經營科學 第12卷 第2號
1995年 8月

A Cost Model for Determining Optimal Audit Timing with Related Considerations for Accounting Data Quality Enhancement*

Kisu Kim**

ABSTRACT

As society's relience on computerized information systems to support a wide range of activities proliferates, the long recognized importance for adequate data quality becomes imperative. Furthermore, current trends in information systems such as dispersal of the data resource together with its management have increased the difficulty of maintaining suitable levels of data integrity. Especially, the importance of adequate accounting (transaction) data quality has been long recognized and many procedures (extensive and often elaborate checks and controls) to prevent errors in accounting systems have been introduced and developed. Nevertheless, over time, even in the best maintained systems, deficiencies in stored data will develop. In order to maintain the accuracy and reliability of accounting data at certain level, periodic internal checks and error corrections (internal audits) are required as a part of internal control system. In this paper we develop a general data quality degradation (error accumulation) and cost model for an account in which we have both error occurrences and error amounts and provide a closed form of optimal audit timing in terms of the number of transactions that should occur before an internal audit should be initiated. This paper also considers the cost-effectiveness of various audit types and different error prevention efforts and suggests how to select the most economical audit type and error prevention method.

---

* This research was supported by the Yeungnam University Grants in 1994.

** Dept. of Business Education, Yeungnam University, Gyongsan 712-749, Korea.

1. Introduction

Information systems designed to respond to a broad spectrum of needs continue to proliferate. However, one of the salient characteristics of such systems is that often a gap develops between user expectations and the performance of the information technology actually implemented. Significant among the factors accounting for this situation is the data quality component [3]. According to [25], "erroneous or falsified input data is the simplest and most common cause of undesirable performance by an application system".

A prerequisite for an information to be effective is the validity of the data from which the information is derived. Information can be effective to users only to the extent that the necessary data possess a level of integrity in line with processing and user requirements [4]. As society's reliance on computerized information systems to support a wide range of activities spreads rapidly, the long recognized importance for adequate data quality becomes imperative. Furthermore, current trends in information systems such as dispersal of the data resource together with its management (distributed computing and/or distributed database systems) have increased the difficulty of maintaining suitable levels of data integrity [21].

It was found that overall error rates in different data sets, even in financial or accounting data sets which usually are assumed to relatively error free, are significantly high [15]. For example, Johnson, Leitch, and Neter [15] found evidence that error rates and amounts tend to differ by type of account, that error rates increase as the number of transactions increase, and that larger firms as well as smaller firms are subject to significant error rates. Therefore, recognizing the importance of adequate level of data validity (quality), procedures and controls for data quality enhancement are necessary for any organization. If these data quality controls are missing, data may be too much corrupted and information based on these data may be misleading.

Errors in accounting (transaction) data may cascade throughout the accounting system and corrupt the information upon which managers make decisions. The function of checking the accuracy and reliability of accounting data in a system is referred to as internal check and obviously compatible with the functions of recording and processing data. This function is an essential part of internal control system [7]. Errors in accounting data can have harmful effects upon the relationship of a business to all the major external parties with which it deals. More importantly, such errors may also damage the effectiveness of internal management, which relies upon accounting information as a basis for decision making. In many respects the maintenance of accurate and reliable records is closely related to the safeguarding of assets because the former will contribute significantly to the latter.

The importance of adequate accounting data quality has been long recognized and many-procedures (extensive and often elaborate checks and controls) to prevent errors in accounting systems have been introduced and developed [7, 9, 12, 19]. Nevertheless, over time, even in the best maintained systems, deficiencies in stored data will develop. In order to maintain the accuracy and reliability of accounting data at certain level, periodic internal checks and error corrections are required as a part of internal control system. Accounting literatures call these periodic activities internal audits. The internal audit discover and correct errors in accounting data. Generally, the shorter the time between interal audits is, the better the average quality of accounting data becomes. However, there are costs associated with internal audits, too. Therefore, internal auditors (controllers) have to decide how often the internal audit should be implemented. The auditor's goal is to audit as little and infrequently as possible, but often enough to minimize deterious economic effects upon the corporation due to significant deterioration of accounting data quality. Therefore the optimal audit frequency should balance the expected losses that occur when an audit does not take place against the cost of performing internal audit itself.

A number of researches have been made that assess the impact of data quality in inforamtion systems and develop procedures and controls effective in assuring data integrity for financial (accounting) systems. A reliability model was proposed by Cushing [6] and extended by Bodner [5], Stratton [24], and Hamlen [12]. Hamlen [12], for example, extended a framework initially formulated by Cushing [7] to develop a model that can be used to design a system of controls that ensures a specified error level. A methodology for evaluating the impact of update and correction procedures on the error rate of the stored data resource is presented by Morey [19]. Ballou and Pazer [3] presented a general model to assess the impact of data and process quality upon the ouputs of multi-user information-decision systems. Issues relating to auditing in a distributed system environment were discussed by Hansen [13]. Ballou and Tayi [4] modeled the data quality maintenance problem as an integer program and developed a heuristic to solve the problem. The goal of the procedure is to determine the most effective distribution of those resources committed to identifying and correcting errors among data sets which have different error rates and the consequence of errors can vary substantially.

