A quantitative optimization model for dynamic risk-based compliance management

S. Müller
C. Supatgiat

The changing nature of regulation forces businesses to continuously reevaluate the measures taken to comply with regulatory requirements. To prepare for compliance audits, businesses must also implement an effective internal inspection policy that identifies and rectifies instances of noncompliance. In this paper, we propose an approach to compliance management based on a quantitative risk-based optimization model. Our model allows dynamic selection of the optimal set of feasible measures for attaining an adequate level of compliance with a given set of regulatory requirements. The model is designed to minimize the expected total cost of compliance, including the costs of implementing a set of measures, the cost of carrying out periodic inspections, and the audit outcome cost for various compliance levels. Our approach is based on dynamic programming and naturally accounts for the dynamic nature of the regulatory environment. Our method can be used either as a scenario-based management support system or, depending on the availability of reliable input data, as a comprehensive tool for optimally selecting the needed compliance measures and inspection policy. We illustrate our approach in a hypothetical case study.

Introduction

Following a number of recent incidents of corporate accounting frauds and theft of consumers' personal data, and with the rising threat of international terrorism, we have seen a surge of new governmental regulations imposed on businesses. Affected businesses must frequently adapt their operations to relevant regulations and periodically demonstrate compliance by submitting to audits. Furthermore, because current regulations such as the Sarbanes–Oxley Act [1] and the USA Patriot Act [2] in the United States carry large penalties (e.g., increased fines and the possibility of imprisonment), they have significantly increased the expected cost of noncompliance. Thus, many corporations are currently spending large amounts of money in their attempt to achieve maximum compliance with current regulations.

Whereas attempting to attain perfect compliance is a worthy goal, due to the complex nature of modern business processes and the possibility of human error it is rarely achievable in practice. There is always a small possibility that some people or systems do not respect relevant legal obligations. To attain a near-perfect degree of compliance, a company would have to continuously inspect its employees, systems, processes, and products. While such a compliance management approach would absorb a considerable amount of financial resources, it would still fall short of providing perfect compliance with certainty. Hence, compliance is an inherently continuous rather than a discrete phenomenon and must be managed in a risk-based way. This must be considered when deciding on the targeted compliance level and selecting and prioritizing compliance activities. The goal of our paper is to show how to manage such compliance risk in a practical manner.

The degree of compliance of an enterprise and the implied compliance risk (i.e., the expected cost of compliance) depend on many factors. Among them are the type, the effectiveness, and the cost of the measures taken to address a specific regulatory requirement; the type, the frequency, and the scrutiny of inspections

©**Copyright** 2007 by International Business Machines Corporation. Copying in printed form for private use is permitted without payment of royalty provided that (1) each reproduction is done without alteration and (2) the *Journal* reference and IBM copyright notice are included on the first page. The title and abstract, but no other portions, of this paper may be copied or distributed royalty free without further permission by computer-based and other information-service systems. Permission to *republish* any other portion of this paper must be obtained from the Editor.

0018-8646/07/$5.00 © 2007 IBM

conducted; and the coverage and the outcome cost of compliance audits. Accordingly, the cost of achieving compliance for an enterprise must be a direct function of the cost and effectiveness of the measures taken, the frequency and the scrutiny of inspections, the likelihood of audits, and the cost of the audit outcomes.

As with any other business activity, compliance-related activities must be financed using scarce resources (time, money, people, etc.) in an economically efficient way. This necessitates a careful analysis, prioritization, and implementation of compliance-related activities with respect to their potential benefits and cost. In particular, the selection, prioritization, and implementation of compliance measures must be managed according to their expected costs.

In this paper, we introduce a dynamic and risk-based approach to compliance management. In our method, deciding what compliance measures to implement and how to perform effective inspections is inherently risk-based. As a result, a business unit employing our approach can manage its targeted compliance risk level by taking into account the cost and effectiveness of operational and future compliance measures, the type and cost of internal inspections, the frequency and scrutiny of audits, the likelihood of audits, and the expected cost of audit outcomes. Our approach can be used as a scenario-based management support system which determines the *optimal portfolio of measures* required to maintain a desired target compliance level and the *optimal inspection policy*. Our model can also be used by regulatory bodies as a policy instrument when new regulation is created.

Risk management has a longstanding tradition in areas such as finance and insurance where it has been used to manage financial risks [3–5], credit risk [6], and recently also operational risk [7–9] and information technology (IT) security risk [10, 11]; however, to our best knowledge, this is the first attempt to address compliance management using a quantitative risk-based approach. Ironically, whereas financial services companies have made risk management one of their core competencies, they apparently have not realized that regulatory compliance can also be addressed using similar techniques.

Compliance measures and inspections

To comply with a given regulatory requirement, a business must not only implement *measures* that ensure compliance but also institute an *inspection policy* to ascertain that the measures taken have the desired effect. Clearly, the business has to assess the costs associated with this undertaking and select the most effective measures for achieving the targeted degree of compliance. Moreover, the inspection policy provides information on the current compliance status and generally leads to an improved compliance level of the enterprise. This also results in a higher likelihood of passing a future audit with a higher satisfaction level and avoiding the (implicit) cost related to passing the audit with lower satisfaction or, in the worst case, failing it. The cost of implementing compliance measures and an inspection policy is often significant. In particular, compliance-related investments represent opportunity costs in that they require funds that could be used for other, more lucrative investments. Furthermore, different types of measures and inspections are not equally effective. That is, depending on the actual set of measures chosen, the compliance level of the enterprise can be increased more or less on the basis of the relative effectiveness of the selected measures.

The terminology used by various governmental regulations is not uniform. The Sarbanes–Oxley Act, for example, recommends that enterprises implement the control framework of the Committee of Sponsoring Organizations of the Treadway Commission (COSO) [12] in order to ensure the accuracy of financial data. COSO refers to compliance measures as "controls" and to the inspection policy as "testing" the controls. Similarly, privacy regulations refer to compliance measures as "access control measures" and to inspections as "testing" the controls.

According to COSO, a control is a process designed to provide reasonable assurance regarding the achievement of objectives in a) effectiveness and efficiency of operations, b) reliability of financial reporting, and c) compliance with applicable laws and regulations. Furthermore, COSO also stresses the important role of people, at every level of an organization, in ensuring that the organization becomes and remains compliant. It recognizes that "internal control can be expected to provide only reasonable assurance, not absolute assurance, to an entity's management and board."

