A Mixed Model for Double Checking Fallible Auditors

Raats, V.M.; Moors, J.J.A.; van der Genugten, B.B.

Publication date:
2004

No. 2004–82

A MIXED MODEL FOR DOUBLE CHECKING FALLIBLE AUDITORS

By V.M. Raats, J.J.A. Moors. B.B. van der Genugten

September 2004

ISSN 0924-7815

A mixed model for double checking fallible auditors

V.M. Raats*, J.J.A. Moors† and B.B. van der Genugten‡
Tilburg University
P.O.Box 90153
5000 LE Tilburg, The Netherlands

Abstract

The paper discusses the problem of a fallible auditor who assesses the values of sampled records, but may make mistakes. To detect these mistakes, a subsample of the checked elements is checked again, now by an infallible expert.

We propose a model for this kind of double check, which takes into account that records are often correct; however, *if* they are incorrect, the errors can take on many different values - as is often the case in audit practice. The model therefore involves error probabilities as well as distributional parameters for error sizes.

We derive maximum likelihood estimators for these model parameters and derive from them an estimator for the mean size of the errors in the population. A simulation study shows that the latter outperforms some other - previously introduced - estimators.

Keywords: audit, inspection errors, maximum likelihood, mixed distribution, monotone missing data

Jel codes: C13, C19

*corresponding author, V.M.Raats@uvt.nl, tel. +44-20-73949618
†J.J.A.Moors@uvt.nl
‡Ben.vdGenugten@uvt.nl

1 Introduction

Statistical modeling and inference of audits[1] is often based on the (implicit) assumption that the auditor does not make mistakes. However, there is no denying that auditors are humans and, as such, fallible. The last couple of years this has been proved only too often by cases like Enron and Worldcom.

These cases show that it is important to take the fallibility of auditors into account. One way to achieve this is by a repeated audit control. In a repeated audit control a fallible auditor checks a random sample of records. A subsample of these (already checked) records is double checked by another more skillful auditor, the so called expert, who is assumed to be infallible. The problem then is, how the information from both the fallible and infallible auditor should be combined to draw the most accurate conclusions. To solve this problem, a suitable statistical model for this problem will be developed here, and estimators for the parameters will be presented.

Tenenbein (1970) introduced a model for a repeated audit control where the sole parameter of interest is the fraction of incorrect records in the population. Hence, the auditor and expert classify the records as either incorrect and correct. Based on both the fallible and infallible classifications, Tenenbein (1970) derived the M(aximum) L(ikelihood) E(stimator)'s. In Tenenbein (1971) other aspects of the model were studied, like the determination of the optimal sample sizes, while Tenenbein (1972) generalized the previous model with dichotomous variables into a model with categorical variables. More recently, Moors *et al.* (2000) proposed a method to determine upper limits for the model with dichotomous variables; Raats and Moors (2003) looked in more detail at the Bayesian approach. Raats *et al.* (2004) generalized the model to include both categorical variables and more than one fallible auditor.

However, we are not only interested in the fraction of incorrect records in the populations, but also in the mean size of the errors in the population. In audit practice the records are often correct (*i.e.* the error is zero), but if they are incorrect the errors can take many different values (see Johnson *et al.* (1981) or Neter *et al.* (1985) *e.g.* for a more detailed discussion). The resulting error hence has a mixed distribution; we therefore will call models for this frequently occurring situation mixed models.

Mixed models for the familiar auditing situation involving one infallible audi-

[1] Throughout this paper we use the term "audit" (and similarly "auditor") in its general meaning of inspections (executed for example by controllers, surveyors or accountants).

tor have been discussed in the literature: Cox and Snell (1979) derived Bayesian estimators and upper limits for a model with non-negative errors with a probability mass in zero. Moors (1983) and Moors and Janssens (1989) expanded on this. Estimators for continuous, but not necessarily positive, errors with a point mass in zero were derived by Fienberg *et al*. (1977), Tamura and Frost (1986), Tamura (1988) and Laws and O'Hagan (2000).

A mixed model for a repeated audit control was given by Barnett *et al*. (2001); first a model for the classification frequencies was presented and MLE's for the parameters derived. Further, based on the size of the observed errors, several estimators for the mean value of the errors in the population were proposed; no relation was specified between the size of the non-zero errors and the (registered) values of the records. In this paper we will assume a normal regression model for the non-zero error sizes; we will derive MLE's for the model parameters and construct an estimator for the mean size of the errors. This last estimator is shown to outperform the estimators of Barnett *et al*. (2001).

Section 2 introduces our mixed model for a repeated audit control. In Section 2.2 the model of Tenenbein (1970) for the classification frequencies is extended slightly; the resulting model is identical to the model of Barnett *et al*. (2001). Conditional on the classification of a record, we specify regression models for the non-zero errors in Section 2.3. These conditional regression models are based on the multivariate linear regression model of Raats *et al*. (2002b) for monotone missing data. Note that this model is applicable here since a repeated audit control can be regarded as a (monotone) missing data problem: the expert's judgement is observed for double checked records, but is missing for the single checked record for which only the (fallible) auditor's assessment and the book value is available.

In Section 3, estimators for the classification frequencies and regression parameters are derived; for the latter we use the estimation techniques of Raats *et al*. (2002b). The relative efficiency of the OLS estimators with respect to the MLE's for the parameters of the conditional regression models are compared both analytically and by means of simulation. Section 4 discusses estimators for the mean value of the errors in the population. We present the MLE for our model and briefly discuss the competing estimators of Barnett *et al*. (2001). All the estimators are compared by means of simulation. The final Section 5 contains our main conclusions and ideas for further research.

2 The model

2.1 Notation

Define the random variable A_0 as the registered value (or the so called book value) of a random record. The random variables A_1 and A_2 are defined as the values of a random record according to the first auditor and the expert, respectively. Since the expert is assumed to be infallible A_2 is the true value. We denote the book and audit values of record t by $A_{t0},\ A_{t1}$, and A_{t2}, respectively.

As before the first auditor checks the records of a random sample (drawn with replacement) of predetermined size n_1; a subsample of size $n_2 \leq n_1$ is checked again by the expert. Now $(A_{t0},\ A_{t1},\ A_{t2})$ are available for the n_2 double checked sample records, while for the $n_1 - n_2$ single checked sample records only $(A_{t0},\ A_{t1})$ are available. Since in practice the book values are known for all records of the population, we will assume that A_{t0} is known for the whole population.

Our mixed model is constructed from an absolute model for the classification frequencies and conditional models for the audit values. First all records are classified into five groups, based on the question whether the two audit values and the book value are identical. In Section 2.2 we give our model for the corresponding classification frequencies. If all three values coincide, no further steps are necessary. In the four other cases, we need to specify models for one of the audit values, or both. Section 2.3 describes these conditional regression models.

2.2 Classifications

The parameter π_0 (π_1) is the probability that the auditor classifies a random record as 'incorrect' ('correct'); the quotation marks already indicate that this classification may be wrong. With conditional probability $\pi_{0|0}$ ($\pi_{1|1}$) the 'incorrect' ('correct') record is indeed incorrect (correct). With conditional probability $\pi_{1|0}$ ($\pi_{0|1}$) the 'incorrect' ('correct') record is misclassified by the auditor and is correct (incorrect) after all. Joint probabilities as $\pi_{01} = \pi_0\pi_{1|0}$ (a random record being classified as 'incorrect' by the auditor and as correct by the expert) follow from these.

So far our model for the classification frequencies is identical to the model of Tenenbein (1970) (1971), (1972) and Raats and Moors (2003). However, now we are interested not only in the fraction errors but also in the size of the errors; an additional subdivision is therefore necessary. If the auditor correctly concludes

that a record is in error, two possibilities remain: (s)he is correct about the size of the error, or not. Accordingly, we introduce the conditional probabilities $\pi_{0e|0}$ ($\pi_{0u|0}$) for the events that the error size indicated by the auditor is *e*qual (*u*nequal) to the true error. So $\pi_{0|0} = \pi_{0e|0} + \pi_{0u|0}$ and $\pi_{00} = \pi_{00e} + \pi_{00u}$.

The foregoing classifications and probabilities can be expressed in terms of book and audit values. For example

$$\pi_{0u|0} = Pr(A_0 \neq A_2, A_1 \neq A_2 | A_0 \neq A_1).$$

The simultaneous probabilities follow at once. Table 2.1 gives an overview of the five possible classifications and their probabilities.

Classification	**Probability**
1. $A_0 = A_1$, $A_0 = A_2$	π_{11}
2. $A_0 = A_1$, $A_0 \neq A_2$	π_{10}
3. $A_0 \neq A_1$, $A_0 = A_2$	π_{01}
4. $A_0 \neq A_1$, $A_0 \neq A_2$, $A_1 = A_2$	π_{00e}
5. $A_0 \neq A_1$, $A_0 \neq A_2$, $A_1 \neq A_2$	π_{00u}

Table 2.1: Classifications and probabilities

We denote the sample classification frequencies by the symbol C with the same subindices as the corresponding probabilities π in Table 2.1. Figure 2.1 gives an overview of the sample frequencies and probabilities.