Most of these researches focused on procedures designed to prevent the storage of erroneous data in computer systems or on models to assess the impact of data quality. Maintaining a satisfactory level of integrity for the stored data resourse, however, is an ongoing responsibility that requires a continual infusion of resources even in the systems with best error prevention procedures [9, p. 611-612]. A review of literature reveals only a few researches directly related to the optimal timing of audits as an integral part of an organization's internal control system for the maintenance of data quality. Hughes [14] constructed a general decision model appropriate to the determination of the optimal timing of internal audits. He solved the problem by using the back ward induction method common to dynamic programming problems. However, his model (formulation) is too complicated to implement and needs too many parameters to be estimated, and the solution method consumes a lot of computing time especially when the decision period becomes large. Boritz and Broca [6] used a deterministic model to determine the optimal fixed time interval for auditing. Morey and Dittman [20] provided a closed form solution technique for determining the optimal timing for an audit in which the required frequency for a given type of audit is driven by managerial objectives which relate to the maximum accumulated dollar discrepancy in the account that is considered tolerable between audits and a level of assurance that this dollar threshold will not be exceeded.

More recently, Liepens [18], Fox, et al. [10], Knight [16], and Redman [22] stated the data quality problem and the importance of accurate data in modern information systems. It has been found that many managers are unaware of the quality of data they use and perhaps assume that information technology ensures that data are perfect. Although poor quality appears to be the norm, rather than the exception, they have largely ignored the issue of quality. Redman [22] described a process AT & T uses to recognize poor data and improve their quality. He proposed a three-step method for identifying data-quality problems, treating data as an important assets, and applying quality systems to the processes that create data.

In this paper, first of all, we are interested in determining the optimal timing of a given type of internal audit for accounting (transaction) data quality enhancement. We provide a closed form of optimal frequencey in terms of the number of financial transactions that should occur before an internal audit should be initiated. The required frequency of internal audit is driven by the cost consequences of errors of certain sizes and the cost of internal audit of a given type. The goal is to determine the audit frequency (timing) that minimizes the total cost of performing the audits required and holding errors accumulated in an account data set in which we have both error occurrences and error amounts. We concentrate on a single account which consists of homogeneous line items (transactions) with the same error distribution, transaction rate, error holding cost, and audit cost.

In addition, we demonstrate how combining our knowledge of optimal timing by audit type, together with the cost of each audit type, can shed some useful insights as to the relative cost effectiveness of various types of audits and the selection of the most economical audit type. Similarly, it is also shown that how the knowledge of the effectiveness and costs of different types of error prevention methods (corrective actions) which effect the error rate and/or amount introduced at each transaction processing point can be used for the selection of the most cost-efficient error prevention method.

Although we concentrate on accounting data quality enhancement in this paper our model and results could be applied to any quantitative data set if the quality of a data set is assumed to be determined by the sum of erorr amounts in each transaction.

The organization of this paper is as follows. Section 2 develops a model to derive optimal audit policies for accounting data quality enhancement. In Section 3, we analyze the model developed in Section 2 and derive the optimal audit timing in terms of the number of transactions. Numerical examples and analysis are presented in Section 4. Section 5 considers the cost-effectiveness of various types of audits and error prevention methods. Error control (data quality enhancement) problem is discussed in a broder perspective. The last section provides an overview of the results, discusses several related issues, and suggests some further research.

2. The Model

We shall consider a firm which conducts internal audits to reduce the errors that may exist between the value recorded on the company's transaction files, book value, and the true value for a particular account, e.g., account receivable file. Errors may be introduced each period (weekly or daily) through keypunching, errors of omission or commission on original documents, errors in pricing, mathematical errors in extensions, etc. The number of errors committed and hence their magnitude may well be related to the number of transactions in each period. In order to attempt to capture the interrelationship between the number of transactions and the size of accumulated errors resulting, we assume that error process is represented by a compound renewal process where the amount of errors accumulated in a given period is sum of the errors associated with each transaction that occurs during the period.

Other error mechanisms are possible, for example, random errors could be introduced between transactions each period, independent of the number of transactions that occur. The appeal for focusing on this particular error mechanism is that the empirical studies [for example, [15]] have shown the size and variability of errors in a period are related to the number of transactions occurring in the period. Additionally the fact that most errors are introduced when data are updated and most updates in accounting (transaction) data take place when transactions occur appeals to use our error mechanism. Our error mechanism also makes it possible for us to use the number of transactions that have occurred since the last audit for initiating an audit instead of the elapsed real time since the last audit. The use of this trigger has a distinct advantage over using the fixed real time for initiating internal audits that if the number of transactions per period is much higher for some reason than expected, an audit can be begun earlier. In other words, if the number of transactions in a given period has greater variability for some reason than expected, an audit can be begun unnecessarily early or too late in the latter case.

Let B_i and A_i denote the book value and the true value, respectively, for the ith line item (transaction record) of an account. Then the error amount of the ith line item is defined to be

$$D_i = B_i - A_i.$$

D_i's can take either negative or positive values depending upon whether the book value is under or overstatement of the true value and zero means no error.