This characterization of the notion of control reflects the risk-based nature of compliance-related measures by defining them as providing merely relative and not absolute assurance. Furthermore, it is recognized that compliance is an ongoing concern and, hence, must be continuously monitored and adjusted. Our compliance management model incorporates these principles.

Audits and the cost of compliance

In some cases, regulations provide direct economic value to affected enterprises (as in the case of privacy regulation in financial services). In other cases, regulations constrain the scope for conducting business, thereby inducing direct and opportunity costs on enterprises and requiring them to implement costly compliance measures. In such cases, businesses

implement compliance measures and conduct inspections on their effectiveness because they are required to submit to periodic external audits. An (external) audit is conducted by outside auditors and involves the evaluation of a firm's systems, processes, or products. The purpose of an audit is to verify that an enterprise operates according to a set of relevant regulatory requirements. Some audits are voluntary, such as audits to certify that the operation of the enterprise complies with a certain standard (e.g., standards issued by the International Standards Organization). In this paper, we focus on audits of compliance with governmental regulations.

Audits can be passed with varying degrees of auditor satisfaction. Depending on the outcome of an audit, an enterprise may have to implement a number of recommendations, which range from substantial, such as additional or stronger compliance measures, to minor, such as minor corrective actions. In any case, the enterprise absorbs the costs associated with these tasks. If the enterprise fails to take the steps that ensure an adequate compliance level, it may fail a subsequent audit and face heavy penalties. In this paper, we consider an audit to have failed only if the auditors found that the measures implemented by the enterprise did not lead to an adequate compliance level.

When an audit results in a high level of auditor satisfaction, the audited company is granted a grace period for implementing the corrective action required for compliance, but no additional audit is required before the next regular audit. When an audit results in a low level of auditor satisfaction, in addition to requiring that the identified deficiencies be corrected, the auditor may alert the responsible governmental agency and fines may be imposed on the enterprise. Then, the costs to the enterprise include both the cost of correcting the identified deficiencies and the fine. The enterprise is granted a grace period within which it must be able to attain and demonstrate a satisfactory compliance level. The enterprise may also be required to pass a follow-up audit, during which the auditors investigate whether the identified deficiencies have been corrected.

The outcome of an audit depends on two factors: the compliance level of the enterprise and the coverage of the audit, that is, the scope of the audit and the thoroughness of the auditors. If the enterprise invests heavily in measures to attain a high degree of compliance, even extensive audit coverage is not very likely to determine that the compliance is inadequate. However, if the overall compliance level is relatively low, broad audit coverage (i.e., a high probability of detection) is likely to reveal that the enterprise is marginally compliant, or in the worst case is not compliant and thus fails the audit.

Over time, enterprises accumulate a track record of performance in compliance audits that influences the behavior of its auditors. The audit is likely to be less thorough when the enterprise has an outstanding track record. In contrast, a company whose track record is poor is likely to be subjected to closer scrutiny.

The expected cost of compliance includes both explicit and implicit costs. Explicit costs include

- The costs of implementing the auditor's recommendations.
- For companies that are found noncompliant, the opportunity cost of lost revenue or lost market share due to the burden of becoming compliant.
- For a company whose product is found noncompliant, the opportunity cost of lost revenue due to the prohibition of selling the product.
- Monetary fines imposed on the enterprise for noncompliance.
- Personal and criminal liability for company executives, auditors, and board members for transgressions related to noncompliance.

Implicit penalties include

- Decline in product demand, loss of reputation, decline in customer goodwill, and decline in stock price due to negative publicity.
- Decline in product demand due to the decline of consumer confidence in the safety quality of products.
- Higher stock price volatility due to the company's uncertain future.

Not all auditor recommendations are equally expensive to implement. Likewise, not every regulation imposes equally severe consequences in case of inadequate compliance. As a result, different compliance levels with respect to different regulations translate to different expected compliance costs. In combination with the scarcity of a business's resources, it is thus important that the business prioritize compliance activities across relevant regulations. The audit coverage, the enterprise's own compliance level, and its audit track record are also important factors when contemplating the expected cost of possible compliance activities and audit outcomes.

The remainder of this paper is structured as follows. In the next section we introduce a mathematical model for dynamic and risk-based compliance management. The model leads to the formulation of a total expected cost-to-go function, which is minimized using dynamic programming. In the following section we present a hypothetical case study and demonstrate how our approach can be used to determine the optimal portfolio

of compliance measures and the associated inspection policy that minimize the expected cost of compliance. We conclude with some final comments.

Model

Let τ be total number of time periods in the decision-making horizon. At the beginning of each period, the compliance manager, the person responsible for the company's compliance with a given set of regulatory requirements, performs a risk assessment and determines the compliance measures to be implemented and the type and the frequency of inspections to be conducted.

Compliance measures and inspections

Let M be the number of possible types of compliance measures. Each measure has two cost components: a one-time fixed cost and a periodic maintenance cost. Measure i costs c_i^M to implement and an additional y_i^M to maintain in each period. If there is no maintenance cost, y_i^M is set to zero. When implementing a measure i, it takes r_i^M periods for the measure to be successfully implemented and effective.

The manager may decide to stop maintaining an operational compliance measure. The measure that is no longer maintained ceases to have any impact on the company's compliance level in the period after its maintenance cost is stopped. If such a measure is to be reactivated at a later time, the cost of its activation is equal to its entire implementation cost (as if the measure had never been implemented), and it takes the same number r_i^M of periods to complete the activation.

There are I possible types of inspections, each with its own cost (that is, inspection j costs c_j^I per inspection) and effectiveness. Without loss of generality, only one inspection type is allowed in any given period. Indeed, if inspection types a and b can be conducted in the same period, we can just define a new inspection type, say c, representing the combined cost and effect of inspection types a and b. Therefore, conducting inspection type c yields the same effect as conducting both inspection types a and b.

We let an integer vector $V_t = (v_1, \cdots, v_M)$ be the historical measure implementation vector. Its ith component represents the number of periods from period t until measure i becomes effective. If measure i is already in effect, then $v_i = 0$. If measure i has never been implemented, then $v_i = -1$.

There are J requirements to fulfill. Accordingly, the measures can be classified into J classes. Each class corresponds to one of the requirements. The primary aim of all measures in that class is to address that requirement. We represent the effectiveness of measure i to address requirement j by e_i^j. The effectiveness has a value between 0 and 1, with 0 denoting no effect and 1 encoding perfect effectiveness.