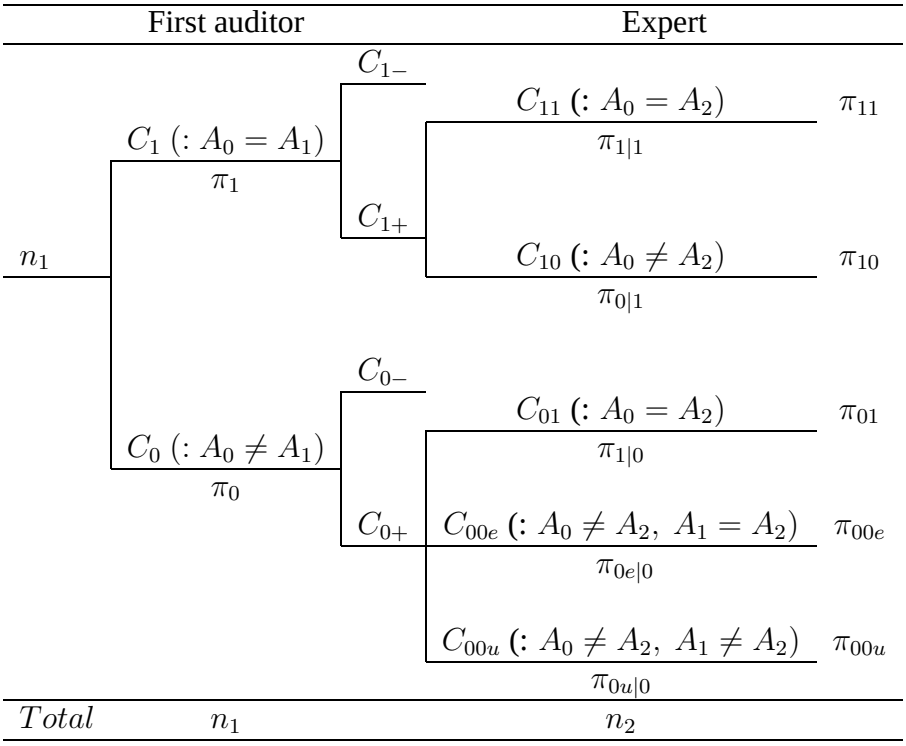


Figure 2.1: Classification frequencies and probabilities

Under the assumption of random sampling with replacement, all random variables in the model have (conditional) multinomial distributions with the (conditional) classification probabilities as parameters:

$$\begin{cases} \mathcal{L}(C_1, C_0) & = & M(n_1; \pi_1, \pi_0) \\ \mathcal{L}(C_{1+}, C_{0+} | C_0 = c_0, C_1 = c_1) & = & M(n_2; c_1/n_1, c_0/n_1) \\ \mathcal{L}(C_{11}, C_{10} | C_{1+} = c_{1+}) & = & M(c_{1+}; \pi_{1|1}, \pi_{0|1}) \\ \mathcal{L}(C_{01}, C_{00e}, C_{00u} | C_{0+} = c_{0+}) & = & M(c_{0+}; \pi_{1|0}, \pi_{0e|0}, \pi_{0u|0}). \end{cases} \tag{2.1}$$

This model for the classification frequencies is identical to the model of Barnett *et al.* (2001)

2.3 Conditional regression

Since the book value is available for each record, it is only necessary to specify a conditional model for A_{t1} given $A_{t1} \neq A_{t0}$. Whether this is the case follows from the classification of record t. If the book value and audit value do not coincide, it

seems reasonable to assume that the book value influences the audit value. So we assume

$$A_{t1} = \beta_0' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \varepsilon_t, \text{ given } A_{t0} \neq A_{t1},$$

with $E(\varepsilon_t|A_{t0}) = 0$ for some (regression) coefficient $\beta_0 \in \mathbb{R}^2$. Note that we omit in our notation (for the expectation and variance) the condition $A_{t0} \neq A_{t1}$. Moreover, we assume a constant variance ($V(\varepsilon_t|A_{t0}) = \sigma_0^2$) and no correlation between records.

We only need to specify a model for A_{t2} if the true value does not coincide with the book value or previous audit value. This is the case for the classifications 2 and 5 in Table 2.1. For both classifications we assume linear regression models, which are not necessary identical: after all, the first auditor missing an error might indicate that the error is small while the first auditor finding an error (but not the true one) might indicate a large or complicated error. We assume

$$A_{t2} = \beta_1' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \varepsilon_t, \text{ given } \begin{cases} A_{t0} = A_{t1} \\ A_{t0} \neq A_{t2} \end{cases},$$

with $E(\varepsilon_t|A_{t0}) = 0$ for some (regression) coefficient $\beta_1 \in \mathbb{R}^2$. Again we assume that the variance of the error terms is constant ($V(\varepsilon_t|A_{t0}) = \sigma_1^2$) and that there is no correlation between records.

Similarly, we assume

$$A_{t2} = \beta_{0u}' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \varepsilon_t, \text{ given } \begin{cases} A_{t0} \neq A_{t1} \\ A_{t0} \neq A_{t2} \\ A_{t1} \neq A_{t2} \end{cases},$$

with $E(\varepsilon_t|A_{t0}) = 0$ for some (regression) coefficient $\beta_{0u} \in \mathbb{R}^2$. Although we assume again a constant variance ($V(\varepsilon_t|A_{t0}) = \sigma_{0u}^2$) and no correlation between different records, we do not impose restrictions on the correlation between the audit and true value per record (or equivalently, the covariance σ_{12}).

Table 2.2 gives an overview of the explanatory and dependent variables of these three conditional regression models.

Parameters	β_1, σ_1^2	β_0, σ_0^2	$\beta_{0u}, \sigma_{0u}^2, \sigma_{12}$
conditions	$\begin{cases} A_{t0} = A_{t1} \\ A_{t0} \neq A_{t2} \end{cases}$	$A_{t0} \neq A_{t1}$	$\begin{cases} A_{t0} \neq A_{t1} \\ A_{t0} \neq A_{t2} \\ A_{t1} \neq A_{t2} \end{cases}$
dependent variables	A_{t2}	A_{t1}	A_{t2}
explanatory variables	$[1 \quad A_{t0}]$	$[1 \quad A_{t0}]$	$[1 \quad A_{t0}]$
error terms previous variables	-	-	$A_{t1} - \beta_0' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix}$
number of observations	C_{10}	C_0	C_{00u}

Table 2.2: Explanatory and dependent variables

In all our conditional regression models, the explanatory variables consist of the constant and the book value. The conditional model given $A_{t0} = A_{t1}$, has the true value as dependent variable. The other two conditional models (given $A_{t0} \neq A_{t1}$) form a bivariate regression model with monotone missing observations: for the first dependent variable (the value according to the first auditor) C_0 observations are available, while for the second dependent variable (the true value) only C_{00u} observations are available. In the regression model with monotone missing observations the error terms of the preceding dependent variables are used for deriving the MLE (see for more details Raats *et al*. (2002b)); the error terms needed for our model are given in the last column of Table 2.2.

Table 2.3 gives an overview of the conditional regression models for all classifications. This overview will be especially useful for the estimation of the mean true value in Section 4.

Classification	Conditional regression model
$A_0 = A_1$,$A_0 = A_2$	-
$A_0 = A_1$, $A_0 \neq A_2$	$A_{t2} = \beta_1' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \varepsilon_t$, $E(\varepsilon_t \mid A_{t0}) = 0$, $Cov(\varepsilon_t \mid A_{t0}) = \sigma_1^2$,
$A_0 \neq A_1$, $A_0 = A_2$	$A_{t1} = \beta_0' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \varepsilon_t$ $E(\varepsilon_t \mid A_{t0}) = 0$, $Cov(\varepsilon_t \mid A_{t0}) = \sigma_0^2$,
$A_0 \neq A_1$, $A_0 \neq A_2$, $A_1 = A_2$	$A_{t1} = \beta_0' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \varepsilon_t$, $E(\varepsilon_t \mid A_{t0}) = 0$, $Cov(\varepsilon_t \mid A_{t0}) = \sigma_0^2$,
$A_0 \neq A_1$, $A_0 \neq A_2$, $A_1 \neq A_2$	$\begin{bmatrix} A_{t1} \\ A_{t2} \end{bmatrix} = \begin{bmatrix} \beta_0' \\ \beta_{0u}' \end{bmatrix} \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \varepsilon_t$, $E(\varepsilon_t \mid A_{t0}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$, $Cov(\varepsilon_t \mid A_{t0}) = \begin{bmatrix} \sigma_0^2 & \sigma_{12} \\ \sigma_{12} & \sigma_{0u}^2 \end{bmatrix}$

Table 2.3: Conditional regression models

3 Estimation of the model parameters

3.1 Classification probabilities

The classification frequencies have binomial and multinomial distributions (see (2.1)). Hence, the MLE's for the classification probabilities are the sample fractions:

$$\begin{cases} \widehat{\Pi}_1 = \dfrac{C_1}{n_1}, \qquad \widehat{\Pi}_0 = \dfrac{C_0}{n_1} \\ \widehat{\Pi}_{1|1} = \dfrac{C_{11}}{C_{1+}}, \quad \widehat{\Pi}_{0|1} = \dfrac{C_{10}}{C_{1+}} \\ \widehat{\Pi}_{1|0} = \dfrac{C_{01}}{C_{0+}}, \quad \widehat{\Pi}_{0e|0} = \dfrac{C_{00e}}{C_{0+}}, \quad \widehat{\Pi}_{0u|0} = \dfrac{C_{00u}}{C_{0+}}. \end{cases} \tag{3.1}$$

These MLE's can be found in Barnett *et al.* (2001) as well.

