STATISTICA, anno LXXIV, n. 1, 2014

ON A TEST OF HYPOTHESIS TO VERIFY THE OPERATING RISK DUE TO ACCOUNTANCY ERRORS

Paola M. Chiodini

Dipartimento di Statistica e Metodi Quantitativi, Università degli Studi di Milano – Bicocca, Via Bicocca degli Arcimboldi 8, 20126 Milano, Italia

Silvia Facchinetti

Dipartimento di Scienze Statistiche, Università Cattolica del S. Cuore, Largo Gemelli 1, 20123 Milano, Italia

1. INTRODUCTION AND BACKGROUND

In the last years the decision making process had to face a growing "dynamism" of the events: market changes, the use of technical innovations and communication means, industry specializations, which lead to follow the different operations made by the firm, to underline the irregularities that can generate crisis, economic losses or reductions in market share (Chiodini and Magagnoli, 2004; Low, 2004).

Internal control includes systematic measures adopted by an organization to conduct business in an efficient manner, to safeguard assets and resources, to check the accuracy and reliability of the accounting data, to promote operative efficiency, to produce reliable and timely financial and management informations, to encourage adherence to prescribed managerial policies and to detect errors and frauds. Therefore, a system of internal control extends beyond the matters which relate directly to the functions of the accounting and financial departments. Generally, controls can be of three types: a preventive control, designed to discourage from occurring errors or irregularities; a corrective control, designed to correct errors or irregularities that have been detected; a detective control, designed to find errors or irregularities after they have occurred (Guy *et al*., 2002).

The demands of the firms of a control extended to all phases of its organization have led to a wide use of statistical analysis procedures in order to locate the system irregularities, to assess the market and to take operational and strategic decisions in conditions of uncertainty (Ashton and Ashton, 1988; Kriens and Veenstra 1985).

The aim of this study is to examine a new suitable procedure for internal auditing in order to define the error risk that could imply distortions and wrong decisions by the management. The audit system develops in different steps: some are not susceptible to sampling procedures, while others may be held using sampling techniques. In usual sampling techniques adopted in auditing, sampling plans are used to estimate the amount of the accountancy during time (i.e. one year), with an inference about the series of transactions that is assumed as the "statistical population" (Arens and Loebbecke, 1981; Smith 1976). Such assumption is denoted "static" or "ex post". In this study the same informations are used to follow the data development during time (Brown and Rozeff, 1979; Caprara, 1988) and to estimate their behaviour from a "dynamic" point of view. In particular, we introduce a statistical test of hypothesis and irregularity signal, that can be connected with the ones applied in the production processes known as "control charts".

A control chart is an on-line statistical tool used in statistical quality control to detect quickly the occurrence of assignable causes or process shifts so that investigation of the process and corrective actions may be undertaken (Montgomery, 2005; Kanji, 2004). Control charts are used to monitor a process for some quality characteristic that can be measured and expressed numerically such as thickness, weight and defective fractions (control charts for variables) or that are attributes and expressed categorically, for example "conforming" or "non-conforming", "defective" or "non-defective" (control charts for attributes).

In auditing, both internal and external, the concept of risk is very complex (Teitlebaum and Robinson, 1975; Libby *et al.*, 1985; Houston *et al.*, 1999; Cleary and Thibodeau, 2005), with reference to the different kinds of fraud and acceptability that have a degree of subjectivity. For a review on statistical fraud detections see Bolton and Hand (2002). Therefore the study assumes as risk measure the probability of "no signal" when some irregularities are happening and hence in terms of the probability of the second type β in function of the shift of the accountancy from the standard conditions. This represents the probability to not report a situation of irregularity in the accounts in the event of removal of the accounting process and therefore it is considered right in terms of regularity. The aim is to monitor the possibility of intervening with a comprehensive analysis of the accounts when it does not meet the conditions required under the null hypothesis (regularity of the accounting entries) and to reduce the probability of not reporting a situation of irregularity in the accounts in terms of number of transactions affected by errors in the sample that overcome a given threshold (natural error rate), and the mean amount of monetary errors found in incorrect records. Therefore we compute a mixed structure, for variables and for attribute different on the standard control procedure.

The paper is organized as follows. The next section explains the marginal and the joint test of hypotheses to verify the regularity of the accounting system. In section 3 the decisional procedure consisting of two hypothesis systems is proposed. In particular, the first system is referred to the analysis of the first marginal test of hypothesis relative to the frequency of accounting errors p, with particular attention to the differences between *non-randomized* or *randomized* test. While the second system is referred to the analysis of the second marginal test of hypothesis relative to the mean (or median) of the book errors conditioned from the first one. Here the distinction is between *unilateral* and *bilateral* test. Moreover, the joint operative characteristic function of the test is calculated. Defined the decision-making procedure that allows to accept or reject the null hypothesis, in section 4, a sensitivity analysis of the procedure is described. Finally, in section 5 some future developments are presented.

2. MODEL AND TEST OF HYPOTHESES

Within the audit accounting, the risk concept is highly complex and difficult to apply in relation to different types of possible frauds and to their acceptability degree.

The given check procedure is a multiple test of hypothesis, iteratively applied to the accountancy records that belongs to homogeneous classes and related to a short time $I_t = (t - \Delta t, t]$, with $\Delta t = 1$, (i.e.: one day), where for the N_t accountancy records available, n_t of them are analysed.

In audit context two quantitative elements play a role of particular interest. The frequency of a book error p and the mean (or median) θ_L of the conditional distribution of the error random variable $X = Vc - Vr$, given by the difference between the real book value Vc and the recorded value Vr, that follows a known law $F(x; \theta_L, \theta_D)$ depending on a location parameter θ_L and on a dispersion parameter θ_D (that is considered constant).

In other words, we can refer to the following two systems of hypotheses:

1. a system referred to the frequency of accounting errors p, $X \neq 0$ or $\left(X^+ > 0, X^- < 0\right)$, in I_t period:

$$\begin{cases} H_0^{'} : p \leq p_0 \\ H_1^{'} : p > p_0 \end{cases} \tag{1}$$

and we accept the null hypotheses if the error fraction p is below the value p_0 (few errors);

2. a system referred to the mean (or median) of the book errors, $X \neq 0$ or $\left(X^+ > 0, X^- < 0\right)$, in I_t period. This may be *unilateral*, by checking the only positive (or negative) errors:

$$\begin{cases} H_0^{''} : \theta_L \leq \theta_{L0} \\ H_1^{''} : \theta_L > \theta_{L0} \end{cases} \tag{2}$$

and we accept the null hypotheses if the mean value of the accountancy errors θ_L is below the value θ_{L0} (small errors); or *bilateral*, by checking the presence of errors both positive and negative

$$\begin{cases} H_0^{''} : |\theta_L| \leq \theta_{L0} \\ H_1^{''} : |\theta_L| > \theta_{L0} \end{cases} \tag{3}$$

The values p_0 and θ_{L0} are boundary values in the acceptable conditions of the accounting. The operative choice of these values depends on the sample size that is linked to the number of the observed units N in the considered period of time and on the significance level α.

Consequently it is possible to set up a complex system of hypotheses to verify the regularity of the accounting system when $p \leq p_0$ or $\theta_L \leq \theta_{L0}$ or of the both, in the situation of the *unilateral* test.

Then the joint system of hypotheses is

$$\begin{cases} H_0 : p \leq p_0 \wedge \theta_L \leq \theta_{L0} \\ H_1 : p > p_0 \vee \theta_L > \theta_{L0} \end{cases} \tag{4}$$

and similarly in the situation of the *bilateral* test.

