and included time series of daily returns for three UK banks, namely, Barclays, HSBC and Lloyds. A subsequent portfolio was constructed from the three stocks. Using a degree of qualitative judgement, a DAG was constructed using the evidence presented in the literature with regards factors being important in relation to their impact on stock returns. In this instance, the market and the liquidity factors were selected with the latter being represented by the 3-month LIBOR versus OIS spread.

The DAG suggested conditional dependencies between the factors and stock returns, subsequently verified by conditional independence tests and partial correlations. Whilst low levels of significance were indicated for the liquidity factor, it did still exist and the linear regression models were specified for the returns of each stock. The latter were subsequently used to simulate time series of returns. Summary statistics and quantiles were compared for the actual returns and the simulated returns. Whilst the simulated returns underestimated minimum values, the quantiles were comparable at the 5% and 10% levels. The latter suggests that the underlying Gaussian BN (GBN) could be applied in modelling stock returns and could be further used to estimate quantiles and VaR cut offs. Although it does assume normality, and may be regarded as over-simplifying the modelling issues, its comparable estimations are a positive. An objective in this instance was to suggest a workable alternative to the RiskMetrics approach in deriving VaR. As suggested by Scutari and Denis (2015), a more complex specification may be preferred but relatively simple models often perform better. Indeed, the widely used RiskMetrics approach is convenient to apply and well understood but, I suggest that the GBN is as intuitive and, furthermore, appears to provide prudent estimates for the quantiles used as the cut-offs in VaR calculations. Given that losses were underestimated in the 2008 financial crisis applying VaR techniques of the time, a model resulting in a potential 1% increase in regulatory capital would be an improvement. Based on a portfolio with a notional value of £1 billion, it would result in at least an additional £10 million in regulatory capital.

There are, of course, certain limitations with this technique, not least of which is determining the DAG structure in the first instance. Subsequently, if the structure is ascertained, there may be issues with data being readily available representing the components of the DAG – for example sourcing regular data in relation to balance sheet indicators such as levels of indebtedness or rising levels of delinquencies amongst bank customers. Furthermore, although the network can be altered or updated for new components, as the number of variables grows, the simulation methods may produce less reliable estimations.

References

Allen, F., and Babus, A., (2009), "Networks in Finance," Wharton Financial Institutions Center Working Paper No. 08-07

Anisa, K.N., and Lin, S.W., (2017), "Effect of Socioeconomic Status on Lung Cancer Survival: A Mediation Analysis Based on Bayesian Network Approach." 2017 IEEE International Conference on Industrial Engineering and Engineering Management.

Aquaro, V., Bardoscia, M., Belotti, R., Consiglio, A., De Carlo, F., and Ferri, G., (2010), "A Bayesian Networks Approach to Operational Risk." Physica A: Statistical Mechanics and its Applications, Volume 389, Issue No. 8, pp. 1721-1728.

Barabasi, A.l., Gulbahce, N., and Loscalzo, J., (2011), "Network Medicine: a network-based approach to human disease." Nature Reviews Genetics, Vol. 12, pp. 56-68.

Battiston, S., Puliga, M., Kaushik, R., Tasca, P., and Caldarelli, G., (2012), "DebtRank: Too Central to Fail? Financial Networks, the FED and Systemic Risk." Scientific Reports, Vol. 2, Article number: 541.

Bilio, M., Getmansky, M., Lo, A.W., and Pelizzon, L., (2012), "Econometric Measures of Systemic Risk in the Finance and Insurance Sectors." Journal of Financial Economics, 104, pp. 535-559.

Bisias, D., Flood, M., and Lo, A.W., (2012). "A Survey of Systemic Risk Analytics." Annual Review of Financial Economics, Volume 4, pp. 255-296.

Chan-Lau, J., (2009), "Co-risk measures to assess systemic financial linkages." IMF Working Paper.

Chan-Lau, J., Espinosa, M., and Sole, J., (2009), "On the Use of Network Analysis to Assess Systemic Financial Linkages." Working Paper. International Monetary Fund. (Forthcoming)

Chong, C., and Kluppelberg, C., (2017), "Contagion in Financial Systems." Working Paper, Cornell University Library.

Cowell, R.G., Verrall, R.J and Yoon, Y.K., (2007), "Modelling Operational Risk with Bayesian Networks." Journal of Risk and Insurance, December 2007, Vol. 74, Issue. No. 4, pp. 795-827.

Diao, X., Li, W., and Yeldan, E., (2000), "How the Asian Crisis Affected the World Economy: A General Equilibrium Perspective." Federal Reserve Bank of Richmond, Economic Quarterly Volume 86/2 Spring 2000, pp. 37-38.