If C_{0+} or C_{1+} is zero, not all MLE's in (3.1) are defined; see Raats *et al.* (2004) Section 3.3 for a more detailed discussion of this situation and possible solutions.

3.2 Regression parameters

We use the following notation for sample averages and variances:

$$\overline{A}_g^{(C_i)} = \frac{1}{C_i}\sum^{C_i} A_{tg}, \tag{3.2}$$

$$\overline{S}_{gh}^{(C_i)} = \frac{1}{C_i}\sum^{C_i}(A_{tg} - \overline{A}_g^{(C_i)})(A_{th} - \overline{A}_h^{(C_i)}), \tag{3.3}$$

where $g, h = 0, 1, 2$ and C_i is either a classification frequency such as C_0 or C_{00u}, or a sample size such as n_2.

We substitute the Greek letters for the model parameters by the Arabic letters to denote the OLS estimators; ML estimators are denoted by original (Greek) symbol and an additional $\widehat{\ }$. The estimators for the regression parameters of the conditional regression models in Section 2.3 can be determined by means of the estimation procedures in Raats *et al.* (2002b) Section 3. The OLS estimators were are acquired by the orthogonal projections of the dependent variables onto the column space of the explanatory variables; the MLE's are acquired by the orthogonal projections of the dependent variables onto the column space of the explanatory variables and the residuals of the preceding dependent variables. Further details are omitted here.

Table 2.2 gives an overview of the dependent and explanatory variables for the parameters in our conditional regression models. The described procedure results

in the following OLS estimators

$$b_1 = \begin{bmatrix} \overline{A}_2^{(C_{10})} - (S_{00}^{(C_{10})})^{-1} S_{02}^{(C_{10})} \overline{A}_0^{(C_{10})} \\ (S_{00}^{(C_{10})})^{-1} S_{02}^{(C_{10})} \end{bmatrix}, \quad s_1^2 = \frac{\sum^{C_{10}} (A_{t2} - b_1' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix})^2}{C_{10} - 2}$$

$$b_0 = \begin{bmatrix} \overline{A}_1^{(C_0)} - (S_{00}^{(C_0)})^{-1} S_{01}^{(C_0)} \overline{A}_0^{(C_0)} \\ (S_{00}^{(C_0)})^{-1} S_{01}^{(C_0)} \end{bmatrix}, \qquad s_0^2 = \frac{\sum^{C_0} (A_{t1} - b_0' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix})^2}{C_0 - 2}$$

$$b_{0u} = \begin{bmatrix} \overline{A}_2^{(C_{00u})} - (S_{00}^{(C_{00u})})^{-1} S_{02}^{(C_{00u})} \overline{A}_0^{(C_{00u})} \\ (S_{00}^{(C_{00u})})^{-1} S_{02}^{(C_{00u})} \end{bmatrix}$$

$$s_{0u}^2 = \frac{\sum^{C_{00u}} (A_{t2} - b_{0u}' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix})^2}{C_{00u} - 2}$$

$$s_{12} = \frac{1}{C_{00u} - 2} \sum^{C_{00u}} (A_{t1} - b_0' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix})(A_{t2} - b_{0u}' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix}),$$

and the following MLE's

$$\widehat{\beta}_1 = b_1, \qquad \widehat{\sigma}_1^2 = \frac{C_{10} - 2}{C_{10}} s_1^2, \qquad \widehat{\beta}_0 = b_0, \qquad \widehat{\sigma}_0^2 = \frac{C_0 - 2}{C_0} s_0^2$$

$$\begin{bmatrix} \widehat{\beta}_{0u} \\ \widehat{\alpha}_{0u} \end{bmatrix} = \begin{bmatrix} C_{00u} & \sum^{C_{00u}} A_{t0} & \sum^{C_{00u}} \widehat{\varepsilon}_{t1} \\ \sum^{C_{00u}} A_{t0} & \sum^{C_{00u}} A_{t0}^2 & \sum^{C_{00u}} A_{t0} \widehat{\varepsilon}_{t1} \\ \sum^{C_{00u}} \widehat{\varepsilon}_{t1} & \sum^{C_{00u}} A_{t0} \widehat{\varepsilon}_{t1} & \sum^{C_{00u}} \widehat{\varepsilon}_{t1}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum^{C_{00u}} A_{t2} \\ \sum^{C_{00u}} A_{t0} A_{t2} \\ \sum^{C_{00u}} \widehat{\varepsilon}_{t1} A_{t2} \end{bmatrix}$$

$$\widehat{\sigma}_{0u}^2 = \widehat{\sigma}_0^2 \widehat{\alpha}_{0u}^2 + \frac{1}{C_{00u}} \sum^{C_{00u}} (A_{t2} - \widehat{\beta}_{0u}' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} - \alpha_{0u} \widehat{\varepsilon}_{t1})^2, \quad \widehat{\sigma}_{12} = \widehat{\sigma}_0^2 \widehat{\alpha}_{0u},$$

where $\widehat{\varepsilon}_{t1} = A_{t1} - \widehat{\beta}_0' \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix}$.

The MLE's for β_1 and β_0 coincide with the OLS estimators. The MLE's for σ_1^2 and σ_0^2 differ from the OLS estimators solely by the denominator: the number of observations versus the degrees of freedom. Only with respect to β_{0u}, σ_{0u}^2 and σ_{12} the MLE's differ essentially from the OLS estimators. In the practical example we study the relative efficiency of the OLS estimators and MLE's for these parameters by simulation.

3.3 Practical example

As in Raats *et al*. (2004), the practical example concerns the Dutch social security payments. However, now we consider another case study where also error sizes are observed. The population consists of 587 social security payments with mean 9.0418 and standard deviation 8.5726 (both in 1000's of Dutch guilders). An internal auditor checks all 587 social security payments; an external auditor (the expert) checks a subsample of size 60 once more. We will assume here that the 587 payments checked by the first auditor constitute a sample from a large population. In this context, the social security payment which actually has been paid is the book value A_0; A_1 (A_2) is the social security payment which should have been paid according to the first auditor (expert). Table 3.1 contains the classification frequencies of this case.

	Total	Single checked sample	Double checked sample Expert		
First auditor			Total	correct	incorrect
'correct'	$c_1 = 551$	$c_{1-} = 493$	$c_{1+} = 58$	$c_{11} = 55$	$c_{10} = 3$
'incorrect'	$c_0 = 36$	$c_{0-} = 34$	$c_{0+} = 2$	$c_{01} = 0$	$c_{00e} = 2$
Total	$n_1 = 587$	$n_1 - n_2 = 527$	$n_2 = 60$	$c_{+1} = 55$	$c_{+0} = 5$

Table 3.1: CTSV example

Clearly, in the double checked sample the first auditor did not make up errors, missed three errors and found two (true) errors; the expert confirmed the size of the latter errors.

For these classification frequencies, (3.1) results in the ML estimates

$$\widehat{\pi}_{11} = 0.8901, \quad \widehat{\pi}_{10} = 0.0486, \quad \widehat{\pi}_{01} = 0, \quad \widehat{\pi}_{00e} = 0.0613, \quad \widehat{\pi}_{00u} = 0.$$

The ML estimates for the regression parameters are determined from the sample observations of A_{t0}, A_{t1} and A_{t2}. Since there are no sample records with $\{A_{t0} \neq A_{t1}, A_{t0} \neq A_{t1}, A_{t1} \neq A_{t2}\}$ (*i.e.* $c_{00u} = 0$), the parameters β_{0u}, σ^2_{0u} and σ_{12} can not be estimated. The ML estimates for the other regression parameters are

$$\widehat{\beta}_1 = \begin{bmatrix} -14.7107 \\ -0.8275 \end{bmatrix}, \quad \widehat{\sigma}^2_1 = 53.5911, \quad \widehat{\beta}_0 = \begin{bmatrix} -0.6807 \\ 0.8808 \end{bmatrix}, \quad \widehat{\sigma}^2_0 = 17.3533.$$