3. THE DECISIONAL PROCEDURE

As said before, the proposed procedure consists of two hypothesis systems. The first system regards the verify of the presence of units characterised by accounting errors, the second one is about the verification of the mean (or median) value on the observed non correct units r. In other words, the second hypothesis system is conditioned to the first one.

In fact, in order to verify conjointly the null hypothesis of the systems (1) and (2) or (3) which form the hypothesis H_0 of the system (4), we have to formalize the decisional rules belonging to each single hypothesis, so that to have a unique decisional rule D_0 to accept the null hypothesis obtained as the intersection of the two marginal decisional rules.

3.1. *The first test of hypothesis*

In the test of hypothesis (1) we verify the presence of units characterised by book errors.

Let consider a simple random sample of size n (with $n << N$), and let r be the number of units characterised by a book error in the sample. The number r is assumed distributed as a Poisson random variable with parameter $\lambda = np$ (considering acceptable the conditions of the asymptotic approximation of hypergeometric and binomial distribution to the Poisson distribution). The hypothesis system (1) becomes

$$\begin{cases} H_0' : p \le p_0 \\ H_1' : p > p_0 \end{cases} \rightarrow \begin{cases} H_0' : \lambda \le \lambda_0 \\ H_1' : \lambda > \lambda_0 \end{cases} \tag{5}$$

where $\lambda_0 = np_0$.

The test may be non-randomized or randomized.

NON-RANDOMIZED TEST. If we consider a non-randomized test, the critical function to reject the null hypothesis is

$$\Psi'(r;\lambda) = \begin{cases} 0 & \text{if} \quad r \le r_c \rightarrow \text{decision} \quad D_0' \\ 1 & \text{if} \quad r > r_c \rightarrow \text{decision} \quad D_1' \end{cases}$$

where the critical value r_c depends on the significance level of the first test α' and D_0' is the acceptance of H_0' while D_1' is the acceptance of H_1', respectively.

Let $g(r;\lambda)$ be the probability distribution function and $G(r;\lambda)$ the cumulative distribution function of a Poisson random variable, respectively. We have

$$G(r_c - 1;\lambda_0) < 1 - \alpha' \le G(r_c;\lambda_0)$$

so that:

$$G(r_c;\lambda_0) - g(r_c;\lambda_0) < 1 - \alpha' \le G(r_c;\lambda_0)$$

This equation allows, assigned α' and λ_0, to obtain the critical value r_c of the procedure.

In this case, the real significance level is equal to:

$$\alpha_*' = \left[1 - G(r_c;\lambda_0)\right] < \alpha'$$

The operative characteristic function of the *non-randomized* test $\beta'(\lambda) = 1 - \Pi'(\lambda)$ for

$\lambda > \lambda_0$ gives the probability of the second type error for the hypothesis system (1) $\Pi^{'}(\lambda)$ is the power function of the test). This function, for assigned values of n, $\alpha^{'}$ and p_0 (from which r_c derived), can be expressed as a function of $\lambda = np = n(p_0 + \Delta)$, where Δ is the distance from p to p_0.

So we have

$$\beta^{'}(\lambda) = \sum_{r=0}^{r_c} g(r;\lambda) = G(r_c;\lambda)$$

In particular, for $\lambda = \lambda_0$ and so $\Delta = 0$, $\beta^{'}(\lambda) \equiv \beta^{'}(\Delta, n)$, we have

$$\beta^{'}(\lambda_0) = (1-\alpha_*^{'}) \geq (1-\alpha^{'})$$

RANDOMIZED TEST. As the *non-randomized* test does not have an exact significant level equal to $\alpha^{'}$ we consider a *randomized* test. In this case the critical function to reject the null hypothesis is

$$\Psi^{'}(r;\lambda) = \begin{cases} 0 \ \textit{if} \ \ r < r_c \\ \psi \ \textit{if} \ \ r = r_c \\ 1 \ \textit{if} \ \ r > r_c \end{cases},$$

where

$$\psi = \frac{G(r_c;\lambda_0) - (1-\alpha^{'})}{g(r_c;\lambda_0)} = \frac{\alpha^{'} - \alpha_*^{'}}{g(r_c;\lambda_0)}$$

is the randomness probability of the decisional procedure for $r = r_c$.

We observe that for $\lambda = \lambda_0$ he expected value of the random variable $\Psi^{'}(r;\lambda)$ is $\alpha^{'}$:

$$\begin{aligned} E\{\Psi^{'}(r;\lambda_0)\} &= 0 \cdot P\{r < r_c\} + \psi \cdot P\{r = r_c\} + 1 \cdot P\{r > r_c\} = \\ &= \frac{G(r_c;\lambda_0) - (1-\alpha^{'})}{g(r_c;\lambda_0)} g(r_c;\lambda_0) + (1 - G(r_c;\lambda_0)) = . \\ &= \left[G(r_c;\lambda_0) - (1-\alpha^{'})\right] + \left[1 - G(r_c;\lambda_0)\right] = \alpha^{'} \end{aligned}$$

The function $E\{\Psi^{'}(r;\lambda)\}$, considered as function of λ, is the power function of the test. Consequently, the operative characteristic function of the *randomized* test is

$$\beta^{'}(\lambda) = \sum_{r=0}^{r_c} g(r;\lambda) - \psi \cdot g(r_c;\lambda).$$

In particular, for $\lambda = \lambda_0$ and so $\Delta = 0$, $\beta^{'}(\lambda) \equiv \beta^{'}(\Delta, n, p_0)$, we have

$$\beta^{'}(\lambda_0) = \left(1 - \alpha^{'}\right).$$

3.2. *The second test of hypothesis*

In the situation of non-systematic absence of abnormal behavior in the recording of the book values, the assumption of normal distribution of the accounting error is justifiable both from theoretical and applicative point of view. Moreover, because in this context it makes no sense to consider small sample sizes, considering the sample mean as test statistic the assumption of normality is approximately (for the central limit theorem) even if the assumption of normal distribution for X is removed. Consequently the *unilateral* and *bilateral* hypotheses systems of the second test of hypothesis regards the random variable $X \sim N(\mu, \sigma_0^2)$, with σ_0^2 assigned on the basis of the experience.

The verification of the mean value is conducted on the observed units r characterised by a book error in the sample. In other words, the second hypothesis system is conditioned to the first one. In fact it is taken into account only if it was observed in the first test $r \in [1, r_c]$.

The test may be *unilateral* or *bilateral*.

UNILATERAL TEST. If we consider the *unilateral* hypotheses systems with $x_i > 0;\ i = 1,\ 2, \ldots, r$, the decisional rule to accept the null hypothesis is

$$\begin{cases} D_0^{''} : \overline{x} \le x_c \\ D_1^{''} : \overline{x} > x_c \end{cases}$$

where $\overline{x} = \sum_{i=1}^{r} x_i / r$ is the sample mean and x_c is the critical value obtained from the significance level $\alpha^{''}$:

$$x_c = \mu_0 + \frac{\sigma_0}{\sqrt{r}} z_{1-\alpha^{''}}.$$

As known, $z_{1-\alpha^{''}}$ is the percentage point $\left(1 - \alpha^{'}\right)100$ of the standard normal distribution variable such that $\phi\left(z_{1-\alpha^{''}}\right) = 1 - \alpha^{''}$, where ϕ is the cumulative distribution function of the standard normal distribution.

