INFORMS TutORials in Operations Research

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Total Quality Management and Operational Risk in the Service Industries

Marcelo CruzMichael Pinedo

TUTORIALS IN OPERATIONS RESEARCH
INFORMS 2008

 | ISBN 978-1-877640-23-0
DOI 10.1287/educ.1080.0053

Total Quality Management and Operational Risk in the Service Industries

Marcelo Cruz
Chief Risk Officer, Aviva, London, EC3P 3DQ United Kingdom, marcelo.g.cruz@gmail.com

Michael Pinedo
Stern School of Business, New York University, New York, New York 10012, mpinedo@stern.nyu.edu

Abstract This tutorial focuses on the relationship between total quality management (TQM) and operational risk in the service industries. TQM and operational risk have received an enormous amount of attention from academicians as well as from practitioners. However, these topics have been kept somewhat separate in the literature. We consider TQM and operational risk in four types of service industries, namely, transportation, healthcare, financial services, and hospitality industries. We compare the types of losses that can be incurred in each type of industry, the measurements of the losses, and the risk mitigation processes and show how TQM should be seen as part of a robust operational risk management process.

Keywords total quality management; operational risk; service industries; aviation; healthcare; financial services; hospitality services

1. Introduction

Total quality management (TQM) has received an enormous amount of attention in the manufacturing literature. Practitioners in industry as well as researchers in academia have developed an extensive literature on this subject. This is mostly due to the fact that quality in manufacturing has a direct connection to the financial bottom line; i.e., the less rework due to errors on the manufacturing line, the less the final cost of the product, thus increasing the profit margin. As an example, the cost of manufacturing in the microelectronics industry is directly proportional to the yield of the production process, and this yield is a direct function of the quality control exercised throughout the manufacturing process. This need for quality control is true in other industries as well, e.g., the paper-manufacturing industry. A higher quality also increases the value of a product or a brand in the eyes of consumer. For example, in automobile manufacturing, customer satisfaction depends heavily on the actual as well as on the perceived quality of the vehicles produced. A number of car manufacturers pride themselves on the quality of their cars, which has a positive impact on their brands and the value of their products. Manufacturing companies, therefore, pay rigorous attention to quality control as a key aspect of their business strategy.

In service industries, however, the importance of quality control is often not as clearly defined. The quality of a service is typically more closely related to consumer satisfaction and costumer loyalty than to the financial bottom line; however, service quality may still have a major impact on the bottom line of a firm. If a major operational or processing error occurs, then correcting such an error may turn out to be expensive. The firm may also be liable and could incur huge legal expenses. The service industries considered in this tutorial include:

(i) transportation industries,
(ii) healthcare industries,

(iii) financial services, and
(iv) hospitality industries.

We discuss TQM and operational risk management in each one of these four industries and compare these four industries to one another. It is clear that these industries have many aspects that overlap. It turns out that some of their most basic characteristics are actually quite similar, whereas other characteristics are different. Quality control in the hospitality industries has aspects that are similar to certain aspects in aviation. All the service industries considered rely to some extent on customer contact (call) centers. The quality control characteristics of the call centers in the different industries tend to be fairly similar as well.

Over the years, numerous metrics have been developed to gauge the quality of operations in service industries, from a managerial perspective as well as from a customer perspective. A significant amount of research has also been done on the so-called gap model, which considers the difference (gap) between service quality specifications and how quality is perceived by the customers. In most cases, quality control metrics developed in a services industry relate to events that cause limited immediate material harm to the company. Such metrics typically track what could be referred to as minor operational failures. In the aviation (transportation) industry, for example, such metrics include the percentage of on-time arrivals, the number of luggage pieces lost, and the number of near misses in flight. An example in the healthcare industry is the number of wrong diagnoses. Examples in the financial services industry include the number of checks processed incorrectly and the number of amendments or exceptions made on trades booked. Examples in the hospitality industries include the number of complaints registered by guests at a hotel or passengers on a cruise liner and the number of criminal activities recorded in a hotel or on a cruise.

Most service industries are also susceptible to operational failures that can cause major damage (e.g., airline accidents, botched surgeries). Fortunately, such events do not occur often. However, if they do happen, they can cause serious damage. In the transportation industries and in the healthcare industries, a catastrophic failure may result in loss of life. Catastrophic events can happen in the financial services industries as well. Mainly because of some serious operational risk events, large financial institutions, such as Barings in the United Kingdom and Daiwa Bank in Japan, have gone into bankruptcy or near bankruptcy. Management of operational risk has become a high priority on the agenda at large service corporations.

Operational risk may be regarded as the probability of the occurrence of an operational failure (e.g., due to a human and/or system error) that causes harm and/or financial damage. For service firms, particularly financial services firms dealing with complex derivatives, the management of operational risk, which typically starts with improving the quality of the operations, has become an important business mission. Considering the lack of a quality control culture, this new focus on quality represents a serious challenge for management.

This tutorial differs in various aspects from the many treatises on TQM and operational risk. Most treatises on operational risk discuss the topic within a framework of overall risk management and compare it to credit risk and market risk, and they limit themselves to the financial services industries. This tutorial focuses on the interplay between TQM and operational risk and the effects on service industries in general. It is organized as follows: In the next section, we relate the concepts of TQM to the concepts of operational risk. In the third section, we describe the types of operational failures that can occur in the different service industries. In the fourth section, we discuss modeling and data analysis issues. In the fifth section we consider catastrophic events and extreme value theory (EVT). In the last section, we conclude with a comparison of the various different service industries and describe their similarities and differences as far as loss potential, risk measurements, and risk mitigation procedures are concerned.

