AUDIT RISK STUDIES: SAMPLING DESIGN AND BAYESIAN BASED MODELS

Carlos N. Bouza[1], Universidad de La Habana

ABSTRACT
Commonly audit risks are studied by using a sample of items from an account. The estimate of the Total Auditing Error is the objective of the inquiry. An Eclectic Bayesian approach is used as an alternative to popular methods as Dollar Unit Sampling. Jacknife is used for computing the standard deviation of one of the predictors. The procedures are evaluated through Monte Carlo experiments.

Key words: Total Auditing Error, superpopulation, Jacknife, Monte Carlo experiments.

RESUMEN
Comúnmente los riesgos de auditoría son estudiados utilizando una muestra de items de una cuenta. Estimar el Error Total de Auditoría es el objetivo de la encuesta. Un enfoque Bayesiano Ecléctico es utilizado como alternativa respecto a métodos populares como el del Muestreo de la Unidad Dólar. Jacknife es utilizado para computar la desviación típica de uno de los predictores. El procedimiento es evaluado a través de experimentos de Monte Carlo.

Palabras clave: Error Total de Auditoría, métodos populares, Jacknife, esperimentos de Monte Carlo.

MSC: 62D05

1. INTRODUCTION

An important problem in auditing is to estimate the total amount of error. The error is measured in terms of the so-called `tainted dollar units'

$$t_i = 100 \text{ Error of item } i/\text{Book Value}$$

It is commonly expected that the overestimation or underestimation must not exceed the 100 %. Hence, after rounding we have $I = \{T_{-100}, T_{-99},....,T_0,....,T_{100}\}$ as the set of possible values of the tainting. Take $|T_j| = N_j$ as the number of items in the account with tainting equal to `j'. The parameter of interest for the auditor is

$$\theta = [RBV/100]\sum_{i\in I} iP_i = \lambda\sum_{i\in I} P_i$$

It is the Total Auditing Error [TAE]. $P_j=N_j/N$ is the unknown proportion of individiual tainting j = -100,....,100. $N = N_{-100}+....,N_{100}$, hence $P_{-100} +....+ P_{100} = 1$. RBV is the reported Book Value which is known. If $t_i \in]j-1, j]$ we assign an observation to T_j. The auditor estimates θ and usually he/she is interested in establishing a confidence region or in testing a certain hypothesis as $H_0:\theta \geq \theta_0$. θ_0 is a critical value of the TAE that fixes a critical state of the account. A normal approximation is used for inferential purposes.

The auditor selects a sample of items and the use of the likelihood $L = n!\Pi_{i\in I} P_i^{\ ni} /\Pi_{i\in I} n_i!$ seems to be a good approach for estimating the P_i's. It is adequate only is simple random sampling with replacement is used for selecting a sample s of size n from the population of items in the account. A very popular methods is to use the so called `Dollar Unit Sampling'[DSU] . It assigns a larger probability of inclusion to the items with a larger recorded value. These values are known and a certain value in dollars is attached to each of them. A detailed description of these methods can be obtained in different books and papers. See for example Kraft [1986]. The corresponding theory is briefly presented in Section 2. DSU is analyzed within the frame of Unequal Probability Sampling. Bayesian principles allow to estimate TAE, but the elicitation of an adequate prior poses an important difficulty, see Crosby [1980], [1981] and Solomon [1982]. A solution is to use Quasi-Bayes audit principles, introduced by Mac Cray [1984]. It depends on one parameter only. Hernández-Vázquez [1997] obtained a theoretical justification of the use of Maximum Likelihood principle. Section 3 presents some necessary results. The objective of the auditor is supposed to be the determination of an upper and a lower bound of θ. The use of confidence intervals seem to be a solution. Using some additional

[1]**Email**: bouza@matcom.uh.cu

modeling the bounds can be derived from a likelihood, see Vázquez-Polo and Hernández-Bastida [1995], Hernández-Bastida and Vázquez-Polo [1997]. We investigate the use of the classic normal distribution of the approximations. The Eclectic Bayesian [EB] approach of modeling through a Superpopulation Approach is developed. Two models are proposed and three predictors are analyzed. Jacknife is used for computing the standard deviation of one of the predictors.