Considering the standardized parameter $\delta = \left(\mu - \mu_0\right) / \sigma_0$ with $\delta \ge 0$ instead of $\mu > \mu_0$, the characteristic function of the test is

$$\beta''(\delta;r) = P\{D_0'': \mu = \mu_0 + \delta\sigma_0\} = 1 - \Pi''(\delta;r) = \phi\left(z_{1-\alpha''} - \sqrt{r}\delta\right). \tag{6}$$

For $\delta = 0$, such function takes the value $\beta''(0;r) = 1 - \alpha''$, and for $\delta \to \infty$, $lim_{\delta\to\infty}\beta''(\delta;\ r) = 0$.

The function $\phi(z)$ can be approximated to zero for $z \le -4$, so we can limit the determination of the function $\beta''(\delta;\ r)$ for values $\delta \in (0,\ (4 + z_{1-\alpha''})/\sqrt{r})$. For example, for $\alpha'' = 5\%$ and $r = 9$ we have $\delta \in (0, \approx 2)$.

BILATERAL TEST. If we consider the *bilateral* case, the hypotheses systems can be write as:

$$\begin{cases} H_0'': |\mu| \le \mu_0 \\ H_1'': |\mu| > \mu_0 \end{cases} \to \begin{cases} H_0'': -\mu_0 \le \mu \le \mu_0 \\ H_1'': (\mu < -\mu_0) \cup (\mu > \mu_0) \end{cases} \quad for\ \ \mu_0 \ge 0$$

The decisional rule based on the sample mean of the *r* observations $x_i \ne 0$; $i =\ 1,\ 2,\ \ldots, r$, is

$$\begin{cases} D_0'': -x_c \le \overline{x} \le x_c \\ D_1'': (\overline{x} < -x_c) \cup (\overline{x} > x_c) \end{cases} \quad for\ \ x_c > 0$$

To determine the critical value x_c, we consider $\mu = \mu_0$ and $\alpha''(< 0.5)$ such that $\alpha'' = \alpha_L'' + \alpha_U''$ where $P\{\overline{X} < -x_c\} = \alpha_L''$ and $P\{\overline{X} > x_c\} = \alpha_U''$ with $\alpha_L'', \alpha_U'' \ge 0$.

Being for hypothesis $\overline{X} \approx N\left(\mu_0, \sigma_0^2 / r\right)$, we have

$$\begin{cases} \alpha_L'' = 1 - \varphi\left(\dfrac{x_c + \mu_0}{\sigma_0/\sqrt{r}}\right) \\ \alpha_U'' = 1 - \varphi\left(\dfrac{x_c - \mu_0}{\sigma_0/\sqrt{r}}\right) \end{cases} \to \begin{cases} z_{1-\alpha_L''} = \dfrac{\sqrt{r}}{\sigma_0}(x_c + \mu_0) \\ z_{1-\alpha_U''} = \dfrac{\sqrt{r}}{\sigma_0}(x_c - \mu_0) \end{cases} \to \begin{cases} x_c = \dfrac{\sigma_0}{\sqrt{r}}\dfrac{\left(z_{1-\alpha_L''} + z_{1-\alpha_U''}\right)}{2} \\ \mu_0 = \dfrac{\sigma_0}{\sqrt{r}}\dfrac{\left(z_{1-\alpha_L''} - z_{1-\alpha_U''}\right)}{2} \end{cases} \tag{7}$$

The values α_L'' and α_U'' derive from the second equations of the system (7) using the relation:

$$z_{1-\alpha_L''} - z_{1-\alpha_U''} = 2\frac{\sqrt{r}}{\sigma_0}\mu_0$$

where $z_{1-\alpha_L''}$ and $z_{1-\alpha_U''}$ are the percentage points $\left(1-\alpha_L''\right)100$ and $\left(1-\alpha_U''\right)100$ of the standard normal distribution.

From the previous equations it is possible to determine the critical value x_c for $\mu_0 \ge 0$.

In particular, for $\mu_0 = 0$ we have $\alpha_L'' = \alpha_U'' = \alpha'' / 2$ and so:

$$x_c = \frac{\sigma_0}{\sqrt{r}} z_{1-\alpha''/2} .$$

The function $\phi(z)$ can be approximated to one for $z \geq 4$, so the relation $\phi\left(\frac{\sqrt{r}}{\sigma_0}(x_c + \mu_0)\right) = 1 - \alpha_L''$ for $\frac{\sqrt{r}}{\sigma_0}(x_c + \mu_0) \geq 4$ implies $1 - \alpha_L'' \cong 1$ and so $\alpha_L'' \cong 0$ and $\alpha_U'' \cong \alpha''$.

If $\mu_0 \geq \frac{\sigma_0}{\sqrt{r}}\left(\frac{4 - z_{1-\alpha''}}{2}\right)$, the approximate critical value is $x_c = \mu_0 + \frac{\sigma_0}{\sqrt{r}} z_{1-\alpha''}$. In such case x_c is the same of the *unilateral* case.

If $0 < \mu_0 < \frac{\sigma_0}{\sqrt{r}}\left(\frac{4 - z_{1-\alpha''}}{2}\right)$ we have to calculate α_L'', α_U'', and x_c using the equations (7).

Once defined x_c, we can calculate the operative characteristic function of the conditioned test:

$$\beta''(\delta; r) = P\left\{\left|\bar{X}\right| \leq x_c : \mu_0 + \delta\sigma_0\right\}$$

where the mean of the random variable X is indicated as $\mu = \mu_0 + \delta\sigma_0$ with $\delta \geq 0$.

As $E\left\{\bar{X}\right\} = \mu_0 + \delta\sigma_0$ and $Var\left\{\bar{X}\right\} = \sigma_0^2 / r$, we have

$$\beta''(\delta; r) = \phi\left(z_{1-\alpha_U''} - \sqrt{r}\delta\right) - \phi\left(z_{\alpha_L''} - \sqrt{r}\delta\right) = \phi\left(z_{1-\alpha_U''} - \sqrt{r}\delta\right) - \phi\left(z_{\alpha'' - \alpha_U''} - \sqrt{r}\delta\right) \quad (8)$$

For a fixed value α'' this is a monotonically increasing function with respect to δ and r.
In particular, for $\delta = 0$:

$$\beta''(0; r) = \phi\left(z_{1-\alpha_U''}\right) - \phi\left(z_{\alpha_L''}\right) = 1 - \alpha_U'' - \alpha_L'' = 1 - \alpha'' \to \Pi''(0; r) = \alpha''$$

As noted before, if $\mu_0 = 0$ we have $\alpha_L'' = \alpha_U'' = \alpha'' / 2$, so:

$$\beta''(\delta; r) = \phi\left(z_{1-\alpha''/2} - \sqrt{r}\delta\right) - \phi\left(z_{\alpha''/2} - \sqrt{r}\delta\right) \quad (9)$$

For $\mu_0 \geq \frac{\sigma_0}{\sqrt{r}}\left(\frac{4 - z_{1-\alpha''}}{2}\right)$ we have $\alpha_L'' \cong 0$ and $\alpha_U'' \cong \alpha''$, consequently

$$\beta''(\delta; r) = \phi\left(z_{1-\alpha''} - \sqrt{r}\delta\right) \quad (10)$$

and the approximation of the *unilateral* case (6) is still valid.

As indicated in the preceding paragraphs it is possible to use as location index the median (or another quantile point) as an alternative to the mean. The benefit would be that these indexes are robust. This property is especially useful when we suspect that some modalities very large or very small are abnormal.

