© *The Journal of Risk and Insurance*, 2007, Vol. 74, No. 4, 795-827

Modeling Operational Risk With Bayesian Networks

R. G. Cowell
R. J. Verrall
Y. K. Yoon

Abstract

Bayesian networks is an emerging tool for a wide range of risk management applications, one of which is the modeling of operational risk. This comes at a time when changes in the supervision of financial institutions have resulted in increased scrutiny on the risk management of banks and insurance companies, thus giving the industry an impetus to measure and manage operational risk. The more established methods for risk quantification are linear models such as time series models, econometric models, empirical actuarial models, and extreme value theory. Due to data limitations and complex interaction between operational risk variables, various nonlinear methods have been proposed, one of which is the focus of this article: Bayesian networks. Using an idealized example of a fictitious on line business, we construct a Bayesian network that models various risk factors and their combination into an overall loss distribution. Using this model, we show how established Bayesian network methodology can be applied to: (1) form posterior marginal distributions of variables based on evidence, (2) simulate scenarios, (3) update the parameters of the model using data, and (4) quantify in real-time how well the model predictions compare to actual data. A specific example of Bayesian networks application to operational risk in an insurance setting is then suggested.

Introduction

Bayesian networks (BNs) have recently been explored as a potential tool for various risk management applications. Its main features of combining subjective opinion with observed data and modeling cause-and-effects make it especially well suited for investigating and capturing the workings of financial institutions. Although its usage has thus far been limited to specific areas (e.g., it has been used for credit risk scoring

R. G. Cowell and R. J. Verrall are associated with Faculty of Actuarial Science and Insurance, Sir John Cass Business School, City University, 106 Bunhill Row, London, EC1Y 8TZ, UK. Y. K. Yoon is with the Insurance Risk Specialist Team, Insurance Supervision Department, Bank Negara Malaysia, 15th Floor, No. 4 Jalan Sultan Sulaiman, 50000 Kuala Lumpur, Malaysia. The first author can be contacted via e-mail: r.g.cowell@city.ac.uk

by banks) its application to wider enterprise risks is being increasingly documented, especially in the area of operational risk (OR).

Chapter 14 of Alexander (2003) provides a brief introduction to modeling OR using BNs via a banking example. Marshall (2001), Cruz (2002), and Hoffman (2002) give brief overviews of BNs and where they fit into the whole framework of OR modeling. There is also an illustrative albeit high-level discussion on causal modeling using BNs via a banking example in King (1999).

The main purpose of this article is to consider two aspects of the application of BNs to OR in greater detail than has so far appeared in the literature. These are the theory and techniques of model updating, and the subject of model assessment. OR takes place in a dynamic setting, with more information becoming available as time progresses. Hence, it is useful to update the models used for OR to take account of this flow of information. This is feasible within the setting of BNs and was briefly mentioned in King (1999); the section on "Updating the Probabilities With New Data"of this article gives more details of how this can be implemented. In any modeling exercise, it is essential to check that the model used provides a reasonable representation of the actual experience. Again, this is best done dynamically in the OR setting, as more information arrives; this is covered in the section on "Model Assessment" of this article.

The article is set out as follows. In the section on "Changes to Supervisory Regimes as a Driver for Operational Risk Modeling," we describe recent developments in the supervision of financial institutions and how this has encouraged greater efforts in OR modeling. In the "Current Approaches to Modeling OR" section, we give a brief introduction to modeling approaches that have been used in the context of OR. The approach used in this article is that of BNs, which are introduced in the next section and applied in a general risk management context in the subsequent section. In the sections on "Updating the Probabilities With New Data" and "Model Assessment," we consider two aspects of BNs and OR that have not been covered in any detail in the literature to date: updating the models and monitoring the appropriateness of the model. In the section on "Insurance Fraud Risk Example," we briefly mention an example of how BNs can be used specifically for OR modeling in an insurance setting. The concluding section contains a discussion of the issues raised in this article.

CHANGES TO SUPERVISORY REGIMES AS A DRIVER FOR OPERATIONAL RISK MODELING

Developments in the quantification of OR has, to a significant extent, been driven by changes in the supervisory regimes for financial institutions. These changes have increased the level of supervisory scrutiny on OR and how it is managed by financial institutions—a reflection of the centrality of OR in high-profiled corporate failures in recent decades. This has become an impetus for financial institutions to develop OR models as a means to demonstrate good management and financial strength.

The banking sector paved the way when the Basel Committee for Banking Supervision (BCBS) of the Bank for International Settlements (BIS) began work in 1999 on a new framework for capital adequacy of banks,[1] also known as "Basel 2." Basel 2 comprises

[1] BCBS (1999).

three "pillars." Pillar I delineates the requirements for a minimum level of capital that banks need to hold depending on their specific risk exposure. Pillar II outlines measures for supervisory review of banks to ensure that the level of capital held and risk management is commensurate with each bank's risk profile. Pillar III provides a framework for market discipline via disclosure.

Under Pillar I, explicit charges to supervisory capital were introduced for banks' exposure to credit risk and OR. These are to be either calculated as a percentage of gross income or an internal model could be used provided certain criteria have been met. It is hoped that potential savings in capital required by the supervisor would provide an incentive for banks to develop internal models with accompanying improvements in risk management.[2]

The insurance sector is also overhauling its capital adequacy framework to be more reflective of insurance companies' risk exposure. The European Commission (EC) issued a consultation for a review of its rules on prudential supervision of insurance in 1999. Also known as "Solvency II," this review aims to produce a system that would "establish a solvency margin requirement that is better matched to the true risks."[3] Solvency II will follow the basic three-pillar structure of Basel 2 albeit modified to suit the insurance industry.[4] Similar to Basel 2, Solvency II will have a solvency capital requirement (SCR), which is intended as a "target capital": a prudent level of capital that will capture all material risks and allow supervisors sufficient time to intervene in adverse circumstances. Whereas the method to arrive at the SCR is still being deliberated, the discussions show that OR will be a key component.[5]

The International Association of Insurance Supervisors (IAIS), the insurance counterpart of BIS, has issued some instructive guidance to its members on supervision of capital adequacy and solvency. According to the IAIS, solvency regimes should have regard for OR.[6] Supervisors are to intervene when an insurance company's capital breaches a predetermined "solvency control level,"[7] whose sensitivity to risk may be based on stress tests.[8] With regards to these stress tests, insurers should be able to demonstrate that they have sufficiently considered OR.[9]

We have seen that OR will feature prominently in emerging capital adequacy regimes. The increased complexity of such regimes would require more intense supervision, not least in OR measurement and management. Pillar II of Basel 2 stipulates that "Supervisors should review and evaluate banks' internal capital adequacy assessments and strategies, as well as their ability to monitor and ensure their compliance with regulatory capital ratios."[10] This would also serve to assess that banks meet

[2] See §46 and §677 of BCBS (2004) and chapter 2 of van den Brink (2002).
[3] See §2.2 of EC (1999).
[4] See §7 of EC (2003).
[5] See §19 of EC (2005b) and §10.45 of Committee of European Insurance and Occupational Pension Supervisors (CEIOPS) (2005).
[6] See §24 of IAIS(2002).
[7] See §30 of IAIS(2002).
[8] See §13 of IAIS(2003a).
[9] See §43 of IAIS(2003b).
[10] See p162 of BCBS (2004).

the criteria for usage of the Standardized Approach or Advanced Measurement Approach[11] for calculating OR capital. Findings from a 2002 report[12] (also known as the "Sharma report") commissioned by the EC as part of the Solvency II review showed that most insurer failures resulted from problems with management. This led to inadequate controls, which left insurers vulnerable to external events. Thus, the predominant root cause of insurer failures is operational. The report recommends that the Solvency II framework requires supervisors to address the underlying problems before they occur via supervisory tools such as assessment of risk management and corporate governance of insurers.[13] To facilitate this, Solvency II will broaden insurance supervisors' powers to enable them to carry out such assessments and take certain actions (such as imposing capital add-ons) where necessary.[14] Principle 13 of the IAIS Principles of Capital Adequacy and Solvency requires supervisors to consider adequacy of companies' internal risk assessment and management, alongside more established aspects of solvency assessment such as accuracy of valuations and compliance with minimum levels.[15] This has lead banks and insurers to develop more sophisticated quantitative models that serve not only to satisfy supervisory authorities of the adequacy of their capital but also to demonstrate strong risk management.

Whereas models have been developed for the purpose of internal management, it is clear from the references made to the new supervisory regimes by current literature on OR that there has not been a shortage of OR models proposed in response to such supervisory requirements. This article adds to the existing corpus by discussing the use of BNs as a framework for modeling OR. We shall see that this is an efficient and intuitively coherent methodology for incorporating expert input. In addition, BNs are useful for capturing causal dependence. This satisfies a vital requirement of any OR modeling framework: the ability to model causation (this is discussed further below). Developments in the field of graphical models (of which BNs are an example) have led to the development of several user-friendly software tools that automatically carry out the complex calculations for making inference from complex networks of causal relationships. Underlying BNs is the Bayesian statistical framework that enables the combination of subjective input with empirical observations. This lends itself very well to situations with a high degree of uncertainty, and where data are costly or sparse—two intrinsic features of OR.

CURRENT APPROACHES TO MODELING OR

Although this section is not intended as a complete survey of current models available for modeling OR, it would be instructive to briefly mention main categories of model to see where BNs fit in the spectrum.

Value-at-Risk and Ruin Theory

The broad approach to the setting of economic capital at the moment revolves around obtaining a single figure that is often defined as a high quantile of a loss distribution

[11] See §660–676 of BCBS (2004).
[12] See §1.3 of Conference of Insurance Supervisory Services of the Member States of the European Union (2002) (Sharma report).
[13] See chapter 6 of the Sharma report.
[14] See §2.1 of EC (2005a), §1 of EC (2005b) and §10.153 of CEIOPS (2005).
[15] See §46 of IAIS (2002).

that is projected over a required period based on past data. This has its origins in Value-at-Risk (VaR) used by banks to measure their exposure to market risk. Thus, for example, BCBS requires that "a bank must demonstrate that its operational risk measure meets a soundness standard comparable to that of the internal ratings based approach for credit risk, (i.e., comparable to a 1-year holding period and a 99.9th percentile confidence interval)."[16] A whole body of literature exists detailing the workings of VaR in setting economic capital for market risks and credit risks (see, e.g., Jorion, 2001; Dowd, 1998).