2. Total Quality Management and Operational Risk

TQM has been defined in many different ways; see, for example, the definitions in Chase et al. [4], Evans and Lindsay [10], Gitlow et al. [13], and Tapiero [18]. Chase et al. [4, p. 274] define TQM as "managing the entire organization so that it excels on all dimensions of products and services that are important to the customer." This definition does not make a distinction between manufacturing companies and services organizations. There are also many definitions for operational risk; see, for example, the definitions in Arthur Andersen [1], Chernobai et al. [5], Cruz [7], and Cruz et al. [8]. Most of these definitions came from the financial services industries. Arthur Andersen [1, pp. 3–5] defines operational risk as the risk that deficiencies in the service process design, in the information system, or in the internal controls will result in unexpected losses. Operational risk is associated with human error, system failures, and inadequate processes and controls.

The definition of TQM refers to a broad-based managerial approach. The TQM concept originated in the manufacturing industries and may currently still be more prevalent in manufacturing than in services. The definition of operational risk, however, focuses more on the probabilities of errors and on the associated financial losses. However, TQM also recognizes that there are costs associated with quality management, as shown in Table 1.

External failure costs are often the most serious ones. An example of an external failure cost is a recall in the automotive industry. Suppose a car company recalls 200,000 cars to adjust or replace one component. If the servicing cost is, say, $100 per car, then the total cost to the company would be $20 million. Even though this is a sizable loss, it is not catastrophic—although such a loss is never good for the reputation of a brand. Some losses caused by external failures have been considerably higher. When Ford and Firestone recalled millions of tires that had been mounted on the Ford Explorer, the losses for the two companies ran in the billions. It turns out that the aspects of TQM that are most closely related to operational risk involve the occurrences of external failures.

The management of quality in a service organization and the quality control that is part of the servicing process are in many ways closely related to the operational risk that is inherent to such a service. However, quality management has a focus that is slightly different from operational risk management. It often measures the comfort levels experienced by the customers being serviced (e.g., the length of the waiting time on the phone in a call center, which may result in a loss of goodwill), whereas operational risk in a servicing environment is concerned with mishaps that involve a monetary loss (e.g., an operator may give a caller wrong information resulting in a financial loss). Operational risk focuses on the risk or the probabilities of procedural failures and the expected losses that are a result of such failures. The concept of operational risk was considered very early on in commercial aviation. However, it was only in the 1990s that it started to receive attention in the healthcare industry and in the financial services industries.

In service organizations, several basic, more strategic, approaches are used to improve the (see Behara and Chase [2], Fitzsimmons and Fitzsimmons [12]) service quality and reduce the exposure to operational risk. One such approach to reduce the exposure to operational

Table 1. Costs associated with quality management.

Type of cost	Includes
(i) Appraisal costs	The cost of measurements
(ii) Prevention costs	Investments that have to be made in improve the quality
(iii) Internal failure costs	Quality control costs and the necessary repair costs that occur before the product is delivered
(iv) External failure costs	Costs incurred after a defective product has been shipped and purchased by a customer

risk is based on the design of the process in which at various points all parameters of the process are checked and rechecked. The service design process typically is based on service blueprint methods. The service blueprint methods attempt to isolate fail points (which may lead to building in redundancies).

A second approach is based on an increase in the capacity of the service system, e.g., adding more personnel or operators. For example, one way to improve the service quality in a contact center (e.g., call center) or in any other environment in which there are a number of operators working side by side and customers have to wait in line is to increase the number of operators available. The capacity increase has two effects: first, the customer experiences a shorter wait; second, because the operators are under less pressure, the likelihood of an error is reduced. TQM is often more interested in the first effect, whereas operational risk management is often more interested in the second effect.

A third approach is based on the building of an optimal level of system redundancy, i.e., to have a backup when one system goes down. Determining the optimal level of redundancy, of course, depends on the probability of a system going down (and the distribution of the down time) on one hand and the cost of building the redundancy on the other hand. This is an important aspect of reliability theory about which a significant amount of research exists. Charles Schwab built two identical information processing centers in Arizona in such a way that each one is capable of dealing with the total customer demand on its own; the two centers are built in different flood zones and are connected to different power grids.

Numerous operational, more tactical, procedures are basic to TQM and are relevant to the reduction of operational risk in the service industries. Three of the most widely used methodologies are

(i) six-sigma quality,
(ii) fail-safe procedures, and
(iii) external benchmarking.

The six-sigma methodology is one of the best-known TQM processes; see Gitlow et al. [13]. This methodology is based on a combination of statistical process control and continuous improvement. The objective is to develop large-scale production processes or servicing processes that operate with an extremely low number of defects. Among the large North American companies, General Electric (GE) in particular has been instrumental in the development and implementation of this methodology; GE has applied this methodology even to its finance arm. That this methodology is important for financial services is clear: Bank of America processes 30,000,000 checks per day, and it wants to do so with as few defects as possible. The application of the six-sigma methodology is quite appropriate here.

The fail-safe process design is also referred to as the Shingo system. These procedures are based on a distinction between errors and defects; see Shingo [17]. Clearly, human errors can never be totally avoided. However, to avoid having such an error resulting in a defect, it is important to design the processes in such a way that the human receives immediate feedback after making an error. This imposes, for example, many requirements on the user interfaces of the computer systems that are used by operators and traders. Whenever an operator or trader enters information that appears to be inconsistent with the current knowledge in the system, the operator or trader receives immediate feedback asking whether the information entered is correct. Such user interfaces may require continuous and up-to-date feeds of all relevant market information and customer information.

External benchmarking is a widely used technique in quality improvement. Whenever a process has been identified that may need improvement, it is important to collect relevant data with regard to lapses in the quality. The analysis of such data is usually based on comparisons to corresponding data at other firms in the same industry or corresponding data industry-wide. In the financial services industries, a consortium of 41 members shares their

Table 2. Comparison between TQM and operational risk management.