In this way it would be possible to use a non-parametric version of the procedure that does not require the assumption of normality of the data.

3.3. *The joint operative characteristic function*

We now define the joint operative characteristic function $\beta(\Delta;\delta)$, where Δ is the distance from p to p_0 (first systems of hypotheses), and $\delta = (\mu - \mu_0)/\sigma_0$ (second systems of hypotheses), for the cases examined in sections 3.1 and 3.2.

NON-RANDOMIZED TEST. If we consider the *unilateral* and *non-randomized* test, the joint operative characteristic function is:

$$\beta(\Delta;\ \delta) = P(D_0 : \Delta, \delta) = g(r = 0;\ \lambda) + \sum_{r=1}^{r_c} g(r;\ \lambda)\beta''(\delta;\ r) \tag{11}$$

where $\beta''(\delta; r)$ for $\delta \geq 0$ is defined in equation (6).

We precise that if $r = 0$ we accept the null hypothesis of the joint test of hypotheses, while for $r > r_c$ we refuse it. For $1 < r < r_c$ the decision derives from the second test (see paragraph 3.2).

For the bi*lateral* and *non-randomized* test, we consider in the previous equation defined in equation (8).

We observe that for $\Delta = 0$ and $\delta = 0$ we have

$$\begin{aligned} \beta(0;0) &= g(r = 0; \lambda_0 = np_0) + \sum_{r=1}^{r_c} g(r; \lambda_0)\beta''(0; r) = \\ &= g(0; \lambda_0) + (1-\alpha'')\sum_{r=1}^{r_c} g(r; \lambda_0) \pm \alpha'' g(0; \lambda_0) = \\ &= (1-\alpha'')\sum_{r=0}^{r_c} g(r; \lambda_0) + \alpha'' g(0; \lambda_0) = \\ &= (1-\alpha'')(1-\alpha'_*) + \alpha'' g(0; \lambda_0) = 1-\alpha \end{aligned} \tag{12}$$

where α is the exact significance level of the joint test of hypothesis.

In equation (12) $\beta''(0;\ r) = 1-\alpha''$, $g(0; \lambda_0) = e^{-\lambda_0} = e^{-np_0}$ and $\sum_{r=0}^{r_c} g(r;\ \lambda_0) = 1-\alpha'_* \geq 1-\alpha'$, so

$$(1-\alpha'_*)(1-\alpha'') = 1-(\alpha'_* + \alpha'') + \alpha'_*\alpha'' > 1-\alpha \rightarrow \alpha \cong (\alpha'_* + \alpha'') - (\alpha'_* + g(0; \lambda_0))\alpha''$$

If we do not consider the term $(\alpha'_* + g(0;\ \lambda_0))\alpha''$, we obtain $\alpha = (\alpha'_* + \alpha'') \leq \alpha' + \alpha''$.

Operatively, we can proceed by assigning a small value to α (1%, 5%, 10%) and build the test assuming $\alpha' = \alpha'' = \alpha/2$.

RANDOMIZED TEST. If we consider the procedure of the first system of hypothesis using a *randomized* test, the joint operative characteristic function of the test $\beta(\Delta;\delta)$ can be written as

$$\begin{aligned}\beta(\Delta;\delta) &= g(0;\lambda)+\sum_{r=1}^{r_c-1} g(r;\lambda)\beta''(\delta;r)+(1-\psi)g(r_c;\lambda)\beta''(\delta;r_c)= \\ &= g(0;\lambda)+\sum_{r=1}^{r_c} g(r;\lambda)\beta''(\delta;r)-\psi g(r_c;\lambda)\beta''(\delta;r_c).\end{aligned} \tag{13}$$

We note that for $\Delta = 0$ and $\delta = 0$ we have

$$\begin{aligned}\beta(0;0) &= g(0;\lambda_0)+\sum_{r=1}^{r_c} g(r;\lambda_0)\beta''(0;r)-\psi g(r_c;\lambda_0)\beta''(0;r_c)= \\ &= g(0;\lambda_0)+\sum_{r=1}^{r_c} g(r;\lambda_0)(1-\alpha'')-\psi g(r_c;\lambda_0)(1-\alpha'')\pm\alpha'' g(0;\lambda_0)= \\ &= (1-\alpha'')\left[\sum_{r=0}^{r_c} g(r;\lambda_0)-\psi g(r_c;\lambda_0)\right]+\alpha'' g(0;\lambda_0)= \\ &= (1-\alpha')(1-\alpha'')+\alpha'' g(0;\lambda_0)=1-\alpha\end{aligned} \tag{14}$$

where $\beta''(0;\,r)=1-\alpha''$, $g(0;\lambda_0)=e^{-\lambda 0}=e^{-np0}$ and $\sum_{r=0}^{r_c} g(r;\lambda_0)-\psi g(r_c;\lambda_0)=1-\alpha'$.

We may observe that in equations (11), (12), (13) and (14) the terms $g(0;\lambda_0)$ or $g(0;\lambda)$ can be disregarded if λ_0 or λ are greater than 7, as their contribution would be lower than 0,001. In this way to consider the joint procedure is equivalent to consider the two joint tests stochastically independent.

4. SENSITIVITY ANALYSIS OF THE PROCEDURE

Defined the decision-making procedure that allows on the basis of sample data to accept or reject the null hypothesis using the critical values (r_c and x_c), consistent with the level of quality required for the correctness accounting and in accordance with the parameters (p_0 and μ_0) and significance levels of the two tests (α' and α''), it is interesting to evaluate the level of risk considered acceptable in a real situation incompatible with the null hypothesis in terms of probabilities. This probability is expressed by the operational characteristic function $\beta(\Delta;\delta)$ where δ and Δ are appropriate measures of the distances of the real values from the parameters p and μ from p_0 and μ_0. This function is called the "risk of error in the decision-making procedure sample" and is an indicator of general use for the choice of parameters that establish the decision-making procedure. It has to be noted that it also depends on the type of procedure of randomness (*random* or *non-random*) as regards the first test, and on the type of hypothesis (*unilateral* or *bilateral*) as regards the second test, conditional on the first.

To evaluate the sensitivity of the parameters on the operative characteristic function, and thus to direct operatively the conductors of the investigation of auditing in the choice of the parameters, the next Table 1 presents the critical values r_c, the real significance level of the first test $\alpha_*^{'}$ and the randomness probability ψ of the decisional procedure for $r = r_c$. In particular we have referred to: $n = 100,\ 250,\ 500$; $p_0 = 0.01,\ 0.05,\ 0.10$ and $\alpha^{'} = 1\%,\ 2.5\%,\ 5\%$.

TABLE 1

Critical values r_c, real significance level $\alpha_*^{'}$ and randomness probability ψ of the decisional procedure for $r=r_c$.