Section 4 is devoted to the comparison of the classic standard deviation estimations and the Superpopulation predictors through Monte Carlo experiments. The use of Jacknife has the best performance. As a result we recommend to elicit a superpopulation model and to apply the Jacknife methodology for ensuring the normality and to use it in the inferences.

2. SAMPLING DESIGNS

A sample of n items is selected from the set U, of N items that conforms an account, for estimating θ. The use of a frequentist model establishes that the 201 parameters P_i, $i \in I$ must be estimated by computing adequate estimates.

A popular sampling design is to select the items proportionally to the number of dollar units assigned to each of them. Taking

$$X_j = \text{Record Value in dollar units of item j of the account} = RVD[j]$$

the use of an unequal probability sampling design with inclusion probabilities set

$$\pi^* = \{\pi_j \mid j \in U,\ \pi_j = nX_j / \textstyle\sum_{i \in U} X_i\},$$

where

$$\text{Record Book Value} = RBV = \textstyle\sum_{i \in U} X_i,$$

implements DSU. Different schemes may be used for deriving the particular selection procedures. Chaudhuri-Voos [1988] analyzed and discussed almost completely the behavior of the existing sampling designs. Therefore DSU is implemented by selecting n units from {1,.2,...,RBV}. Item j is observed whenever at least one of the n randomly selected numbers belongs to $[\sum_{i<i} X_j, \sum_{i \leq j} X_i]$ taking $\sum_{i<i} X_j = 0$ if $j = 1$.

The evaluation of a selected item generates a vector $\underline{\mathbf{Y}}_j = [Y_{-100[j]}, .., Y_{100[j]}]$ such that

$$Y_{j[i]} = 1 \text{ if } j \in T_i \ [= 0 \text{ otherwise}].$$

Hence

$$P_i = \textstyle\sum_{j \in I} Y_{j[i]} / N = \text{Number of counts of } T_i/N = N_i/N$$

Using the information provided by the sample s we can compute

$$n_i = \textstyle\sum_{j \in s} Y_{j[i]}$$

The assignment is made by evaluating if the error of the tainted dollar-unit has error $i \cong [100/RBV]$ $[X_j$ - Audited value of item j] = $[100/RBV][X_j - Z_j]$.

The use of the likelihood function

$$L[\underline{\mathbf{P}}^{\mathbf{t}} = (P_{-100}, ..., P_{100}) \mid \underline{\mathbf{n}}^{t} = (n_{-100}, ..., n_{100})] = n!\Pi_{i \in I} P^{n_i}/n_i!$$

is valid when the n_i's are independent and simple random sampling with replacement is used for selecting the book. In general the use of a complex sampling design is incompatible with inferences based on this likelihood because of its flatness, see Chaudhuri-Voos [1988].

In any case, as θ is a linear function, an unbiased estimator is derived when unbiased estimators p_i's are obtained and

$$\theta_d = \lambda \textstyle\sum_{i \in I} i p_i$$

is computed for sample selected by means of the sampling design d.

When d is simple random sampling with replacement we have that

$$p_i = n_i/n = \sum_{j\in s} Y_{j[i]}/n$$

is unbiased for P_i and

$$\theta_{srs} = \lambda\sum_{i\in I} i\sum_{j\in s} Y_{j[i]}/n$$

It is easily derived using that property that

$$Var[\theta_{srs}] = V[\theta_{srs}] = \lambda^2\sum_{i\in I} i^2 P_i[1 - P_i]/n$$

is its sampling error.