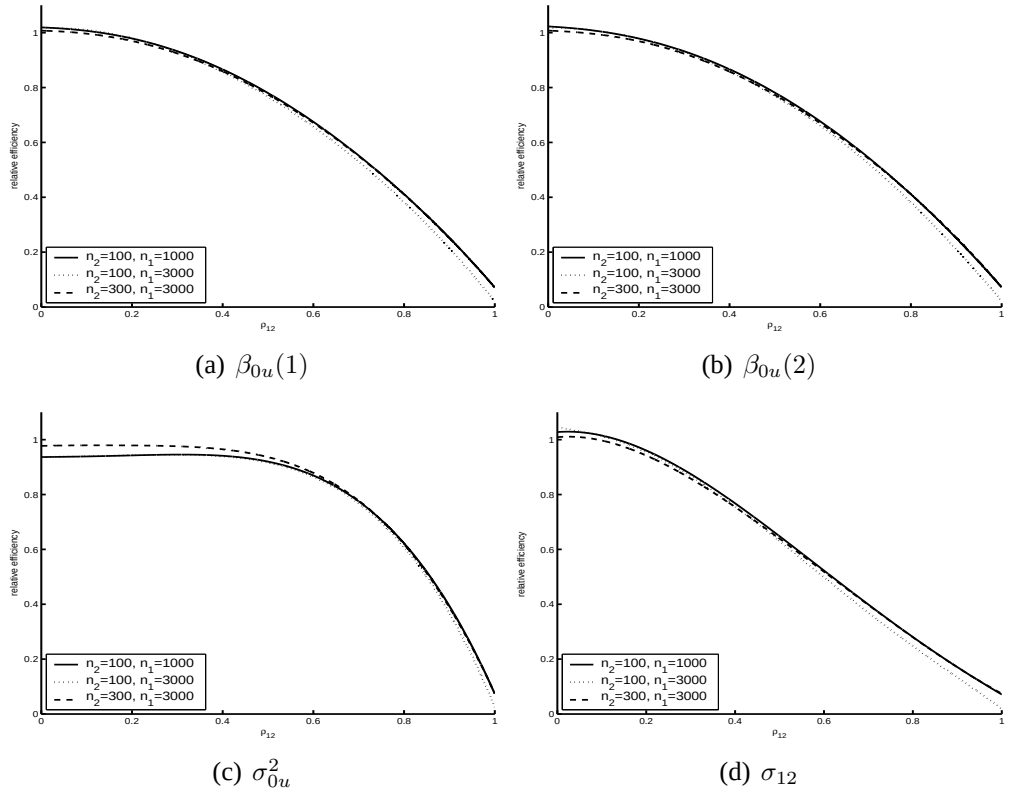
These ML estimates are used in our simulations to study the relative efficiency of the OLS estimators and MLE's for β_{0u}, σ^2_{0u} and σ_{12}.

The difference between OLS and ML estimation mainly stems from the treatment of the C_{00u} observations where the auditor correctly identifies an error, but errs in its size. Hence in the simulation study, we use a classification probability π_{00u} which is unlikely to lead to zero observations in this category:

$$\pi_{11} = \pi_{10} = \pi_{01} = \pi_{00e} = 0.1, \quad \pi_{00u} = 0.6.$$

We take the regression parameters equal to the corresponding ML estimates of the practical example; in addition we assume that β_{0u} (σ^2_{0u}) is equal to β_0 (σ^2_0). Since we expect the correlation between A_{t1} and A_{t2} (given $\{A_{t0} \neq A_{t1}, A_{t0} \neq A_{t1}, A_{t1} \neq A_{t2}\}$) to be important for the relative efficiency, we look at different values for the correlation coefficient (ρ_{12}); this determines as well the covariance $\sigma_{12} = \rho_{12}\sigma_0\sigma_{0u}$.

We simulate the book values from a normal distribution with mean 9.0418 and standard deviation 8.5726 from the practical example. The audit values are also drawn from (multi)normal distributions. To determine the effect of the sample sizes, we have simulated data (each with runsize 10,000) for three different situations: (a) $n_2 = 100, n_1 = 1000$, (b) $n_2 = 100, n_1 = 3000$ and (c) $n_2 = 300, n_1 = 3000$. Figure 3.1 contains the smoothed curves of the relative efficiency (*i.e.* the ratio of the mean squared errors of the OLS and ML estimators) for the different parameters as function of ρ_{12}. Note that each graph contains three curves, which however often partly coincide.



(a) $\beta_{0u}(1)$ (b) $\beta_{0u}(2)$

(c) σ^2_{0u} (d) σ_{12}

Figure 3.1: Relative efficiency of OLS in relation to ML

The first and second graph show the relative efficiency for the first and second component of β_{0u}, respectively. In Appendix 6.1 we derive analytical expression for the relative efficiency of the LS and GLS estimators for the regression coefficients. The first two graphs show the same pattern as Figure 6.1 in Appendix 6.1 and hence confirm these findings. For low values of the correlation coefficient, there is hardly any difference in efficiency between the two estimators; for high values, $\widehat{\beta}_{0u}$ is much more efficient than b_{0u}. This difference in efficiency increases with the missing data ratio. Note that the difference seems not to depend on the absolute sample sizes themselves, only on this ratio $1 - n_2/n_1$.

The third and fourth graph, for σ^2_{0u} and σ_{12}, show a similar picture as the first two. This is understandable since the MLE's $\widehat{\sigma}^2_{0u}$ and $\widehat{\sigma}_{12}$ are functions of $\widehat{\sigma}^2_0$ which is based on all n_1 observations.

4 Estimation of the mean true value

4.1 Notation

In a repeated audit control, the main parameter of interest often is the mean true value in the population or equivalently the total true value in the population. The mean population error size is the difference between the mean population book value μ_0 and the mean population true value, μ_2: $\mu_0 - \mu_2$. Since we assume that the book values are available for all population elements, the estimator for the mean error size is obtained by subtracting the estimator for μ_2 from the known parameter μ_0.

In Section 4.2 we propose an estimator for μ_2 based on our model. Section 4.3 discusses several estimators of Barnett *et al*. (2001). All four estimators are compared by simulation in Section 4.4.

In addition to (3.2) and (3.3), we use the following comparable notation

$$\widehat{\alpha}_{gh}^{(C_i)} = \frac{\sum^{C_i}(A_{tg} - \overline{A}_g^{(C_i)})(A_{th} - \overline{A}_h^{(C_i)})}{\sum^{C_i}\left(A_{tg} - \overline{A}_g^{(C_i)}\right)^2}.$$

The symbol θ will denote all model parameters, *i.e.* all classification probabilities and regression parameters; the MLE for θ is denoted by $\widehat{\theta}$.

4.2 A new estimator

Our estimator for μ_2 is the average of the observed and predicted true values of all population elements:

$$\widehat{\mu}_2 = \frac{1}{n_p}\sum_{t=1}^{n_p}\widehat{A}_{t2}, \tag{4.1}$$

with

$$\widehat{A}_{t2} = \begin{cases} A_{t2}, & \text{if } t = 1,\ldots,n_2 \\ E\{A_{t2}|A_{t0}, A_{t1}, A_{t0} = A_{t1}, \widehat{\theta}\}, & \text{if } t = n_2+1,\ldots,n_1 \text{ and } A_{t0} = A_{t1} \\ E\{A_{t2}|A_{t0}, A_{t1}, A_{t0} \neq A_{t1}, \widehat{\theta}\}, & \text{if } t = n_2+1,\ldots,n_1 \text{ and } A_{t0} \neq A_{t1} \\ E\{A_{t2}|A_{t0}, \widehat{\theta}\}, & \text{else.} \end{cases}$$

Each missing A_{t2} is estimated by its conditional expectation (under the normality assumption) given the observations and the (estimated) parameter values. The

conditional expectations differ per classification (see Table 2.3) and are given in Appendix 6.2.

The advantage of this estimator is that it distinguishes the different classifications, while using all available sample and population information. It also has some nice (asymptotic) properties.

4.3 Estimators Barnett

Although Barnett *et al*. (2001) did not specify a relation between the size of the non-zero errors and the book values, several estimators for μ_2 (or $\mu_0 - \mu_2$) were proposed: the regression estimator, the post-stratification estimator and the estimator from non-overlapping samples.