				$\alpha^{'} = 1\%$			$\alpha^{'} = 2.5\%$			$\alpha^{'} = 5\%$		
n	p_0	λ_0	$g(0;\lambda_0)$	r_c	$\alpha_*^{'}(\%)$	ψ	r_c	$\alpha_*^{'}(\%)$	ψ	r_c	$\alpha_*^{'}(\%)$	ψ
	0.01	1.0	0.3679	4	0.37	0.4136	3	1.9.0	0.0981	3	1.90	0.5058
100	0.05	5.0	0.0067	11	0.55	0.5517	10	1.37	0.6234	9	3.18	0.5011
	0.10	10.0	~0	18	0.72	0.3968	17	1.43	0.84	15	4.87	0.0363
	0.01	2.5	0.0821	7	0.42	0.5788	6	1.42	0.3885	5	4.20	0.1194
250	0.05	12.5	~0	21	0.94	0.0759	20	1.73	0.5789	19	3.06	0.9129
	0.10	25.0	~0	37	0.92	0.1478	35	2.25	0.2233	33	4.98	0.0101
	0.01	5.0	0.0067	11	0.55	0.5517	10	1.37	0.6234	9	3.18	0.5011
500	0.05	25.0	~0	37	0.92	0.1478	35	2.25	0.2233	33	4.98	0.0101
	0.10	50.0	~0	67	0.89	0.3128	64	2.36	0.1695	62	4.24	0.5725

In the Table 1 we observe that the values of the function $g(0;\lambda_0)$ show the probability that there are no accounting errors. Considering the critical values r_c we observe that in some situations such values are very small and, therefore, the means calculated at the second test result lacking in precision. In particular, already for the smallest value of $\alpha^{'} = (1\%)$ in some situations we observe values of r_c very small (i.e. 4, 7, 11) and decreasing when $\alpha^{'}$ increases. To have sense the use of the test on the mean you need to consider combinations of p_0 and n such that ensure a value of $\lambda_0 = np_0 \geq 10$.

Moreover the real significance level of the first test $\alpha_*^{'}$ is that of the case of *non-randomized* test: $\alpha_*^{'} < \alpha^{'}$, while for the randomized test is equal to the nominal level $\alpha^{'}$.

To complete the observations arising from Table 1, the Figure 1 presents the probability to accept the null hypothesis of the first test (operative characteristic function) as a function of Δ for the *randomized* and *non-randomized* test. In particular we have referred to: $p_0 = 0.05;\ n = 100,\ 250$ and $\alpha^{'}$ $\alpha^{'} = 2.5\%,\ 5\%$. We note that in the graph the thick lines are referred to the *randomized* test, while the fine lines are referred to the *non-randomized* test.

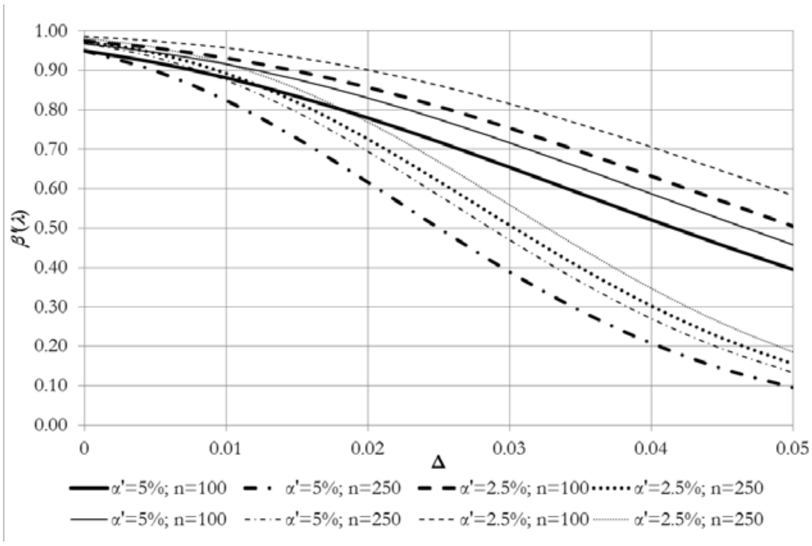

Figure 1 – Probability to accept the null hypothesis of the first test for the *randomized* (thick lines) and *non-randomized* (fine lines) test.

The figure shows that the risk function is very sensitive to sample size. In particular, it is much more sensitive the higher is the value of n. It is generally argued that the test presents a good discriminating ability.

Once defined the conditions of the first test about the frequency of accounting errors p, it is possible to carry out the second one relative to the mean of those errors. We emphasize that the second test is conditioned to the first one.

For this test we have chosen $\alpha'' = \alpha'$. Denoting by

$$u_0 = \frac{\mu_0}{\sigma_0/\sqrt{r}} \quad \text{and} \quad u_c = \frac{x_c}{\sigma_0/\sqrt{r}}$$

the standardized values of μ_0 and x_c respectively, Table 1 contains, for the *bilateral* test, the values u_0 and the differences $u_c - u_0$. We note that the values u_c and u_0 are standardized with respect to r and are obtained as function of α''_U by an iterative procedure for α'' equal to 1%, 2.5%, 5%. From those we can obtain $x_c = ku_c$ and $\mu_0 = ku_0$ with $k = \sigma_0/\sqrt{r}$.

TABLE 2
Standardized values u_o, differences u_c - u_o and probabilities $\alpha_U^{''}(\%)$.

$\alpha^{''} = 1\%$			$\alpha^{''} = 2.5\%$			$\alpha^{''} = 5\%$		
u_0	$\alpha_U^{''}(\%)$	$u_c - u_0$	u_0	$\alpha_U^{''}(\%)$	$u_c - u_0$	u_0	$\alpha_U^{''}(\%)$	$u_c - u_0$
0	0.5	2.5758	0	1.25	2.2414	0	2.5	1.96
0.0418	0.56	2.5362	0.051	1.414	2.1933	0.0589	2.842	1.9045
0.0837	0.619	2.5011	0.102	1.574	2.151	0.1177	3.174	1.8558
0.1255	0.675	2.4703	0.153	1.725	2.1143	0.1766	3.487	1.8136
0.1673	0.727	2.4437	0.204	1.863	2.0829	0.2354	3.773	1.7777
0.2092	0.774	2.421	0.255	1.986	2.0566	0.2942	4.026	1.7477
0.251	0.816	2.4018	0.306	2.094	2.0348	0.3532	4.244	1.7231
0.2928	0.852	2.3859	0.3569	2.184	2.0171	0.4113	4.423	1.7035
0.3346	0.883	2.3729	0.4078	2.259	2.003	0.4695	4.569	1.6881
0.3765	0.9	2.3623	0.4591	2.319	1.9919	0.5269	4.684	1.6763
0.4183	0.929	2.3539	0.51	2.366	1.9833	0.5845	4.772	1.6674
0.4602	0.946	2.3472	0.561	2.403	1.9769	0.6399	4.836	1.6609
0.5021	0.959	2.342	0.612	2.43	1.972	0.7063	4.892	1.6554
0.5439	0.969	2.338	0.6629	2.451	1.9685	0.7651	4.927	1.652
0.5857	0.977	2.335	0.7139	2.466	1.9659	0.8239	4.951	1.6496
0.6274	0.983	2.3327	0.7645	2.476	1.9641	0.8808	4.967	1.648
0.6694	0.988	2.3309	0.816	2.484	1.9627	0.9368	4.978	1.6469
0.7106	0.991	2.3297	0.867	2.489	1.9618	0.9984	4.987	1.6462
0.7529	0.994	2.3287	0.9179	2.493	1.9612	1.0595	4.992	1.6457
0.7943	0.996	2.328	0.9857	2.496	1.9607	1.1183	4.995	1.6454
0.8361	0.997	2.3275	1.0195	2.497	1.9605	1.1754	4.997	1.6452
>0.8368	1	2.3263	>1.0200	2.5	1.96	>1.1776	5	1.6449

In the Table 2 we observe that for $u_0 = 0$ (first line highlighted in the table), the values of the probability $\alpha_U^{''} = P\{\overline{x} > x_c\}$ are exactly the half of the values of $\alpha^{''}$. In this case the *bilateral* test is perfectly symmetric. Instead, for u_0 greater than a calculated value (last line highlighted in the table), we have $\alpha_U^{''} = \alpha^{''}$. In this situation it is possible to not consider the lower tail (and so $\alpha_L^{''}$), therefore the *bilateral* test tends to that *unilateral*. For example for $\alpha^{''} = 2.5$ and $u_0 > 1.0200$ we have $\alpha_U^{''} = \alpha^{''} = 2.5\%$.