Six-sigma phases	Six sigma	Operational risk
1. Define	• Design a high-level map of the process. • A clear statement of the intended improvement and a list of what is important to the (internal or external) customer.	• Risk mapping, identify risks within the process. • List of identified risks in order of their importance.
2. Measure	• A more focused problem statement. • Baseline data on current process performance. • Data collection.	• Collect data (loss experience), key metrics, develop scenario analysis, etc., for the risks identified in the process. • Measure operational risk.
3. Analyze	• Define/identify the root cause(s) and confirm them with data.	• Define/identify root cause(s) and confirm them with data. • Analyze the impact of certain risk factors on the firm's capital requirements.
4. Improve	• Try and implement solutions that address root causes.	• Try and implement solutions that address these root causes. • Mitigate financial risks in the bottom line caused by errors in this process.
5. Control	• Design a monitoring system. • Before and after analysis—show benefits. • Completed documentation of results, learning, and recommendations.	• Develop a monitoring system. • Develop reports to senior management and regulators. • Develop and maintain documentation on the entire process (from risk mapping to reports).

information about losses due to operational risk with one another and keeps all members' information in a common database.

Overall, the methodologies basic to TQM and the techniques used in the management of operational risk have many similarities. For example, the various different phases within six sigma ("DMAIC"—Define, Measure, Analyze, Improve, and Control) are very similar to the methods banks are using (and regulators are suggesting) in their management of operational risk. These similarities can be seen in Table 2.

From Table 2 we can easily see how the two management philosophies relate to one another. For instance, in a six-sigma project, the first task is the mapping of the process. At this stage, operational risk managers would map the risks in the process. It is clear that operational risk managers, by improving the design of their processes, can have a significant impact on the risk levels in their institution. The same applies for any other phase of the project.

Although there are obviously synergies between six-sigma and operational risk, only two financial institutions, Bank of America and JPMorgan Chase, have gone public in their annual reports with an explicit mentioning of their six-sigma initiatives. They reported gains in the billions of dollars through quality improvements in the operations of just a few of their business units (namely, credit cards at Bank of America and equity operations at JPMorgan Chase). Although the overall impact of such TQM efforts on operational risk management has not been evaluated yet, it is only logical to imagine that a better control environment in operations must lead to a reduction in the operational losses profile and subsequently to an improvement in the statistical distribution of the losses, implying less regulatory operational risk capital.

3. Operational Failures in Service Industries

In service industries in general, there are typically four main sources of operational failures:

(i) people,
(ii) technology,
(iii) process or procedure, and
(iv) external.

All four sources of risk can play a role in virtually every service organization. People risks may be due to incompetence, fatigue, or fraud committed by an employee or manager in a firm. Technology risk events may be due to system failures, telecommunication failures, programming errors, or information risk. The process or procedural risk depends on the standard procedures in place for each activity and how closely they are being followed and monitored. In financial services, process risk may be divided into three subcategories, namely, model risk, transaction risk, and operational control risk. The model risk may be due to a methodology error or a mark-to-model error. The transaction risk may be due to product complexity, execution error, booking error, or settlement error. The operational control risk can be due to such factors as volume risk and security risk. The probabilities of certain operational errors occurring may also depend on external factors or environmental conditions. These factors and conditions may be caused by acts of nature, or by threats from the outside world (lawsuits, criminals, terrorists) in combination with questionable human decision making.

An efficient mitigation framework can only succeed by addressing all sources of risk. People risk can be reduced through better training, improved oversight, and proper staffing. Technology risk can be reduced through improved system design, optimal redundancy, and backup strategies. Process risk can be mitigated, for example, through the implementation and enforcement of proper procedures with independent oversight (neutral reporting of errors).

Many types of events that occur in service industries cause only minor material or economic damage. For example, a customer who feels that he or she has not been treated well may never return, resulting in an economic loss for the firm. The frequencies of such events may be relatively high, but the expected loss incurred each time is small. In any case, the sum of the costs of these minor events can, over a period of time, still be significant. The total cost is often quite predictable (in the expectation sense) and tends to have a relatively small variance (e.g., credit card losses each month due to fraud). However, other types of events in the service industries may occur with a very low frequency, but may cause a huge amount of damage. These events are typically referred to as catastrophic events. The damage may be measured either in economic (monetary) terms or in loss of life.

In the transportation industry, the probability of a catastrophic event occurring is low; however, such an event may entail a major loss of life, which results, of course, automatically in a severe economic loss for the organization responsible. An example can be an airplane crash due to human error (either pilot or air traffic control), resulting in a major loss of life. Another example of a catastrophic event in the transportation industry is an oil tanker running aground and breaking up due to human error with the resulting oil spill causing major environmental damage (e.g., the Exxon Valdez). In the aviation industry, many measures have been implemented to reduce the probability of a catastrophic event occurring (e.g., checklist routines by aircraft personnel, design of human and machine redundancies and backups, rules personnel must adhere to such as maximum number of consecutive hours of duty).

In the healthcare industry, the probabilities of errors are constant due to the diversity of situations healthcare practitioners have to deal with on a regular basis. The occurrence of an error may have as a consequence the death of a patient. Such an event could easily cost several millions of dollars in the form of reparations to the family. An example of a more extreme event would be a situation in which many inpatients in a healthcare institution perish due to an infection that should have been controlled. The reputational consequence of such an error can damage the image of a hospital beyond repair, threatening the continuance of its operations. Over the last decade, the healthcare industry started to implement measures similar to those developed in the aviation industry. The interplay between developing quality processes and the management of operational risk is clear.