Note that if DSU is the design we can assume that

- Prob[j,j'$\in$s] = $\pi_{jj'} > 0$ is compatible
- $\pi_{jj'} \leq n\pi_j\, \pi_{j'}$

Hence, if the auditor fixes the sample size DSU is a πPX sampling design. The Horvitz Thompson estimator is highly recommended for such designs, see Ardilly [1992] for details. Taking

$$p_{i[HT]} = \sum_{j\in s} Y_{j[i]}/N\pi_j = N_{i[HT]}/N$$

is unbiased. As a result

$$E[\theta_{HT}] = E[\lambda\sum_{i\in I} i p_{i[HT]}] = \theta$$

and under the hypothesis of independence is easily derived that

$$V[\theta_{HT}] = [\lambda^2/N^2][\sum_{i\in I} i^2[\sum_{j\in I} Y_{j[i]}{}^2(1 - \pi_j)\pi_j^{-1} + \sum_{j\neq j'\in U} Y_{j[i]} Y_{j'[i]}(\pi_{jj'} - \pi_j\pi_{j'})(\pi_j\pi_{j'})^{-1}]$$

is the sampling error of the estimator.

These results permit to establish the estimators. They do not incorpore the experience of the auditor except in the selection of the d. We will analyze Bayesian based alternatives which permit to use the information provided by the auditor on the prior distribution of the parameters and/or varibles involved.

3. BAYESIAN BASED APPROACHES

The auditories are repeated over time frequently. Hence the auditor can provide a prior distribution and classic Bayesian procedures can be used for estimating the TAE. In practice to solve the involved problem may be complicated because of:

- The difficulty of dealing with statistically-minded well-informed auditors.
- The techniques to be developed require of the use of sophisticated computing tools.

Following Hernández-Bastida and Vázquez-Polo [1997] a solution is to look not for priors of the 201 involved parameters but to work with a prior for θ. Supposing that θ $\Theta = \{\theta_1,...., \theta_H\}$ the auditor can elicit the mass function

$$\xi(\theta_t) = \mathrm{Prob}\{\theta_t = \theta\}$$

Defining

$$\varphi^{-1}(\theta) = \{\underline{\mathbf{P}}\in[0,1]^{201} \mid \sum_{j\in I} P_j = 1\}$$

and

$$L^*[\underline{\mathbf{P}}] = \sum_{\theta_0 \in \Theta} \mathrm{Sup}_{\varphi} - 1_{(\theta)}\, L[\mathrm{P}]\, \mathrm{I}_{\varphi} - 1_{(\theta)}\, [\underline{\mathbf{P}}]$$

with

$$\mathrm{I}_{\varphi}\text{-}1_{(\theta)}[\underline{\mathbf{P}}] = 1 \text{ if } \underline{\mathbf{P}} \in \varphi^{-1}(\theta) \ [= 0 \text{ otherwise}].$$

The Most Likely Posterior Curve is

$$\xi(\theta_0 \mid \underline{\mathbf{X}}) = L^*[\underline{\mathbf{P}}]\xi(\theta_0)/\sum_{\theta\in\Theta} L^*[\underline{\mathbf{P}}]\xi(\theta)$$

Its use see Hernández-Bastida, Martel-Escobar , and Vázquez-Polo [1998], has the following advantages:

- The posterior distribution does not need of the existence of a previous model of the tainting which permits to fix the likelihood.

- The auditor works only with a model on θ.

As quoted by Hernández-Bastida, Martel-Escobar, and Vázquez-Polo [1998] it is necessary to "elicit a complete prior distribution and it can be a very difficult task for auditors".

Each θ_t can be the result of obtaining a vector from the set $P_t = \{\underline{\mathbf{P}}^* \in [0,1]^{201} \mid \lambda\sum_{i\in I} i P_i^* = \theta_t\}$. The corresponding Maximum Likelihood estimate of θ is the solution of the optimization problem

$$\alpha(\theta_t) = \mathrm{Max}_{\{\underline{\mathbf{P}}^* \in P_t\}}\{\boldsymbol{L}(\mathrm{P}_{-100},\ldots,\mathrm{P}_{100} \mid n_{-100},\ldots,n_{100})\}$$

Therefore we have transformed the classic Likelihood looking for the compatibility of the prior distribution suggested by the auditor. The Quasi-Bayesian likelihood requires again of the determination of L_t and $\underline{\mathbf{P}}^*$.