To complete the analysis of the second test, the Figure 2 represents the operative characteristic function of the conditioned test $\beta^{''}(\delta; r)$ expressed in terms of $\sqrt{r}\delta$ for the *unilateral* test in equation (6) and for the two extreme situations for the *bilateral* test in equations (9) and (10). Here we have considered the values $\alpha^{''} = 1\%,\ 2.5\%,\ 5\%$. We note that in the graph the thick lines are referred to *unilateral* test, while the fine lines are referred to the *bilateral* test.

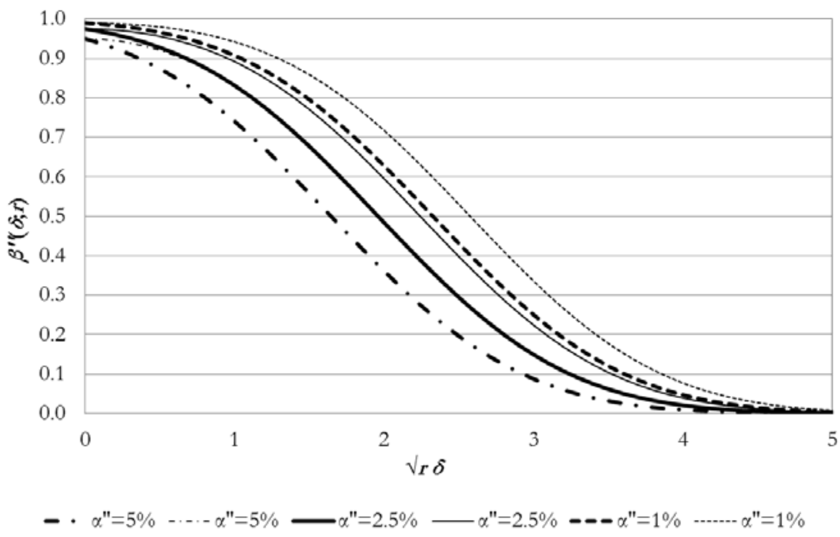

Figure 2 – Characteristic function of the conditioned test $\beta^{''}(\delta; r)$.

The figure shows the influence of the values $\alpha^{''}$ on the operative characteristic function. In particular the higher is the value of $\alpha^{''}$ than $\beta^{''}(\delta; r)$ is much more sensitive. Moreover for the *unilateral* test the curves drawn on the graph identify the single values of the characteristic function. Instead, for the *bilateral* test instead you get a range of values for this function making able, as shown in the table above, the *bilateral* test to be perfectly symmetric or to be reduced to the *unilateral* case.

For the joint test of hypothesis, Table 3 contains the global significance level α given by equations (11) and (13) for the *non-randomized* test and *randomized* test respectively with respect to the conditions of the null hypothesis of the first and the second test together. We remember that in this case, for $\Delta = 0$ and $\delta = 0$ we have $\alpha \leq \alpha^{'} + \alpha^{''}$ and the equality is valid only in the situation of stochastic independence of the two systems of hypotheses. In the table the probability α is here obtained as function of $n = 100,\ 250,\ 500$; $p_0 = 0.01,\ 0.05,\ 0.10$ and $\alpha^{'} = \alpha^{''} = 2.5\%,\ 5\%$.

TABLE 3
Global significance level α.

n	p_0	λ_0	$g(0;\lambda_0)$	$\alpha' = \alpha'' = 2.5\%$ *non-randomized* test	$\alpha' = \alpha'' = 2.5\%$ *randomized* test	$\alpha' = \alpha'' = 5\%$ *non-randomized* test	$\alpha' = \alpha'' = 5\%$ *random-ized* test
	0.01	1	0.3679	3.43	4.02	4.96	7.91
100	0.05	5	0.0067	3.82	4.92	7.99	9.72
	0.10	10	$\tilde{}$ 0	3.89	4.94	9.63	9.75
	0.01	2.5	0.0821	3.68	4.73	8.58	9.34
250	0.05	12.5	$\tilde{}$ 0	4.19	4.94	7.91	9.75
	0.10	25	$\tilde{}$ 0	4.69	4.94	9.73	9.75
	0.01	5	0.0067	3.82	4.92	7.99	9.72
500	0.05	25	$\tilde{}$ 0	4.69	4.94	9.73	9.75
	0.10	50	$\tilde{}$ 0	4.8	4.94	9.03	9.75
Global significance level in the situation of stochastic independence of the two systems of hypotheses				4.94		9.75	

From Table 3 we observe that smaller is the sample size n much more visible are the differences of the probability values α obtained with *randomized* and *non-randomized* procedures. The last line in the table contains the global significance level in the situation of stochastic independence of the two systems of hypotheses. This situation is verified only when the probability that there are no accounting errors is $g(0;\lambda_0) \tilde{} \ 0$. In these cases we note that the probability obtained with the randomized test is exactly that of the situation of independence. For the situations where $g(0;\lambda_0) > 0$ the differences are due to the fact that the second test is conditioned to the result of the first one.

To evaluate the sensitivity of the parameters Δ and δ on the operational characteristic function $\beta(\Delta;\delta)$, we present in the next figures the level curves (β probabilities) for $\beta = 20\%, 40\%, 60\%, 80\%$ for the *unilateral* test to vary of Δ and δ. We remember that as indicated in paragraph 3.2, the level curves for the *unilateral* test express also the extreme situation for the *bilateral* test. The considered cases are for $\alpha' = \alpha'' = 2.5\%$; $n = 100,\ 250$ and $p_0 = 0.01,\ 0.05,\ 0.10$.

In particular, for $n = 100$ we have the following level curves for the *non-randomized* and for the *randomized* test.