Presently, if VaR is calculated for OR, the approach commonly ranges from the residual approach (firmwide VaR minus market VaR and credit VaR) to setting the VaR as a multiple of the standard error. Some of the more advanced approaches use models that essentially try to fit a loss distribution from which a VaR figure can be extracted.[17] With market VaR and credit VaR this is more straightforward as an underlying normality is assumed to exist in the data. However, OR losses are nonnormal, positively skewed, and have fat tails. The tails of loss distributions are the main concern in allocating economic capital. In modeling the severity of losses, a combination of approaches has been used, including the use of extreme value theory.

An analogue of VaR to be found in insurance is ruin theory (now often used as part of dynamic financial analysis).[18] In ruin theory, simulations are conducted using compound distributions of claims and expense payouts to arrive at ruin probabilities, given an initial level of "surplus"—an equivalent term for capital in insurance business. These could then be used to derive initial surplus levels that would not be depleted (a situation described as "ruin") in, say, 99.9 percent of final outcomes (perhaps year-end surplus levels). The usage of compound distributions arguably offers greater flexibility compared with the VaR approach.

Causal Modeling of OR

The literature reviewed suggest that OR needs to be dealt with in terms of causes rather than effects (i.e., the loss event), as a financial loss may have various underlying causes, which may or may not be operational. This can be seen from the definition of OR by BCBS ("Operational risk is defined as the risk of loss *resulting from* inadequate or failed internal processes, people and systems or from external events."[19]) and the proposed loss event type classification.[20] The Operational Risk Working Party of the Institute of Actuaries proposed a framework for analysis of OR based on cause and consequence where they stressed the importance of working in terms of causes rather than consequence to avoid double-counting or omissions.[21] The Sharma report

[16] See §667 of BCBS (2004).
[17] Several recent texts on OR give descriptions of the methodology, for example, Cruz (2002), Hoffman (2002), and King (1999). See also chapter 19 of Jorion (2001) and p. 198 of Dowd (1998).
[18] An example of application to OR can be found in the section on "Model Assessment" of Tripp et al. (2004).
[19] See §644 of BCBS (2004). Emphasis added.
[20] See Annex 7 of BCBS (2004).
[21] See §3.5.5 of Tripp et al. (2004).

mentioned earlier also highlighted the importance of analyzing the full causal chain of insurer failures using cause–effect methodologies.[22]

In addition to ensuring appropriate risk classification, causal modeling of OR is vital for the understanding of the how the risk of OR losses arise within the structure and operations of the organization. It also provides a basis on which management may intervene to achieve the desired alteration in risk profile. Techniques currently used for causal modeling in risk management include *time-series analysis* and *econometric* modeling, for example an *autoregressive* time series model with *conditional heteroskedasticy* (ARCH). *Factor analysis* is also used to decompose uncertainty in profit and loss figures into various causal factors of manageable sizes. Chapter 8 of Cruz (2002) provides a brief overview of these methods. Nonlinear and nonparametric models, including neural networks, fuzzy logic, and BNs, are increasingly popular for causal modeling as they offer greater flexibility.

Expert Input and Nonlinear Methodologies

A major problem with any attempt to model OR losses is the lack of data. Internal data of low frequency events are rarely sufficient to model the loss distributions to the required accuracy. The use of parametric loss distributions requires parameter estimation based on the existing data. Even for extreme value theory, the small sample sizes result in the shape of the tails being very sensitive to inclusions or exclusions of single events, implying a greater degree of subjectivity than may at first be apparent.

In an attempt to overcome the lack of internal data, the use of external data has been suggested as a solution. Cruz (2002) describes a method for pooling data from different locations called "frequency analysis" and Frachot and Roncalli (2002) propose using linear credibility. There have also been some efforts made at an industry-wide level to collect data in order to provide a shared source of external data. Merging internal and external data are not a straightforward exercise—especially in the case of operational loss data. There are qualitative problems such as quality of data from external parties, the dissimilarity of OR management practices across firms, lack of detailed breakdown of data and lack of up-to-date data that may pose difficulties to using external data to set internal capital requirements (which are essentially prospective).

The lack of data and complexity of operations suggest the inclusion of *expert* input. An expert in this case would be anyone whose knowledge and expertise enables him/her to make sufficiently credible conjectures about how company operations affect the company's risk profile. Such input can be used as a proxy for data and may yield valuable information about the company operations that is difficult to capture from data alone. The challenge of the modeler, then, is to incorporate such input into the overall OR modeling framework.

It has been found that qualitative information, such as management decisions, competencies, and preferences, can be better incorporated into a measurable (and hence qualitative) framework using nonlinear methods. Some of these include *fuzzy logic, neural networks, system dynamics*, and *BNs*.

[22] See §1.3 and §3.2 of the Sharma report.

Fuzzy logic uses a multivariate logical set that recognizes that human decisions are often not binary (e.g., Yes/No, Hot/Cold) by allowing gradations in its formulations (e.g., rather hot, very hot). Hoffman (2002) has a brief case study on how this has been used in a bank in its OR management.[23]

Although not strictly a method that is used for incorporation of qualitative opinion, *neural networks* are useful for modeling complex relationships between variables that would be difficult to do using linear methods. The network consists of nodes with values of input, output, and intermediate variables. Data mining techniques are used to "train" the model by using complex algorithms that learn the relationships between the variables. The model is then calibrated such that its output is as close to the actual data output as possible. A drawback of this approach is its heavy reliance on the availability of data. A brief example of its application to OR can be found in Hoffman (2002), although this involves a nonfinancial institution.[24]

System dynamics was developed by Jay Forrester of the Massachusetts Institute of Technology and has been suggested by Miccolis and Shah (2000) and Shah (2001) as an approach to modeling OR. This approach involves using expert input to map a network of cause-and-effect relationships between variables affecting the OR of a business unit. The relationship between each cause-and-effect set of variables is then quantified by combining data and expert input to obtain a plot on two axes (one for each of the cause-and-effect variables).

BAYESIAN NETWORKS AND GRAPHICAL MODELS

Graphical models are a combination of probability theory and graph theory that in recent years have become popular (in particular in the field of artificial intelligence) in the statistical modeling of complex interactions between random variables. The 1980s saw the development of computational algorithms that exploited the conditional independences of the graphical models, enabling the efficient computation of, among other quantities, posterior marginal distributions. BNs are a special class of graphical model that may be used to model *causal* dependencies between random variables. To illustrate the use of graphical models, consider the three graphs in Figure 1.

A and B are called *nodes* and represent random variables A and B. In Figure 1(a), the directed edge from A to B models a causal relationship between A and B. What we mean by this is that a change in what is known about A *causes* a change in what is known about B. This change is usually the result of new information arriving about A. This new information is called *evidence*. When variables are connected in this way, we call the variable from which the edge originates the *parent* and the variable to which the edge leads the *child*. When the edge between the nodes are not directed, as illustrated in Figure 1(b), then no causation is implied but rather that some "weaker" form of association (e.g., correlation) exists between A and B. Using the same sort of descriptive language, A and B are called *neighbors*. Finally if, as in Figure 1(c), no edges exist between A and B, then A and B are *independent* that is, knowledge about the state of A is not informative about the state of B, and vice versa.

---

[23] See pp. 326–330.

[24] More specifically, the National Aeronautic and Space Administration (NASA) in the United States. See pp. 299–300.

FIGURE 1
Three Simple Graphical Models

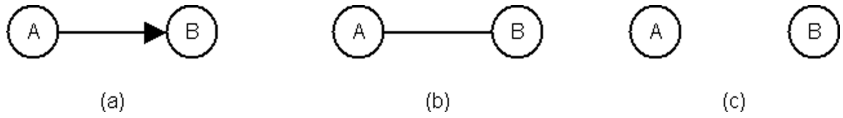


FIGURE 2
A Directed Acyclic Graph on Four Nodes

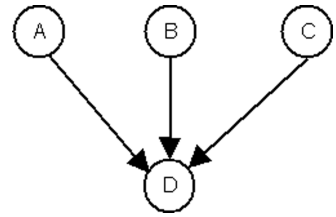


Figure 1(a and b) represents a joint probability of A and B, but express this joint probability differently. In Figure 1(a), the causal relationship that exists between A and B means that the joint distribution can be expressed as a product of the probability of A and the probability of B conditional on A or simply written: $P(A)P(B \mid A)$. Since no such relationship is defined in Figure 1(b) this graph only expresses the joint distribution itself: $P(A, B)$. Finally, Figure 1(c) represents a joint distribution that factorizes: $P(A, B) = P(A)P(B)$.

In general, graphical models comprise a network of such nodes with edges to connect variables that have some form of relationship, whether of correlation or of causation. Causal BNs on the other hand express causal relationships between random variables and involve nodes connected by directed edges. The lack of an edge between two nodes represents a *conditional independence* relationship between them, though not necessarily *independence*; for example, two unjoined variables might be dependent because they have a common neighbor or parent.

BNs belong to a subset of graphical models that are known as *directed acyclic graphs* (DAGs). DAGs are constructed with relationships such as those in Figure 1(a) as its basic building block. These building blocks are arranged in such a way that the variables are not cyclic, that is, starting from any node and moving along the edges in the directions implied, it is impossible to return to a node previously visited. Hence, the term "acyclic." This acyclic property captures an intuitive feature of causal modeling (for example, the future does not influence the past).

Associated with each node of a DAG that has at least one parent is a set of conditional probabilities. These describe the behavior of the node *conditioned* on all its parents. This may be written as $P(X \mid pa(X))$ where $pa(X)$ represents the parents of X. For example, the parent set for node D in Figure 2 is the set of nodes $\{A, B, C\}$.

From the graphical structure of the BN one can read of the factorization of the joint distribution in terms of this set of conditional distributions: if $V = \{X_i : i = 1, \ldots, n\}$

represents the set of n random variables in the BN then the factorization of the joint distribution $P(V)$ is given by

$$P(V) = \prod_{i=1}^{n} P(X_i \mid pa(X_i)).$$

This factorization expresses the conditional independence properties inherent in the structure of DAGs, and is exploited in the efficient inference algorithms mentioned earlier, which reexpress the factorization on the DAG in terms of a factorization on another graphical structure called a *junction tree*. Typically for modern BN software the user only sees the DAG, the marginal densities of each node and how they change as evidence is entered—the technicalities behind the calculations are kept hidden and the user need not worry about them. For further theoretical background, detailed proofs, extensions of the methods, and examples, see Cowell et al. (1999).

USING BAYESIAN NETWORKS FOR CAUSAL MODELING AND CAPITAL ALLOCATION

In this section we illustrate how BNs may be used for causal modeling and capital allocation by using a hypothetical example of an Internet on-line business. The next two sections use the same example to demonstrate how the model parameters can be updated to take into account new data and some criteria for model selection. The methodology set out here is applicable wherever the risk management context requires similar model features. One such context is OR and an OR-specific model is suggested in the "Insurance Fraud Risk Example" section.

