

Mgr inż. Grzegorz Sitek

Summary of a doctoral dissertation  
**„Methods of statistical inference in financial audit with the use of mixtures of probability distributions”**

written under the scientific guidance of prof. dr hab. Janusz Wywiół.

There is a trend to increase auditing in socio-economic activity, it is believed that it will increase the efficiency of enterprise control systems, and will prevent fraud. In particular, audit methods in IT are developing dynamically, which increases the quality of IT systems operation. Similar effects are expected from an audit in the field of improving management systems of enterprises, especially their finances. Qualitative and formal-legal methods are one of the directions of audit research, on the other hand, there is a need to control large populations of accounting entries and documents, which extends the time and increases the cost of control tests.

The indicated needs inspire economists to look for new methods of efficient audit or improvement of existing procedures. Therefore, the importance of statistical methods in control studies is expected to increase. As you know, they allow you to save money on research and time spent on it. In this field, there is a need to adapt the existing statistical methods to audit research and to propose new procedures. The use of statistical inference methods allows statutory auditors to generalize the results of the examination of a random sample to the entire population and may protect auditors from the allegation of bias. At the same time, these methods allow the experts to control the quality of these generalizations. During the study, the auditor has the opportunity to use readily available additional information, which may significantly increase the accuracy of the statistical methods used in the audit.

Computerization of accounting processes gives great opportunities to analyze the economic side of business activity. This also applies to the conduct of audit studies. From the point of view of the representative method, computerization provides convenient sampling frames. These frames also contain data on additional features of all objects that make up the studied population. These additional features (e.g. the observed nominal values of documents) greatly contribute to increasing the accuracy of statistical inference. This in effect reduces the risk of an audit.

The subject of the research is the population of documents and accounting entries. Moreover, usability testing of appropriate statistical methods in the audit, in particular the analysis of the usefulness of classical statistical inference and simulation methods: Monte Carlo procedures, bootstrap. The considered procedures will be tested on authentic data from the warehouse control of a large concern. The data will also be obtained by computer simulations according to assumed theoretical probability distributions, so that their value structure reflects typical error characteristics in accounting entries.

The paper presents methods of statistical inference in financial audit. Those methods which are based on mixtures of probability distributions for modeling the generation of book values were considered in more detail. The problem of the application of statistical methods in financial audit was conducted mainly in the context of the verification of statistical hypotheses. The hypotheses

were verified taking into account the risks involved in the audit and their connection with the probabilities of making type I and II errors.

The main part of the work consists of an introduction, three chapters numbered with consecutive natural numbers, and an ending.

The first chapter presents the necessary definitions in the field of statistical methods in financial audit. Apart from the basic definitions in the audit, selected statistical procedures used in the financial audit known in the literature are presented.

The second chapter presents the basic information on the mixtures of probability distributions that were used to model the generation of book values. This chapter describes the basic methods of estimating the parameters of probability distributions: the maximum likelihood method, the moments method and the MCMC method. The EM algorithm was used to estimate the parameters of the Poisson distribution mixture. The ECM algorithm, i.e. the extended version of the EM algorithm, was used to estimate the parameters of the mixture of gamma distributions. Then, the second chapter presents some modifications to the models of generating accounting values using mixtures of probability distributions. The model of generating book values based on the mix of Poisson distributions was described. For this model, the formula for the average value of the audit error depending on the parameters of the distribution mixture was presented. Then, for this model, inference was made on the basis of the moments from the test. Another model described was the model of generating book values based on a mixture of gamma distributions. For this model, the formula for the average value of the audit error depending on the parameters of the distribution mixture was presented and inference was made based on the moments from the sample. The last model assumed that accounting values are generated by the sum of normal and gamma random variables, where the values of the normally distributed variable represent the values of audit errors.

In the third chapter, inference was made on the average value of audit errors. In subsection 3.1, inference based on the likelihood function was carried out. For the models of generating book values based on the mixture of two Poisson distributions and the mixture of two gamma distributions with the same scale parameter, the power of the likelihood ratio test was determined by simulation. For the mixture of Poisson distributions, exact formulas for the estimators of the mixture parameters obtained with the use of the EM algorithm were derived. The simulation studies were carried out for the sample size ranging from 2% to 15% of the population size. The average error value from the  $H_1$  hypothesis took the values from  $1.01\tau_0$  to  $1.3\tau_0$ , where  $\tau_0$  is the average audit error specified by the  $H_0$  hypothesis. In financial audit, it is particularly important to control the risk of making a type II error, therefore the data was generated for the determined value of the average error from the  $H_1$  hypothesis. For the mixture of gamma distributions, formulas for the estimators of the mixture parameters obtained with the use of the ECM algorithm were derived. The results obtained in subsection 3.1 for the mixture of gamma probability distributions were the reason for the simulation tests that were carried out in subsection 3.3. The main purpose of these tests was to determine the expected sample size by means of an appropriate simulation procedure. The results obtained in subsection 3.3 are consistent with those presented in subsection 3.1. The expected sample sizes for the mixture of gamma distributions were much larger than the sizes

considered in the test power studies in Section 3.1. The results obtained in subsection 3.3 may form the basis for further research, the purpose of which is to determine the expected sample size in sequential tests based on the likelihood ratio. In subsection 3.2, simulation tests of the bootstrap power of the version of the mean error significance test were performed. The power of this bootstrap test was, as expected, lower than that of the likelihood-ratio test. Empirical research and all simulations were carried out in the R statistical environment using proprietary calculation scripts.

The most important results of the dissertation include:

- review of the basic selected statistical methods in financial audit: chapter 1.
- analysis of the properties of models of generating book values using mixtures of two or three probability distributions: subchapter 2.2.
- adaptation of the methods of moments, maximum likelihood and its extensions in the form of EM, ECM procedures for estimating mixtures of gamma or Poisson distributions: subsection 2.1 and paragraphs 3.1.5 and 3.1.6.
- a proposal to use the sums of gamma and normal distributions to model the generation of book values: subchapter 2.7.
- analytical determination of the expected sample size to test the hypothesis of mean error in modeling the accounting value generation based on a mixture of two Poisson distributions: paragraph 3.3.2.
- simulation determination of the expected sample size for the sequential testing of the mean error hypothesis: paragraph 3.3.3.
- simulation test of the power of the likelihood ratio test and the bootstrap test.
- creating proprietary computer programs used during simulation tests.

**The theoretical aim** of the work was to analyze selected statistical methods used in audit research. This goal was achieved in subsequent chapters of the work, starting with the first chapter, where selected statistical procedures used in financial audit were considered. The second and third chapters present the audit test procedure based on the mix of probability distributions. Some of the methods previously proposed in the literature were verified during simulation tests. The third chapter proposes a method of estimating the parameters of the likelihood function based on the EM algorithm.

**The practical aim** of the work was to analyze the quality of operation of the proposed statistical procedures, to verify the hypotheses regarding the audit by means of appropriately constructed simulation studies with the use of proprietary computer programs. The results of the simulation test of the power of tests on the mean value of the audit errors in the case of generating book values using mixtures of probability distributions are moderately encouraging for their practical use. In order to verify the hypotheses about the average audit error, in practice, bootstrap tests can be used, which do not require meeting the assumptions of classic tests and allow you to control the power on an ongoing basis. If the power is too low, the sample size should be gradually increased.

The scope of the considerations presented in the paper does not fully exhaust the issues of statistical inference in a financial audit with the use of mixtures of probability distributions. It seems worthwhile to continue them, especially towards a wider use of bootstrap procedures in financial auditing. Great possibilities of use in practice appear before sequential tests.