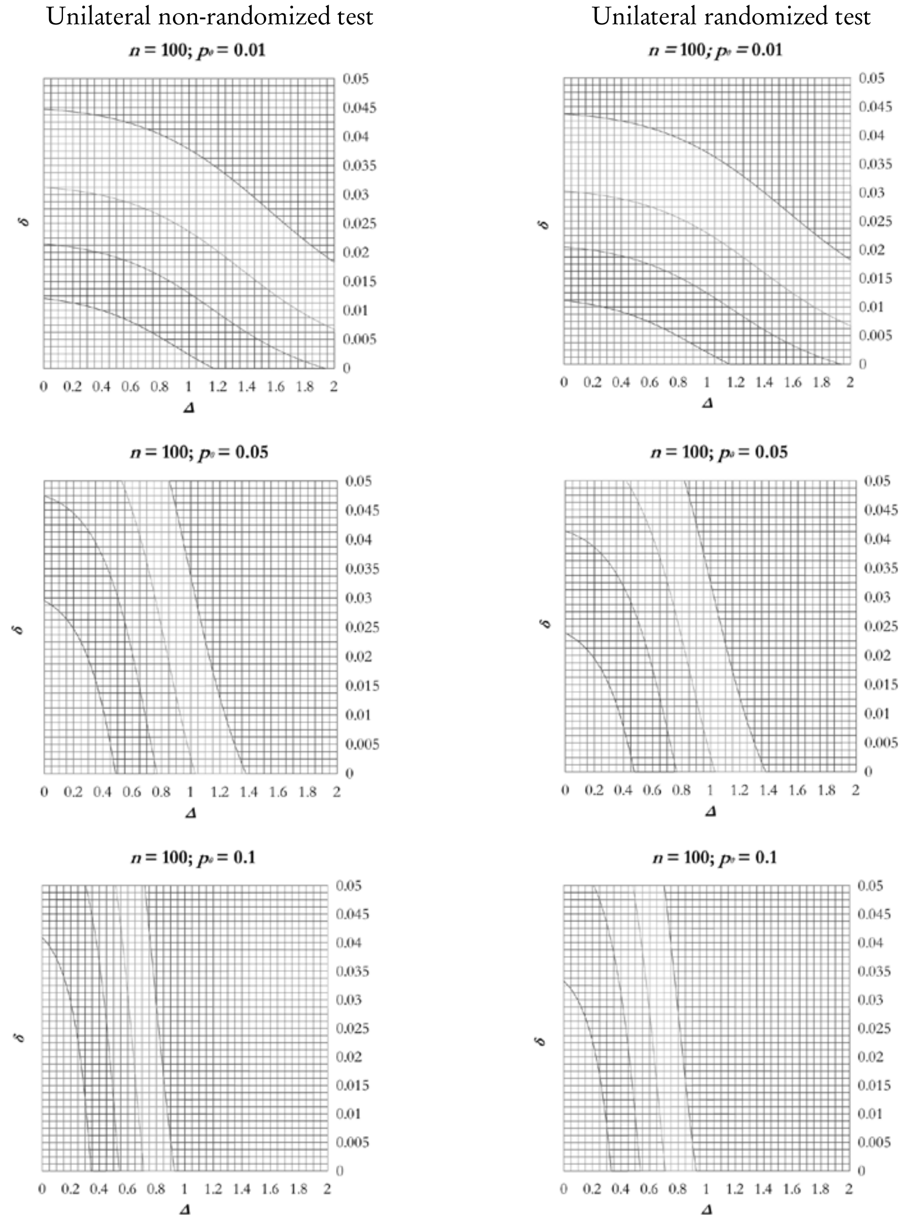

Figure 3 – Level curves for the *unilateral non-randomized* and *randomized* test for $\alpha^{'} = \alpha^{''} = 2.5\%$; $n = 100$ and $p_0 = 0.01,\ 0.05,\ 0.10$.

For $n = 250$ we have the following level curves for the *non-randomized* and for the *randomized* test.

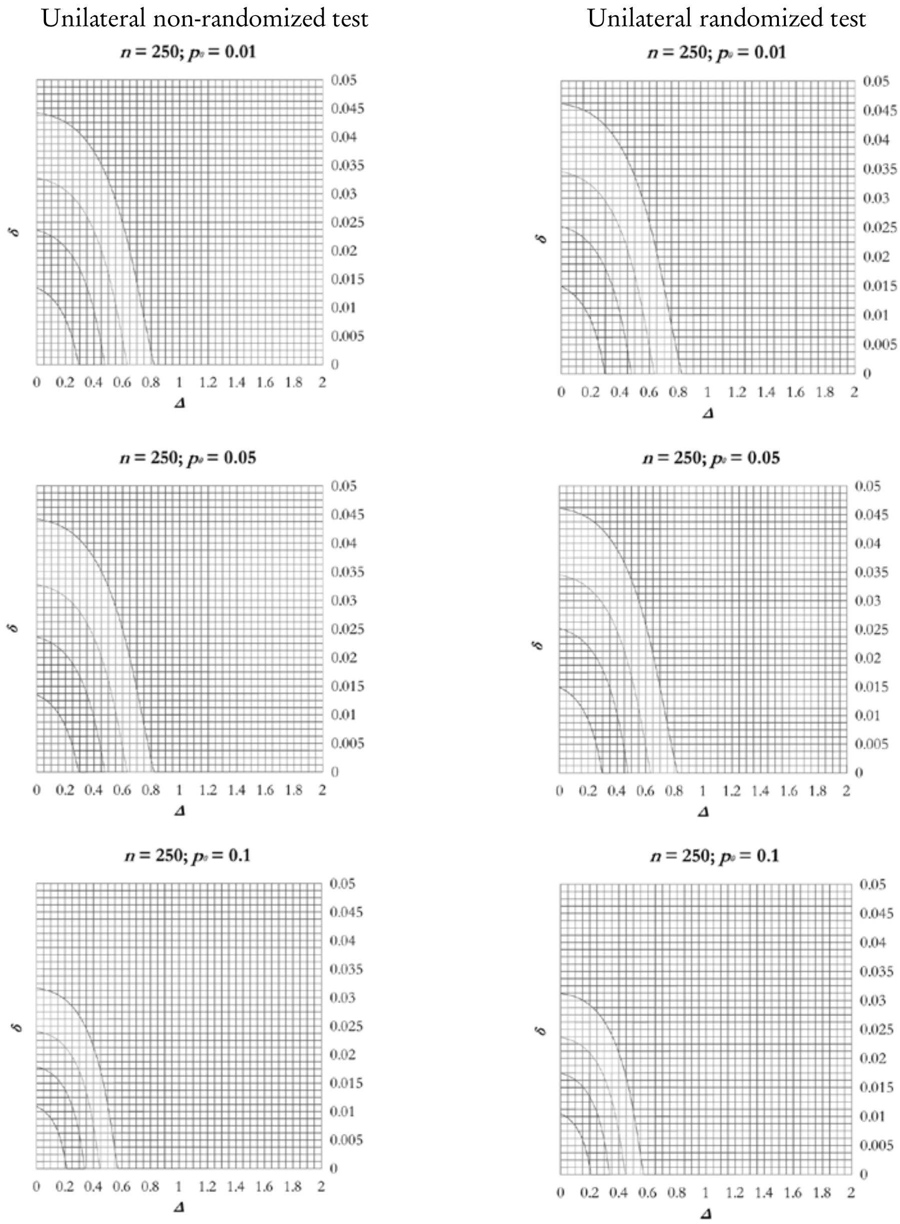

Figure 4 – Level curves for the *unilateral non-randomized* and *randomized* test for $\alpha^{'} = \alpha^{''} = 2.5\%$; $n = 250$ and $p_0 = 0.01,\ 0.05,\ 0.10$.

Observing the figures 3 and 4 we note that there is not a great difference in the level curves between *non-randomized* and *randomized* tests, for a fixed value of the sample size n. While in those figures we observe some differences of the level curves when p_0 varies, for a fixed n, or n varies, for a fixed p_0.

The following table shows the values of $\beta(\Delta;\delta)$ for a particular couple of Δ and δ. In particular here we have considered $\Delta = 0.01$ and $\delta = 0.5$ for $\alpha' = \alpha'' = 2.5\%,\ 5\%$.

TABLE 4
Global Percentage values of $\beta(\Delta;\ \delta)$ *for* $\Delta = 0.01$ *and* $\delta = 0.5$.