A fictitious medium-sized insurance company (BayeSure Insurance Co. or BSI) has decided to set up an on line business (called BSNet) so that customers may purchase insurance via the Internet. After a year of operating BSNet, the management decided to put the IT managers, marketing personnel, and various BSNet user departments together with some statisticians from Risk Management Unit to set up a BN to serve a two-pronged strategy:

- Identify causation of events such as transaction downtime, server downtime, and application failure, and
- Help the management to decide how much risk capital to allocate to cover the loss from such events in all but the most extreme of scenarios.

Defining the Structure of the Bayesian Network

After some deliberation the working committee was able to put together a structure that looks like Figure 3. A list of the variables employed in this model is given in Table 1. The set of values for each variable is also given along with the abbreviations to be used throughout this article. We now describe the causal structure of the BN and how it relates to the business environment.

The three main causes of loss are transaction downtime (TD), data loss (DL), and server downtime (SD). These can have varying degrees of severity but can be grouped into half a day or a full day to get a server or network up and running, or in the case of data loss, either 50 percent or complete data loss. Based on data gathered in the first

FIGURE 3
DAG of On line Business Example

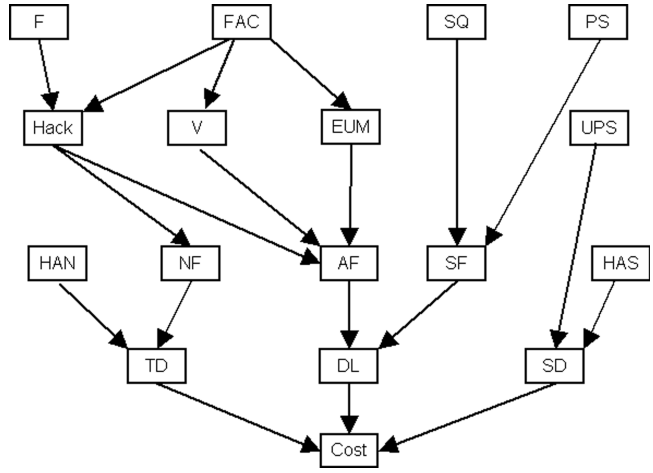


TABLE 1
The Random Variables of the On line Business Example

No.	Description	Values	Abbreviation
1.	Application failure	Application corruption, Lockup, OK(No failure)	AF
2.	Cost of losses from network risk (Cost)	0.0 m, 0.5 m, 1.0 m 1.5 m, 2.0 m, 2.5 m	Cost
3.	Data loss	0%, 50%, 100%	DL
4.	End user modification	Yes, No	EUM
5.	Firewall	Application Proxy, Packet Filter	F
6.	File access control	High, Low	FAC
7.	High-availability network	Yes, No	HAN
8.	High-availability server	Yes, No	HAS
9.	Hacker attack	Yes, No	Hack
10.	Network failure	Yes, No	NF
11.	Power surge	Yes, No	PS
12.	Server downtime	0 day, 0.5 day, 1 day	SD
13.	Server failure	Yes, No	SF
14.	Server hardware quality	High, Low	SQ
15.	Transaction downtime	0 day, 0.5 day, 1 day	TD
16.	Uninterrupted power supply	Yes, No	UPS
17.	Virus attack	Yes, No	V

year of operation with some additional expert input by the marketing department the total cost as a result of various combinations of such events was estimated. This would include, for example, the cost of repairs, lost business opportunities during the downtime, wages paid to idle staff, and costs to recover loss data. All this is combined into one variable, cost, representing the bottom-line effect.

The cause of TD is network failure (NF) but there is a mitigating factor: whether or not a high-availability network (HAN) was employed in the running of BSNet. SD is caused by server failure (SF), which in turn is due to power surges (PS) and the server quality (SQ). Again there are mitigating factors here: the availability of uninterrupted power supply (UPS)—perhaps an internal generator—and the use of a high-availability server (HAS).

DL can be caused by either SF or application failure (AF). AF has three main causes: (1) modifications made by end users in BSI to the system, either intentionally or accidentally (EUM); (2) virus attacks (V); and (3) malicious hacking by external parties (Hack). These three events are largely controlled by the level of file access given to various parties (FAC). The type of firewall (F) used will also affect the ease with which malicious hackers can access the system. Malicious hacking not only causes AF but NF as well.

This model has been defined to reflect a "holding period" of 1 week. Thus, for example, the marginal distribution for the variable cost would be the probability distribution of NR costs incurred by the company over the period of 1 week. The choice of holding period is arbitrary as far as the mechanics of the model is concerned. The only consideration in this case would be the availability of data. As the system has been operating for only 1 year, it makes more sense to use 52 sets of weekly data than one set of annual data.

Prior Specification

At this point the DAG will need to be populated with the prior probabilities at each node. These would be the unconditional prior distribution for nodes without parents and conditional prior distributions for child nodes. For each prior, we would need the probabilities for each configuration of the combination of states of variables involved.

These can be determined by:

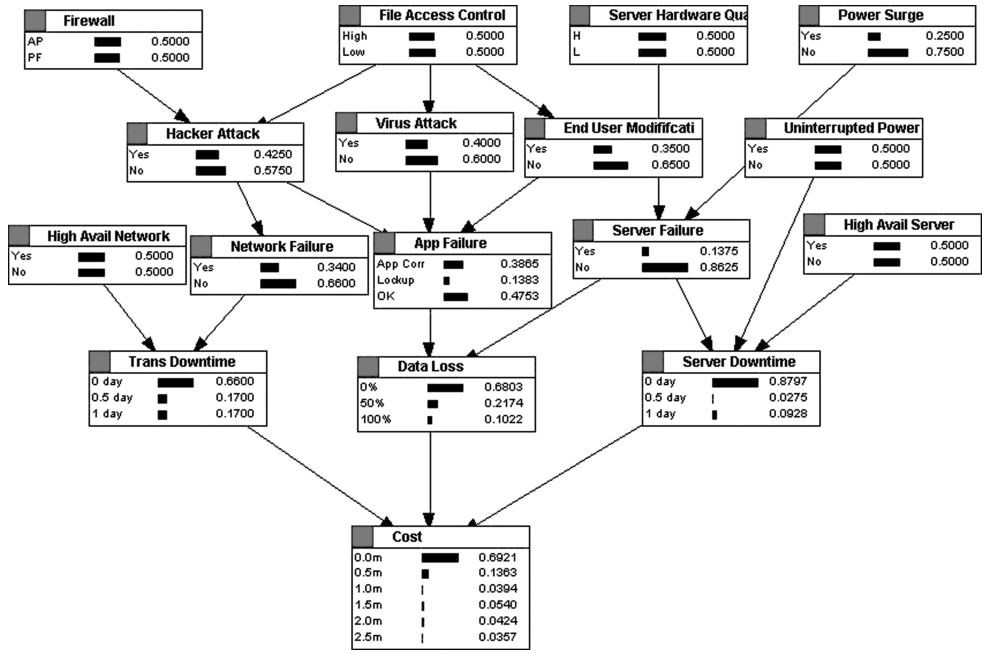
- *Subjective opinion of the expert.* Experts are interviewed in a series of questionnaires to arrive at quantified conclusions of the probabilities. Of course, sufficient confidence in the accuracy of the expert's advice is a prerequisite to use this method.
- *Maximum likelihood estimation with complete data.* For unconditional priors, this would simply be the proportion of occurrence between the various states of the variable. For conditional priors this method would entail taking a ratio of frequency of the event to the frequency of the parent configuration, for example:

$$P(SF_y \mid SQ_h, PS_y) = \frac{n(SF_y, SQ_h, PS_y)}{n(SQ_h, PS_y)}.$$

- *Maximum likelihood estimation with incomplete data.* The EM algorithm may be used to estimate the conditional and unconditional priors (Lauritzen, 1995).

In practice, there will not be a clear dichotomy between these methods as the experts will also rely on past data but tempered with experience and knowledge regarding their applicability for future events. The section on "Updating the Probabilities With New Data" discusses the revision of the network probabilities in the light of new data.

FIGURE 4
Prior Marginal Distributions



Having performed these exercises, the company arrives at the probabilities detailed in Appendix A where, for flexibility in the model, various unconditional priors have been assigned uniform distributions. These will be treated as input variables to reflect the actual state of BSNet once "evidence" as been entered. Power surge, obviously, is outside of the control of the company; thus, past data have been used to arrive at the probabilities and it will not be treated as an input variable. Theoretically, any node can be treated as an input variable. To distinguish the types of input, when evidence is entered in the other nodes it will be treated as a stress or scenario test input.

Inference in DAGs commonly involves the incorporation of evidence entered at some nodes of the graph, and "propagating" this evidence through the network to the other nodes and evaluating their marginal posterior distributions. If no evidence is entered, this procedure yields prior marginals for the individual variables: Figure 4 shows the prior marginals for all nodes in the on line business example.[25] Typically in BN software evidence is entered on a node by use of a mouse click to select a state (corresponding to an observation of the variable in that state). For example, to enter FAC = "High," usually all that is needed is a click on the bar representing the marginal probability of the state "High" at the FAC node chart. This automatically sets the bar to 1 and posterior distributions may then be evaluated and displayed, usually automatically.

[25] Figures 4–7 were produced using the software XBAIES, written by one of the authors (RGC), and may be downloaded from his home page at http://www.staff.city.ac.uk/~rgc.

Risk Capital Allocation

We now come to one of the main applications of BNs in the modeling of risks. Having set up the model as above, the company would like to decide on the amount of financial capital to allocate toward protecting the company from all but the most extreme cases of the identified events.

At this point the model can be configured to reflect the actual known state of certain variables to reflect the actual position of the company. For example, the IT Department may provide the information that Application Proxies are used 24 hours as the firewall for the network and that High Availability Networks have been purchased and implemented for round-the-clock availability. Such information is incorporated by inserting evidence at the relevant nodes (F and HAN). Once this is done, the rest of the propagation then carries on as before. The marginal probabilities obtained now will reflect the impact of the evidence entered.

Suppose the set of evidence in Table 2 represents the status of BSI's BSNet system. When this evidence is incorporated into the network and propagated, this yields the set of posterior marginals shown in Figure 5. No evidence was introduced for power surge, as this remains uncertain for a future period.

It is now fairly straightforward to determine the risk capital to be allocated to cover NR in BSNet for 95 percent of cases over a holding period of 1 week. This is simply calculated by linearly interpolating to obtain the 95th percentile of the probability distribution for cost. In this case, the result is 0.32 million.