Compliance level

The major component in compliance-level modeling is a *target compliance level*, denoted by $T(V_t)$. It is defined as

$$T(V_t) = \sum_{j=1}^{J} v_j \underbrace{\left(\underbrace{1 - \prod_{l=1}^{J} \underbrace{\left[1 - \max_{i|(V_t)_i = 0 \wedge i \in \Theta_l} (e_i^j) \right]}_{A}}_{B} \right)}_{C},$$

where the symbol $(V)_i$ represents the ith element of vector V and the set Θ_l represents the class of measures whose primary aim is to address requirement l.

The target compliance level represents the maximally achievable compliance level given all measures that are currently in effect. It is computed as a weighted average of the individual compliance levels with respect to the J requirements. The weights v_j are assigned according to the degree to which each requirement contributes to the total regulatory exposure. They should sum to one (i.e., $\sum_{j=1}^{J} v_j = 1$).

In the above formula, the term A represents the minimal degree of noncompliance attained by implementing a set of measures from the same class addressing the jth requirement. The term B represents the total noncompliance level resulting from implementing measures from different classes. Finally, the term C yields the total degree of compliance with respect to the jth requirement.

From the target compliance-level formula, we see that if two or more measures from the same class are implemented together, only the one with the higher effectiveness for the corresponding requirement will affect the target compliance level. If measures belonging to different classes are implemented together, their combined effectiveness will define the target compliance level. The target compliance level is the maximal compliance level that can be obtained given the set of implemented measures. To increase the target compliance level, more measures or measures with higher effectiveness must be implemented.

We denote the compliance level of a company with respect to a particular set of regulatory requirements at the beginning of period t by a number b_t. The compliance level b_t is an indication of the company's current internal compliance level with respect to the set of relevant

regulatory requirements. It takes a value between 0 and 1, with $b_t = 1$ denoting the highest compliance level.

If there is no inspection in period t, the compliance level at the beginning of period $t+1$ is a function of a) the measures taken and the numbers of periods before they take effect, V_t and V_{t+1}, at the beginning of periods t and $t+1$, and b) the compliance level in the last period b_t.

When an inspection is conducted in period t, the compliance level is normally increased. The inspection effectiveness varies depending on the type of inspection chosen. Let a_t^I be an integer from 0 to I representing the type of inspection that is conducted in period t. The value of -1 means that no inspection is conducted. For inspection type i, the improvement is denoted by O_i, which is a factor ranging from 0 to 1 that indicates the increase in the compliance level b_{t+1}. A value of 1 means full improvement with respect to the original level achievable by the implemented measures, while a value of 0 stands for no improvement in the compliance level. The improvement O_i is assumed to be a random variable with probability distribution F_i^O. In any case, no matter how effective the inspection, the maximum compliance level after any inspection is limited to the target compliance level corresponding to the measures that are in effect.

The compliance level can decrease with the passage of time, for example, because the employees become more relaxed over time and do not adhere to the implemented measures as much. We define the decay factor ρ, with a value between 0 and 1, as a multiplier to the current-period compliance level to obtain the next-period compliance level. The higher the decay factor, the more rapidly the compliance level drops.

The compliance level b_{t+1} is a function f of $(V_t,\ V_{t+1},\ a_t^I,\ b_t)$ and is defined as follows:

$$b_{t+1} \equiv f(V_t,\ V_{t+1},\ a_t^I,\ b_t) = \begin{cases} \dfrac{b_t \rho T(V_{t+1})}{T(V_t)} & \text{if } a_t^I = 0, \\ \dfrac{b_t \rho T(V_{t+1})}{T(V_t)} + O_i\left[T(V_{t+1}) - \dfrac{b_t \rho T(V_{t+1})}{T(V_t)}\right] & \text{otherwise.} \end{cases}$$

In the first case, there is no inspection in period t. If there is no change to the effective measures, i.e., $V_t = V_{t+1}$, the next-period compliance level b_{t+1} is just $b_t\rho$ (i.e., the current compliance level with one period decay). If there is a change in the effective measures, the new compliance level is equal to the ratio $b_t\rho/T(V_t)$ of the new target level $T(V_{t+1})$.

If there is an inspection in period t, the compliance level is improved by the amount $O_i[T(V_{t+1}) - b_t\rho T(V_{t+1})/T(V_t)]$. Note that the term $[T(V_{t+1}) - b_t\rho T(V_{t+1})/T(V_t)]$ represents the gap between the target compliance level and the actual compliance level under no inspection. The improvement factor O_i is multiplied with this gap to determine the improvement in the compliance level due to inspection i. If O_i, is 100%, the compliance level is equal to the target compliance level. If O_i is 0%, there is no improvement. Since O_i is a random variable, the compliance level b_{t+1} is also a random variable.

Auditing

We assume that auditing takes place every fixed interval. The inter-auditing interval is denoted by T_A. For example, when the time period represents one week and auditing occurs twice per year, auditing occurs every 26 periods, or $T_A = 26$.

There are N possible outcomes of an audit, ranging from outcome 1 (passed with 100% satisfaction) to outcome N (failed with 0% satisfaction). We let K be the number of past audit outcomes sufficient to determine the audit outcome cost; that is, it is sufficient to calculate the audit outcome cost if we know only the past K audit outcomes. We let an integer vector $H_t = [h_1, \cdots, h_K]$ be the historical audit outcome vector. Its ith component represents the audit outcome, which is a number from 1 to N, at the ith-last audit since period t.

The audit outcome cost at time t is a function of the historical outcome vector. For example, the cost will be high after a series of consecutive poor audit outcomes. On the other hand, it will be low if a number of previous audits were passed with high auditor satisfaction. A company with a good auditing track record may only risk a warning or incur low costs if it passes an audit with lower satisfaction, whereas a company with a poor track record will incur a significant cost (e.g., due to the implementation of many auditor recommendations). Furthermore, a high number of consecutively bad audits will also lead to more auditor scrutiny, yielding a higher probability of detection. If an audit is conducted in period t, the actual auditing coverage is denoted by q_t. The value of q_t is between 0 and 1, where $q_t = 1$ means 100% coverage. We model it as a function of the historical audit outcomes H_t,

$$q_t = g(H_t).$$

It is sensible to assume that audit coverage q_t will be high when the past audit outcomes have been poor and will be lower when the past audit outcomes have been good. This is because the auditors tend to put extra focus on companies with poor records.