				$\alpha' = \alpha'' = 2.5\%$				$\alpha' = \alpha'' = 5\%$			
n	p_0	λ_0	$g(0;\lambda_0)$	*unilateral non-randomized* test	*unilateral randomized* test	*bilateral non-randomized* test	*bilateral randomized* test	*unilateral non-randomized* test	*unilateral randomized* test	*bilateral non-randomized* test	*bilateral randomized* test
	0	1	0.3679	78.44	76.91	81.25	79.63	73.65	66.51	78.11	70.25
100	0.1	5	0.0067	74.45	72.79	81.38	79.46	62.59	60.67	71.67	69.33
	0.1	10	$\tilde{}$ 0	60.89	59.97	70.34	69.2	47.14	47.07	58	57.9
	0	2.5	0.0821	63.32	58.95	67.87	63.07	46.84	45.37	51.95	50.27
250	0.1	12.5	$\tilde{}$ 0	47.08	47.77	58.51	57.29	36.38	34.87	47.07	44.97
	0.1	25	$\tilde{}$ 0	25.53	25.43	34.67	34.53	16.36	16.36	24.54	24.53
	0	5	0.0067	41.54	36.49	46.58	40.76	28.33	24.83	33.41	29.17
500	0.1	25	$\tilde{}$ 0	20.84	20.68	28.81	28.57	12.58	12.58	19.22	19.21
	0.1	50	$\tilde{}$ 0	4.42	4.41	7.56	7.54	2.2	2.18	4.31	4.26

The table confirms as observed in the previous figures that there are modest differences between *randomized* and *non-randomized* test. These differences are greater for $p_0 = 0.01$. The differences between unilateral and bilateral tests are quite reduced, too. The analysis confirms that the influence on the level curves is given mainly by n .

5. FUTURE DEVELOPEMENTS

It is of fundamental importance to adapt the decision making of corporate management to changing conditions due to vary the change of the behaviour of customers, suppliers and the production system itself, for the employment of innovative technological tools and communication instruments, that compel to follow the multiple transactions that the company carries out. Therefore the statistical tools are increasingly made suitable for a dynamic vision.

The proposed procedure provides a tool of quantitative measure in terms of probability error that detects the anomalous circumstances in the accountancy book. The choice of values and parameters has an impact on the economic consequences and on the efficiency of the auditing, but also provides to the statistical analyst an easy to implement and to use procedure. In particular, it is of great importance the choice of the starting parameters p_0, μ_0, n and of the probabilities α' and α'' to evaluate the parameters influence on the selective ability β of the test. Therefore it should be interesting to conduct in the future a more analitical study concerning the selection of these parameters.

Moreover in this analysis we have made some assumptions regarding the distribution laws of some random variables. In particular in the first hypothesis system we have supposed that the number r of units characterised by a book error in the sample is distributed as a Poisson random variable (considered acceptable the conditions of the asymptotic approximation of hypergeometric and binomial distribution to the Poisson distribution). While in the second hypothesis system we have assumed that the sample mean follows a normal distribution. Given the validity of these assumptions in standard situations, it should be interesting carry out further analysis to assess the changes in the sensitivity of the procedure. In particular, it should be interesting to consider some distributions for the error random variable X different from normal distribution, for example lognormal, chi-square and Weibull distributions.

In addiction as location index it is possible to use the median (or another quantile point) as an alternative to the mean so it should be possible to use a non-parametric version of the procedure.

Finally, a further consideration reguards the type of sampling carried out. In the present work we have considered a simple random sample, but it should be interesting to see how the procedure changes by performing a sampling whithout replacement.

ACKNOWLEDGEMENTS

The authors gratefully acknowledge Professor U. Magagnoli for his suggestions and supervision on the paper.

REFERENCES

A.A. ARENS, J.K. LOEBBECKE (2002). *Applications of Statistical Sampling in Auditing*. Englewood Cliffs, Prentice Hall.

A.H. ASHTON, R.H. ASHTON (1988). *Sequential Belief Revision in Auditing.* The Accounting Review, Vol. LXIII, 4, 623-641.

AU Section 350. Audit Sampling. AICPA.
http://www.aicpa.org/Research/Standards/AuditAttest/DownloadableDocuments/AU-00350.pdf.
Retrieved 1 December 2012.

R.J. BOLTON, D.J. HAND (2002). *Statistical Fraud Detection: A Review.* Statistical Science, Vol. 17, 3, 235-255.

L.D. BROWN, M.S. ROZEFF (1979). *Univariate Time-Series Models of Quarterly Accounting Earnings per Share: A Proposed Model.* Journal of Accounting Research, Vol. 1, 17, 179-189.

G. CAPRARA (1988). *Stato e prospettive dell'impiego della statistica e dell'informatica nella moderna gestione aziendale: il problema della formazione professionale.* Statistica, Vol. XLVIII, 649-665.

P.M. CHIODINI, U. MAGAGNOLI (2004). *L'impiego dei metodi statistici nel controllo dinamico della gestione aziendale.* Relation presented to the Conference MTISD "Metodi, Modelli e Tecnologie dell'Informazione a Supporto delle Decisioni", Università del Sannio, Benevento, 24-26 June 2004, 1-4.

R. CLEARY, J.C. THIBODEAU (2005). *Applying Digital Analysis Using Benford's law to detect Fraud: The Dangers of Type I Errors.* Auditing: A Journal of Practice & Theory, Vol. 24, May, 77-81.

D.M. GUY, D.R. CARMICHAEL, R. WHITTINGTON, (2002). *Audit Sampling: an introduction*. John Wiley & Sons, New York.

R.W. HOUSTON, M.F. PETERS, J.H. PRATT (1999). *The Audit Risk Model, Business Risk and Audit-Planning Decisions*. The Accounting Review, Vol. 74, 3, 281-298.

G.K. KANJI (2004). *Total Quality Management and Statistical Understanding.* Total Quality Management, 5, 105-114.

J. KRIENS, R.H. VEENSTRA (1985). *Statistical Sampling in Internal Control by Using the AOQL-System.* The Statistician, 34, 383-390.

R. LIBBY, J.T. ARTMAN, J.J. WILLINGHAM (1985). *Process Susceptibility, Control Risk, and Audit-Planning*. The Accounting Review, Vol. 60, 4, 212-230.

K.-Y. LOW (2004). *The effects of Industry Specialization on Audit Risk Assessment and Audit-Planning Decisions*. The Accounting Review, Vol. 79, 1, 201-219.

D.C. MONTGOMERY (2005). *Introduction to Statistical Quality Control, 5th Edition.* John Wiley & Sons, New York.

T.M.F. SMITH (1976). *Statistical Sampling for Accountants.* Haymarket, London.

A.D. TEITLEBAUM, C.F. ROBINSON (1975). *The Real Risks in Audit Sampling.* Journal of Accounting Research, 13, Studies on Statistical Methodology in Auditing, 70-91.

SUMMARY

On a Test of Hypothesis to Verify the Operating Risk Due to Accountancy Errors

According to the Statement on Auditing Standards (SAS) No. 39 (AU 350.01), audit sampling is defined as “the application of an audit procedure to less than 100 % of the items within an account balance or class of transactions for the purpose of evaluating some characteristic of the balance or class”. The audit system develops in different steps: some are not susceptible to sampling procedures, while others may be held using sampling techniques. The auditor may also be interested in two types of accounting error: the number of incorrect records in the sample that overcome a given threshold (natural error rate), which may be indicative of possible fraud, and the mean amount of monetary errors found in incorrect records. The aim of this study is to monitor jointly both types of errors through an appropriate system of hypotheses, with particular attention to the second type error that indicates the risk of non-reporting errors overcoming the upper precision limits.

Keywords: auditing; error risk; non-randomized test; operative curve.