Let $(T_1, T_2, \cdots)$ be the sequence of time between transaction occurrences. We assume that T_i's and D_i's are independent and identically distributed random variables. T_i's take nonnegative values and the expected value of $1/\lambda$, which means the mean number of transactions per period (unit time) is λ. Then, the amount of errors introduced in an individual account until time t, $D(t)$, can be expressed as

$$D(t) = \sum_{i=1}^{N(t)} D_i$$

where $N(t)$ is the number of transactions occurred until time t.

Let the probability density function of D_i be $f(x)$ and assume that it can be estimated by "scrubbing" randomly selected groups of transactions. This would lead to a periodic updating of the error density function and accordingly the audit frequencies and possibly the type of audit. A number of authors have reported and analyzed empirical evidence on the characteristics of errors in accounting data [11, 17]. Most of them found that the shape of error amount (D_i's) distributions are far from normal but frequently highly positively skewed with concentrations in a relatively small portion of the entire range.

In general, the population of error amounts, D_i's, consists of two subpopulations. One has all zeros, which represents correct records, and the other has all nonzeros, which represents records in error [15]. Then the probability density function (pdf) for the error amount of each transaction, D_i, can be represented by

$$f(x) = (1-p) \cdot f_1(x) + p \cdot f_2(x)$$

where p is the probability that x is from the second population and $f_1(x)$ and $f_2(x)$ are pdf's for the two subpopulations. $f_1(x)$ is a spike at zero and $f_2(x)$ is the pdf of the nonzero errors. p is the proportion of nonzero errors or error rate. The mean and standard deviation of the combined population are

$$\mu_t = p \cdot \mu \text{ and } \sigma_t = p \cdot \sigma$$

respectively, where μ and σ are the mean and standard deviation of the error subpopulation, respectively.

Most systems make some kind of efforts such as increased error detection efforts when they process each transaction to prevent or reduce errors from being stored in transaction file. The error accumulation rate and, thus, $f(x)$ may depend upon the error prevention method used in the system. The more error prevention efforts is made, the lower the error accumulation rate would be, but the more cost would be required. The cost-effectiveness of different error prevention methods will be discussed later in this paper.

Errors, in general, will cost the system not exactly proportional to the error amount. Often while small error amounts may cost negligible and be ignored, error amounts above a certain level may cause serious problems and incur a lot of cost. Let $L(x)$ be the cost of holding error amount of x in each transaction per unit time. $L(x)$ is, in general, a nondecreasing function of x and is determined specifically by the managements' objective and the pattern of impacts of error amounts on the system. Typically, $L(x)$ may take the following form:

$$L(x) = \begin{cases} a_1 + a_2 x, & x > M^+ \geq 0 \\ b_1 + b_2 x, & 0 < x \leq M^+ \\ c_1 - c_2 x, & M^- \leq x < 0 \\ d_1 - d_2 x, & x < M^- \leq 0 \end{cases} \tag{1}$$

where a_1, a_2, b_1, b_2, c_1, c_2, d_1, and d_2 are nonnegative constants and M^+ and M^- are threshold error levels for over and understatements, respectively.

Managements and their internal auditors may have different objectives related to any discrepancies of each transaction (line item) that may be found due to audit. The auditor might be interested in assuring himself that no material error exists from the standpoint of overstatement, for example account receivable, sales, inventory, etc., in this case we may set c_i and d_i equal to zero. If the auditor would be interested in the absolute amount of error of each transaction irrespective of whether it is over or understated, then we may set a_i equal to c_i and b_i equal to d_i. If the managerial impact of overstatement is greater than that of understatement, a_i and b_i may be set greater than c_i and d_i, respectively. When an error is introduced in an account the error will not be corrected until an audit is taken place and the cost of holding the error in the system depends upon the error amount (negative or positive) and the time until the error is corrected.

When an audit takes place, costs associated with the audit itself, $CA(n) = CS + cc \cdot n$, accrue if an account is audited after n transactions occurred since the last audit. These costs include the fixed cost of an audit (CS) and the variable costs (cc) of finding and correcting errors which may depend upon the number of trasactions occurred before an audit. We assume that each audit take place only at the time of a transaction arrival not in between transaction arrivals. The audit might not perfectly discover and correct all errors in an account. A random residual error for the ith line item, denoted R_i, remains in an account even after any audit adjustment has been made. This will be the case if any type of sampling audit scheme is used. We assume that R_i has a general probability density function $g(x)$. The perfect audit case can be treated as a special case with $R_i = 0$ for all i.

Audit costs and the residual error may depend upon the type of audit, i. e., different types of audits may accure different audit related costs and residual errors. More through audits may need longer audit time and/or larger sample size and may result in more costs and more error free data set, in general. Accordingly, we may assume that the expected value and the variance of R_i decrease (close to zero) as the cost of audit increases. This will be discussed later in this paper with the cost-effectiveness of various types of audits.