The probability of detection from an audit depends directly on the audit coverage and the compliance level of the company. Broader audit coverage is associated with a higher probability of detection. A lower compliance level implicitly reflects a higher number of less-compliant parts (i.e., components, systems, or processes) of the company or a moderate number of highly noncompliant parts. Hence,

a lower compliance level is assumed to be associated with a higher detection probability. We have that

$$P(\text{detect when audit in period } t) = q_t(1 - b_t).$$

In line with our definition above, auditors may reveal $N - 1$ possible noncompliant states of the enterprise. Given the current compliance level b_t, the current audit outcome h_0 can still be uncertain and depend on uncontrollable factors outside the model. It is a random function of the current compliance level, i.e.,

$$h_0 = U(b_t),$$

where U is a random function with distribution F^U.

Let d_t define the audit outcome cost incurred after the compliance level is audited in period t. We assume that the cost d_t is the result of a function z mapping the current audit outcome h_0 and historical audit outcomes H_t to a positive real number. That is,

$$d_t = z(h_0, H_t).$$

We assume that the cost d_t is higher for a worse audit outcome h_0. Furthermore, it is also possible that the auditors may impose additional penalties on companies with poor track records.

In the next subsection we formulate our multi-period decision problem as a dynamic programming model [13, 14].

Dynamic programming model

There are three types of uncertainties in our model: the uncertainty of the inspection effectiveness (F^O), the uncertainty of detecting noncompliant behavior in an audit [P(detect when audit in period t)], and the uncertainty of the auditing outcome after a noncompliant event is detected (F^U). At the beginning of each period, the compliance manager decides which measures to implement and which type of inspection to conduct. We denote the actions in period t by A_t^M and a_t^I. The vector A_t^M is a binary vector of M elements, with its ith element representing whether measure i is implemented or maintained in period t. The value of 1 means that it is implemented or maintained in period t, while 0 represents a measure that is not implemented. Action a_t^I is an integer from 0 to I representing the type of inspection conducted in period t. The value of 0 means that no inspection is conducted.

The state of the model at the beginning of period t, denoted by S_t, consists of three components: H_t, V_t, and b_t.

Single-period cost

The cost incurred in period t, denoted by C_t, consists of three components: measure cost (implementation and maintenance costs), inspection cost, and audit outcome cost. In a non-auditing period t, when the manager decides to take actions A_t^M and a_t^I, the cost incurred is

$$\begin{aligned} C_t = & \sum_{i=1}^{M} c_i^M \,\mathrm{Ind}\Big((A_t^M)_i (V_t)_i = -1\Big)_i \\ & + \sum_{i=1}^{M} y^M \,\mathrm{Ind}\Big((V_t)_i = 0 \text{ and } (A_t^M)_i = 1\Big) \\ & + \sum_{j=1}^{I} c_j^I \,\mathrm{Ind}(a_t^I = j), \end{aligned}$$

where $\mathrm{Ind}(x)$ is an indicator function that yields value 1 if condition x is true and 0 otherwise. The first term in the above equation represents the aggregate implementation cost. The second term represents the aggregate maintenance cost, while the third term represents the inspection cost.

In an auditing period t, the audit outcome cost incurred is random and depends on the auditing result. The expected cost in period t is

$$\begin{aligned} E[C_t] = & \sum_{i=1}^{M} c_i^M \,\mathrm{Ind}\Big((A_t^M)_i (V_t)_i = -1\Big)_i \\ & + \sum_{i=1}^{M} y^M \,\mathrm{Ind}\Big((V_t)_i = 0 \text{ and } (A_t^M)_i = 1\Big) \\ & + \sum_{j=1}^{I} c_j^I \,\mathrm{Ind}(a_t^I = j) \\ & + [g(H_t)(1 - b_t)] E_{F^U}\{z[U(b_t), H_t]\}. \end{aligned}$$

The fourth term in the above equation represents the expected audit outcome cost, i.e., the cost induced by the given compliance level in t.

Recursion

We define a cost-to-go function $L_t(H_t, V_t, b_t)$ as the expected present value of the cost from period t to the end of the horizon τ, when the manager optimally manages the compliance risk, and when the current state at the beginning of period t is (H_t, V_t, b_t). We denote the one-period discount factor as γ. The dynamic programming recursion can be written as follows. In a non-auditing period t, the cost-to-go function is

$$\begin{aligned} & L_t(H_t, V_t, b_t) \\ & = \min_{A_t^M, a_t^I} \left\{ \begin{array}{l} \displaystyle\sum_{i=1}^{M} c_i^M \,\mathrm{Ind}\Big((A_t^M)_i (V_t)_i = -1\Big)_i \\ \quad \displaystyle + \sum_{i=1}^{M} y^M \,\mathrm{Ind}\Big((V_t)_i = 0 \text{ and } (A_t^M)_i = 1\Big) \\ \quad \displaystyle + \sum_{j=1}^{I} c_j^I \,\mathrm{Ind}(a_t^I = j) \\ \quad \displaystyle + \gamma E_{F_i^O}\Big\{ L_{t+1}\Big[H_t, V_{t+1}, f(V_t, V_{t+1}, a_t^I, b_t)\Big]\Big\} \end{array} \right\}, \end{aligned}$$

where

$$(V_{t+1})_i = \max \begin{cases} r_i^M & \text{if } (A_t^M)_i (V_t)_i = -1, \\ 0, (V_t)_i - 1 & \text{if } (A_t^M)_i = 1 \text{ and } (V_t)_i \neq -1, \\ -1 & \text{if } (A_t^M)_i = 0. \end{cases}$$

In an auditing period t, the cost-to-go function becomes

$$L_t(H_t, V_t, b_t) = \min_{A_t^M, a_t^I} \begin{cases} \sum_{i=1}^{M} c_i^M \text{Ind}\left((A_t^M)_i (V_t)_i = -1\right)_i \\ \quad + \sum_{i=1}^{M} y^M \text{Ind}\left((V_t)_i = 0 \text{ and } (A_t^M)_i = 1\right) \\ \quad + \sum_{j=1}^{I} c_j^I \text{Ind}(a_t^I = j) \\ \quad + \gamma\{1 - [g(H_t)(1 - b_t)]\} \\ \quad \times E_{F_i^\rho}\left\{L_{t+1}[H_t \cup 1, V_{t+1}, f(V_t, V_{t+1}, a_t^I, b_t)]\right\} \\ \quad + \gamma[g(H_t)(1 - b_t)] \times E_{F^U, F_i^\rho}\{z[U(b_t), H_t] \\ \quad + L_{t+1}[H_t \cup U(b_t), V_{t+1}, f(V_t, V_{t+1}, a_t^I, b_t)]\} \end{cases},$$

where

$$H_t \cup x = [x, (H_t)_1, \cdots, (H_t)_{K-1}].$$

The boundary condition of the program is $L_{\tau+1}(H_t, V_t, b_t) = 0$.