Similar to (4.1), the regression estimator for μ_2 is the average of the observed and predicted A_{t2} of all population elements. However, the predictions for the A_{t2} differ from ours. The regression estimator $\widehat{\mu}_{2r}$, used by Barnett *et al*. (2001) equation (17), equals

$$\widehat{\mu}_{2r} = \overline{A}_2^{(n_2)} + (\overline{A}_1^{(n_1)} - \overline{A}_1^{(n_2)})\widehat{\alpha}_{12}^{(n_2)} + (\mu_0 - \overline{A}_0^{(n_1)})\widehat{\alpha}_{01}^{(n_1)}\widehat{\alpha}_{12}^{(n_2)}. \tag{4.2}$$

This estimator is quite logical in case of the following model:

$$\begin{bmatrix} A_{t0} \\ A_{t1} \\ A_{t2} \end{bmatrix} = \beta' + \varepsilon_t, \text{ with } E\{\varepsilon_t\} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad Var\{\varepsilon_t\} = \begin{bmatrix} \sigma_{00} & \sigma_{01} & 0 \\ \sigma_{01} & \sigma_{11} & \sigma_{12} \\ 0 & \sigma_{12} & \sigma_{22} \end{bmatrix}.$$

Note however, that this model contradicts the model for the classification probabilities, since it does not distinguish the different classifications. This in contrast to the post-stratification estimator for μ_2 (see Barnett *et al*. (2001) equation (21))

$$\widehat{\pi}_{11}\mu_0 + \widehat{\pi}_{10}\overline{A}_2^{(n_2)} + \widehat{\pi}_{01}\mu_0 + \widehat{\pi}_{00e}\overline{A}_1^{(n_1)} + \widehat{\pi}_{00u}\overline{A}_2^{(n_2)}.$$

This estimator is the sum of the MLE's for the classification probabilities times the estimator for the mean true value of elements with that classification. The disadvantage of this estimator is that the estimators for the mean values per classification can be quite biased. An alternative estimator $\widehat{\mu}_{2p}$ with the same structure but with different estimators for the stratum means is

$$\widehat{\mu}_{2p} = \widehat{\pi}_{11}\overline{A}_2^{(C_{11})} + \widehat{\pi}_{10}\overline{A}_2^{(C_{10})} + \widehat{\pi}_{01}\overline{A}_2^{(C_{01})} + \widehat{\pi}_{00e}\overline{A}_2^{(C_{00e})} + \widehat{\pi}_{00u}\overline{A}_2^{(C_{00u})}. \tag{4.3}$$

Since our simulation results with this estimator agree with Barnett's, we assume that this is the estimator which Barnett *et al*. (2001) actually used in their simulations. The disadvantage of this post-stratification estimator is that it uses the sample information of the single checked elements solely for the estimation of the classification probabilities; the estimation of the stratum means is only based on the double checked sample.

The last estimator $\widehat{\mu}_{2w}$ uses information from both single and double checked sample elements (see Barnett *et al*. (2001) equation (25))

$$\begin{aligned}\widehat{\mu}_{2w} &= \mu_0 - \frac{n_2}{n_1}(\overline{A}_0^{(n_2)} - \overline{A}_2^{(n_2)}) \\ &\quad - \frac{n_1 - n_2}{n_1}\frac{C_{0-}\widehat{\pi}_{0|0} + C_{1-}\widehat{\pi}_{0|1}}{C_{0-}}(\overline{A}_0^{(n_1-n_2)} - \overline{A}_1^{(n_1-n_2)}). \end{aligned} \tag{4.4}$$

This estimator is μ_0 minus the weighted average of the mean error size of the double checked elements, minus the mean error size of the single checked sample elements according to the auditor (multiplied by a correction factor for the misclassifications). Theorem 4.1 shows that $\widehat{\mu}_{2w}$ is not always consistent.

Theorem 4.1. *In case of random sampling* $\widehat{\mu}_{2w} \xrightarrow{P} \mu_2$ *if and only if* $E\{A_{t0} - A_{t1}|A_{t0} \neq A_{t1}\} = E\{A_{t0} - A_{t2}|A_{t0} \neq A_{t2}\}$.

Proof. We denote the fraction incorrect elements in the population by $p_0 (= \pi_{10} + \pi_{00})$. Since sample means converge to their expectations in case of random sampling, it follows that

$$\overline{A}_0^{(n_2)} - \overline{A}_2^{(n_2)} \xrightarrow{P} \mu_0 - \mu_2, \qquad \overline{A}_0^{(n_1-n_2)} - \overline{A}_1^{(n_1-n_2)} \xrightarrow{P} \mu_0 - \mu_1,$$

$$\frac{C_{0-}\widehat{\pi}_{0|0} + C_{1-}\widehat{\pi}_{0|1}}{C_{0-}} = \frac{\frac{C_{0-}}{n_1-n_2}\widehat{\pi}_{0|0} + \frac{C_{1-}}{n_1-n_2}\widehat{\pi}_{0|1}}{\frac{C_{0-}}{n_1-n_2}} \xrightarrow{P} \frac{\pi_0\pi_{0|0} + \pi_1\pi_{0|1}}{\pi_0} = \frac{p_0}{\pi_0}.$$

From this and $\mu_0 - \mu_1 = \pi_0 E\{A_{t0} - A_{t1}|A_{t0} \neq A_{t1})$, it follows that

$$\widehat{\mu}_{2w} \xrightarrow{P} \mu_0 - \frac{n_2}{n_1}(\mu_0 - \mu_2) - \frac{n_1 - n_2}{n_1}p_0 E\{A_{t0} - A_{t1}|A_{t0} \neq A_{t1}).$$

Only if $E\{A_{t0} - A_{t1}|A_{t0} \neq A_{t1}\} = E\{A_{t0} - A_{t2}|A_{t0} \neq A_{t2}\}$, we have $p_0 E\{A_{t0} - A_{t1}|A_{t0} \neq A_{t1}) = p_0 E\{A_{t0} - A_{t2}|A_{t0} \neq A_{t2}\} = \mu_0 - \mu_2$ and hence $\widehat{\mu}_{2w} \xrightarrow{P} \mu_2$. □

4.4 A simulation study

We compare the performance of the estimators of this section by simulation. The simulation procedure we use is almost identical to the one of Barnett *et al.* (2001) Section 5.

The simulations (runsize 10,000) are performed for several sets of given classification probabilities and sample sizes; see Table 4.1. The n_1 book values are drawn from the following distribution:

book value	100	500	1000	2000	5000
probability	0.9	0.05	0.03	0.015	0.005

.

The classifications of the items are drawn from multinomial distributions. The fractional error sizes have the following uniform distributions:

$$\begin{aligned}
\frac{A_{t0} - A_{t1}}{A_{t0}} &\sim U(0,1), \qquad \text{if } A_{t0} \neq A_{t1}, \\
\frac{A_{t0} - A_{t2}}{A_{t0}} &\sim U(0,1), \qquad \text{if } A_{t0} = A_{t1}, A_{t0} \neq A_{t2}, \\
\frac{A_{t0} - A_{t2}}{A_{t0}} &= 1 - \frac{A_{t1}}{A_{t0}}, \qquad \text{if } A_{t0} \neq A_{t1}, A_{t0} \neq A_{t2}, A_{t1} \neq A_{t2}.
\end{aligned}$$

So far the simulation procedure is identical to the one of Barnett *et al.* (2001). However, to avoid undefined MLE's, we select in the second round at least one "incorrect" ("correct") element if $C_0 > 0$ (if $C_1 > 0$); this does not alter the MLE's (see Raats *et al.* (2004) Section 3.3 for a more detailed discussion).

From the described simulation procedure, the mean population error size can be determined analytically for each set of classification probabilities. In each simulation run $\mu_0 - \mu_2$ is estimated using the four discussed estimators. Note that $E\{A_{t0} - A_{t1} | A_{t0} \neq A_{t1}\} = E\{A_{t0} - A_{t2} | A_{t0} \neq A_{t2}\}$ in the described simulation procedure. Table 4.1 contains the results of the simulations.

From the four studied estimators, $\widehat{\mu}_{2r}$ has the largest bias; the other three estimators have a small bias (if any at all). The bias of $\widehat{\mu}_{2w}$ (never exceeding 0.1) is caused by the fact that $E\{A_{t0} - A_{t1} | A_{t0} \neq A_{t1}\} = E\{A_{t0} - A_{t2} | A_{t0} \neq A_{t2}\}$ for the simulated data.

Higher sample sizes in the first and second round lead to a lower variance for all estimators except $\widehat{\mu}_{2p}$; the variance of $\widehat{\mu}_{2p}$ decreases for higher n_2, but n_1 hardly seems to have an impact. See for example the first entry of the second half of the table: the standard deviation of $\widehat{\mu}_{2p}$ is 11.9, 12.0 and 7.0 for (n_1, n_2) equal to (1000,100), (3000,100) and (3000,300), respectively.