Now we may compute the long-run average total cost per unit time as a function of the number of transaction, n. Since the times between transaction arrivals are assumed independent and identically distributed and audits occur only just after a transaction arrival not in between transaction arrivals, the times between audits, $X_1, X_2, \cdots$, are a sequence of independent and identically distributed random variables. Hence it follows that $X_1, X_2, \cdots$ constitude the arrival times of a renewal process to which the Renewal Reward Theorem [17] may be applied, and the identical probabilistic characteristics (regenerative process) of each audit cycle allow us to consider only a typical cycle. Assuming the expected cost and the expected length of each cycle are finite, the Renewal Reward Theorem states that the long-run average cost per unit time of the renewal sequence $X_1, X_2, \cdots$ is, with probability of one, equal to the expected cost of a cycle divided by the expected length of a cycle [17]. Thus if we compute both the expected cost of a cycle and the expected length of a cycle, we can compute the long-run average cost per unit time. These expected values are a function of the number of transactions, n. Hence the value of n that minimizes the long-run average cost per unit time can be found by the classical optimization method. In the next section we compute the long-run average cost per unit time and derive the optimal value of n.

3. The Optimal Audit Timing

The objective of this section is to derive the optimal audit timing in terms of the number of transactions, n, that minimizes the long-run average cost per unit time. We begin by computing the expected cost per cycle, the numerator of the long-run average cost per unit time. It consists of the cumulative expected error cost during a cycle and the expected audit cost. The former can be expressed as

$$\begin{aligned} CH(n) &= \sum_{k=1}^{n-1}(k/\lambda)\int_{-\infty}^{\infty} L(x)\, f(x)dx+(n/\lambda)n\int_{-\infty}^{\infty} L(x)\, g(x)dx \\ &= \left(\int_{-\infty}^{\infty} L(x)\, f(x)dx\right)\sum_{k=1}^{n-1}(k/\lambda)+(n/\lambda)n\int_{-\infty}^{\infty} L(x)\, g(x)dx \\ &= \left(\int_{-\infty}^{\infty} L(x)\, f(x)dx\right)(n(n-1)/2\lambda)+(n/\lambda)n\int_{-\infty}^{\infty} L(x)\, g(x)dx \end{aligned}$$

and the latter is just $CA(n) = CS + cc \cdot n$. The expected length of a cycle, the denominator of the long-run average cost per unit time, $T(n)$, is n/λ since it is the time until n transactions arrive. Therefore the expected long-run average cost per unit time, $C(n)$, is

$$\begin{aligned} C(n) &= (CH(n)+CA(n))/T(n) \\ &= (\lambda \cdot CS/n)+\lambda \cdot cc+((n-1)/2)\int_{-\infty}^{\infty} L(x)\, f(x)dx+n\int_{-\infty}^{\infty} L(x)\, f(x)dx. \end{aligned} \tag{2}$$

We find the optimal value of n by temporarily assuming that n is a continuous variable. First take the derivative of the $C(n)$ with respect to n. This yields

$$\frac{dC(n)}{dn} = -\lambda \cdot \frac{CS}{n^2}+\frac{1}{2}\int_{-\infty}^{\infty} L(x)\, f(x)dx+\int_{-\infty}^{\infty} L(x)g(x)dx.$$

Setting this derivative equal to zero, we see that n^*, the optimal value of n, satisfies

$$-\lambda \cdot \frac{CS}{n^2}+\frac{1}{2}\int_{-\infty}^{\infty} L(x)f(x)dx + \int_{-\infty}^{\infty} L(x)\, g(x)dx=0. \tag{3}$$

Since $\frac{d^2C(n)}{dn^2} = \frac{\lambda \cdot CS}{n^3}>0$ for all C, λ, $n>0$, the positive root of (3) is the unique minimizing value of n. This yields $n^*=\sqrt{\frac{\lambda \cdot CS}{K}}$ where $K = \frac{1}{2}\int_{-\infty}^{\infty} L(x)g(x)dx + \int_{-\infty}^{\infty} L(x)g(x)dx$. For a specific $L(x)$ given in (1), K becomes as follows

$$\begin{aligned} K = \frac{1}{2}\Big\{ & d_1\int_{-\infty}^{M^-} f(x)dx-d_2\int_{-\infty}^{M^-} x\, f(x)dx+c_1\int_{M^-}^{0} f(x)dx-c_2\int_{M^-}^{0} xf(x)dx \\ & +b_1\int_{0}^{M^+} f(x)dx+b_2\int_{0}^{M^+} x\, f(x)dx+a_1\int_{M^+}^{\infty} f(x)dx+a_2\int_{M^+}^{\infty} x\, f(x)dx \Big\} \\ & +d_1\int_{-\infty}^{M^-} g(x)dx-d_2\int_{-\infty}^{M^-} x\, g(x)dx+c_1\int_{M^-}^{0} g(x)dx-c_2\int_{M^-}^{0} x\, g(x)dx \\ & +b_1\int_{0}^{M^+} g(x)dx+b_2\int_{0}^{M^+} x\, g(x)dx+a_1\int_{M^+}^{\infty} g(x)dx+a_2\int_{M^+}^{\infty} x\, g(x)dx. \end{aligned}$$

Then the long-run average cost per unit time when internal audits are taken place every time n^* transactions have been processed since the last internal audit, $C(n^*)$, is simplified as

$$C(n^*)=2\sqrt{\lambda \cdot CS \cdot K} + \lambda \cdot cc - \frac{1}{2}\int_{-\infty}^{\infty} L(x)\, f(x)dx. \qquad (4)$$

If n^* is not an integer, compute $C(n)$ for $n = [n^*]$ and $n = [n^*] + 1$, where $[n^*]$ represents the greatest integer less than or equal to n^*. The value of n that yields the smaller value for $C(n)$ is the optimal integer value of n. Note that we can assume any form of error holding cost, $L(x)$, and any probability density functions, $f(x)$ and $g(x)$, for the error amount in each tranction, D_i, and the residual error amount after each internal audit, R_i, respectively. This makes our model very general so that it can be applied to any specific real situation.