Solution

We have solved this dynamic program using a backward induction algorithm implemented in Java**. In the following section, we present a simple hypothetical case study and show the optimal compliance management policy obtained from our dynamic programming model. We first introduce a number of assumptions, then present our results.

Case study

Assumptions

JustStarted, Inc. is a fictitious medium-sized financial institution whose characteristics are listed in **Table 1**.

The company is subject to a new privacy regulation that includes the following two requirements:

1. *Implement role-based access control to protect and ensure the integrity of electronic data and thus protect the customers' privacy.*
2. *Implement mechanisms to ensure that customer data are of high quality and up-to-date.*

Table 1 Characteristics of the financial institution JustStarted, Inc.

Name	JustStarted, Inc.
Location	Switzerland
Company size	100
Number of customer account managers	10
Number of transaction-handling managers	20
Customer base	100,000

The compliance measures listed in **Table 2** are available to address the above requirements. We assume that all compliance measures considered provide at least an adequate compliance level. Each measure is associated with implementation and maintenance costs, a maximally achievable compliance level, and a certain implementation time. The costs are given in Swiss francs (CHF). The values in Table 2 are estimates based on experience and the characteristics of JustStarted, Inc. given in Table 1.

Measures 1 and 2 primarily address requirement 1, whereas measures 3 and 4 primarily address requirement 2. Measure 5 is of a special type in that it does not affect either of the two requirements if it is implemented alone. However, if it is implemented together with measures 3 and 4, their combined effectiveness on requirement 2 is increased. The increased effectiveness is shown in the last column.

Depending on the actual measure selected, implementation costs may include the following:

- IT implementation cost, initial user training.
- Cost of preparing handbooks and other training materials for the use of employees.
- Customer training in using the system.
- Loss of customers due to the change to a more cumbersome user interface.

Similarly, costs for maintaining the compliance measures may include costs for the following:

- Manpower for customer service.
- Manpower for fixing bugs in the IT systems.
- Ongoing customer training in the use of the system.
- Administration of user passwords.

To monitor compliance and to evaluate the effectiveness of the implemented measures, JustStarted, Inc. may perform internal inspections. There are three inspection types: Type 0, based on sampling with 5%

Table 2 Compliance measures considered by JustStarted, Inc. [Costs given in Swiss francs (CHF)].

Measure no.	*Measures*	*Implementation cost*	*Monthly maintenance cost*	*Implementation period* (months)	*Effectiveness on Requirement 1, on the first day if only one implemented* (%)	*Effectiveness on Requirement 2, on the first day if only one implemented* (%)	*Effectiveness on Requirement 2 with Measure 5* (%)
1	Six-letter password for every individual user. Three-month forced change	35,000	4,000	0	50	10	10
2	Fingerprint reader access	150,000	700	1	99	20	20
3	Manual plausibility checks/review of data	96,000	34,000	1	0	65	89.50
4	Update data per customer mail request (letter with signature)	843,333	20,833.00	1	0	80	94.00
5	Address change verification letter (sent to old address)	15,000	2,400.00	0	0	N/A	0

Table 3 Types of internal inspection performed by JustStarted, Inc.

Inspection no.	*Inspection type*	*Cost per inspection* (CHF)	*Improvement* (%)
−1	No inspection	0	0
0	Sampling with 5% coverage	400	25
1	Sampling with 50% coverage	3,500	70 w/probability 0.5, 75 w/probability 0.5
2	Full inspection (100% coverage)	6,000	95 w/probability 0.4, 100 w/probability 0.6

coverage, Type 1, based on sampling with 50% coverage, and Type 2, based on sampling with 100% coverage. The inspection types, together with associated costs and resulting improvements, are summarized in **Table 3**. Whereas the improvement factor of the inspection type 0 is assumed to be certain at 25%, the improvement factors of the other inspection types are statistically distributed random variables. For example, inspection type 1 has two possible improvement factors, 70% or 75%, each with a 0.5 probability. It is also possible not to inspect at all, which results in zero inspection cost and yields no improvement.

JustStarted, Inc. is audited every *four* periods. There are four possible outcomes, which correspond roughly to the categories used to evaluate operational effectiveness for Sarbanes–Oxley [15].

0. *Full compliance testified.* Auditors have testified that the enterprise is fully compliant with all relevant regulatory requirements. The implemented measures address and fulfill all requirements to the fullest satisfaction of the auditors.
1. *Minor deficiency detected.* Auditors have identified minor deficiencies in the way that requirements have been implemented or with respect to the effectiveness of the implemented measures. This can indicate that a necessary measure is missing, an existing measure is not properly designed, or a properly implemented measure does not operate as designed.
2. *Significant deficiency detected.* Auditors have detected a significant deficiency, which can be a minor deficiency in a significant measure or an aggregation of such deficiencies that could result in

a violation of a relevant requirement that is more than inconsequential.

3. *Material weakness found.* A material weakness is a significant deficiency or an aggregation of significant deficiencies that preclude the implemented measures from providing reasonable assurance that compliance with regulatory requirements can be achieved. The inability to provide such reasonable assurance results from one or more significant deficiencies. The existence of a material weakness precludes the responsible party from concluding that the implemented measures are effective.

The outcome of the audit depends on the compliance level at the time of the audit. Because auditing involves some uncontrollable degree of uncertainty (e.g., the auditor's subjectivity), the audit outcome for each compliance level is described by the discrete probability distributions in **Table 4**. We note that there are two degenerate cases: perfect compliance (i.e., $b_t = 1$) and perfect noncompliance (i.e., $b_t = 0$). Table 4 shows, for every compliance level, the probability prob(i) of audit outcome i ($i = 0, 1, 2, 3$).