The presented approaches are based on the knowledge of a convenient prior distribution or their properties hold only if the sampling design is simple random sampling with replacement. Assuming that we can relate the value of each Y_i with a known variable X_i, $i \in I$, we can denote it by

$Q[Y_{-100},\ldots,Y_{100} \mid X_{-100},\ldots,X_{100}) = Q(\underline{\mathbf{Y}} \mid \underline{\mathbf{X}})$. For example $\underline{\mathbf{X}}$ may be the vector of the taintings at the previous

analysis of the account. If μ is a Lebesgue measure and d[s] is a sampling design we can use the posterior derived by Scott [1977]

$$Q_d\, [Y_{-100},\ldots,Y_{100} \mid X_{-100},\ldots,X_{100}) = d[s \mid \underline{\mathbf{X}}]\, Q(\underline{\mathbf{Y}} \mid \underline{\mathbf{X}}) / \int d[s \mid \underline{\mathbf{X}}]\, Q(\underline{\mathbf{Y}} \mid \underline{\mathbf{X}})\, d\mu(\underline{\mathbf{Y}} \mid \underline{\mathbf{X}}).$$

which is also design independent because $Q_d(\underline{\mathbf{Y}} \mid \underline{\mathbf{X}})$ is proportional to $Q(\underline{\mathbf{Y}} \mid \underline{\mathbf{X}})$. This fact is very attractive for the auditors because it means that the inferences should be based only on his/her knowledge obtained from the observed auxiliary variable $\underline{\mathbf{X}}$ and expressed by $Q(\underline{\mathbf{Y}} \mid \underline{\mathbf{X}})$. This approach suggests that an auditor can use a known variable for describing a relationship with the unknown taintings. Assuming that a process at the item level generates the tainting we can use a superpopulation approach as an Eclectic Bayesian procedure.

As each item produces a tainting that is the result of sampling an infinite population the

Location-superpopulation model m

$$Y_{j[i]} = P_i \varepsilon_j \qquad [3.1]$$

is a possible model. We will analyze its meaning in the contex of audit risks studies.

[3.1] describes the behavior of the item j with respect to the i-th tainting. The auditor assumes that ε_j is an unobservable random error with zero model-mean $[E_m(\varepsilon_j) = 0]$ and $V_m[\varepsilon_j] = \sigma_i^2$. Hence $Em[Y_j^i \mid j \in T_i] = P_i$ and $Vm\,[Y_j^i \mid j \in T_i] = \sigma_i^2$. The number of items in the i-th tainting in a sample s is

$$n_i = \sum_{j\in s} Y_{j[i]}$$

It is a Binomial random variable with $E_m[n_i] = nP_i$ and $V_m[n_i] = n\sigma_i^2$. Therefore $p_i = n_i/n$ is also an m-unbiased predictor. Therefore

$$\theta_m = \lambda\sum_{i\in I} ip_i,\ \lambda = RBV/100$$

is also m-unbiased for θ and its error

$$V_m[\theta_m] = \lambda^2\sum_{i\in I} i^2\sigma_i^2/n$$

is readibly derived. It is estimated unbiasedly by

$$v_m[\theta_m] = \lambda^2\sum_{i\in I} i^2 p_i[1 - p_i]/n$$

The traditional inferences, tests of hypothesis, interval estimation, etc , can be made using the corresponding well known Limit Theorems which relate the Binomial and the Normal distributions.

Another modeling approach is to infer conditioning on the sample. Consider that we observe a set of taintings. A tainting is observed if $s \cap T_i = s_i \neq \phi$. Then we may compute the predictor

$$\theta_m^* = \lambda\sum_{\{i \mid |si| > 0\}} ip_i = \sum_{\{i \mid |si| > 0\}} in_i/n$$

For convenience we will define the weights

$$W_i = i/\sum_{\{i \mid |si| > 0\}} i = i/W$$

and rewrite

$$\theta^*_m = W\lambda\sum_{i\in I} W_i p_i = W\,\theta_m^{*'}.$$

It predicts which is the sample value of the TAE under the conditions described by the superpopulation model. As they are quite general the value of θ_m^* permits to infer on the results of the auditories developed under the same conditions for the account. This is the objective in many cases: to determine what is expected to happen in the auditories if changes are not introduced.