Probabilities					$n_1 = 1000$ and $n_2 = 100$				$n_1 = 3000$ and $n_2 = 100$				$n_1 = 3000$ and $n_2 = 300$			
π_{11}	π_{10}	π_{01}	π_{00e}	π_{00u}	$\widehat{\mu}_2$	$\widehat{\mu}_{2r}$	$\widehat{\mu}_{2p}$	$\widehat{\mu}_{2w}$	$\widehat{\mu}_2$	$\widehat{\mu}_{2r}$	$\widehat{\mu}_{2p}$	$\widehat{\mu}_{2w}$	$\widehat{\mu}_2$	$\widehat{\mu}_{2r}$	$\widehat{\mu}_{2p}$	$\widehat{\mu}_{2w}$
Mean error size = 10																
.89	.02	.01	.06	.02	10.1	10.1	9.9	10.0	10.0	10.1	9.9	10.0	10.0	10.1	10.1	10.0
					(3.1)	(6.3)	(8.7)	(3.4)	(2.4)	(6.0)	(8.5)	(2.4)	(1.7)	(3.8)	(5.0)	(1.9)
.89	.06	.01	.02	.02	10.0	10.0	10.0	9.9	10.0	10.2	10.1	10.0	10.0	10.2	10.1	10.0
					(3.8)	(7.9)	(9.1)	(4.5)	(3.5)	(8.2)	(9.1)	(3.4)	(2.3)	(4.9)	(5.0)	(2.6)
.87	.02	.03	.06	.02	10.0	10.2	9.9	10.0	10.0	10.3	9.9	10.0	10.0	10.1	9.9	10.0
					(3.1)	(6.8)	(8.7)	(3.4)	(2.7)	(6.8)	(8.8)	(2.6)	(1.8)	(4.2)	(5.0)	(1.9)
.87	.06	.03	.02	.02	10.1	10.3	10.0	10.0	10.1	10.3	10.0	10.0	10.0	10.1	9.9	10.0
					(3.8)	(8.9)	(8.8)	(4.2)	(3.5)	(8.6)	(8.7)	(3.3)	(2.4)	(5.4)	(5.0)	(2.4)
.85	.02	.05	.06	.02	10.1	10.4	10.1	10.0	10.0	10.3	10.0	10.0	10.0	10.2	10.0	10.0
					(3.3)	(7.7)	(9.0)	(3.4)	(2.8)	(7.6)	(8.8)	(2.7)	(1.8)	(4.7)	(5.1)	(1.9)
.85	.06	.05	.02	.02	10.0	10.3	10.0	10.0	10.0	10.4	10.1	10.0	10.0	10.4	10.1	10.0
					(3.8)	(9.4)	(8.8)	(4.0)	(3.6)	(9.5)	(9.0)	(3.3)	(2.4)	(5.8)	(5.1)	(2.3)
Mean error size = 20																
.78	.04	.02	.12	.04	20.0	20.2	19.8	20.0	20.1	20.2	19.9	20.1	20.0	20.1	19.9	20.0
					(4.1)	(8.8)	(11.9)	(4.7)	(3.3)	(8.5)	(12.0)	(3.3)	(2.5)	(5.3)	(7.0)	(2.7)
.78	.12	.02	.04	.04	20.0	20.3	20.1	20.0	20.0	20.3	20.0	20.0	20.1	20.2	20.1	20.0
					(5.4)	(11.5)	(12.2)	(6.3)	(5.1)	(11.4)	(12.4)	(4.7)	(3.5)	(7.0)	(7.2)	(3.6)
.74	.04	.06	.12	.04	19.9	20.2	19.9	20.0	20.0	20.4	20.1	20.0	20.0	20.3	20.0	20.0
					(4.2)	(9.9)	(12.2)	(4.6)	(3.6)	(10.1)	(12.5)	(3.6)	(2.6)	(6.1)	(6.9)	(2.7)
.74	.12	.06	.04	.04	20.0	20.4	20.0	20.0	20.0	20.5	19.9	20.0	20.0	20.3	20.0	20.0
					(5.5)	(12.5)	(12.3)	(5.9)	(5.2)	(12.3)	(12.5)	(4.6)	(3.6)	(7.6)	(7.2)	(3.3)
.70	.04	.06	.12	.04	20.0	20.6	20.1	20.0	20.0	20.4	19.9	20.0	20.0	20.4	20.0	20.0
					(4.4)	(10.9)	(12.6)	(4.7)	(3.8)	(10.7)	(12.2)	(3.7)	(2.6)	(6.8)	(7.1)	(2.7)
.74	.12	.06	.04	.04	20.0	20.5	19.9	20.0	20.1	20.9	20.2	20.0	20.0	20.4	20.0	20.0
					(5.5)	(13.7)	(12.3)	(5.6)	(5.3)	(13.5)	(12.7)	(4.6)	(3.5)	(8.1)	(7.2)	(3.2)

Table 4.1: Simulated means (and standard deviations) of the estimators

We see that the variances of all estimators are lower for the small mean error size (10) than for the high mean error size (20). For example, for $n_1 = 1000$ and $n_2 = 100$ the standard deviation of $\widehat{\mu}_2$ is 3.1 for the first set of probability parameters with $\mu_0 - \mu_2 = 10$; for the first set of parameter values with $\mu_0 - \mu_2 = 20$ the standard deviation is 4.1.

In every second line of the table the probability of an auditor missing an error is higher, and the probability of an auditor finding the right size of an error is lower than in the previous line. Comparing two subsequent lines, we see that a higher π_{10} and a lower π_{00e} cause an increase in the variance of the estimators. For example, in the first two lines of the table the standard deviation of $\widehat{\mu}_2$ increases from 3.1 to 3.8 for $n_1 = 1000$ and $n_2 = 100$.

Based on the results of Table 4.1, we can conclude that estimators $\widehat{\mu}_2$ and $\widehat{\mu}_{2w}$ have comparable variances and outperform $\widehat{\mu}_{2r}$ and $\widehat{\mu}_{2p}$ (in terms of variance). The simulations in this section were constructed such that $E\{A_{t0} - A_{t1} | A_{t0} \neq A_{t1}\} = E\{A_{t0} - A_{t2} | A_{t0} \neq A_{t2}\}$, which is a necessary condition for consistency of $\widehat{\mu}_{2w}$. This is not an essential condition for the consistency of $\widehat{\mu}_2$. Moreover, $\widehat{\mu}_{2w}$ does not outperform $\widehat{\mu}_2$ even under this condition and with a model for the simulated data which deviates from our model in Section 2. Hence, our new estimator $\widehat{\mu}_2$ seems to be the preferable estimator.

5 Final remarks and conclusions

We introduced a mixed model for a repeated audit control with two rounds. This model consists of a model for the absolute classification frequencies and submodels in terms of conditional regression for the audit values. As main results we derived MLE's for the model parameters and based on these we constructed a new estimator for the mean size of the errors. This last estimator was shown to outperform the estimators of Barnett *et al.* (2001), although the underlying model of the simulation study differed from our model in Section 2.

The generalization to a repeated audit control with k rounds ($k-1$ fallible auditors and the final infallible expert) is quite straightforward. The basic variables of the general model are $A_0, A_1, \ldots, A_k$, where A_i $(i = 1, \ldots, k)$ is the value according to auditor i of a random record. The records can be classified based on the question whether some of the k audit values and book values coincide; note that the number of classifications increases sharply in k. Next, similar to Section 2.3, conditional regression models can be specified for the audit values which do not coincide with the book value or previous audit values according to

the classification.

As mentioned previously, repeated audit controls can be regarded as a missing data problem (or more specific: as a monotone missing data problem). In the missing data literature, Olkin and Tate (1961) have already introduced a model with a mixture of both categorical and continuous variables: the general location model. In this model, K categorical variables are classified, and the M continuous variables have a (M-variate) normal distribution conditional on this classification. The model in this chapter differs essentially from the general location model: the classifications are not based on separate categorical variables but on the equality of the continuous variables, and the dimensionality of the conditional models may be lower than M. For example, the conditional regression models in Table 2.2 are uni- and bivariate.

In the model discussed the sample size in the second round. n_2 was assumed to be fixed. However, it is possible to define the number of sample elements (even per classification) as a function of the results of the first auditor. This does not alter the ML estimators (see Raats *et al*. (2004) Section 3.3 for a more detailed discussion).

So far we have only discussed point estimators for the parameters, but confidence limits are at least as important in auditing practice. In auditing practice, selection with probabilities proportional to the recorded value ('monetary unit sampling') is applied frequently instead of the discussed sampling techniques. It will be interesting to investigate this sampling method as well. We leave these topics for further research.

6 Appendices

6.1 Relative efficiency

6.1.1 Model and notation

In this appendix we study the relative efficiency in the general model for multivariate linear regression with monotone missing observations of the dependent variables. This model has been studied extensively in Raats *et al*. (2002b). In this section we only give the model and results as far as needed for the derivation of the the relative efficiency. The notation is taken from Raats *et al*. (2002b) and deviates at some points from the notation of this paper.

Consider the multivariate linear regression model with M dependent variables and k (deterministic) explanatory variables; observations are gathered for N cases.

Let $X_{tj} \in I\!R$ be the observed value of the j^{th} explanatory variable ($j = 1, \ldots, k$) for the t^{th} case; complete data are available for the explanatory variables, so $t = 1, \ldots, N$ for all j.

The observations of the dependent variables are incomplete; the dependent variables are ordered such that later added variables come last. So their data are divided into r ordered groups according to the pattern of increasingly missing data. Group i contains m_i variables for which exactly the first N_i observations are available:

$$N = N_1 \geq N_2 \geq \ldots \geq N_r; \quad M_i = \sum_{j=1}^{i} m_j \quad (i = 1, \ldots, r,\ M_r = M).$$

The vector $Y_{ti} \in I\!R^{m_i}$ contains the values of these m_i dependent variables for case t. So Y_{ti} is observable for $t = 1, \ldots, N_i$ and missing for $t = N_i + 1, \ldots, N$. The special case $N = N_1 = \ldots = N_r$ gives the usual complete model.