4. Numerical Examples and Analysis

To illustrate our model and the optimal audit timing, consider the following scenario for an account receivable file where we define a period (unit time) to be one working day. It is assumed that the distribution of error amount introduced in each transaction, $f(x)$, is a positively skewed gamma-type pdf. Suppose further:

(i) the average number of transactions per day, λ, is 10.

(ii) the random errors introduced into the account have a mean of \$5, i. e., $\mu_t = \$5$.

(iii) the error holding cost function, $L(x)$, takes the form given in (1) and parameters of the function have the values of $M^+ = \$15$, $M^- = -\$15$, $a_1 = d_1 = \$50$, $a_2 = d_2 = 1$, $b_1 = c_1 = \$0$, and $b_2 = c_2 = \$0$.

(iv) the audit cost of a given audit type has a fixed cost of \$6000 and a variable cost of \$2 per transaction, i. e., $CS = \$6000$ and $cc = \$2$.

(v) the residual error amount of a particular audit type, R_i, is exponentially (normally) distributed and has the mean of \$2 and standard deviation of \$2.

Then n^* and $C(n^*)$ could be calculated numerically with an estimated gamma pdf $f(x) = (x/c)^{c-1} [exp(-x/b)]/b\, \Gamma(c)$ where $\mu_t = bc = 5$ for the error amount distribution, and an estimated normal residual error distribution with a pdf of $g(x) = \frac{1}{\sigma(2\pi)^{\frac{1}{2}}} \exp\left[-(x-\mu)^2/2\sigma^2\right]$

where $\mu = 2$ and $\sigma = 2$, or exponential residual error distribution with a pdf of $g(x) = (1/b)\, exp(-x/b)$ where $b = 2$. For example, if we assume that the parameters of the error amount pdf are specifically $b = 3$ and $c = 5/3$ and the residual error distribution is an exponential distribution with the mean of 2, then

$$
\begin{aligned}
K = \frac{1}{2} \{ & 50 \int_{15}^{x} (x/(5/3))^{(5/3)-1} [\exp(-x/3)]/3\Gamma(5/3)dx \\
& +0.1 \int_{15}^{x} (x/(5/3))^{(5/3)-1} [\exp(-x/3)]/3\Gamma(5/3)dx \} \\
& +50 \int_{15}^{x} (1/2) \exp(-x/2)dx + 0.1 \int_{15}^{x} x(1/2) \exp(-x/2)dx = 0.667
\end{aligned}
$$

and thus $[n^*] = [\sqrt{\frac{(10)(6000)}{(0.667)}}] = 300$ transactions which means, in the average, audits take place in every 30 working days, and

$$
\begin{aligned}
C(n^*) = & \sqrt{(10)(6000)(0.667)} + (10)(2) \\
& - \frac{1}{2} \{50 \int_{15}^{x} (x/(5/3))^{(5/3)-1} [\exp(-x/3)]/3\Gamma(5/3)dx \\
& +0.1 \int_{15}^{x} x(x/(5/3))^{(5/3)-1} [\exp(-x/3)]/3\Gamma(5/3)dx\} \\
= & \$420.
\end{aligned}
$$

We used a well known computational algebra package called MathCAD [1] to calculate these values. A comparison of the differences in the optimal audit frequencies and the long-run average costs per unit time (LACPUT) for several different values of parameters of the gamma error amount pdf with an exponential residual error pdf and with a normal residual error pdf is given in Table 1 and Table 2, respectively. This comparison is also shown graphically in Figure 1 and Figure 2. We observe that with a fixed $bc = 5$ as b decreases audits could be made less frequently with less LACPUT. That is because decreasing c with a fixed $bc = 5$ means decreasing the standard deviation of the error amount distribution, $\sigma_1 = b\sqrt{c}$, with the fixed mean $\mu_1 = bc = 5$. As the standard deviation of the gamma error amount distribution gets smaller with a fixed mean, the probability of having large error amount also gets smaller and thus the cumulative error holding cost increases slower, which leads to the less frequent audits and in turn results in the less LACPUT. The comparison between the exponential and normal risidual error distributions shows that at the same mean and standard deviation the optimal audit frequency and associated LACPUT in the latter case is always less frequent and thus smaller, respectively, than those in the former case.