Following Pothoff **et al**. [1992] and Bouza [1995] we will reanalyze the inferential procedures. Let us take

$$n^* = W^2/\sum_{i\in I} W_i^2$$

which is called “equivalent sample size”. Using the transformed weights, $q_i = n^*W_i/W$ we have that

$$n^* = \sum_{i\in I} W_i = \sum_{i\in I} W_i^2\ .$$

Then is easily derived that

$$V_m[\theta_m^*] = \lambda^2 \sum_{i\in I} q_i^2\sigma_i^2/nn^*$$

because

$$\theta_m^* = \lambda\sum_{i\in I} q_i p_i/n^*.$$

An estimator of $V_m[\theta_m^*]$ is

$$v_m[\theta_m^*] = \lambda^2\sum_{i\in I} q_i^2 [y_{j[i]} - p_i]^2/n[n^* - 1]$$

As

$$E_m(v_m[\theta^*_m]) = V_m[\theta^*_m] + \lambda^2(\sum_{i\in I}[q_i - q_i^2]\sigma_i^2 + \sum_{i\in I}q_i(p_i - \theta]^2)/n[n^*-1]$$

Then $v_m[\theta^*_m]$ overestimates the error even if the error of the taintings is constant [$\sigma_i = \sigma$] which implies that the second term at the right hand side is zero. It is almost incredible that $p_i = \theta$ for any $i \in I$. Hence a positive bias is generally present.

Note that the defined weights are random but they may be assumed to be directly proportional to the number of items in the tainting divided by the corresponding DSU's. Therefore this is an unimportant problem, see Pothoff **et al**. [1992].

Then the auditor may use for inferences the statistic $T = [\theta^*_m - \theta_0] / v_m[\theta^*_m]^{½}$. It follows, approximately, a T-Student with

$$f = 2[n^* - 1]^2 \left(V([n^* - 1]v_m[\theta^*_m] / v_m[[\theta^*_m])\right]$$

degrees of freedom. Note that generally in auditing, f is sufficiently large for accepting the normal approximation of the T-Student.

Note that each T_i can be considered as a stratum. We do not know which items belong to each of them. As $N_i = |T_i|$ is unknown the selection of the sample s and the classification of the items can be modeled by using poststratification. This procedure is of common use for dealing with different problems as non-responses and in small area estimation. The corresponding results using the procedures discussed above.

Another superpopulation model is given by using m* where Z_j, the audited value of item j, is described by

$$Z_i = \sum_{0\le k\le K} \beta_k X_{ki} + \varepsilon_j$$

This is a Regression-superpopulation model where the $\beta_k^{'}$s are unknown parameters and the $X_{ki}^{'}$s are known values of a variable X_i related with Y. For example it can be the reported book value in the i-th previous month. m* is similar to m , $E[\varepsilon_j] = 0$, $V[\varepsilon_j] = \sigma_j^2$, etc., but it permits to model the behavior of the unobserved items.

Taking the population matrix $X = [X_{ki}]_{N\times K}$, the vector $\beta_{K\times 1}$ and 1^M a vector of M ones

$$Z_t = 1^M \times \beta$$

We will denote by X_s the X-matrix of the sample s and by $\mathbf{y}_s$ the vector of the observed values of Y. Using Least Squares Estimation [LSE] we have that $B = [X_s^T X_s]^{-1} X_s^T y_s$ estimates β. Hence we can predict each Z_j by using

$$Z_j = \sum_{0\le k\le K} B_k X_{ki}$$

and to compute

$$Y^*_{j[i]} = 1 \text{ if } [X_j - z_j]/RBV = i \text{ [= 0 otherwise].}$$

From the observed items we can compute

$$n_i = \sum_{j\in s} Y_{j[i]}$$

Denote by $M_i = N_i - n_i$ the number of unobserved items in the i-th tainting. Its the prediction is