Faust, K., and Wasserman, S., (1994, pp. 7), "Social Network Analysis: Methods and Applications." Cambridge University Press. ISBN: 0521387078.

Gandy, A., and Veraart, L.A.M., (2017), "A Bayesian Methodology for Systemic Risk Assessment in Financial Networks." Management Science, Volume 63, Issue No. 12, pp. 4428-

4446.

Giudici, P., and Spelta, A., (2016), “Graphical Network Models for International Financila Flows.” Journal of Business and Economic Statistics, Volume 34, Issue No. 1, pp. 128-138.

Hager, D., and Andersen, L.B., (2010), “A Knowledge Based Approach to Loss Severity Assessment in Financial Institutions Using Bayesian Networks and Loss Determinants.” European Journal of Operational Research, Volume 207, Issue No. 3, pp. 1635-1644.

Haldane, A.G., and May, R.M., (2011), “Systemic risk in banking ecosystems.” Nature, International Journal of Science, Issue 469, pp. 351-355.

Hu, D., Zhao, J.L., Hua, Z., and Wong, M.C.S., (2012), “Network-Based Modelling and Analysis of Systemic Risk in Banking Systems.” MIS Quarterly, Vol. 36, No. 4, pp. 1269-1291.

Hull, J., and White, A., (2013), “LIBOR vs. OIS: The Derivatives Discounting Dilemma.” The Journal of Investment Management, Vol. 11, No. 3, pp. 14-27.

International Monetary Fund (2009a), “Assessing the Systemic Implications of Financial Linkages,” Global Financial Stability Review, pp. 73-110.

International Monetary Fund (2012), “Systemic Risk from Global Financial Derivatives: A Network Analysis of Contagion and its Mitigation with Super-Spreader Tax.” IMF Working Paper, (2012).

Krause, A., and Giansante, S., (2012), "Interbank Lending and the spread of bank failures: A network model of systemic risk." Journal of Economic Behaviour and Organisation, Vol. 83, Issue No. 3, pp. 583 – 608.

Luke, D.A., and Harris, J.K., (2007), "Network Analysis in Public Health: History, Methods and Applications." Annual Review of Public Health, Vol. 28, pp. 69-93.

Markose, S., Giansante, S., and Shaghaghi, A.R., (2012), "Too Interconnected to fail financial network of US CDS Market: Topological fragility and systemic risk." Journal of Economic Behaviour and Organisation, Vol. 83, Issue 3, pp. 627-646.

Martin, N., Fenton N., and Tailor, M., (2005), "Using Bayesian Networks to Model Expected and Unexpected Operational Losses." Risk Analysis, Volume 25, Issue No. 4, pp. 963-972.

Narayan, S., Nicholls, R.J., Clarke, D., and Simmonds, D., (2018), "A Bayesian Network Model for Assessments of Coastal Inundation Pathways and Probabilities." Journal of Flood Risk Management, January 2018, 11:S233-S250.

Scutari, M., and Denis, J.B., (2015), "Bayesian Networks With Examples in R." CRC Press, Taylor and Francis Group, ISBN: 9781482225600.

Shenoy, C. and Shenoy, P.P., (2000), "Bayesian Network Models of Portfolio Risk and Return." Computational Finance, pp. 87-106, The MIT Press, Cambridge, MA.

Soramaki, K., Bech, M.L., Arnold, J., Glass,R.J., and Beyeler, W.E., (2007), "The Topology of Interbank Payment Flows." Physica A:Statistical Mechanics and its Applications, Volume 379, Issue 1, pp. 317-333.

Soramaki, K., Birch, J., and Pantelous, A.A., (2016), "Analysis of Correlation Based Networks Representing DAX 30 Stock Price Returns." Volume 47, Issue No. 4, pp. 501-525.

Soramaki, K., and Cook, S., (2013), "SinkRank: An Algorithm for Identifying Systemically Important Banks in Payment Systems." Economics, The Open-Access, Open-Assessment E-Journal, Vol. 7, 2013-28.

Soramaki, K., and Langfield, S., (2016), "Interbank Exposure Networks." Computational Economics, Vol. 47, Issue No. 1, pp. 3-17.

Thoni, A., Tjoa., A.M., and Taudes, Q., (2018), "An Information System for Assessing the Likelihood of Child Labour in Supplier Locations Leveraging Bayesian Networks and Text Mining." Journal of Information Systems and e-Business Management, 7th February 2018, pp. 1-34.