Table 1. Audit Frequencies and LACPUT for Different Parameter Values With Exponential Residual Error

b	5	4	3	2	1.25	1
c	1	1.25	5/3	2.5	4	5
*n**	213	242	300	448	818	1071
c(*n**)	582	514	420	287	166	132

Table 2. Audit Frequencies and LACPUT for Different Parameter Values With Normal Residual Error

b	5	4	3	2	1.25	1
c	1	1.25	5/3	2.5	4	5
*n**	215	245	306	472	994	1606
c(*n**)	576	507	411	274	141	95

Fig. 1 Comp. of Opt. Num. of Trans.(n*)

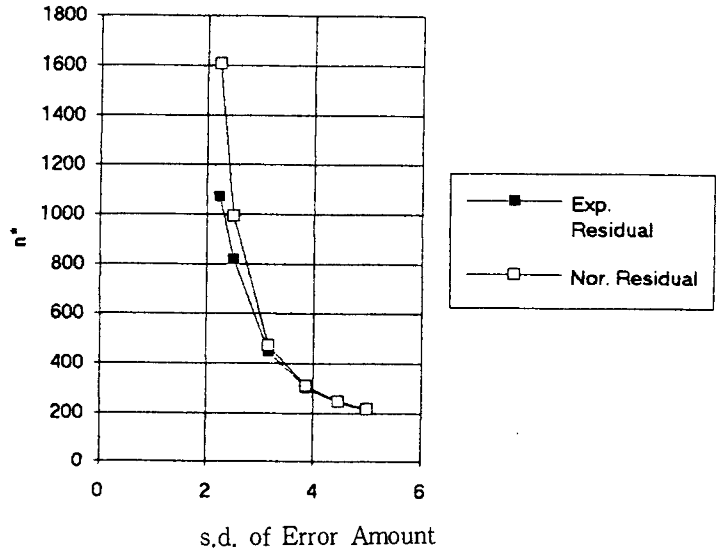
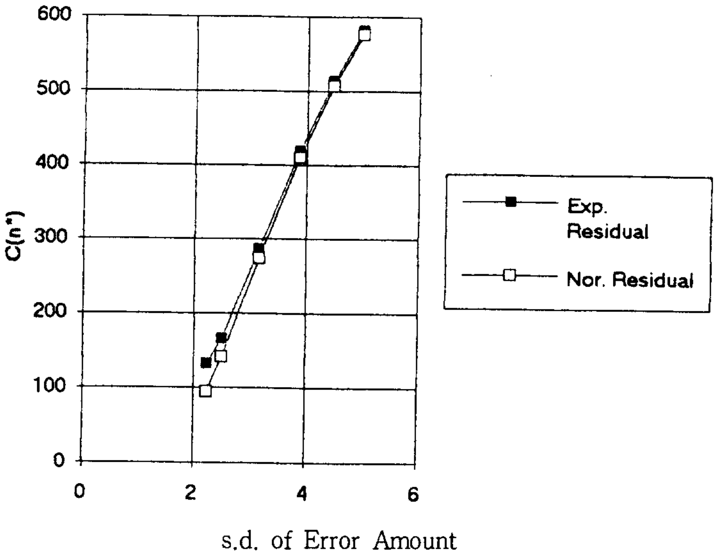


Fig 2. Comp. of Opt. LACPUT(C(n*))



5. Considerations of Relative Cost-Effectiveness of the Various Types of Audits and Error Prevention Methods

Audit costs and the residual error may depend upon the type of audit. More through audits may need longer audit time and/or larger sample size and thus may result in more costs and more error free account data, in general. Accordingly, we may assume that the expected value and the variance of R_i decrease (close to zero) as the cost of audit increases. Of course the shape of probability distribution of the random residual error may also be dependent upon the audit type. Since there is a tradeoff between the cost and effectiveness of different types of audit, management is interested in selecting the most economical audit type over all.

Suppose management has estimated the cost and precision associated with each of N different types of audits where the precision is measured by the mean, standard deviation, and/or the form of the probability distribution function of any residual error R_{ij} ($j = 1, 2, \cdots, N$). Management (Controller) desires to select the most cost-effective audit type. Given the respective imprecision and cost of an imperfect audit of type j ($j = 1, 2, \cdots$, N), we can calculate n_j^* *and* $C(n_j^*)$. Then it is straitforward to select an audit type j which have the smallest $C(n_j^*)$ as the most cost-effective audit type. However, it is a cumbersome computation process and takes a lot of computation time if N is large. We may be able to reduce the computation somewhat by observing the followings.

We know $n_j^* = \sqrt{\frac{\lambda \cdot CS_j}{K_j}}$ and

$$C(n^*_j) = (\lambda \cdot CS/n^*_j)+\lambda \cdot cc+((n^*_j-1)/2)\int_{-\infty}^{\infty} L(x)\, f(x)dx + n^*_j \int_{-\infty}^{\infty} L(x)\, g_j(x)dx$$

$$= 2\sqrt{\lambda \cdot CS_j \cdot K_j} + \lambda \cdot cc_j - \frac{1}{2}\int_{-\infty}^{\infty} L(x)\, f(x)dx$$

from section 3 where $K_j = \frac{1}{2}\int_{-\infty}^{\infty} L(x)\, f(x)dx + \int_{-\infty}^{\infty} L(x)\, g_j(x)dx$ and $g_j(x)$ is the pdf of random residual error for aduit type j. First, K_j and then $C(n_j^*)$ could be calculated without knowing n_j^*, the optimal number of transactions before initiating an audit of type j, for each audit type $j = 1, 2, \cdots, N$, then the most cost-effective audit type would be found, which is the one with the smallest $C(n_j^*)$, i. e., audit type k such that $C(n_k^*))<C(n_j^*)$ for all $j \neq k$. Once the most cost-effective audit type is selected the optimal number of transactions, n_j^*, could be found from the above for $j = k$ with K_k which is already known.