$$m_i = \sum_{j\notin s} Y^*_{j[i]}$$

can be obtained. Our prediction of P_i under this model is

$$p_i^* = N_i^* / N \; p^*_i = [n_i + m_i]/N$$

It is model unbiased because $E_{m*}[Z_j] = Y_{j[i]}$. As a result we recommend the predictor

$$\theta^*_{m*} = \lambda \Sigma_{i \in I} i p_i^*$$

The structure of it determines that its error depends on the model through the prediction of the Y's for the unsampled items. To obtain an analytical expression of it is very difficult because the model and the use of LSE are sources of error correlated in with respect to the sample and the model. A solution to estimate the variance of this predictor is to use an intensive computation method. We decided to use Jacknife because it is less costly in our case than Bootstrap.

Taking $\theta^*_{m*[j]}$ as the predictor computed deleting the sampled j-th item the pseudo value is given by

$$\theta^{*[i]}_{m*} = n\, n\theta^*_{m*} - [n - 1]\, \theta^*_{m*[j]}$$

and the Jacknife predictor is

$$\theta^{*J}_{m*} = \sum_{1 \le j \le n} \theta^{*[j]}{}_{m*}/n.$$

The robustness of the Jacknife method sustains that, because of the smoothness of our predictor, that

$$V^J_{m*}[\theta^*_{m*}] = \sum_{1 \le j \le n} [\theta^{*[j]}_{m*} - \theta^{*J}_{m*}]^2 / n[n - 1]$$

tends to the true variance and that $[\theta^*_{m*} - \theta]\left[V^J_{m*}[\theta^*_{m*}]\right]^{-1/2}$ follows approximately a T-Student distribution.

4. ANALYSIS OF THE BEHAVIOR OF THE DIFFERENT APPROACHES

An analytical comparison of the different approaches can not be made because the expressions of the errors do not share common factors. The comparisons are made by performing Monte Carlo experiments. Each run generated an estimate or a prediction of the TAE. A confidence interval was calculated for each result and the percent or runs in which the true value of θ belonged to it was the final result. Clearly the methods with a percent closer to the prescribed α = 0,05 must be preferred. The robustness of the competitors can be analyzed by establishing which was closer to the results expected.

Three experiments were performed. The variables were considered as standardized with support [-4, 4]. This interval was partitioned into 201 subintervals I_{-100} =]-4, -0,045[, I_{-99} =] -0,045, -0,125[,...., I_{100} =]3,5, 4[. They were representative of the corresponding taintings. As a result

$$P_i = \int_{Ii} f[z]dz.$$

The generation of n items permitted to compute the $n_i^{'}s$ and the estimates or predictors.

In the first experiment a population of 10 000 accounts was generated Bivariate Standard Normal distributions, with correlation coefficients ρ_{xz} = 0,5 0,7 and 0,9, were generated. X is the reported book value and Z the true one to be detected by the auditory. Then any selected j is classified in a tainting by evaluating Z_j. The population size was N = 10 000 and samples of size n = 100, 500 and 1 000 were selected.

The results are given in Table 4.1. Note that the normal approximation is very good for simple random sampling, as expected, when it is the sampling design. For DSU sampling θ_{srs} has a very different behavior and we can not rely in it. The normal inferences based on θ_{HT} seems to be not reliable. θ_m has a good behavior only for n = 1 000 and for θ^*_{m*} they are adequate when n > 100. For DSU sampling θ^*_{m*} and θ^*_m

have a similar behavior but the former is better for simple random sampling. These results seem to be due to the robustness properties of Jacknife procedures.