The financial industry differs from other industries in the risk-taking process. Banks and asset managers have to take risks on a daily basis to make a profit. In this industry, risk taking is not just stimulated, it is actually rewarded. Not surprisingly, the finance world has invested heavily in the development of risk management techniques. Various types of risks surround this industry, namely credit risk, market risk, financial risk, and reputation risk. It is only over the last decade that operational risk has begun to receive attention as well, mainly because of some spectacular failures that attracted worldwide publicity. Operational risk in financial services is somewhat different from operational risk in the transportation or healthcare industries. In the finance industry, a failure typically results only in monetary damage and not in loss of life. Operational risk in finance has another dimension that is not present in aviation or healthcare that increases the risk. A person involved in any transaction may obtain a personal gain from performing transactions in an inappropriate manner. (In the aviation industry, a pilot has a clear incentive to follow regulations because his own life is also at stake.)

In contrast to the other service sectors, the financial services industry historically has not been too concerned with the quality of their transaction processing. Losses due to errors have reached levels that are almost incredible. It is fairly common in financial institutions to see daily error rates in transaction settlements in the order of 20%–50% of the total transaction volume. In other service industries, such error rates would not have been acceptable and would have made those industries not viable. This low quality of processing has only recently become a matter of concern for the financial services industry. The most important reason being the so-called mark-to-market accounting of securities (for details on this accounting process, see Cruz [7]). Because of the mark-to-market method, the front office traders and the sales people were only concerned with market profits and losses in the securities traded and completely ignored errors made in the back offices because such errors were reported in different accounts. Front office personnel, in most cases, were rewarded by the results of their trading accounts and not by the overall results of the firm. Because of this, there has not been any incentive to manage transaction processing quality and operational risks. After the Basel II Accord, banks had a regulatory mandate to identify, assess, and manage this risk (see Chorafas [6]). The cultural changes that the banks since then have gone through are significant.

Operational risk failures in financial services typically include the following types of failures:

(i) transaction errors,
(ii) loss of or damage to assets,
(iii) theft, fraud, and unauthorized activities,
(iv) regulatory, compliance, and taxation penalties, and
(v) legal liabilities.

Transaction errors may include execution errors, booking errors, or settlement errors. Loss of or damage to assets may be due to acts of nature or acts of terrorism. Because of such potential losses, a certain amount of redundancy may have to be built into the system. Fraud and unauthorized activities are fairly common in the finance world; the Securities and Exchange Commission therefore imposes mandatory rules concerning vacations and days that employees have to stay away from the office. (In Europe, such rules are not always enforced, as was evident in the case of the Societe Generale fraud in 2007.) Regulatory penalties may be incurred if insider information is used improperly, which may ultimately lead to legal liabilities as well.

After several spectacularly large operational losses, banks started to take operational risk more seriously. Risk managers quickly realized that operational risk is not just a function of past operational losses, but rather a function of the quality of the control environment within the firm. The lower the quality of the transaction processing, the higher the risk. It is becoming clear that managing the quality of the processes is key to reducing operational risk. TQM, often in the form of six sigma, has reportedly been implemented in various financial institutions with significant success. A number of global banks and insurance companies have initiated six-sigma projects in parts of their organizations.

The Basel Committee for Banking Supervision included, in its latest set of regulations, significant capital reserve requirements for financial institutions to cover operational risks. Financial institutions now have to develop tight controls and metrics for this risk to comply with the new regulations.

4. Data Analysis and Modeling

The history of operational risk measurement is relatively short. The industry where risk measurement historically has received the most attention is the airline industry (Federal Aviation Administration System Safety Handbook [11], Patankar and Taylor [16], Wells and Chadbourne [20]). A significant amount of data has been collected over the years, especially with regard to catastrophic risk. Fortunately, the number of airplane crashes has been very low over the years. Because of the rarity of these events, the data collected (especially in commercial aviation) are not that useful for any form of analysis or forecasting. Because of the low frequency of accidents, statisticians in the airline industry developed the concept of a "near miss." Near misses are well defined and categorized: the occurrences of such events are recorded and well documented. Data have been collected for several decades. It turns out that near misses occur at a frequency that is an order of magnitude higher than the frequency of actual accidents. The data available with regard to these near misses are therefore much more suitable for statistical analysis and forecasting than the data concerning actual accidents.

In the healthcare industry, some work has been done with regard to operational risk (see Caroll [3]), but not as much as in the aviation industry. Estimates have been made concerning human errors in diagnoses and the repercussions. A U.S. Institute of Medicine report suggests that in the United States each year, between 44,000 and 98,000 hospitalized patients die as a result of medical error. During the last decade, the healthcare industry has been following the lead of the aviation industry and has been implementing near-miss reporting systems as well (see Muermann and Oktem [15], Wagner et al. [19]). Data with regard to monetary compensation for medical errors are also available because most medical practitioners are insured and insurance companies have developed their own ways to price this risk. For this reason, most hospitals employ dedicated risk managers to deal with operational risk.

The analysis of operational risk in the financial services industry has started only relatively recently. The first article on the subject of operational risk in financial services was published in 1998 (see Cruz et al. [8]). One of the main challenges in measuring operational risk is the access to good-quality historical data. Because of the disregard that financial institutions have for quality control, most time series of historical losses are incomplete, absolutely nonexistent, or hard to obtain.

Due to this industry-wide lack of interest in collecting operational loss data, regulators now demand from financial institutions that they use four sources of data for the measurement of operational risk and not limit themselves to their own loss experience. These four sources are as follows:

(i) Their own historical losses, in the form of a time series of what happened inside the organization. Firms have to keep at least five years of information for the different types of operational risk following a specific Basel II classification.
(ii) Losses that took place in other firms, usually very large losses that are available in public databases.
(iii) Key risk indicators that would help assess the quality of the control environment at any point in time.
(iv) Scenario analyses. The argument for this type of information is that at many times the first three sources of data are not sufficient to assess the real level of operational risk in a financial institution.