Similarly we may consider the cost-effectiveness of different types of error prevention efforts (methods). Suppose management has estimated the cost and effectiveness associated with each M different error prevention methods where the effectiveness is measured by the probability distribution of the error amount introduced in each transaction, D_i, for $i = 1, 2, \cdots$. Given the respective effectiveness and cost of an error prevention method l ($l = 1, 2, \cdots, M$), we can calculate n_l^* and $C_T(l^*) = C(n_l^*) + CP_l$ where $n_l^* = \sqrt{\frac{\lambda \cdot CS}{K_l}}$ and CP_l is the cost per unit time of the error prevention method l. Here $K_l = \frac{1}{2}\int_{-\infty}^{\infty} L(x)\, f_l(x)dx + \int_{-\infty}^{\infty} L(x)g(x)dx$ *where* $f_l(x)$ is the pdf of the error amount introduced in each transaction under the error prevention method l and

$$C(n_l^*) = (\lambda \cdot CS/n_l^*) + \lambda \cdot cc + ((n_l^*-1)/2 \int_{-\infty}^{\infty} L(x)\, f_l dx + \int_{-\infty}^{\infty} L(x)\, g(x)dx$$

$$= 2\sqrt{\lambda \cdot CS \cdot K_l} + \lambda \cdot cc - \frac{1}{2}\int_{-\infty}^{\infty} L(x)\, f_l(x)dx.$$

Then the error prevention method k such that $C(n_k^*) + CP_k < C(n_l^*) + CP_l$ for all $l \neq k$ is the most cost-effective and n_k^* is the optimal number of transactions before initiating a given type of audit under the most cost-effective error prevention method k. Again for each error prevention method $l = 1, 2, \cdots, M$, K_l and then $C(n_l^*)$ could be calculated before computing n_l^*, then the error prevention method with the smallest value of $C(n_l^*) + CP_l$ for $l = 1, 2, \cdots, M$ would be selected as the most cost-effective one. The optimal number of transactions under this error prevention method, say method K, could be calculated easily with K_k which is already calculated.

For example, suppose management has six alternative error prevention methods whose costs (per unit time) and effectivenesses which are measured by the mean error amount in a line item are estimated as below (in Table 3). Both the distributions of error amount in a line item and the residual error amount in a line item after an internal audit are assumed to be exponential and other costs and parameters are assumed to be the same as those assumed in section 4. The results are summarized in the following Table 3. The error prevention method 4 turned out to be the most cost-effective and the optimal number of transactions before implementing an internal audit, n^*, becomes 213 when this error prevention method is adopped.

Table 3. Effectiveness of Different Error Prevention Methods

Err. Prev. Method	1	2	3	4	5	6
CP_l	100	300	450	600	800	1000
μ_{tl}	10	8	6	5	4	3
$C(n_l^*)$	1203	1001	739	582	411	241
$C(n_l^*)+CP_l$	1303	1301	1189	1182	1211	1241

$n^*=213$

6. Summary and Conclusion

As society's reliance on computerized information systems to support a wide range of activities spreads rapidly, the long recognized importance for adequate data quality becomes imperative. Among data used for managerial decision making in organizations accounting (transaction) data is the most common and fundamental. If the controls on these data quality are missing, errors may cascade throughout the accounting system and corrupt the information upon which managers make important managerial decisions. In order to maintain the accuracy and reliability of accounting data at certain level, periodic interal checks and error corrections (internal audits) are required since, over time, even in the best maintained systems deficiencies in stored data will develop.

In this paper we developed a general data quality degradation (error accumulation) and cost model for an account in which we have both error occurrences and error amounts and provided a closed form of optimal timing in terms of the number of transactions that should occur before an internal audit should be initiated. This paper also considers the cost-effectiveness of various audit types and different error prevention efforts and suggests how to select the most economical audit type and error prevention method. For these works this paper utilized the informations on the mean intertransaction time, the probability distribution function of the error amount in each transaction, the cost function of errors accumulated in an account, the precisions (in terms of the pdf of residual error) and costs of various audits, and the effectivenesses (in terms of the pdf of error amount introduced in each transaction) and costs of different error prevention efforts.

The model and policies developed in this paper assumes the above informations have been available and are reasonably stable. In practice the mean intertransaction time and the pdf of the error amount would be estimated by "scrubbing" randomly selected groups of transactions. This would lead to a periodic updating of the audit frequencies and possibly the type of audit and the error prevention method to be utilized. Determining the cost of errors is somewhat more involved but not too difficult. The analytical models which assess the impact of data quality upon the output information [for example, [3]] could be used for the estimation of the error costs. Estimation of the audit cost and the pdf of residual error may be obtained each time an audit is performed. Similarly the cost of each error prevention effort and its effectiveness could be estimated by observing the system each time a particular error prevention effort is applied.