Scenario Testing and Causal Analysis

The model can be used to test various scenarios to help management in optimizing its risk profile. For example, if BSI wishes to reduce costs by reducing the level of file access control to "Low," the effect of this action can be easily investigated by entering the evidence FAC = "Low" into the network. The evidence is then propagated and the resulting marginals observed as before (this is shown in Figure 6). We can deduce that the trade-off of the lower costs is an increased capital requirement of 0.97 million. We assume that the 95 percent confidence level is maintained.

TABLE 2
Evidence Entered

Variable	Evidence
F	Application proxy
FAC	High
HAN	Yes
HAS	Yes
SQ	High
UPS	Yes

FIGURE 5
Posterior Marginals Based on Evidence Given in Table 2

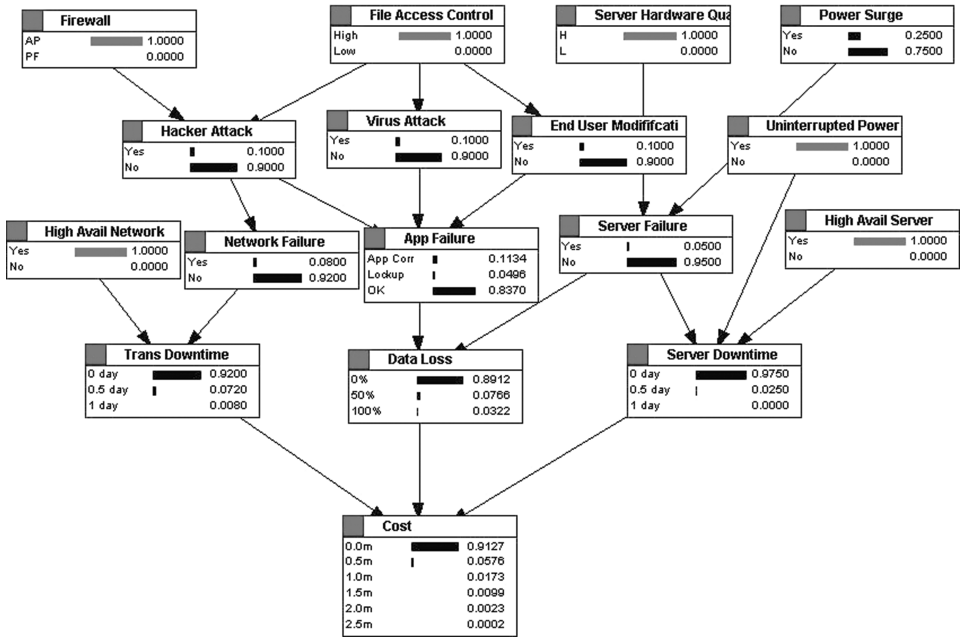


FIGURE 6
Posterior Marginals of Scenario Testing

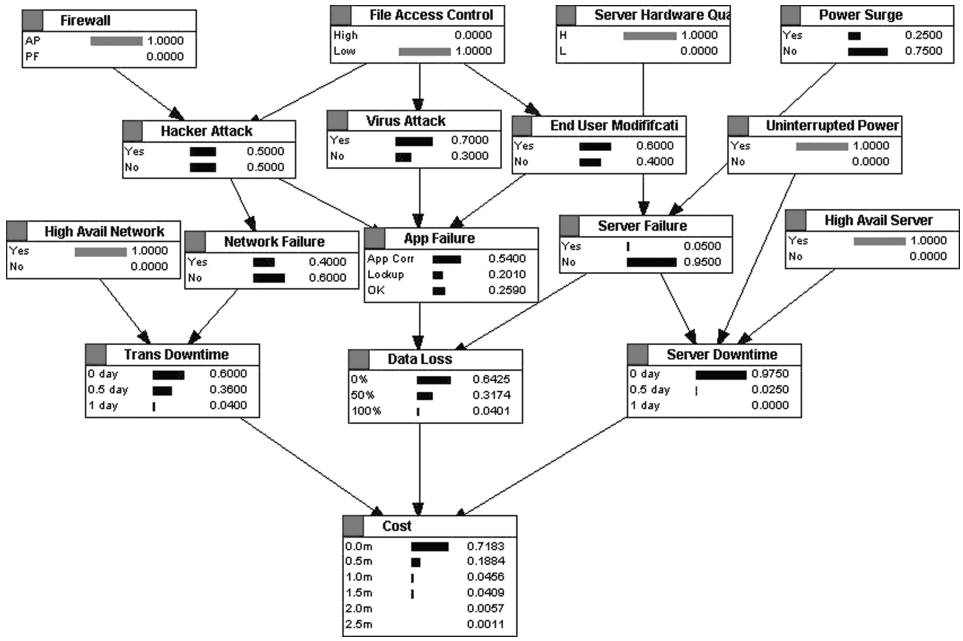
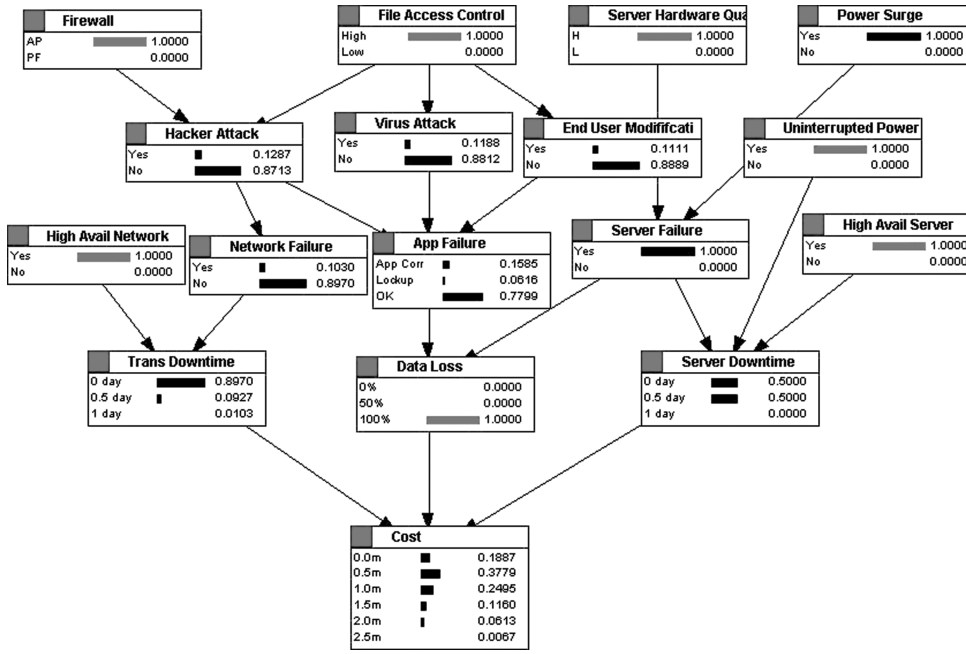


FIGURE 7
Cause and Effect of a 100 Percent Data Loss



The company may wish to perform some stress testing by investigating the impact of adverse events. Supposing the management is interested in examining the effects of situation of complete loss of data. This can be done by setting the node DL to "100 percent." The effect is then propagated again and each node is updated accordingly. The final result can be seen in Figure 7.

Note the shift in the marginal probability distribution of the cost. We can now read off a few statistics from this new distribution. The expected cost is now 0.75 million with standard deviation 0.57 million. The 95th percentile has now increased to 1.64 million.

Note also the changes in the other nodes. For example, we now have $P(SF_{Yes})$ and $P(PS_{Yes})$ equal to 1. Conversely, the probability of "No" has increased significantly for Hack, V, EUM, and NF. Similarly, AF is overwhelmingly "OK." We can read from this that server failure is the definite cause of data loss. Consequently, we also deduced that a power surge was the definite cause of server failure. As a result, other events that might have been the causation of data loss have been *explained away*, giving the increase in "No" probabilities for these nodes.

UPDATING THE PROBABILITIES WITH NEW DATA

The prior probabilities in the on line business example given in Appendix A are *point probabilities* and are not suitable for updating in the light of new data that may become available. However, it turns out that if the expert opinion is quantified (at least to a good approximation) in the form of Dirichlet distributions, then it is possible to perform Bayesian inference efficiently as new data arrives.

TABLE 3
Format of the Table for Recording Mean and Variance of Probability Values From a Prior Elicitation Exercise

	C_y		C_n	
Parent	$E(\theta_y)$	$\text{Var}(\theta_y)$	$E(\theta_n)$	$\text{Var}(\theta_n)$
A_y, B_y				
A_n, B_y				
A_y, B_n				
A_n, B_n				

Specifically, for a node X having k states, an independent Dirichlet distribution on k parameters is associated with each parent configuration in the conditional probability distributions $P(X \mid pa(X))$. Under complete data, standard Bayesian conjugate updating will result in posterior distributions that are also Dirichlet in form. We illustrate the procedure here; further details may be found in chapter 9 of Cowell et al. (1999).

Suppose we are considering a prior for the probability $P(C \mid A, B)$, where all three variables are binary. Four independent Beta distributions will be required to specify $P(C \mid A, B)$. In practice, the Beta distributions would usually not be specified directly by the expert. Instead, their parameters can be specified in the following way. First we gather the expert's opinion on the mean and standard deviation of the probabilities of the values in the variable. Some experts who are not familiar with statistical concepts can be asked to quote a best estimate and the range of most likely values. The best estimate can then be taken as the mean and the range can be taken to encompass two standard deviations about the mean, from which the variance can be easily obtained. The results can be stored as shown in Table 3.

Then, for each parent configuration we have the following pairs of simultaneous equations:

$$C_y: \quad E(\theta_y) = \frac{\alpha_y}{\alpha_y + \alpha_n} \quad \text{and} \quad \text{Var}(\theta_y) = \frac{\alpha_y \alpha_n}{(\alpha_1 + \alpha_2)^2(\alpha_y + \alpha_n + 1)} = \frac{E(\theta_y)(1 - E(\theta_y))}{\alpha_y + \alpha_n + 1}$$

$$C_n: \quad E(\theta_n) = \frac{\alpha_n}{\alpha_y + \alpha_n} \quad \text{and} \quad \text{Var}(\theta_n) = \frac{\alpha_y \alpha_n}{(\alpha_1 + \alpha_2)^2(\alpha_y + \alpha_n + 1)} = \frac{E(\theta_n)(1 - E(\theta_n))}{\alpha_y + \alpha_n + 1}.$$

Either pair may then be solved for α_y and α_n. For a Dirichlet distribution with k parameters, there will be k pairs of equations, typically yielding k different precisions between them. In this case it is more conservative to take the *lowest* precision to be the precision of the Dirichlet distribution (it thus has a higher variance), and from this precision the individual parameters of the Dirichlet distribution by multiplying the precision by the expert prior means.