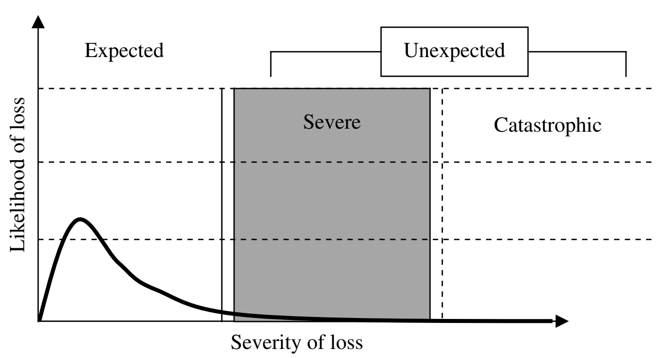
Considering the data challenges, the measurement of operational risk in financial institutions has followed two main methodologies: one, more subjective, using scenario analysis and the other, more analytical, using an actuarial approach. Currently it is believed that most financial institutions use actuarial/statistical methods to measure their operational risk. Basically, they measure separately the frequency and the severity of the losses and then combine these with simulation or fast Fourier transforms.

In financial services, usually the easiest place to find loss information is in processing errors and in the mishandling of transaction settlements. For example, the number of cancellations and amendments to all possible types of financial transactions are commonly tracked very carefully in most large firms. Because of extensive data analysis, many retail banks now know the distribution of the number of mishandling events (i.e., the number of transaction cancellations and amendments) per day as well as the mishandling loss severity distribution. An example of an outcome of such research is that the number of mishandling events as a percentage of daily volume is the highest on Mondays and the lowest on Wednesdays, and is high on Fridays.

From Figure 1, it can be seen that there are three regions in which events can occur. The first region is the region of the minor failures that occur at a very high frequency. It is easy to do a statistical analysis as well as make forecasts of errors and resulting losses in this region. The variance of the total loss in this region is low. The middle region also has a fairly large number of events, but there is already a higher variability in the total expected loss. The last region (which is the most dangerous) is the region of the catastrophic events, which happen rarely.

To measure operational risk, the financial industry has opted for the development of models that are similar to those being used for market and credit risks, the so-called value-at-risk (VaR) models. The VaR type of models, whose development began in the financial industry in the early 1990s, is considered currently a standard measure for market risk and

FIGURE 1. Distribution of operational losses.



has even been extended to credit risk (see Jorion [14]). From a market risk viewpoint, VaR estimates the losses (within certain confidence intervals) in the market value of a portfolio that one can expect until the position is neutralized. Putting it in another way, VaR estimates losses that result from holding a portfolio for a predetermined period using as a measure of volatility the asset prices over the last n days.

A similar logic can be applied to operational risk. However, because the underlying stochastic processes in market and operational risks are different, changes will have to be made in the framework developed for the market VaR. In operational risk, two separate stochastic processes have to be investigated, namely, the frequency of the losses and the severity of the losses. Statistical studies and data analysis have enabled the banks to develop stochastic models for these processes. The results of the data analysis provided a basis for the assumptions underlying these probabilistic models. The frequency distribution is often assumed to be Poisson or negative binomial, whereas the severity distribution is usually assumed to be Lognormal, Weibull, or Pareto.

The operational VaR will be an aggregation of these processes, i.e., the aggregated loss distribution. Putting it simply, VaR = frequency × severity. More practically, the operational VaR model will take as inputs time series of internal loss data and try to use this data set to estimate losses, at several confidence intervals, for different periods ahead.

The severity and frequency of the losses can easily be linked to the evaluation of scenarios, crystallized as potential future events. Their evaluation would involve learning how the frequency and severity would behave in extreme environments. To start the scenario analysis modeling process, it is necessary to develop a clear procedure for generating a representative set of scenarios that takes into account all relevant risk drivers that might influence its control environment, which ultimately will determine the operational risk level. Understanding the relationship between these risk drivers and the frequency, severity, and aggregated operational losses is important in the development of the scenarios because it is then easier to incorporate expert opinions into the model. The risk drivers can then be categorized and give rise to scenario classes (e.g., system crashes). There are quite a few ways to generate these scenarios. Some of them include stressing the parameters or results of the VaR and causal models. In accordance with the operational risk measure, the scenario analysis process must lead to an evaluation of the potential frequency and severity of the financial impact of particular scenarios. The scenario analysis would consider every possible input available such as expert opinion, key risk indicators, historical internal losses, and any relevant external events. The weighting attributed to each would depend on the quality of the information available.

In addition to the use of a pure VaR approach, in which historical losses are used to generate estimates of future losses within specified confidence intervals, for operational risk it would be important and informative to relate directly the inputs, i.e., the control environment factors and/or indicators, to the outputs, i.e., the observed losses (or the VaR figures), to attempt to understand their causal relationship. This is very similar to TQM in manufacturing, where we are trying to understand the factors in a production line that might lead to a defect.

Similarly, it would be of value to do the same when considering operational risk. We can also consider the processes in financial services as some type of "production line." To relate the observable outputs to the inputs (which should be manageable to some degree), we might seek to develop causal models that explain the influence of certain variables in the results. One approach to perform this analysis is through multifactor analysis. The dependent factors can be, for example, "minutes that a system is offline" or "number of transactions processed." Any factor representing an input variable may be used and its importance to the analysis can be even tested by factor or principal component analysis.