Table 4.1. Percentage of Confidence Intervals that included the true TAE in 100 runs
Bivariate Standard Normal Case.
N = 10 000, α = 0,05

		Simple n = 100	Random n = 500	Sampling n = 1000	Dollar n = 1000	Unit n = 500	Sampling n = 1000
ρ_{xz}	0,5	0,12	0,05	0,06	0,14	0,12	0,13
θ_{SRS}	0,7	0,11	0,07	0,05	0,13	0,14	0,13
	0,9	0,12	0,06	0,05	0,15	0,11	0,11
	0,5	0,21	0,24	0,10	0,16	0,13	0,10
θ_{HT}	0,7	0,22	0,22	0,11	0,11	0,15	0,12
	0,9	0,18	0,24	0,19	0,12	0,11	0,09
	0,5	0,19	0,12	0,07	0,22	0,14	0,05
θ_M	0,7	0,21	0,13	0,06	0,21	0,12	0,07
	0,9	0,16	0,12	0,04	0,20	0,10	0,05
	0,5	0,12	0,11	0,08	0,11	0,10	0,07
θ^*_m	0,7	0,14	0,10	0,06	0,14	0,09	0,07
	0,9	0,11	0,10	0,07	0,10	0,12	0,06
	0,5	0,10	0,09	0,06	0,11	0,08	0,06
θ^*_m	0,7	0,13	0,08	0,05	0,12	0,08	0,04
	0,9	0,11	0,09	0,07	0,10	0,10	0,06

In the second experiment m was assumed as representative of the behavior of the account. Using a sampling design n integers were selected from {1,...,N}. If $j \in T_i$ an uniform random variable ε with zero mean and variance $4P_i[1-P_i]$ is generated. If $P_i + \varepsilon$ is closer to one than to zero then $Y_{j[i]} = 1$, else another j* is selected and $Y_{j^*[i^*]} = 1$ if $j^* \in T_{i^*}$.

The results are given in Table 4.2. θ_{srs} is not a good alternative when m is the generating the belonging to the taintings. The use of θ_{HT} is similarly unreliable. θ_m has a better behavior as expected but θ^*_m is considerably better under this superpopulation. This result is supported by the general convergence properties of the class to which it belongs. It is very interesting that $\theta^*_{m^*}$ and θ^*_m have a similar behavior. Again the reliability of the use of $\theta^*_{m^*}$ is higher.

The third experiment implemented the model m* by generating $Z_t = \beta X_t + \varepsilon$. Then the value is determined once X is generated and the correlation coefficient is fixed. The results are highly interesting because $\theta^*_{m^*}$ has also a good behavior. Note in Table 4.3 that θ_{srs} has a behavior similar to those obtained in the second experiment. θ^*_m has results very close to those of θ_m and θ_{HT} is very unreliable.

These results suggest that the best alternative is to use $\theta^*_{m^*}$ for predicting TAE under the unknowledge of the real state of the account and its relations with other variables.

Table 4.2 Percentage of Confidence Intervals that included the true TAE in 100 runs.
m generates Y. N = 10 000, α = 0,05.

		Simple n = 100	Random n = 500	Sampling n = 1000	Dollar n = 100	Unit n = 500	Sampling n = 1000
ρ_{xz}	0,5	0,23	0,18	0,18	0,21	0,18	0,18
θ_{SRS}	0,7	0,24	0,16	0,17	0,17	0,19	0,14
	0,9	0,18	0,17	0,16	0,18	0,12	0,18
	0,5	0,25	0,23	0,20	0,22	0,18	0,14
θ_{HT}	0,7	0,16	0,19	0,18	0,20	0,12	0,15
	0,9	0,19	0,15	0,15	0,17	0,16	0,10
	0,5	0,19	0,08	0,07	0,12	0,08	0,06
θ_M	0,7	0,11	0,10	0,07	0,10	0,07	0,08
	0,9	0,10	0,07	0,06	0,10	0,06	0,06
	0,5	0,12	0,11	0,08	0,10	0,08	0,07
θ^*_m	0,7	0,08	0,07	0,06	0,09	0,09	0,05
	0,9	0,07	0,06	0,05	0,09	0,06	0,04
	0,5	0,07	0,06	0,06	0,10	0,10	0,08
θ^*_m	0,7	0,08	0,09	0,05	0,10	0,09	0,07
	0,9	0,07	0,09	0,04	0,10	0,06	0,07

The results suggest that the best alternative seem to be the use of θ^*_{m*} for predicting TAE when the autor unknowledge the real state of the account and its relations with other variables.