The r (multivariate) regression equations can be written as

$$Y_{ti} = \mu_{ti} + \varepsilon_{ti}, \quad \mu_{ti} = \sum_{j=1}^{k} X_{tj}\beta_{ji}, \quad i = 1, \ldots, r, \quad t = 1, \ldots, N_i, \quad (6.1)$$

where $\beta_{ji} \in I\!R^{m_i}$ denotes a vector of unknown regression coefficients. For the errors we assume

$$E\{\varepsilon_{ti}\} = 0, \quad Cov(\varepsilon_{ti}, \varepsilon_{sj}) = \delta_{ts}\sigma_{ij}, \quad (6.2)$$

with (completely unknown) non-singular $\Sigma = (\sigma_{ij}) \in I\!R^{M \times M}$ not depending on the β_{ji}. We write $\Sigma > 0$ for positive definiteness. If normality of the errors is assumed, it will be mentioned explicitly.

The union of the groups 1 up to i will be denoted by (i), hence $Y_{t(i)} = (Y'_{t1} \ldots Y'_{ti})' \in I\!R^{M_i}$, $i = 1, \ldots, r$ and similarly for $\mu_{t(i)}$ and $\varepsilon_{t(i)}$. The error covariance matrix $\Sigma_{(i)(i)}$ of $\varepsilon_{t(i)}$ can be partitioned as follows

$$\Sigma_{(i)(i)} := Cov(\varepsilon_{t(i)}) = Cov \begin{pmatrix} \varepsilon_{t(i-1)} \\ \varepsilon_{ti} \end{pmatrix} = \begin{bmatrix} \Sigma_{(i-1)(i-1)} & \Sigma_{(i-1)i} \\ \Sigma_{i(i-1)} & \Sigma_{ii} \end{bmatrix}. \quad (6.3)$$

So, $\Sigma_{(i)(i)} \in I\!R^{M_i \times M_i}$, $\Sigma_{(i-1)(i-1)} \in I\!R^{M_{i-1} \times M_{i-1}}$, $\Sigma_{(i-1)i} \in I\!R^{M_{i-1} \times m_i}$ and in particular $\Sigma_{(r)(r)} = \Sigma$ and $\Sigma_{(1)(1)} = \Sigma_{11}$.

6.1.2 Notation

We introduce some column- and matrix-notation for the observed variables and regression coefficients. The index i refers to group i and (i) again to the union of the groups $1, 2, \ldots, i$.

$$X = \begin{bmatrix} \boxed{\begin{matrix} X_{1,1} & X_{1,2} & \cdots & X_{1,k} \\ \vdots & \vdots & & \vdots \\ X_{N_i,1} & X_{N_i,2} & \cdots & X_{N_i,k} \end{matrix}} \\ \begin{matrix} \vdots & \vdots & & \vdots \\ X_{N,1} & X_{N,2} & \cdots & X_{N,k} \end{matrix} \end{bmatrix}$$

$$\begin{bmatrix} \boxed{X_i} \end{bmatrix}$$

$$\beta = \begin{bmatrix} \boxed{\begin{matrix} \beta'_{1,1} & \cdots & \beta'_{1,i-1} \\ \vdots & & \vdots \\ \beta'_{j,1} & \cdots & \beta'_{j,i-1} \\ \vdots & & \vdots \\ \beta'_{k,1} & \cdots & \beta'_{k,i-1} \end{matrix}} & \begin{matrix} \beta'_{1,i} & \cdots & \beta'_{1,r} \\ \vdots & & \vdots \\ \beta'_{j,i} & \cdots & \beta'_{j,r} \\ \vdots & & \vdots \\ \beta'_{k,i} & \cdots & \beta'_{k,r} \end{matrix} \end{bmatrix}$$

$$\begin{bmatrix} \boxed{\beta_{(i-1)}} & \beta_i & \cdots & \beta_r \end{bmatrix}$$

So $X_i \in I\!R^{N_i \times k}$ is the matrix with the first N_i observations of all explanatory variables. The submatrices $\beta_{(i-1)} \in I\!R^{k \times M_{i-1}}$ and $\beta_i \in I\!R^{k \times m_i}$ of $\beta \in I\!R^{k \times M}$ contain the regression coefficients corresponding to groups $(i-1)$ and i of dependent variables, respectively. The Y_{ti} can be grouped in a corresponding way:

$$\begin{bmatrix} \boxed{\begin{matrix} Y'_{1,1} & \cdots & Y'_{1,i-1} \\ \vdots & & \vdots \\ \vdots & & \vdots \\ Y'_{N_i,1} & \cdots & Y'_{N_i,i-1} \end{matrix}} & \begin{matrix} Y'_{1,i} & \cdots & Y'_{1,r} \\ \vdots & \cdots & \vdots \\ \vdots & \cdots & Y'_{N_r,r} \\ Y'_{N_i,i} & & \end{matrix} \\ \begin{matrix} \vdots & \cdots & Y'_{N_{i-1},i-1} \\ Y'_{N,1} & & \end{matrix} & \end{bmatrix}$$

$$\begin{bmatrix} \boxed{Y_{(i-1)}} & Y_i & \cdots & Y_r \end{bmatrix}$$

The matrix $Y_i \in I\!R^{N_i \times m_i}$ contains all observations of group i. But the matrix $Y_{(i-1)} \in I\!R^{N_i \times M_{i-1}}$ contains *only* the first N_i observations of the foregoing groups $(i-1)$ (with $Y_{(0)} = 0$). We use similar definitions for the μ_{ti} and ε_{ti}.

6.1.3 Estimators and relative efficiency

The OLS estimators for the regression coefficients β_i are

$$b_i = G_i X_i' Y_i \quad \text{with } G_i = (X_i' X_i)^-, \tag{6.4}$$

where a g-inverse is denoted by $^-$ (see Raats *et al*. (2002b) equation (3.5)). The GLS estimator for β_i are

$$\breve{\beta}_i = G_i X_i' (Y_i - \breve{\zeta}_i) \tag{6.5}$$

(see Raats *et al*. (2002b) equation (3.15)).

We compare the performance of the discussed LS estimators by means of the relative efficiency of the estimators for the regression coefficients under the normality assumption. The relative efficiency of estimator $\widehat{\theta}_1$ in relation to estimator $\widehat{\theta}_2$ can be expressed as the determinant of the following function of the M(ean) S(quared) E(rror)s:

$$MSE(\widehat{\theta}_1)^{-\frac{1}{2}} MSE(\widehat{\theta}_2) MSE(\widehat{\theta}_1)^{-\frac{1}{2}}, \tag{6.6}$$

other possibilities are the maximum eigenvalue or the trace.

Throughout this section we assume without loss of generality that $m_i = 1$ for all i. In case of normality all LS estimators for the regression coefficients are unbiased and their MSE's coincide with their variances. The variance of OLS estimator b_i follows directly from its definition in (6.4):

$$Var\{b_i\} = \sigma_{ii} (X_i' X_i)^{-1}. \tag{6.7}$$

The variance of the GLS estimator $\breve{\beta}_i$ is more complicated.

Theorem 6.1. *For $i = 2, \ldots, r$,*

$$Var\{\breve{\beta}_i\} = Var\{\breve{\beta}_{(i-1)} \alpha_i\} + (X_i' X_i)^{-1} X_i' \Gamma_{ii} X_i (X_i' X_i)^{-1}. \tag{6.8}$$

Proof. We determine the variance by the relation

$$Var\{\breve{\beta}_i\} = Var\{E\{\breve{\beta}_i|Y_{(i-1)}\}\} + E\{Var\{\breve{\beta}_i|Y_{(i-1)}\}\}.$$

For the variance of the conditional expectation we have

$$\begin{aligned} Var\{E\{\breve{\beta}_i|Y_{(i-1)}\}\} &\stackrel{1}{=} Var\{\beta_i + (X_i'X_i)^{-1}X_i'(\varepsilon_{(i-1)} - \breve{\varepsilon}_{(i-1)})\alpha_i\} \\ &\stackrel{2}{=} Var\{(X_i'X_i)^{-1}X_i'(X_i\breve{\beta}_{(i-1)} - X_i\beta_{(i-1)})\alpha_i\} \\ &\stackrel{3}{=} Var\{\breve{\beta}_{(i-1)}\alpha_i\}. \end{aligned}$$

The first equality follows from (6.5) and $E\{Y_i|Y_{(i-1)}\} = X_i\beta_i + \varepsilon_{(i-1)}\alpha_i$; the second from $\varepsilon_{(i-1)} - \breve{\varepsilon}_{(i-1)} = X_i\breve{\beta}_{(i-1)} - X_i\beta_{(i-1)}$ and $Var\{\beta_i\} = 0$. Rewriting and $Var\{\beta_{(i-1)}\} = 0$ gives the last equality.

For the conditional variance we have

$$\begin{aligned} Var\{\breve{\beta}_i|Y_{(i-1)}\} &= Var\{(X_i'X_i)^{-1}X_i'Y_i|Y_{(i-1)}\} \\ &= (X_i'X_i)^{-1}X_i'\Gamma_{ii}X_i(X_i'X_i)^{-1}, \end{aligned}$$

where the first equality follows from (6.5) and $Var\{\breve{\varepsilon}_{(i-1)}\alpha_i|Y_{(i-1)}\} = 0$; the second one from $Var\{Y_i|Y_{(i-1)}\} = \Gamma_{ii}$ (see Raats *et al.* (2002b) equation (3.9)). □

Corollary. *If* $M_2 = 2$*, then*

$$Var(\breve{\beta}_2) = \rho_{12}^2\sigma_{22}(X_1'X_1)^{-1} + (1-\rho_{12}^2)\sigma_{22}(X_2'X_2)^{-1}, \tag{6.9}$$

where $\rho_{12} = \dfrac{\sigma_{12}}{\sqrt{\sigma_{11}\sigma_{22}}}$.