The form of the model is

$$\alpha_t = \beta_t + \beta_{1_t} X_{1_t} + \cdots + \beta_{n_t} X_{n_t} + \varepsilon_t,$$

where α_t represents the level of operational risk VaR (or operational losses) in the bank at time t in a particular business unit or area, X_{n_t} represents the control environment factors, and the βs are the estimated scaling parameters at the same point in time.

It is important to realize that for this model to be most effective the loss events due to operational risk should be attributed to the operating period in which the materialization of operational risk triggered the loss: in general, operational losses manifest themselves some appreciable time after the failure occurred. In a causal model such as this one, it is important that the state of the control environment is measured at the time when the error/failure occurs so that the modeling process can use the most accurate level of the inputs, leading to a more "true" model. As an example, consider a processing system that crashes for a material length of time and consequently results in some unsettled transactions that, in turn, give rise to interest claims by the counterparties, quite possibly paid on different days. If we wish to identify accurately the state of the control environment when the original error occurred, it will be necessary to identify the relevant time; this is, of course, dependent upon the recognition that appropriate systems are in place so that we are able to recognize that all of the resulting claims had the same root cause. Pursuing the losses individually might be misleading.

A numerical example will help clarify the point. Consider the hypothetical data set in Table 3 in which we have specified four control environment factors to explain the losses:

(i) system downtime (measured in minutes per day),
(ii) employees (number of employees in the back office on a given day),
(iii) data quality (in reality, a key performance indicator—the ratio of the number of transactions with no input errors from the front office to the total number of transactions on a given day), and
(iv) the total number of transactions that day.

Table 3. Transaction processing data set.

Date	Losses	System downtime	No. of employees	No. of transactions	Data quality (%)
2 Sept. 2002	42,043.83	2	13	1,115	93
3 Sept. 2002	34,732.14	1	14	1,250	95
4 Sept. 2002	20,725.78	0	14	999	96
5 Sept. 2002	14,195.16	0	15	1,012	98
6 Sept. 2002	2,213,891.54	20	11	1,512	65
9 Sept. 2002	31,654.90	1	14	1,076	94
10 Sept. 2002	39,948.60	1.5	13	1,003	95
11 Sept. 2002	11,901.87	0	15	855	99
12 Sept. 2002		0	15	915	100
13 Sept. 2002	112,169.21	4	12	1,590	78
16 Sept. 2002	80,816.23	3	13	1,390	90
17 Sept. 2002		0	15	891	100
18 Sept. 2002	65,053.57	2	13	1,422	91
19 Sept. 2002	114,862.81	4	12	1,615	75
20 Sept. 2002		0	15	920	100
23 Sept. 2002	51,006.72	2	13	1,412	90
24 Sept. 2002	24,770.00	1	15	1,215	95
25 Sept. 2002	35,232.53	1	15	1,111	93
26 Sept. 2002	35,285.33	1	15	1,115	93
27 Sept. 2002	18,460.19	0	15	997	97

TABLE 4. ANOVA.

SUMMARY OUTPUT

Regression statistics	
Multiple R	99.651%
R square	99.304%
Adjusted R square	99.118%
Standard error	45,785.94
Observations	20

	df	SS	MS	F	Significance F
ANOVA					
Regression	4	4.48463E+12	1.12116E+12	534.8133074	5.59373E−16
Residual	15	31445290155	2096352677		
Total	19	4.51607E+12			

	Coefficients	Standard error	t stat	P value
Intercept	(1,108,992.08)	543,484.5594	−2.040521771	0.059307281
System downtime	135,910.45	5,830.735382	23.30931554	3.39554E−13
No. of employees	49,712.13	17,739.10049	2.802404203	0.013394591
No. of transactions	(221.02)	106.2911473	−2.079418268	0.055146767
Data quality	578,089.20	442,238.0696	1.307190033	0.210832372

	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	−2,267,402.705	49,418.55378	−2,267,402.705	49,418.55378
System downtime	123,482.5249	148,338.3767	123,482.5249	148,338.3767
No. of employees	11,902.10885	87,522.15069	11,902.10885	87,522.15069
No. of transactions	−447.5781105	5.530603461	−447.5781105	5.530603461
Data quality	−364,519.5149	1,520,697.909	−364,519.5149	1,520,697.909

Note. df, degrees of freedom; SS, sum of squares; MS, mean square.

Using simple ordinary least squares estimation, the multifactor model in this case may be given by

$$\begin{aligned}\text{Losses} &= 1{,}108 + 135.9 \times \text{System Downtime} + 49.7 \times \text{Employees}\\ &\quad + 221.02 \times \text{Transactions} + 578.1 \times \text{Data Quality} + \varepsilon.\end{aligned}$$

Knowing the coefficients in this equation, i.e., the sensitivity of the operational losses to fluctuations in the variables, we can "price" individual units of the variables. For instance, the cost of one more minute of system downtime in any day is \$135,910.45, and therefore, any decision that leads to an improvement in the processing systems may use such figures as a starting point.

The analysis of variance (ANOVA) results are presented in Table 4. The "goodness of fit" of the model in this case is extremely high, over 99%,[1] meaning that we can have trust in the model with that degree of confidence. This level of fit does not happen frequently, but we have reasons to believe that to a very high extent there is a linear correlation between these variables (input) and losses (output).

If, for example, we were to consider that all variables took their mean level in the period 2–27 September 2002 (i.e., 2.18 minutes of downtime, 1,170 transactions, and 14 employees

[1] This good fit can be explained by the short period that the data set covers, i.e., 20 observations. Based on practical experience, more realistic long-run numbers would generally be in the range of 75%–90%, which is still relatively high.

per day), an improvement of 1% in the mean data quality factor (from 92%–93%) would result in a decrease in losses of $5,780.90. As would be expected by considering the mean values, the reduction in volatility (i.e., variability) of the input variables would result in an improved performance because less risk would be present in the system. The decrease seen, however, is an approximation as the coefficients in the equation above are conditioned on the original data set.