The audit outcome cost for low compliance with a given regulatory requirement is a function of the current audit outcome and the historical record of the last k audit outcomes; here we assume that $k = 2$. Hence, only the current outcomes plus the two previous audit outcomes are considered when calculating the outcome cost of the current audit. Formally,

$$d_t = z(h_0, h_1, h_2) = \left\{ f_1(h_0) + f_3(h_0) f_2 \left[\frac{(h_1 + h_2)}{2} \right] \right\} \varphi.$$

The penalty factor φ is a constant, which we set to be 1,000,000 in our example. The functions f_1, f_2, and f_3 are defined for all possible values of h_0, h_1, h_2 and $(h_1 + h_2)/2$, as shown in **Table 5**.

In the same way that the audit outcome cost of a specific regulatory requirement depends on the historical audit outcomes, the scrutiny with which auditors inspect the compliance status of an enterprise depends on the historical audit outcomes. **Table 6** shows the percentage of audit coverage $q_t = g[(2h_1 + h_2)/3]$ as a function of the previous audit outcomes h_1 and h_2.

Results

We have solved the above problem using a Java implementation of our algorithm. Calculating the optimal solution required approximately two hours on an Intel Pentium** 4 machine rated at 3.00 GHz, with 3 GB of RAM and running Microsoft Windows**. Assuming a time horizon τ of 60 periods, with audits every four periods, and a decay factor ρ of 0.98, the program

Table 4 Probability distribution of audit outcomes as a function of compliance level.

Compliance level (%)	*Prob(0)*	*Prob(1)*	*Prob(2)*	*Prob(3)*
100	1	0	0	0
90–99	0.9	0.1	0	0
70–89	0.1	0.5	0.3	0.1
50–69	0.05	0.25	0.5	0.2
30–49	0	0.1	0.4	0.5
1–29	0	0	0.1	0.9
0	0	0	0	1

Table 5 Calculation of audit outcome costs. The penalty factor φ is a constant, here set to 1,000,000.

h_0	$f_1(h_0)$	$f_3(h_0)$
0	0	0
1	10	1
2	30	1.5
3	60	2

$(h_1 + h_2)/2$	$f_2[(h_1 + h_2)/2]$
0	0
0.5	3
1	6
1.5	10
2	13
2.5	16
3	20

Table 6 Percentage of audit coverage as a function of audit outcome history.

$(2h_1 + h_2)/3$	$g[2h_1 + h_2)/3]$ (%)
0.00	5
0.33	10
0.67	15
1.00	20
1.33	25
1.67	30
2.00	35
2.33	40
2.67	45
3.00	50

Table 7 Portfolio of optimal compliance measures for periods 0 through 3 ($t = 3$: audit period).

t	h_1	h_2	v_1	v_2	v_3	v_4	v_5	b_t	c_t	a_t^{m1}	a_t^{m2}	a_t^{m3}	a_t^{m4}	a_t^{m5}	a_t^i
0	0	0	−1	−1	−1	−1	−1	0	1,475,101.6	0	0	0	0	0	−1
1	0	0	−1	−1	−1	−1	−1	0	1,490,001.6	0	1	1	0	0	−1
2	0	0	−1	1	1	−1	−1	0	1,256,567.2	0	1	1	0	1	2
3	0	0	−1	0	0	−1	0	0.94	1,311,975.5	0	1	0	0	1	−1

Table 8 Portfolio of optimal compliance measures for periods 4 and 5, $h_1 = 0$.

t	h_1	h_2	v_1	v_2	v_3	v_4	v_5	b_t	c_t	a_t^{m1}	a_t^{m2}	a_t^{m3}	a_t^{m4}	a_t^{m5}	a_t^i
4	0	0	−1	0	−1	−1	0	0.58	1,318,696	0	1	0	0	1	−1
5	0	0	−1	0	−1	−1	0	0.57	1,328,884.9	0	1	1	0	1	−1

Table 9 Portfolio of optimal compliance measures for periods 4 through 7, $h_1 = 1$.

t	h_1	h_2	v_1	v_2	v_3	v_4	v_5	b_t	c_t	a_t^{m1}	a_t^{m2}	a_t^{m3}	a_t^{m4}	a_t^{m5}	a_t^i
4	1	0	−1	0	−1	−1	0	0.58	1,334,034	0	1	0	0	1	−1
5	1	0	−1	0	−1	−1	0	0.57	1,344,377.8	0	1	1	0	1	−1
6	1	0	−1	0	1	−1	0	0.56	1,255,356.2	0	1	1	0	1	1
7	1	0	−1	0	0	−1	0	0.94	1,326,249	0	1	0	0	1	−1

resulted in four database tables with 93,552 records each (one for each inter-audit period), which we evaluated using Structured Query Language (SQL) queries.

Under the assumptions that the first audit is conducted in the fourth period ($t = 3$) and that no compliance measure had been implemented at $t = 0$, the program calculates the optimal portfolio of measures (a^{m1} through a^{m5}) that must be implemented in the first period. It also determines the optimal inspection type a^i for the given setting. The result is shown in the row $t = 0$ in **Table 7**.

The result in row $t = 0$ informs the management that the optimal portfolio of measures in the starting period contains no measures and thus does not require any inspection ($a_t^i = -1$). Following this recommendation, the management implements no measure, and the company finds itself in the next period ($t = 1$) in the situation depicted in row $t = 1$ of Table 7. Row $t = 1$ now advises the management that measures 2 and 3 ($a_t^{m2} = a_t^{m3} = 1$) must be implemented and that there is still no inspection required ($a_t^i = -1$).

In the next period ($t = 2$), the implementation of measures 2 and 3 has not yet been completed, since both have an implementation period of 1. Row $t = 2$ in Table 7 now requires the implementation of an additional measure, a^{m5}, while measures 2 and 3 are being maintained. In addition, a full inspection is required ($a_t^i = 2$). Because the new measure has an implementation period of 0, all three measures will be effective in the next period and will simultaneously affect the company's new compliance level.

The next period ($t = 3$) is an audit period. All measures implemented by the company to date will affect the compliance level assessed by the auditors. As row $t = 3$ in Table 7 shows, the portfolio of optimal measures that are currently implemented now includes measures 2, 3, and 5, resulting in a compliance level of 0.94. The full inspection ensures that the combined effect of the implemented measures on the compliance level equals the target compliance level of the respective measures.