As an example, consider the conditional probability $P(\text{NF} \mid \text{Hack}_y)$ underlying the node NF. Initially, we have to specify a Dirichlet prior distribution (again a Beta prior

TABLE 4
Estimating Priors of a Beta Distribution

	NF = "Yes"				
Parent	$\mu_{P(NF_y \mid Hack_y)}$	Range	$\sigma^2_{P(NF_y \mid Hack_y)}$	α_y	α_n
Hack = Yes	0.8	0.70–0.90	0.1^2	12	3
	NF = "No"				
Parent	$\mu_{P(NF_n \mid Hack_y)}$	Range	$\sigma^2_{P(NF_n \mid Hack_y)}$	α_y	α_n
Hack = Yes	0.2	0.05–0.35	0.15^2	4.88	1.22

TABLE 5
Bayesian Updating With Data

	NF = "Yes"				
Parent	Prior Mean	Data	α_y	$\alpha_y + n_y$	Posterior Mean
Hack = Yes	0.8	3	4.88	7.88	0.651
			NF = "No"		
Parent	Prior Mean	Data	α_n	$\alpha_n + n_n$	Posterior Mean
Hack = Yes	0.2	3	1.22	4.22	0.349

distribution with two parameters, since this is a binary node). This involves specifying the parameters of the Beta distributions. There will be two possible sets of parameters: one for the case NF = "Yes" and another for the case NF = "No." Table 4 shows an example based on the two pairs of parameters, the second has the lowest precision (at 6.1) and so we take as the prior the Beta (4.88, 1.22) distribution.

Further, suppose that over the course of a year, the system was penetrated by malicious hacker in 6 out of the 52 weeks and half of those resulted in network failure. The Beta parameters then can be updated easily by adding the counts directly to the relevant parameters and the latest probability estimates can be obtained from the mean of the resulting posterior Beta distribution. The results are shown in Table 5.

The results obtained are fairly intuitive: the initial prior estimate of network failure in the event of a hacker attack was reduced from 80 percent to 65 percent when the data, which indicate a 50 percent occurrence, was incorporated. Similar updates can be performed simultaneously for all the nodes with the complete data set. We see that Bayesian updating for BNs given new data is fairly easy to perform and would be cost effective to conduct regularly as new cases arrive. This whole process of Bayesian updating is sometimes also described as *learning*. When data are analyzed using a BN *without* learning the probabilities remain unchanged.

MODEL ASSESSMENT

It is very useful to be able to compare a model against actual data to verify that the model adequately corresponds to reality and to assess its usefulness as a predictive

tool. Very often there will be two or more alternatives with regards to model structures or the sets of probabilities. In such cases, there will also be a need to distinguish the better model. This section considers the use of logarithmic scores to assess and monitor the various aspects of the BNs being applied.

Logarithmic Scores

The *logarithmic score* as a measure of the level of "surprise" at each data point is given by

$$S_m = -\log p_m(y_m),$$

where $p_m(\cdot)$ is the predictive distribution for the event, Y_m, after $m-1$ occurrences of events. If learning is allowed then $p_m(\cdot)$ incorporates all updates resulting from the $m-1$ events. The *LS* is the negative log of the probability of the event in the actual outcome y_m.

To see how the logarithmic score indicates the level of "surprise," suppose an event that occurred was predicted to happen with a probability of 0.1, then the *LS* would be $-\log(0.1) = 2.3$. Conversely, an event expected to occur with probability of 0.9 would carry a score of 0.1. Thus, the less likely an event is predicted to happen, the more "surprising" it is if it did happen. If an event was certain to occur, the score would be $-\log(1) = 0$, that is, no surprise at all. However, if an event considered impossible occurred, this would be infinitely surprising: $-\log(0) = \infty$.

For a series of M events, the *total penalty* is $S = \sum_{m=1}^{M} S_m$. If the probabilities $p_m(\cdot)$ were coherently updated with each subsequent event, then S is invariant to the order in which the y_m's occur:

$$\begin{aligned} -\sum_{m=1}^{M} \log p_m(y_m) &= -\log \prod_{m=1}^{M} p_m(y_m) = -\log \prod_{m=1}^{M} p(y_m \mid y_1, \ldots, y_{m-1}) \\ &= -\log p(y_1, \ldots, y_M). \end{aligned}$$

A lower penalty is always more desirable.

Model Assessment

We now need some criteria by which to accept or reject a model. The total penalty incurred as a single figure is quite arbitrary on its own: it needs to be compared to a standard. This can be done in two ways:

- *Relative standardization.* In this method the model defined by the expert is tested against a reference model, which is a predefined benchmark by which the model is assessed. The total penalty incurred by the model defined by the expert, S, is compared against the penalty incurred (using the same data set) by the reference

model, S_{ref}. The model is rejected if S exceeds S_{ref}. The degree to which the expert model is preferred over the reference model is indicated by

$$\exp(S_{ref} - S) = \frac{P(\text{data} \mid \text{ expert's prior})}{P(\text{data} \mid \text{reference prior})}.$$

This is also known as the *Bayes' factor* in favor of the expert's model.

- *Absolute standardization.* In this method, a test statistic is compiled using the penalties and tested against the null hypothesis that the data arises from the model. We define the expectation and variance of the penalty at each update:

$$E_m = -\sum_{k=1}^{K} p_m(d_k) \log p_m(d_k)$$

$$\text{Var}_m = \sum_{k=1}^{K} p_m(d_k) \log^2 p_m(d_k) - E_m^2,$$

where the d_k's are all the possible states of the node considered.

The sums of the *actual* penalties incurred by the data, S, together with the expectation and variance at each update are used to compute the following test statistic

$$TestStatistic = \frac{S - \sum_{m=1}^{M} E_m}{\sqrt{\sum_{m=1}^{M} \text{Var}_m}}.$$

For large data sets this test statistic will have the standard Normal distribution if the data arise from the model, so if values outside the range $(-2, +2)$ are obtained this could lead one to doubt the adequacy of the model.

Model Monitors

There exists a set of monitors that can be used to diagnose the validity of BNs in the light of data. In general, these monitors are used to obtain the *LS* for each piece of data from which the total penalties can be compiled and tested according to the two criteria laid out above. The thing that differentiates one type of monitor from another is the probability forecast that is being used to evaluate the *LS*. There are three types of monitors: parent–child monitors, node monitors, and the global monitor.

Parent–child monitors measure how well the conditional probabilities in the BN that link parent and child nodes predict the outcome of the child nodes when the parents nodes are observed. For each parent–child set of nodes, we have the following *LS*

$$-\log p_m(x_k \mid X_{pa(k)} = \rho),$$

where $p_m(\cdot \mid X_{pa(k)} = \rho)$ is the probability distribution of the child node k for the parent configuration $X_{pa(k)} = \rho$ after $m - 1$ cases of complete data have arrived and x_k is the

state of the k node on the mth state. This monitor measures how well the conditional probabilities in the BN that link parent and child nodes predict the outcome of the child nodes.

For example, if we are monitoring the parent–child set expressed in the conditional probability $P(C \mid A_y, B_y)$ then as the first piece of data arrives (say, A_y, B_y, C_y) the *LS* can be determined as $S_1 = -\log P_1(C_y \mid A_y, B_y)$ and this is obtained from the expert prior distribution, Beta(α_y, α_n). The data are then incorporated into the conditional probability as illustrated in the section "Updating the Probabilities With New Data" to obtain the posterior $P_2(C \mid A_y, B_y)$ which has the distribution Beta($\alpha_y + 1, \alpha_n$). The process then continues iteratively as each subsequent case arrives.

The total penalty, S, can then be compared with the total penalty from various alternative reference models, S_{ref}. Some examples of reference models could be using the same expert prior without learning or using a reference prior B (0.5, 0.5) with learning.

For absolute standardization, the penalties incurred would be the same. The expectation and variance can be calculated for the first data set as follows:

$$E_1 = -P_1(C_y \mid A_y, B_y)\ \log P_1(C_y \mid A_y, B_y) - P_1(C_n \mid A_y, B_y)\ \log P_1(C_n \mid A_y, B_y)$$

$$\text{Var}_1 = P_1(C_y \mid A_y, B_y)[\ \log P_1(C_y \mid A_y, B_y)]^2 + P_1(C_n \mid A_y, B_y)[\log P_1(C_n \mid A_y, B_y)]^2 - E_1^2.$$

This is similarly performed for all subsequent updates. The test statistic can then be found. If the absolute value goes above 2, then this indicates a possible problem with the choice of parents of the node.

It is often useful to plot the cumulative values of the penalty (for relative standardization) or cumulative values of the test statistic (for absolute standardization) against the data. The graph can be used to compare the different alternative models and their paths give an indication of how well/soon the models adapt to the data. The preferred model is the one where the cumulative penalty increases very little with new data and the cumulative test statistic centers about zero.

Node monitors measure what is happening at the level of the individual nodes. There are two types of node monitors: *unconditional node monitors* and *conditional node monitors*. Unconditional node monitors detect poorly estimated marginal distributions. The *LS* is

$$-\log p_m(x_\upsilon),$$

where $p_m(\cdot)$ is the marginal distribution of the states of the node X_υ after $m-1$ cases. The score expresses the "surprise" at obtaining $X_\upsilon = x_\upsilon$ on the mth case. Note that if a node has no parents, then its unconditional node monitor is the same as its parent–child monitor. Conditional node monitors can be used to detect poor structure. The *LS* is

$$-\log p_m(x_\upsilon \mid \varepsilon_m \backslash X_\upsilon),$$

where $p_m(\cdot \mid \varepsilon_m \backslash X_v)$ is the probability distribution of the node X_v after $m-1$ cases of evidence have been incorporated and the latest set of evidence ε_m has just been propagated throughout the BN *except* the evidence on the node X_v itself. This score then measures the "surprise" at obtaining $X_v = x_v$ given the available evidence on the other nodes.

The global monitor of a BN measures the *LS* of the total evidence entered for the mth case after $m-1$ cases have been entered. This is

$$-\log p_m(\varepsilon_m).$$

The overall global monitor is just a sum of the *LS* for all the cases

$$G = \sum_{m=1}^{M} -\log p_m(\varepsilon_m).$$

In other words it is the negative of the log likelihood of the observed data. The global monitors of two competing models, $G^1$and G^2 are then compared, with the preferred model having the lower value. The Bayes factor in favor of model 2 is $\exp(G^1 - G^2)$.