Another example of the usefulness of a wider application of these types of models is the following: Suppose that the management of a financial institution, due to the profitability of the products traded in this area, decides to increase the daily volume of negotiated transactions by 30%. Consequently, the board would also like to see an assessment of the likely impact on the level of operational risk as a result of pursuing this course of action but with the overriding constraint that there is to be no increase in the number of employees in this area (effectively the exercise of a discretionary option by senior management). The board then asks the operational risk management to examine the operational impact of this decision; however, letting them know that no employees can be hired because the bank wants to reduce headcount. An increase of 30% in the daily volume of transactions over the period sees the mean go up to 1,521. The average of trades during the period 2–27 September was 1,170. Using the above model we realize that, keeping all other variables constant, the daily additional losses, obtained from the equation above, would be at a level of $77,579.34. Because senior management has decreed that no employees can be hired, the management that controls the processes involved in handling transactions should find ways to improve the quality of operations without hiring additional personnel. This could be accomplished, for example, by improving the data quality factor, which is just 92%. If internal quality programs were to be developed and the quality of the input was increased to an average level of 95%, then there would be no net impact arising from the desired growth in the number of transactions. Such analyses, which are widely available in many other industries, are a novelty in the financial services industry.

Operational losses of the type described above, which are highly predictable, should be covered by the business plan. Therefore, the institution should carry the associated risk itself. It is for these types of events, which occur at a moderately high frequency and which result in minor financial losses, where TQM methodologies are most useful. The three approaches (six-sigma, continuous improvement (Shingo systems), and benchmarking) may be very effective in reducing these types of losses over time.

5. Catastrophic Events and Extreme Value Theory

In the previous section, we focused our attention on the analysis of high-frequency events that result in minor damages. Because the total costs in such scenarios are quite predictable, the analysis of these types of events is rather easy. However, the events management is most concerned with are the rare events that result in catastrophic losses and that can bring down the company (see Figure 1). Attempts have been made to draw parallels with rare events in nature such as earthquakes. However, drawing such parallels may not be entirely appropriate. In a service industry, a learning process takes place; i.e., when errors have been made in the past there may be a tendency not to make the same errors again in the future.

Because large/catastrophic events do not happen very frequently, historical loss data alone might not be a good indication of the operational risk level of the firm, at least not in the early stages of the measurement process. Considering these facts, the operational VaR model has limitations in performing robust estimation of large/infrequent events. By including scenario analysis as a supplementary model, decisions on both capital and operational risk management can be based on more information than the initially scarce data collected internally. Scenario analysis can provide useful information about a firm's risk exposure that VaR methods can easily miss, particularly if VaR models focus on the "regular" risk rather than on the risks associated with rare or extreme events. Such information can then be included in strategic planning, capital allocation, hedging, and other major decisions.

As an alternative to scenario analysis for the analysis of rare events associated with catastrophic losses, EVT is often used. EVT extends the central limit theorem (which deals with the distribution of the average of independent and identically distributed variables drawn from an unknown distribution) to the distribution of their tails (see de Haan and Ferreira [9]). One of the most well-known extreme value distributions is the so-called generalized extreme value distribution (GEV). The GEV cumulative distribution function is given by

$$F(x;\mu,\sigma,\xi)=\exp\left\{-\left[1+\xi\left(\frac{x-\mu}{\sigma}\right)\right]^{-1/\xi}\right\}$$

for $1+\xi(x-\mu)/\sigma>0$, where μ is the location parameter, $\sigma>0$ the scale parameter, and ξ the shape parameter.

A critical parameter of the extreme value distribution is the shape parameter ξ that corresponds to the weight in the tail of the distribution. If the data fit an extreme distribution, then the shape parameter probably is significant. Besides the shape parameter, it is also important to determine the scale parameter and the location parameter. Various methods have been designed to estimate these parameters, including the moments method, the probability weighted moments method, and the maximum likelihood method. Cruz [7] applied these methods to fraud data that were collected by a large British bank over the period 1992–1996. This data set contained more than 3,000 observations and gave the real losses of each one of these frauds.

The use of EVT in the analysis of operational risk has its advantages as well as its disadvantages. The main advantage is that it provides a method to analyze and forecast probabilities of catastrophic losses occurring. However, this method may not be particularly accurate.

From a risk management point of view, potential catastrophic losses have to be dealt with through some form of hedging techniques. Hedging these large operational risk events in the financial services industry is difficult. Insurance can be one option, but not always, and usually after one large event the price of the coverage becomes exorbitant. For example, recently, banks were sued for the mishandling of deals with companies such as Enron and Worldcom. All banks have settled these issues for reasonable sums. Because these banks were covered by errors and omissions policies, they were able to claim back these losses against insurers. As a consequence, insurers incurred massive losses and decided not to offer such coverage to the banks any longer. If a similar event would happen today, no financial hedge would be available. Probably the most likely type of hedging for financial firms would be securitization as banks have been doing with credit risk since the 1990s. However, this solution is still years away becuase, unlike credit risk, operational risk in banks is not rated by any one of the rating agencies, making it difficult for the capital markets to price such deals.

6. Conclusions

TQM is often concerned with the tracking and measurement of various statistics and measures concerning the behavior of the servicing system, whereas operational risk is mainly concerned with the probabilities of mishaps occurring as well as the distribution of the size of the resulting financial loss. However, any action that a TQM program comes up with to improve the operating statistics of the servicing process will have, typically, an impact on the probabilities of the mishaps as well as the distribution of the associated losses. Therefore, TQM can be seen as an activity closely connected to operational risk management or even as a mitigating factor.