According to Table 4, with an initial compliance level between 90% and 99%, auditors will attest full compliance (audit outcome 0) with probability 0.9 and will detect minor deficiencies (audit outcome 1) with probability 0.1. Hence, JustStarted, Inc. may end up in either of the two states. As suggested by row $t = 3$ in Table 7, for the following period only measures 2 and 5 have to be maintained, and no inspection is conducted. Applying the decay factor of 0.98 to the target compliance level attained through the implemented measures, JustStarted, Inc. ends up with a compliance level of roughly 0.58, as shown in rows $t = 4$ in **Tables 8** and **9**. Depending on the audit outcome, h_1 is either 0 or 1. Although we cannot

predict the state at $t = 4$, the portfolio of optimal measures stays the same. JustStarted, Inc. should maintain measures 2 and 5; no internal inspection is required whatever the audit outcome.

Assuming an audit outcome h_1 of 0 and maintaining measures 2 and 5 without inspection as suggested by row $t = 4$ in Table 8, in the second period after the audit JustStarted, Inc. reaches the situation in row $t = 5$ of Table 8. With the recommendation to re-implement measure 3 while maintaining measures 2 and 5 without inspection, it is easy to see that JustStarted, Inc. now reaches a compliance state that oscillates between 0.58 and 0.94, as depicted in **Figure 1**. During audit periods measures 2, 3, and 5 are in effect, a full inspection ensures that the compliance level equals the target compliance level, and there is a high likelihood (0.9) that the auditors attest full compliance.

However, assuming an audit outcome h_1 of 1 and maintaining measures 2 and 5 without inspection, in period 5 JustStarted, Inc. reaches the situation indicated in row $t = 5$ in Table 9. In the following period, JustStarted, Inc. still maintains measures 2 and 5, implements measure 3, does not inspect, and finds itself in the situation summarized in row $t = 6$ in Table 9. The next period ($t = 7$) is again an auditing period. By now, the re-implementation of measure 3 has been completed, and JustStarted, Inc. again reaches a compliance level of 0.94 and the state captured in row $t = 7$ in Table 9.

We observe that even in the case of the worse audit outcome (i.e., $h_1 = 1$), it is also optimal to re-implement measure 3 while maintaining measures 2 and 5, and thereby achieve a compliance level of 0.94 again. As long as JustStarted, Inc. manages to attain this compliance level in the auditing periods, it will never experience an audit outcome lower than 1. Given the available measures and other assumptions of this case study, the worst possible audit outcome is that the auditors register minor deficiencies and that JustStarted, Inc. has to implement their recommendations. In such cases, and in general with a track record of subsequent audit outcomes of 1, the cost of compliance is slightly higher than in the case in which JustStarted, Inc. reaches the audit outcome 0 (which is much more likely in any case).

By calculating the evolution of the compliance level and the expected cost of compliance of the fictitious company JustStarted, Inc., we demonstrate that attaining a high level of compliance with regulatory requirements may not only be a moral obligation but may also be economically optimal.

As an additional result, **Figure 2** shows how the optimal inspection type for any given compliance level varies with the audit outcome history in a period just before an audit. With the current target compliance level being 0.58 and the previous audit outcome being 1,

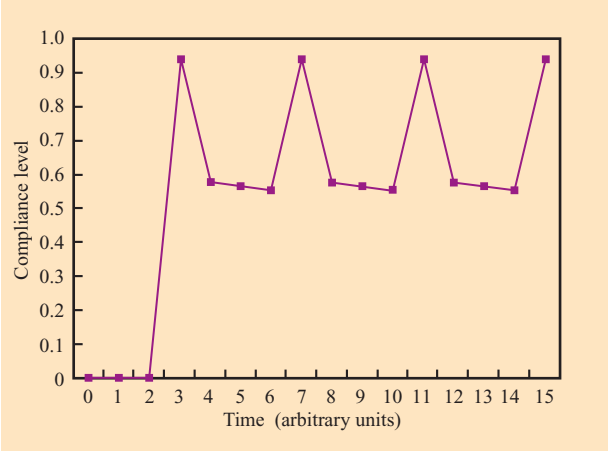

Figure 1

Evolution of compliance level over time.

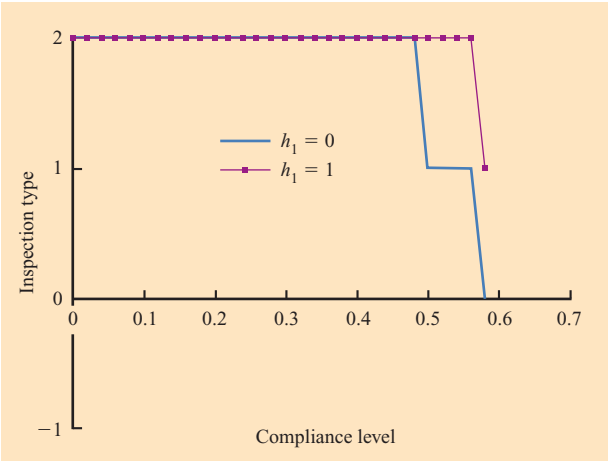

Figure 2

Post-audit optimal inspection type as function of compliance level for $h_1 = 0$, $h_1 = 1$ ($V_t = [1, 0, 1, 1, 0]$).

JustStarted, Inc. will almost always conduct a full inspection (i.e., $a_t^i = 2$) to raise its compliance level to the target level. Only in cases in which the current compliance level is already close to the target compliance level will a partial inspection with 50% coverage (i.e., $a_t^i = 1$) suffice. In the case in which the previous audit outcome was 0, JustStarted, Inc. does not require an equally stringent inspection strategy. For example, given a compliance level of 0.5, JustStarted, Inc. will conduct an inspection with only 50% coverage, as opposed to the full inspection if the audit outcome was 1.

Conclusion

Attaining perfect compliance with all regulatory requirements is idealistic but close to impossible, especially for large enterprises. Attempting to reach this lofty goal potentially consumes more resources than is economically optimal. In this paper, we describe a quantitative optimization model for dynamic and risk-based compliance management based on dynamic programming. Given a set of available measures that can be used to address compliance concerns, our approach determines the optimal portfolio of such measures that must be implemented and the optimal type and frequency of internal inspections to be instituted.