Example

To illustrate how the model may be assessed, we consider the application of the parent–child monitor at the NF node of the on-line business example. The conditional probability underlying this node is $P(\text{NF} \mid \text{Hack})$. We investigate the particular case where malicious hacking has occurred. Suppose we have data for 16 consecutive weeks. We can see how well the model fits the data by investigating how well the model adapts to simulated cases drawn from a slightly different distribution. The differences used are shown in Table 6, which is quite different from the prior of the model.

Two alternatives were compared: one where the probabilities were sequentially updated as each new case arrives (as discussed in the section on "Updating the Probabilities With New Data"), and the other where the probabilities were left as originally specified. At each update, the following statistics were compiled:

- The logarithmic score (or penalty),
- The expectation of penalty, and
- The variance of penalty.

TABLE 6
Probability Distribution Used in Model Assessment Simulation

Distribution	$P(\text{NF} = \text{Yes} \mid \text{Hack} = \text{Yes})$	$P(\text{NF} = \text{No} \mid \text{Hack} = \text{Yes})$
Prior	0.8	0.2
Simulated	0.5	0.5

FIGURE 8
Parent–Child Monitor for NF Node: Relative Standardization

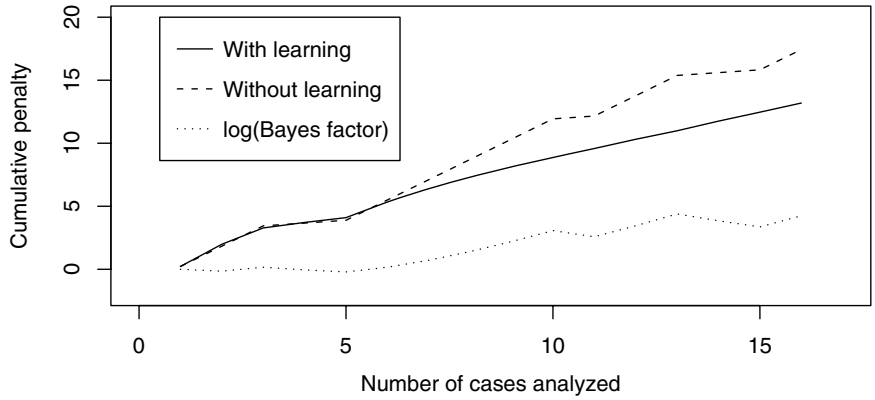
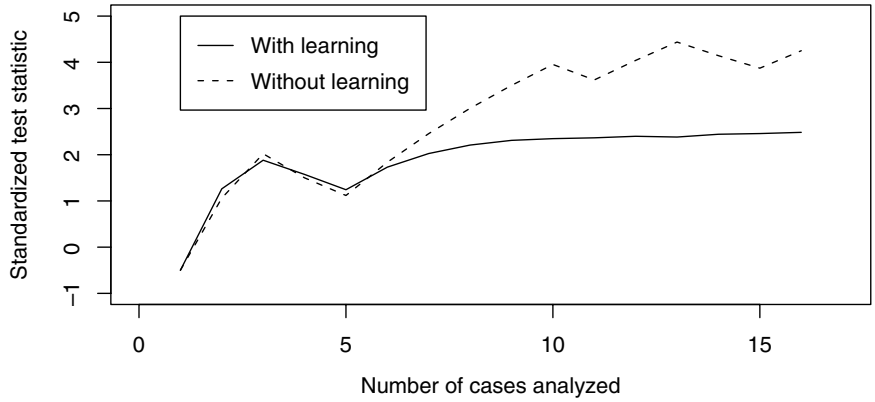


FIGURE 9
Parent–Child Monitor for NF Node: Absolute Standardization



The cumulative penalties and the absolute standardization test statistics were calculated (see Appendix B) and plotted in Figures 8 and 9.

Figure 8 shows a plot of the cumulative penalties both with and without learning, and their difference (the log of the Bayes factor). A lower penalty is desirable, so clearly the model with learning is performing better, with a log (Bayes factor) of 4.24 in its favor after 16 data points. This translates into an odds of $\exp(4.24) : 1 = 69 : 1$ in its favor.

Figure 9 shows the plots of the absolute standardized test statistics of the two models (note that the difference is not plotted in this case, it is not a log Bayes factor). A plot such as this can be used to highlight problematic nodes in the network. This is usually done without a formal hypothesis test since this would be based on asymptotic theory. If there were a large sample, then the hypothesis that the model is satisfactory would be rejected at the 5 percent level if the standardized test statistic exceeded 1.96 in absolute value. In this case we have a relatively small sample and the standardized

test statistic for the model with learning plotted in Figure 9, while giving cause for some concern, is probably still acceptable. In contrast, the model without learning, with a standardized test statistic of 4.25 after 16 data points, is unacceptable, and this would lead to rejecting this part of the model.

Similar plots can be made for node monitors and used to form an opinion on the viability of parts of the model. Alternative structures can be tested in parallel with the results plotted on the same graph for comparison. In the plots of the penalties (relative standardization) models with lower lying curves are preferred. In the plots of the test statistics (absolute standardization) models whose curves deviate outside the range $(-2, 2)$ should be treated with caution or perhaps even rejected. Plots of penalties for global models may also be prepared for alternative models, but unless the data are complete, standardizing a global monitor to evaluate its test statistic can be computationally expensive and is not usually done.

INSURANCE FRAUD RISK EXAMPLE

We now consider the application of BNs to model a specific OR event common to insurance companies: fraudulent claims. In this example, we model the cost of fraudulent claims arising over a week from a commercial fire insurance portfolio with annual gross premium of 1 m. Fraudulent claims may involve cases of arson by the insured or artificially inflated claim amounts.

The main hypothesis is that the level of control in underwriting new policies together with the stage of the economic cycle is responsible for explaining the incidence of fraudulent claims. The level of control in underwriting will determine the number of policyholders with a higher propensity to make fraudulent claims accepted onto the books of the insurer. During times of economic downturn these policyholders may resort to arson or artificially inflating claim figures as a means of economic gain. Policies that offer compensation for loss of profits, as many commercial fire policies do, are especially vulnerable to this sort of abuse. The level of control in underwriting in turn depends on:

1. Experience of the underwriter—Senior underwriters are better than junior underwriters at assessing which proposals are more likely to result in a fraudulent claim.
2. Business volume—High volume of business puts pressure on the system and reduces the quality of underwriting.
3. Reliance on branch—Underwriting for less complex risks may be partially outsourced to staff of branches to reduce costs and increase efficiency. However, this also increases the risk of fraudulent claims due to less control over the quality of underwriting at the branch level.

The claims handling department has the duty to detect fraudulent claims. The likelihood of detection depends on the level of control in claims management. If they do detect fraud, they can then act to mitigate the loss by launching investigations to uncover evidence that can be used to refuse or reduce the payment. These efforts are costly and may not always succeed in eliminating fraudulent claims payouts. The level of control in claims management in turn depends on:

1. Experience of the claims assessor—The reasoning is similar to the experience of the underwriter: more experienced claims assessor are more able to detect fraudulent claims.
2. Engage the services of a loss adjuster—Loss adjusters specialize in assessing the amount of loss suffered in a fire. This helps to mitigate artificially inflated claims. They are also able to ascertain the causes of a fire, especially if arson is a possibility.
3. Random checks—The company may carry out random checks on claim files for more detailed assessment before approval of payment.

Finally, the cost of insurance will depend on both the rates of incidence and detection. It should be noted that even with detection, the cost is not totally eliminated due to the inability to refute all suspicious claims and the cost of investigations. A possible structure to model such a description of fraudulent claims cost is shown in Figure 10. The random variables used in this example are shown in Table 7.

Using the same method of prior elicitation explained in the "Updating the Probabilities With New Data" section, the prior distributions at each node are shown in Appendix C. Finally, the prior marginals of each node are shown in Figure 11.

This model can now be used for stress testing, causal modeling, and capital allocation as shown for the on line business example.

FIGURE 10
DAG of Insurance Fraud Example

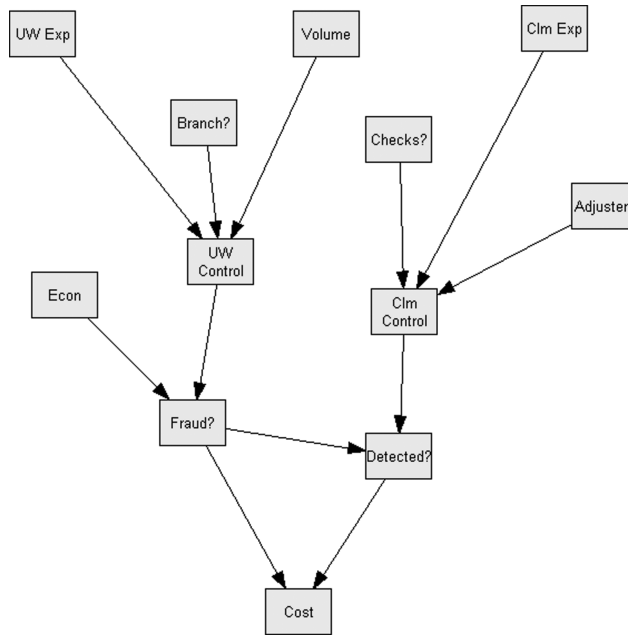
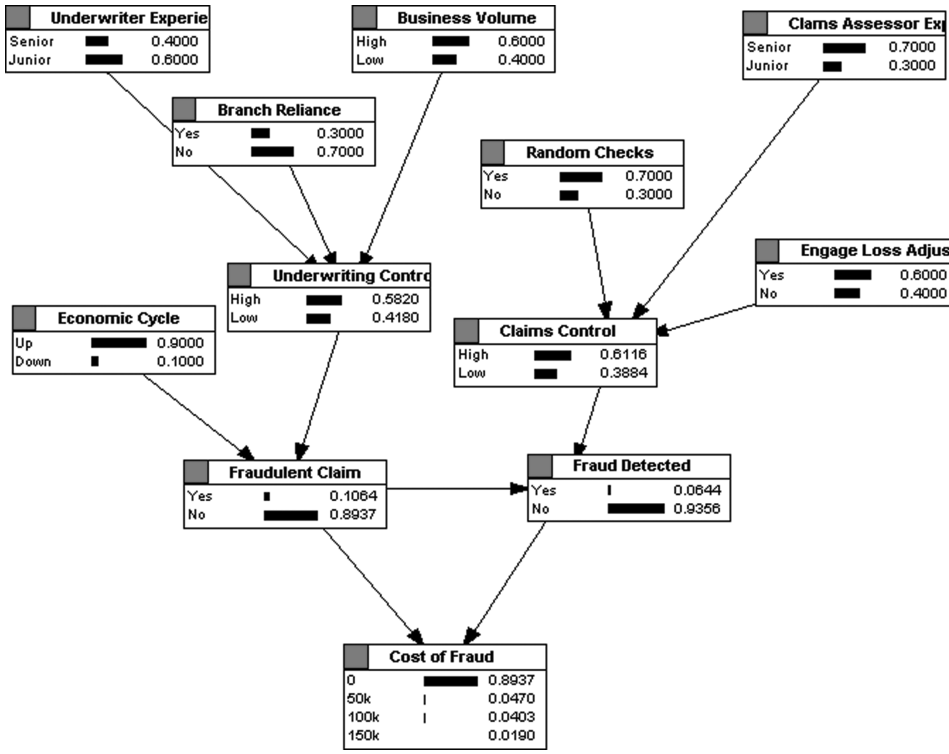