This corollary follows from Theorem 6.1, $\breve{\beta}_1 = b_1$ and (6.7).

We look into more detail at the relative efficiency for the frequently occurring situation $M_2 = 2$. Substituting (6.7) and (6.9) into (6.6) gives the relative efficiency of b_2 in relation to $\breve{\beta}_2$

$$(1-\rho_{12}^2) + \rho_{12}^2(X_2'X_2)^{\frac{1}{2}}(X_1'X_1)^{-1}(X_2'X_2)^{\frac{1}{2}}. \tag{6.10}$$

It is clear that (6.10) is always smaller (or equal) to one, *i.e.* $\breve{\beta}_2$ always outperforms b_2 in terms of variance (as can be expected). GLS is relatively more efficient for high values of ρ_{12} and small $(X_2'X_2)(X_1'X_1)^{-1}$; the latter usually corresponds

with a high fraction of missing observations, *i.e.* N_2/N_1 is small. This seems to be quite a logical result: GLS makes use of the sample information of preceding dependent variables in contrast to OLS. If there is relatively a lot of additional information available (*i.e.* n_1/N_2) is high) and the preceding dependent variable is highly correlated with the current one, the additional information concerning the preceding dependent variable will result in more accurate estimates. Figure 6.1 plots the relative efficiency of b_2 in relation to $\breve{\beta}_2$ as function of ρ_{12} for several combinations of N_1/N_2 (under the assumption that $(X_2'X_2)(X_1'X_1)^{-1}$ is equivalent to N_2/N_1).

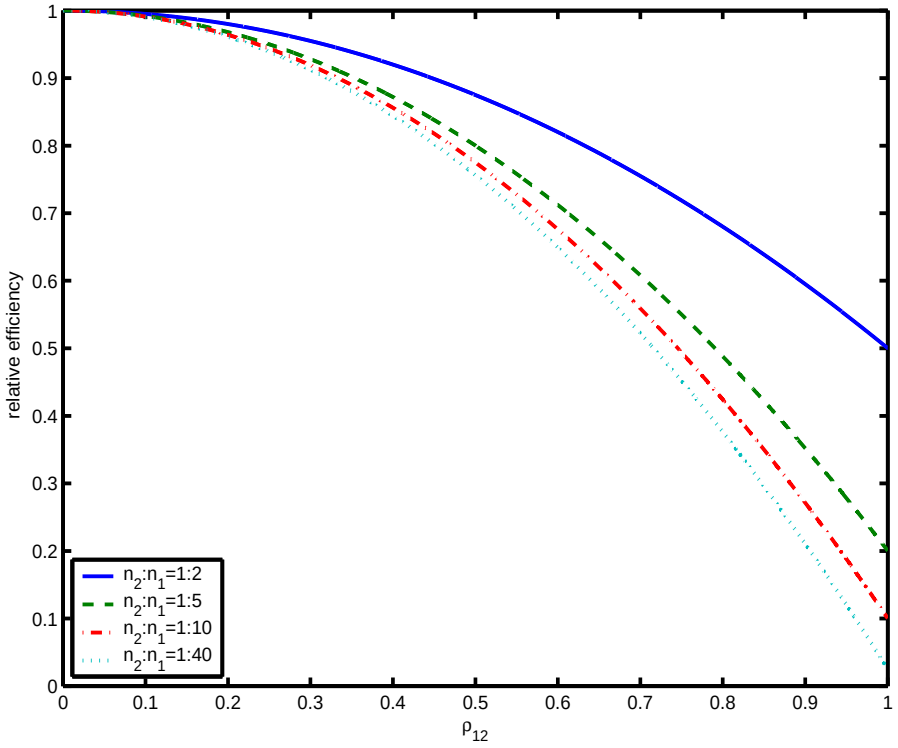


Figure 6.1: Relative efficiency of b_2 in relation to $\tilde{\beta}_2$

It is quite hard to derive a closed form expression for $Var\{\widehat{\beta}_i\}$. However, (6.8) will give a good approximation for large sample sizes since EGLS is asymptotically equivalent to GLS.

6.2 Conditional expectations

$$E\{A_{t2}|A_{t0}, A_{t1}, A_{t0} = A_{t1}, \widehat{\theta}\} = \widehat{\pi}_{1|1}A_{t0} + \widehat{\pi}_{0|1}\widehat{\beta}'_0 \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix}$$

$$E\{A_{t2}|A_{t0}, A_{t1}, A_{t0} \neq A_{t1}, \widehat{\theta}\} = \widehat{\pi}_{1|0}A_{t0} + \widehat{\pi}_{0e|0}A_{t1} + \widehat{\pi}_{0u|0}(\widehat{\beta}'_{0u} \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \widehat{\alpha}_{0u}\widehat{\varepsilon}_{t1})$$

$$E\{A_{t2}|A_{t0}, \widehat{\theta}\} = (\widehat{\pi}_{11} + \widehat{\pi}_{01})A_{t0} + \widehat{\pi}_{10}\widehat{\beta}'_1 \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \widehat{\pi}_{00e}\widehat{\beta}'_0 \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix} + \widehat{\pi}_{00u}\widehat{\beta}'_{0u} \begin{bmatrix} 1 \\ A_{t0} \end{bmatrix}$$

References

Barnett, V., J. Haworth, and T.M.F. Smith (2001). A Two-Phase Sampling Scheme with Applications to Auditing or Sed Quis Custodiet Ipsos Custodes. *Journal of the Royal Statistical Society A*, **164**, 407–422.

Cox, D.R. and E.J. Snell (1979). On Sampling and the Estimation of Rare Errors. *Biometrika*, **66**, 125–132.

Fienberg, S.E., J. Neter, and R.A. Leitch (1977). Estimating the Total Overstatement Error in Accounting Populations. *Journal of the American Statistical Association*, **72**, 295–302.

Johnson, J.R., R.A. Leitch, and J. Neter (1981). Characteristics of Errors in Accounts Receivable and Inventory Audits. *The Accounting Review*, **56**, 270–293.

Laws, D.J. and A. O'Hagan (2000). Bayesian Inference for Rare Errors in Populations with Unequal Unit Sizes. *Applied Statistics*, **49**, 577–590.

Moors, J.J.A. (1983). Bayes' Estimation in Sampling for Auditing. *Statistician*, **32**, 281–288.

Moors, J.J.A., B.B. van der Genugten, and L.W.G. Strijbosch (2000). Repeated Audit Controls. *Statistica Neerlandica*, **54**, 3–13.

Moors, J.J.A. and M.J.B.T Janssens (1989). Exact Distributions of Bayesian Cox-Snell Bounds in Auditing. *Journal of Accounting Research*, **27**, 135–144.

Neter, J., J.R. Johnson, and R.A. Leitch (1985). Characteristic of Dollar-Unit Taints and Error Rates in Accounts Receivable and Inventory. *The Accounting Review*, **60**, 488–499.

Olkin, I. and R.F. Tate (1961). Multivariate Correlation Models with Mixed Discrete and Continuous Variables. *Annals of Mathematical Statistics*, **32**, 448–465.

Raats, V.M., B.B. van der Genugten, and J.J.A. Moors (2002a). A General Model for Repeated Audit Controls Using Monotone Subsampling. CentER Discussion Paper 10, Tilburg University.

Raats, V.M., B.B. van der Genugten, and J.J.A. Moors (2002b). Multivariate Regression with Monotone Missing Observations of the Dependent Variables. CentER Discussion Paper 63, Tilburg University.

Raats, V.M., B.B. van der Genugten, and J.J.A. Moors (2004). A General Model for Repeated Audit Controls Using Monotone Subsampling. *Communications in Statistics: Theory and Methods*, **33**, 949–977.

Raats, V.M. and J.J.A. Moors (2003). Double Checking Auditors: a Bayesian Approach. *Statistician*, **52**, 1–15.

Tamura, H. (1988). Estimation of Rare Errors Using Expert Judgement. *Biometrika*, **75**, 1–9.

Tamura, H. and P.A. Frost (1986). Tightening CAV (DUS) Bounds by Using a Parametric Model. *Journal of Accounting Research*, **24**, 364–371.

Tenenbein, A. (1970). A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications. *Journal of the American Statistical Association*, **65**, 1350–1361.

Tenenbein, A. (1971). A Double Sampling Scheme for Estimating from Binomial Data with Misclassifications; Sample Size Determination. *Biometrics*, **27**, 935–944.

Tenenbein, A. (1972). A Double Sampling Scheme for Estimating from Inspection. *Technometrics*, **14**, 187–202.