Our main contribution lies in a novel way of formulating and solving the compliance management problem. We have stressed the notion of compliance as a continuous rather than a binary phenomenon. We have shown that compliance is best managed by a risk-based approach by which we optimally select, prioritize, and implement appropriate compliance measures and determine the optimal inspection policy.

Our model demonstrates that the optimal investment for ensuring the company's compliance is calculable. In our case study, we have shown that striving to attain compliance with regulations may be not only a civic duty but also an economically optimal use of a company's resources. Furthermore, for the data used in the case study, we have found that the practice of performing comprehensive internal inspections shortly before an expected audit is economically justified.

We are aware that it may be difficult to populate our model with meaningful data. Whereas it may often be impossible to derive precise estimates of various input parameters or function definitions on the basis of solid empirical data, our tool still lends itself nicely to sensitivity analysis and scenario-based decision evaluation. The tool might therefore prove to be a valuable decision support system for managing enterprise compliance. In this context, a certain level of imprecision when estimating individual model parameters may well be tolerable, but more research is needed to ascertain this hypothesis.

We are optimistic that today's enterprises will improve further with respect to data integration through standards, harmonization, and simplification. We also observe that more and more IT systems are being instrumented to allow for event monitoring. Over time, we thus expect to see enterprises evolving toward a point at which continuous monitoring and assurance are within reach and our quantitative model can be populated with more reliable data. One can also conceive of enterprises belonging to the same industry sharing compliance risk-relevant input data (e.g., data on audit outcome cost, measure effectiveness, and audit coverage) in an anonymous form, similarly to the way in which members of a consortium of financial institutions, Operational Riskdata eXchange Association (ORX) [16], share operational risk data anonymously.

Being interested in effective regulation, governmental institutions and standards organizations might also take advantage of our model. Using our approach and assuming reliable input data, lawmakers could better evaluate whether a new regulation can be effectively enforced by simulating enterprise behavior in the face of new regulation. Governments would thus be in a position to minimize bureaucratic overhead by avoiding ineffective regulation and to induce economically efficient compliance by setting suitable incentives.

Acknowledgments

We are grateful to the anonymous reviewers for their helpful comments.

**Trademark, service mark, or registered trademark of Sun Microsystems, Inc., Intel Corporation, or Microsoft Corporation in the United States, other countries, or both.

References

1. "Sarbanes–Oxley Act of 2002," Public Law 107-204, 116 Stat 745, United States Code (2002).
2. "USA Patriot Act of 2001," Public Law 107-56, 115 Stat 272, United States Code (2001).
3. P. L. Bernstein, *Against the Gods—The Remarkable Story of Risk*, John Wiley and Sons, New York, 1996.
4. A. J. McNeil, R. Frey, and P. Embrechts, *Quantitative Risk Management: Concepts, Techniques, and Tools*, Princeton University Press, Princeton, NJ, 2005.
5. J. C. Hull, *Options, Futures and Other Derivative Securities*, 2nd Edition, Prentice-Hall, Englewood Cliffs, NJ, 1993.
6. P. J. Schönbucher, *Credit Derivatives Pricing Models: Models, Pricing, Implementation*, John Wiley and Sons, New York, 2003.
7. G. E. G. Beroggi and W. A. Wallace, "Operational Risk Management—A New Paradigm for Decision-Making," *IEEE Trans. Syst., Man & Cybernet.* **24**, No. 10, 1450–1457 (1994).
8. M. Leippold and P. Vanini, "The Quantification of Operational Risk," *J. Risk* **8**, No. 1, 59–85 (2005).
9. C. Supatgiat, C. Kenyon, and L. Heusler, "Cause-to-Effect Operational Risk Quantification and Management," *Risk Manage.* **8**, 16–42 (2006).
10. A. Gehani and G. Kedem, "RheoStat: Real-Time Risk Management," *Proceedings of the 7th International Symposium on Recent Advances in Intrusion Detection* (RAID 2004), Sophia Antipolis, France, September 15–17, 2004, pp. 296–314.
11. L.-F. Kwok and D. Longley, "Security Modelling for Risk Analysis," *Proceedings of the IFIP 18th World Computer Congress (SEC 2004)*, Toulouse, France, August 22–27, 2004, pp. 29–46.
12. "Internal Control—Integrated Framework," Committee of Sponsoring Organizations of the Treadway Commission (COSO), American Institute of Certified Public Accountants (AICPA), Jersey City, NJ, 1992.
13. R. Bellman, *Dynamic Programming*, Princeton University Press, Princeton, NJ, 1957.
14. E. Denardo, *Dynamic Programming: Models and Applications*, Dover Publications, New York, 2003.
15. "Control Objectives for Information and Related Technology (COBIT)," Version 4.0, IT Governance Institute, 2005; see *http://www.isaca.org/cobit*.

16. Operational Riskdata eXchange Association (ORX); see *http://www.orx.org/*.

Received August 9, 2006; accepted for publication December 10, 2006; Internet publication May 15, 2007

Samuel Müller *IBM Research Division, Zurich Research Laboratory, Säumerstrasse 4, 8803 Rüschlikon, Switzerland (sml@zurich.ibm.com)*. Mr. Müller received an M.S. degree in computer science and an M.A. degree in economics, both from the University of Zurich. In 2004 he joined IBM Research in Zurich, where he is currently doing research in the area of risk and compliance. In parallel, he is working toward his doctorate degree as an external Ph.D. student at the Swiss Federal Institute of Technology (ETH) Zurich, where he is a member of the Information Security group. Mr. Müller's research interests include modal logics, formal methods and modeling, risk and compliance management, game theory, and economics.

Chonawee Supatgiat *Research Group, Thales Fund Management, 140 Broadway, New York, New York 10005 (chs@thales.com)*. The work described in this paper was performed while Dr. Supatgiat was a Research Staff Member in the Business Optimization group at the IBM Zurich Research Laboratory, Switzerland. Dr. Supatgiat received a B.S.E. degree in industrial engineering from Chulalongkorn University, Thailand, in 1993, and M.S.E. and Ph.D. degrees in industrial and operations engineering from the University of Michigan at Ann Arbor in 1996 and 1999, respectively. Before joining IBM in 2004, he had worked in quantitative research functions at Enron, Tractebel/Suez, and RWE Group. He joined Thales in 2006. Dr. Supatgiat's research interests include sequential decision-making processes, large-scale stochastic optimization, financial engineering, and game theory.