Table 5. Comparison of the different service industries.

Industry	Loss potential	Risk measurement	Risk mitigation procedures
Transportation (aviation; shipping)	Major loss of life; environmental damage	Near-miss reporting systems	Checklists; redundancies
Healthcare (hospitals; nursing homes)	Loss of life	Success rate of surgeries	Second opinions; knowledge system software
Financial services (retail banks; investment banks)	Major financial losses	Losses can be measured precisely (relatively high probability of catastrophic loss)	Redundancies; hedging; insurance; securitization
Hospitality industries (hotels; cruise ships)	Limited financial losses (thefts; accidents)	Surveys; losses cannot be measured easily (low probability of catastrophic loss)	Security systems; training of personnel

The various industries considered in this paper have basically very different characteristics, which greatly affect the manner in which they apply TQM and manage their operational risk. Table 5 presents a comparison among the various different service industries.

As stated before, the measurement and the mitigation of operational risk historically have received the most attention in the aviation industry. The near-miss reporting systems that are common in the aviation industry show clearly how seriously operational risk is taken. Near misses are well defined (e.g., two planes flying within a given number of feet of one another), and the frequencies of near misses are significantly higher than the frequencies of actual accidents. Also, the causes of the near-miss events are easier to analyze than the causes of accidents because there are witnesses.

Over the last decade, the healthcare industry has begun to establish near-miss reporting systems in hospitals as well as in nursing homes; see Muermann and Oktem [15] and Wagner et al. [19]. In financial services, however, near misses are rarely recorded. There are various barriers in the financial services industries, such as fear of job loss.

The monetary losses in the financial services industries can be measured much more accurately than the losses in the other industries. However, it is clear that more can be done in the financial services industries about the mitigation of operational risk. First, one could introduce the concept of near-miss reporting systems with regard to catastrophic events. Second, the analysis of noncatastrophic events can be done more thoroughly. Such an event often could have actually resulted (with the same probability) in a real catastrophic loss (for example, a typo in the execution of a trade may result in a certain loss; however, if the notional of that trade would have been higher, the resulting loss would have been higher as well). Third, the pursuit of continuous improvement through the use of TQM principles (six-sigma, Shingo systems, benchmarking) could be pursued more vigorously as is being done in the microelectronics industry and the automotive industry.

References

[1] Arthur Andersen. *Operational Risk and Financial Institutions*. Risk Books, London, 1998.

[2] R. S. Behara and R. B. Chase. Service quality deployment: Quality service by design. R. K. Sarin, ed. *Perspectives in Operations Management: Essays in Honor of Elwood S. Buffa*. Kluwer Academic Publishers, Norwell, MA, 87–99, 1993.

[3] R. Caroll. *Risk Management Handbook for Health Care Organizations*, 5th ed. Jossey-Bass, San Francisco, 2006.

[4] R. B. Chase, F. R. Jacobs, and N. J. Aquilano. *Operations Management for Competitive Advantage*, 10th ed. McGraw-Hill/Irwin, Burr Ridge, IL, 2004.

[5] A. S. Chernobai, S. T. Rachev, and F. J. Fabozzi. *Operational Risk—A Guide to Basel II Capital Requirements, Models, and Analysis*. John Wiley & Sons, Hoboken, NJ, 2007.

[6] D. Chorafas. *Operational Risk Control with Basel II—Basic Principles and Capital Requirements*. Elsevier Butterworth-Heinemann, Oxford, UK, 2004.

[7] M. Cruz. *Modeling, Measuring, and Hedging Operational Risk*. John Wiley & Sons, New York, 2002.

[8] M. Cruz, R. Coleman, and G. Salkin. Measuring and modeling operational risk. *Journal of Risk* 1:63–72, 1998.

[9] L. de Haan and A. Ferreira. *Extreme Value Theory—An Introduction*. Springer Series in Operations Research and Financial Engineering, Springer, New York, 2000.

[10] J. R. Evans and W. M. Lindsay. *The Management and Control of Quality*. South-Western/Thomson Learning, Cincinnati, 2002.

[11] Federal Aviation Administration. Operational Risk Management. *FAA System Safety Handbook*, Chapter 15. http://www.faa.gov/library/manuals/aviation/risk_management/ss_handbook, 2000.

[12] J. A. Fitzsimmons and M. J. Fitzsimmons. *Service Management—Operations, Strategy, and Information Technology*, 5th ed. Irwin/McGraw-Hill, Burr Ridge, IL, 2005.

[13] H. Gitlow, A. Oppenheim, and R. Oppenheim. *Quality Management: Tools and Methods for Improvement*, 2nd ed. Irwin/McGraw-Hill, Burr Ridge, IL, 1995.

[14] P. Jorion. *Value at Risk*, 3rd ed. McGraw-Hill, New York, 2006.

[15] A. Muermann and U. Oktem. The near-miss management of operational risk. Technical report, The Wharton School, University of Pennsylvania, Philadelphia, 2002.

[16] M. S. Patankar and J. C. Taylor. *Risk Management and Error Reduction in Aviation Maintenance*. Ashgate Publishing Limited, Aldershot, UK, 2004.

[17] S. Shingo. *Zero Quality Control: Source Inspection and the Poka-Yoke System*. Productivity Press, Stamford, CT, 1986.

[18] C. Tapiero. *The Management of Quality and Its Control*. Chapman and Hall, London, 1992.

[19] L. M. Wagner, E. Capezuti, and J. G. Ouslander. Reporting near-miss events in nursing homes. *Nursing Outlook* 54:85–93, 2006.

[20] A. T. Wells and B. D. Chadbourne. *Introduction to Aviation Insurance and Risk Management*, 2nd ed. Krieger Publishing Company, Malabar, FL, 2000.