Though our approach is not the most efficient and clearly puts the burden on management (controller) to estimate several things, it does represent useful relationships among the error process, the error costs, and the audit costs and also provide a powerful decision aid to help assess the cost impacts of various audit types and error prevention efforts. Importantly, we can assume any form of error holding cost, $L(x)$, and any probability density functions for the error amount in each transaction, $f(x)$, and the residual error, $g(x)$. This makes our model very general so that it can be applied to any specific real situation. Also we allow to consider both overstated and understated errors.

This paper concentrates on the data audits of a single account and ignores the possible interdependencies between accounts. For example, we do not address the cost savings that could result from obtaining information from the audit of an accout, e. g., account receivable, which might provide information concerning the account balances in another related account, e. g., sales. In the future research we would develop a more general model and derive optimal policies considering this situation (problem).

REFERENCES

[1] Anderson, R. B., ***The Student Edition of MathCAD, Verson 2.0***, Addison-Wesley, Reading, M. A., 1989.

[2] Baily, R. W., ***Human Error in Computer Systems***, Prentice-Hall, Englewood Cliffs, N. J., 1983.

[3] Ballou, D. P. and Pazer, H. L., "Modeling Data and Process Quality in Multi-input, Multi-output Information Systems, ***Management Science***, Vol. 31, No. 2 (1985), pp. 150-162.

[4] Ballou, D. P. and Tayi, G. K., "Methodology for Allocating Resources for Data Quality Enhancement", ***Communications of the ACM***, Vol. 32, No. 3 (1989), pp. 320-329.

[5] Bodner, G., "Reliability Modeling of Internal Control Systems", ***Accounting Review***, Vol. 50, No. 4 (1975), pp. 747-757.

[6] Boritz, J. E. and Broca D. S., "Scheduling Internal Audit Activities", ***Auditing: A Journal of Practice & Theory***, Vol. 6, No. 1 (1986), pp. 1-19.

[7] Cushing, B. E., "A Mathematical Approach to the Analysis and Design of Internal Control Systems", ***Accounting Review***, Vol. 49, No. 1 (1974), pp. 24-41.

[8] Cushing, B. E. and Romney, M. B., ***Accounting Information Systems and Business Organizations***, 4th Ed, Addison-Wesley, Reading, M. A.,1987.

[9] Davis, G. B. and Olson, M. H., ***Management Information Systems: Conceptual Foundation, Structure and Development***, McGraw-Hill, New York, 1985.

[10] Fox, C. J., Levitin, A. V., and Redman, T. C., "The Notion of Data and Its Quality Demensions", ***Information Processing and Management***, Vol. 30 (1994), pp. 9-19.

[11] Ham, J., Losell, D., and Smieliauskas, W., "An Empirical Study of Error Characteristics in Accounting Populations", ***Accounting Review***, Vol. 60, No. 3 (1985), pp. 387-406.

[12] Hamlen, S. S., "A Chance Constrained Mixed Integer Programming Model for Internal Control Design", ***Accounting Review***, Vol. 55, No. 4 (1980), pp. 578-593.

[13] Hansen, J. V., "Audit Considerations in Distributed Processing Systems", ***Communications of the ACM***, Vol. 26, No. 8 (1983), pp. 562-568.

[14] Hughes, J. S., "Optimal Internal Audit Timing", ***The Accounting Review***, Vol. 12,No. 1 (1977), pp. 56-68.

[15] Johnson J. R., Leitch, R. A., and Neter, J., "Characteristics of Errors in Accounts Receivables and Inventory Audits", ***Accounting Review***, Vol. 56, No. 2 (1981), pp. 270-293.

[16] Knight, B., "The Data Pollution Problem", ***Computerworld***, Vol. 28, September (1992).

[17] Kreutzfeldt, R. W. and Wallace, W. A., "Error Characteristics in Audit Populations: Their Profile and Relationship to Environmental Factors", ***Auditing: A Journal of Practice & Theory***, Vol. 5, No. 1 (1986), pp. 20-43.

[18] Liepens, G. E., "Sound Data Are a Sound Investment", ***Quality Progress***, Semptember 1989, pp. 61-64.

[19] Morey, R. C., "Estimating and Improving the Quality of Information in a MIS", ***Communications of the ACM***, Vol. 25, No. 5 (1982), pp. 337-342.

[20] Morey, R. C. and Dittman, D. A., "Optimal Timing of Account Audits in Internal Control",

Management Science, Vol. 32, No. 3 (1986), pp. 272-282.

[21] Nesbit, I. S., "On Thin Ice: Micros and Data Integrity", *Datamation*, Vol. 31, No. 21 (1985), pp. 80-85.

[22] Redman, T. C., "Improve Data Quality for Competitive Advantage", *Sloan Management Review*, Winter 1995, pp. 99-107.

[23] Ross, S. M., *Introduction to Probability Models*, Academic Press, New York, 1981.

[24] Stratton, W. O., "The Reliability Approach to Internal Control Evaluation", *Decision Science*, Vol. 12 (1981), 51-67.

[25] U. S. National Bureau of Standards, *Guideline for Automatic Data Processing Risk Analysis*, FIPS Publication 65, U. S. Dept. of Commerce, Springfield, V. A., 1979.