TABLE 7
The Random Variables of the Insurance Fraud Example

No.	Description	Values	Abbreviation
1.	Underwriter experience	Senior, Junior	UW Exp
2.	Branch reliance	Yes, No	Branch?
3.	Business volume	High, Low	Volume
4.	Claims assessor experience	Senior Junior	Clm Exp
5.	Random checks	Yes, No	Checks?
6.	Engage loss adjuster	Yes, No	Adjuster?
7.	Underwriting control	High, Low	UW Control
8.	Claims control	High, Low	Clm Control
9.	Economic cycle	Up, Down	Econ
10.	Fraudulent claim	Yes, No	Fraud?
11.	Fraud detected	Yes, No	Detected?
12.	Cost of fraud	0, 50k, 100k, 150k	Cost

FIGURE 11
Prior Marginal Distributions



DISCUSSION

We have used a fictitious on line business and an insurance fraud example to illustrate how a BN can be set up using a combination of past data and expert input. In these

illustrations we have also demonstrated the application of BNs in the areas of setting of capital for OR and scenario testing for causal analysis—two important components of supervisory regimes of financial institutions. The model has been shown to be easily adaptable to incorporate new input, and techniques for assessing the suitability of the model have also been demonstrated.

The main advantages of using BN for modeling OR is the facility to incorporate expert opinion through:

- Choice of the variables of interest;
- Definition of the structure of the model via the causal dependencies, and
- Specification of the prior distributions and the conditional probabilities at each node.

Bayesian statistical methodology ensures that the model can quickly adapt to new input and incorporate it with prior expert opinion in a mathematically tractable manner. Monitors are also available to enable the efficacy of this process to be observed in real time, thus facilitating informed model criticism and choice.

We have demonstrated the use of BNs to model a loss distribution, on which decisions could be based, particularly in the allocation of risk capital. Thus, there is great potential here for an internal model for the setting of supervisory capital for OR, such as is required by the emerging supervisory regimes. Stress and scenario testing is often a feature of early warning systems in supervisory regimes, and how these can be done fairly easily on BNs was illustrated.

Finally, the graphical presentation of BNs aids the understanding of the causal structure and presents the risk profile of the company in an intuitive way—improving management understanding of, and hence participation in, the management of OR.

Having emphasized the positive aspects of the use of BNs, there are still various challenges to overcome in the use of BN to model OR. Firstly, the model can get very complex with many nodes to specify, especially if the nodes have many parents. In such cases, there can be many conditional probabilities to specify, which will require a significant volume of data if the maximum likelihood method of estimating probabilities is used, thus reducing one of the main advantages of using Bayesian methods. The alternative to that would be a rigorous exercise in prior elicitation from experts through costly methods, which may involve many rounds of questionnaires. A main challenge in this area is in dealing with experts who may not be comfortable thinking in terms of frequencies, although one would hope that this is not the case in financial institutions.

There is also the issue of the nonuniqueness of the causal structure of the model: choosing a suitable model structure can be as much an art as a science and is often subject to debate. Thus, a logical progression is to move toward structural learning, that is, inferring both the network structure and its probability tables from the data. This is currently an active area of research; a summary may be found in chapter 11 of Cowell et al. (1999).

Although a BN is cheap and easy to run, the whole process of setting one up can be costly, resource consuming, and potentially politically messy if many business units/cost centers are involved. In many cases, a fairly complex BN is required to capture all the necessary variables. In addition, accuracy of results might be pursued at the expense of model parsimony especially when business decisions are at stake.

The main advantage of BNs, the ability to incorporate subjective knowledge, can be a disadvantage when it comes to meeting certain criteria for an internal model for solvency capital. Supervisors who require an objective standard to approve of internal models may find it difficult to find a standard for acceptance of BNs due to its high subjective content. Supervisors might need to specify rules on the process of model specification and prior elicitation to reduce the subjectivity.

In the on line business example, the distributions have been expressed as discrete probabilities for simplicity. More advanced modeling can be used to deal with continuous distributions in BNs. These are also known as conditional Gaussian distributions (see chapter 7 of Cowell et al., 1999; Lauritzen and Jensen, 2001).

APPENDIX A

Prior Distributions for Network Risk Example

FAC

High	Low
0.5	0.5

HAN

Yes	No
0.5	0.5

HAS

Yes	No
0.5	0.5

F

AP	PF
0.5	0.5

PS

Yes	No
0.25	0.75

SQ

High	Low
0.5	0.5

UPS

Yes	No
0.5	0.5

SF|PS, SQ

		SF	
SQ	PS	Y	N
H	Y	0.2	0.8
	N	0	1
L	Y	0.9	0.1
	N	0	1

Hack|FAC, F

		Hack	
F	FAC	Y	N
AP	High	0.1	0.9
	Low	0.5	0.5
PF	High	0.3	0.7
	Low	0.8	0.2

TD|HAN, NF

		TD		
HAN	NF	0 day	0.5 day	1 day
Y	Y	0	0.9	0.1
	N	1	0	0
N	Y	0	0.1	0.9
	N	1	0	0

AF|V, EUM, Hack

			AF		
Hack	EUM	V	App Corr	Lockup	OK
Y	Y	Y	0.9	0.1	0
		N	0.6	0.4	0
	N	Y	0.7	0.3	0
		N	0.6	0.2	0.2
N	Y	Y	0.4	0.3	0.3
		N	0.1	0.3	0.6
	N	Y	0.5	0	0.5
		N	0	0	1

SD|UPS, SF, HAS

			SD		
HAS	SF	UPS	0 day	0.5 day	1 day
Y	Y	Y	0.5	0.5	0
		N	0	0.2	0.8
	N	Y	1	0	0
		N	1	0	0
N	Y	Y	0	0.1	0.9
		N	0	0	1
	N	Y	1	0	0
		N	1	0	0

V|FAC

	V	
FAC	Y	N
High	0.1	0.9
Low	0.7	0.3

EUM|FAC

	EUM	
FAC	Y	N
High	0.1	0.9
Low	0.6	0.4

NF|Hack

	NF	
Hack	Y	N
Y	0.8	0.2
N	0	1

DL|AF,SF

		DL		
SF	AF	0%	50%	100%
Y	App Corr	0.1	0	0.9
	Lockup	0.1	0.1	0.8
	OK	0.2	0.2	0.6
N	App Corr	0.5	0.5	0
	Lockup	0.7	0.3	0
	OK	1	0	0

Cost|SD, TD, DL

			Cost					
DL	**TD**	**SD**	**0.0m**	**0.5m**	**1.0m**	**1.5m**	**2.0m**	**2.5m**
0%	0 day	0 day	1	0	0	0	0	0
		0.5 day	0	0.6	0.4	0	0	0
		1 day	0	0.1	0.2	0.3	0.2	0.2
	0.5 day	0 day	0.7	0.3	0	0	0	0
		0.5 day	0	0.5	0.3	0.2	0	0
		1 day	0	0	0.2	0.4	0.2	0.2
	1 day	0 day	0.6	0.4	0	0	0	0
		0.5 day	0	0.4	0.3	0.2	0.1	0
		1 day	0	0	0.1	0.2	0.4	0.3
50%	0 day	0 day	0.6	0.4	0	0	0	0
		0.5 day	0	0.5	0.4	0.1	0	0
		1 day	0	0	0.1	0.3	0.4	0.2
	0.5 day	0 day	0.3	0.3	0.2	0.2	0	0
		0.5 day	0	0.3	0.3	0.2	0.2	0
		1 day	0	0	0.1	0.2	0.5	0.2
	1 day	0 day	0.3	0.2	0.2	0.2	0.1	0
		0.5 day	0	0.3	0.2	0.2	0.2	0.1
		1 day	0	0	0	0.1	0.4	0.5
100%	0 day	0 day	0.4	0.4	0.2	0	0	0
		0.5 day	0	0.4	0.3	0.2	0.1	0
		1 day	0	0	0	0.3	0.4	0.3
	0.5 day	0 day	0.2	0.2	0.3	0.2	0.1	0
		0.5 day	0	0.2	0.2	0.3	0.2	0.1
		1 day	0	0	0	0.2	0.4	0.4
	1 day	0 day	0	0.1	0.2	0.3	0.2	0.2
		0.5 day	0	0	0.2	0.3	0.3	0.2
		1 day	0	0	0	0	0.2	0.8

APPENDIX B

Calculations for Parent–Child Monitor for NF Node

This appendix contains tables of the calculated values for the parent–child monitor of $P(NF \mid Hack = Yes)$ used in Figures 8 and 9. Define $\theta_y = P(NF = Yes \mid Hack = Yes)$ and $\theta_n = P(NF = No \mid Hack = Yes)$