Table 4.3 Percent of Confidence Intervals that included the true TAE in 100 runs.
m* generates Z. N=10 000, α=0,05

		Simple n = 100	Random n = 500	Sampling n = 1000	Dollar n = 100	Unit n = 500	Sampling n = 1000
ρ_{xz}	0,5	0,22	0,15	0,13	0,24	0,19	0,17
θ_{SRS}	0,7	0,21	0,15	0,15	0,20	0,18	0,16
	0,9	0,15	0,16	0,16	0,15	0,16	0,17
	0,5	0,26	0,23	0,20	0,22	0,18	0,23
θ_{HT}	0,7	0,24	0,18	0,23	0,20	0,22	0,21
	0,9	0,22	0,25	0,21	0,22	0,20	0,21
	0,5	0,19	0,08	0,07	0,12	0,08	0,06
θ_M	0,7	0,13	0,13	0,11	0,12	0,08	0,08
	0,9	0,11	0,12	0,07	0,09	0,08	0,09
	0,5	0,07	0,06	0,08	0,11	0,08	0,06
θ^*_m	0,7	0,07	0,07	0,07	0,08	0,07	0,06
	0,9	0,06	0,05	0,05	0,05	0,05	0,04
	0,5	0,06	0,05	0,06	0,07	0,06	0,06
θ^*_m	0,7	0,06	0,05	0,06	0,06	0,05	0,05
	0,9	0,04	0,04	0,05	0,05	0,06	0,04

ACKNOWLEDGEMENTS

This paper was partially written while the author was visiting Prof. I. Schubertt at the FH-Konstanz supported by a postdoctor fellowship of DAAD.

REFERENCES

ARDILLY, P. (1992): **Les Techniques de Sondage**, Editions Technip. Paris.

BOUZA, C. (1995): "Linear rank test derived from superpopulation model", **Biometrical J**., 37, 497-506.

CHAUDHURI, A. and J.W.E. VOOS (1988): **Unified Theory and Strategies of Survey Sampling**, N. Holland, Amsterdam.

CROSBY, A. (1980): "Implications of prior probability elicitation on auditor sampling size decisions", **J. of Accounting Res**. 18, 585-593.

__________ (1981): "Bayesian statistics in auditing: a comparison of probability elicitation techniques", **The Accounting Rev**. 56, 355-365.

HERNANDEZ BASTIDA, A.; M.C. MARTEL ESCOBAR and F.J. VAZQUEZ POLO (1998): "On maximum entropy priors and a most likely likelihood in auditing", **Questiio** 22, 231-242.

HERNANDEZ BASTIDA, A. and F.J. VAZQUEZ POLO (1997): "A note on the Quasi-Bayesian audit risk model for dollar unit sampling", **The European Accounting Review**, 6, 501-507.

JURECKOVA J. and P. K. SEN (1996): **Robust Statistical Procedures: Asymptotics and Interrelations**, J. Wiley. New York.

KRAFT, W. (1986): "Statistical sampling for auditors: a new look", **The J. of Accounting**, Autumn, 49-56.

Mc CRAY, J.H. (1984): "A Quasi Bayesian audit risk model for Dollar Unit Sampling and multiple population", **The Accounting Rev**. 54, 35-51.

POTHOFF. R.R.; M.A. WOODBURY and K.G. MARTON 1992): "Equivalent sample size and equivalent degrees of freedom refinements for inference using survey weights under superpopulation models", **J. American Stat. Ass**. 383-396.

SCOTT, A.J. (1977): "On the problem of randomization in survey sampling", **Sankhya-C**, 39, 1-9.

SITTER, R.R. (1992): "A resampling procedure for complex survey data", **J. of the American Stat**. **Ass**. 87, 755-765.

SOLOMON, Y. (1982): "Probability assessment by individual auditors and audit teams: an empirical investigation", **J. of Accounting Res**. 20, 689-710.

VAZQUEZ POLO, F.J. and A. HERNANDEZ BASTIDA [1995]: "Behavior of the posterior error rate with partial prior information in auditing", **J. of Applied St**. 22, 469-476.