Monitor Values With Learning:

m	Data NF	α_y	α_n	$E(\theta_y)$	$E(\theta_n)$	- log $E(\theta_y)$	- log $E(\theta_n)$	S_m	S	E_m	Var_m	$\sum E_m$	$\sum Var_m$	Test Statistic
0	Prior	4.88	1.22	0.800	0.200	0.223	1.609							
1	Yes	5.88	1.22	0.828	0.172	0.189	1.761	0.223	0.223	0.500	0.307	0.500	0.307	-0.500
2	No	5.88	2.22	0.726	0.274	0.320	1.294	1.761	1.984	0.459	0.352	0.959	0.659	1.262
3	No	5.88	3.22	0.646	0.354	0.437	1.039	1.294	3.279	0.587	0.189	1.546	0.848	1.881
4	Yes	6.88	3.22	0.681	0.319	0.384	1.143	0.437	3.715	0.650	0.083	2.196	0.931	1.574
5	Yes	7.88	3.22	0.710	0.290	0.343	1.238	0.384	4.099	0.626	0.125	2.822	1.056	1.243
6	No	7.88	4.22	0.651	0.349	0.429	1.053	1.238	5.337	0.602	0.165	3.424	1.221	1.731
7	No	7.88	5.22	0.602	0.398	0.508	0.920	1.053	6.390	0.647	0.089	4.071	1.310	2.026
8	No	7.88	6.22	0.559	0.441	0.582	0.818	0.920	7.310	0.672	0.041	4.744	1.350	2.209
9	No	7.88	7.22	0.522	0.478	0.650	0.738	0.818	8.129	0.686	0.014	5.430	1.364	2.311
10	No	7.88	8.22	0.489	0.511	0.714	0.672	0.738	8.867	0.692	0.002	6.122	1.366	2.348
11	Yes	8.88	8.22	0.519	0.481	0.655	0.733	0.714	9.581	0.693	0.000	6.815	1.367	2.366
12	No	8.88	9.22	0.491	0.509	0.712	0.675	0.733	10.314	0.692	0.001	7.507	1.368	2.399
13	No	8.88	10.22	0.465	0.535	0.766	0.625	0.675	10.988	0.693	0.000	8.200	1.368	2.383
14	Yes	9.88	10.22	0.492	0.508	0.710	0.676	0.766	11.754	0.691	0.005	8.891	1.373	2.443
15	Yes	10.88	10.22	0.516	0.484	0.662	0.725	0.710	12.464	0.693	0.000	9.584	1.374	2.458
16	No	10.88	11.22	0.492	0.508	0.709	0.678	0.725	13.189	0.693	0.001	10.277	1.375	2.484

Monitor Values Without Learning:

m	Data NF	$E(\theta_y)$	$E(\theta_n)$	- log $E(\theta_y)$	- log $E(\theta_n)$	S_m	S	E_m	Var_m	$\sum E_m$	$\sum Var_m$	Test Statistic
0	Prior	0.800	0.200	0.223	1.609							
1	Yes	0.800	0.200	0.223	1.609	0.223	0.223	0.500	0.307	0.500	0.307	- 0.500
2	No	0.800	0.200	0.223	1.609	1.609	1.833	0.500	0.307	1.001	0.615	1.061
3	No	0.800	0.200	0.223	1.609	1.609	3.442	0.500	0.307	1.501	0.922	2.021
4	Yes	0.800	0.200	0.223	1.609	0.223	3.665	0.500	0.307	2.002	1.230	1.500
5	Yes	0.800	0.200	0.223	1.609	0.223	3.888	0.500	0.307	2.502	1.537	1.118
6	No	0.800	0.200	0.223	1.609	1.609	5.498	0.500	0.307	3.002	1.845	1.837
7	No	0.800	0.200	0.223	1.609	1.609	7.107	0.500	0.307	3.503	2.152	2.457
8	No	0.800	0.200	0.223	1.609	1.609	8.717	0.500	0.307	4.003	2.460	3.005
9	No	0.800	0.200	0.223	1.609	1.609	10.326	0.500	0.307	4.504	2.767	3.500
10	No	0.800	0.200	0.223	1.609	1.609	11.935	0.500	0.307	5.004	3.075	3.953
11	Yes	0.800	0.200	0.223	1.609	0.223	12.159	0.500	0.307	5.504	3.382	3.618
12	No	0.800	0.200	0.223	1.609	1.609	13.768	0.500	0.307	6.005	3.690	4.041
13	No	0.800	0.200	0.223	1.609	1.609	15.378	0.500	0.307	6.505	3.997	4.438
14	Yes	0.800	0.200	0.223	1.609	0.223	15.601	0.500	0.307	7.006	4.305	4.143
15	Yes	0.800	0.200	0.223	1.609	0.223	15.824	0.500	0.307	7.506	4.612	3.873
16	No	0.800	0.200	0.223	1.609	1.609	17.433	0.500	0.307	8.006	4.920	4.250

Comparison of Model Alternatives

m	*S*		log(Bayes Factor)	Test Statistic	
	With learning	Without learning		With learning	Without learning
1	0.223	0.223	0.000	-0.500	-0.500
2	1.984	1.833	-0.151	1.262	1.061
3	3.279	3.442	0.163	1.881	2.021
4	3.715	3.665	-0.050	1.574	1.500
5	4.099	3.888	-0.211	1.243	1.118
6	5.337	5.498	0.161	1.731	1.837
7	6.390	7.107	0.717	2.026	2.457
8	7.310	8.717	1.407	2.209	3.005
9	8.129	10.326	2.197	2.311	3.500
10	8.867	11.935	3.068	2.348	3.953
11	9.581	12.159	2.578	2.366	3.618
12	10.314	13.768	3.454	2.399	4.041
13	10.988	15.378	4.390	2.383	4.438
14	11.754	15.601	3.847	2.443	4.143
15	12.464	15.824	3.360	2.458	3.873
16	13.189	17.433	4.244	2.484	4.250

APPENDIX C

Prior Distributions for Insurance Fraud Example

UW Exp

Senior	Junior
0.4	0.6

Clm Exp

Senior	Junior
0.7	0.3

Adjuster?

Yes	No
0.6	0.4

Checks?

Yes	No
0.7	0.3

Volume

High	Low
0.6	0.4

Branch?

Yes	No
0.3	0.7

Econ

Up	Down
0.9	0.1

UW Control | Volume, Branch?, UW Exp

Volume	Branch?	UW Exp	UW Control	
			High	Low
High	Yes	Senior	0.5	0.5
		Junior	0.1	0.9
	No	Senior	0.7	0.3
		Junior	0.4	0.6
Low	Yes	Senior	0.8	0.2
		Junior	0.6	0.4
	No	Senior	0.9	0.1
		Junior	0.8	0.2

Clm Control | Adjuster?, Checks?, Clm Exp

Adjuster?	Checks?	Clm Exp	Clm Control	
			High	Low
Yes	Yes	Senior	0.9	0.1
		Junior	0.6	0.4
	No	Senior	0.8	0.2
		Junior	0.5	0.5
No	Yes	Senior	0.5	0.5
		Junior	0.2	0.8
	No	Senior	0.3	0.7
		Junior	0.1	0.9

Fraud? | UW Control, Econ

UW Control	Econ	Fraud?	
		Yes	No
High	Up	0.05	0.95
	Down	0.3	0.7
Low	Up	0.1	0.9
	Down	0.6	0.4

Detected? | Clm Control, Fraud?

Clm Control	Fraud?	Detected?	
		Yes	No
High	Yes	0.7	0.3
	No	0	1
Low	Yes	0.3	0.7
	No	0	1

Cost | Fraud?, Detected?

Detected?	Fraud?	Cost			
		0	50k	100k	150k
Yes	Yes	0	0.6	0.3	0.1
	No	1	0	0	0
No	Yes	0	0.2	0.5	0.3
	No	1	0	0	0

REFERENCES

Alexander, C., 2003, Operational Risk: Regulation, Analysis, and Management. Prentice Hall. Basel Committee on Banking Supervision (1999), A New Capital Adequacy Framework. Bank for International Settlements.

Basel Committee on Banking Supervision, 2004, International Convergence of Capital Management and Capital Standards. Bank for International Settlements.

Committee of European Insurance and Occupational Pension Supervisors, 2005, Draft Answers to the European Commission on the 'second wave' of Calls for Advice in the framework of the Solvency II project. Available at http://www.ceiops.org/

Conference of the Insurance Supervisory Services of the Member States of the European union, 2002, Prudential Supervision of Insurance Undertakings. Available at http://europa.eu.int/comm/internal_market/insurance/solvency/solvency2-conference_en.htm

Cowell, R. G., A. P. Dawid, S. L. Lauritzen, and D. J. Spiegelhalter, 1999, *Probabilistic Networks and Expert Systems* (New York: Springer-Verlag).

Cruz, M. G., 2002, *Modeling, Measuring and Hedging Operational Risk*. (New York: Wiley).

Dowd, K., 1998, *Beyond Value at Risk: The New Science of Risk Management*. (New York: Wiley).

European Commission, 1999, The Review of the Overall Financial Position of an Insurance Undertaking. (MARKT/2095/99).

European Commission, 2003, Design of a Future Prudential Supervisory System in the EU—Recommendations by the Commission Services. (MARKT/2509/03).

European Commission, 2005a, Specific Calls for Advice from CEIOPS (Third Wave). (MARKT/2501/05).

European Commission, 2005b, Framework for Consultation on Solvency II. (Annex to MARKT/2505/05).

Frachot, A., and T. Roncalli, 2002, Mixing Internal and External Data for Managing Operational Risk, Groupe de Recherche Operationelle, Credit Lyonnais.

Hoffman, D., 2002, *Managing Operational Risk: 20 Firmwide Best Practice Strategies*. Wiley.

International Association of Insurance Supervisors, 2002, Principles on Capital Adequacy and Solvency. Available at http://www.iaisweb.org/02solvency.pdf

International Association of Insurance Supervisors, 2003a, Solvency Control Levels Guidance Paper. Available at http://www.iaisweb.org/185solvency03.pdf

International Association of Insurance Supervisors, 2003b, Stress Testing Guidance Paper. Available at http://www.iaisweb.org/185stresstesting03.pdf

Jorion, P., 2001, *Value at Risk: The New Benchmark for Managing Financial Risk*. (New York: McGraw-Hill).

King, J. L., 1999, *Operational Risk*. (New York: Wiley).

Lauritzen, S. L., 1995, The EM Algorithm for Graphical Association Models with Missing Data, *Computational Statistics and Data Analysis*, 19: 191-201.

Lauritzen, S. L., and F. Jensen, 2001, Stable Local Computation with Conditional Gaussian Distributions, *Statistics and Computing*, 11: 1098-1108.

Marshall, C., 2001, Measuring and Managing Operational Risks in Financial Institutions. (New York: Wiley).

Miccolis, J. A., and S. Shah, 2000, Getting a Handle on Operational Risks, *Emphasis 2000/1*, Tillinghast-Towers Perrin.

Shah, S., 2001, Operational Risk Management, *Casualty Actuarial Society 2001 Seminar on Understanding the Enterprise Risk Management Process*, Powerpoint slides available from Casualty Actuarial Society at http://www.casact.org

Tripp, M. H., et al., 2004, Quantifying Operational Risk in General Insurance Companies, *British Actuarial Journal*, 10: 919-1026.

Van Den Brink, G. J., 2002, Operational Risk: The New Challenge for Banks. (New York: Palgrave).