A hierarchical Bayes model for multilocation auditing

David J. Laws and Anthony O’Hagan
University of Sheffield, UK

[Received September 2001. Revised June 2002]

Summary. The paper provides a Bayesian analysis of a practical problem in auditing in which substantial prior information needs to be combined with limited sample data. The specific context of the paper is a multilocation audit in which auditors take a two-stage sample of transactions from different sites within an organization. The hierarchical model that we propose has the flexibility to deal with this sort of stratified population in which some strata are not sampled and in which strata are not exchangeable. A careful, thorough elicitation of auditors’ prior information is another key feature of our approach. Inference from our model is by direct Monte Carlo simulation. The example data set is given in the context of an audit carried out by the UK’s National Audit Office who funded this research.

Keywords: Dirichlet process; Elicitation; Multilocation inference; Nonparametrics; Rejection sampling; Simulation; Statistical auditing; Stratified populations

1. Introduction

1.1. The audit problem
The financial audit problem is a decision-making problem in which the auditors must choose between giving an unqualified opinion that the financial statements produced by an organization are ‘true and fair’, and giving a qualified opinion on the statements, indicating that they do not show a ‘true and fair’ view. If relevant evidence is not made available, the auditors may qualify their opinion through a scope limitation, stating that sufficient evidence could not be obtained.

The auditors will qualify their opinion if they find evidence that material error is present in the financial statements. Declaring an error as material is an expression of the relative significance or importance of a particular matter in the context of the financial statements as a whole. Errors can be material by virtue of their nature (e.g. if transactions were carried out illegally), context (e.g. if errors change a surplus into a deficit) or value. There is material error by value if the total error in the financial statements is larger than would be acceptable to ‘addressees of the auditors’ report’. These include shareholders for the audit of a company and Parliament for the audit of organizations that are funded by central Government. Auditors must use their judgment in deciding what is acceptable. Within this paper we are concerned with helping the auditor to judge whether the evidence collected indicates that there is material error by value in a financial statement.

Typically the auditors do not expect to find errors that will lead them to qualify the account, so they usually look to obtain sufficient audit evidence that material error is not present. This
evidence can include the results of audit tests on a small sample of transactions within the financial statements. These transactions will be examined carefully and the auditor will determine whether there is an error in the recorded value of each transaction and the monetary value of any errors. Then the errors in this sample may be extrapolated to obtain a point estimate of the total error in the financial statements.

However, the information in the audit sample of transactions is typically poor. In most audits, errors are rare. Also, the cost of sampling is such that sample sizes are necessarily small. Error rates of 1% or 2%, coupled with sample sizes of under 200, are common. With very few observed errors, there is little information concerning the frequency of errors, and still less useful data about the magnitudes of errors. Therefore the standard error of any point estimator of total error based on this sample will be very large.

Nevertheless, the auditor can be satisfied with a point estimate based on such a small amount of sample information because he has available other sources of evidence about the likely errors in transactions. As well as the auditors’ own experience and knowledge concerning likely transaction errors in this and similar organizations, other available relevant evidence may include

(a) information from previous audits of the same organization,
(b) an assessment of the quality of management, accounting and internal audit procedures,
(c) analytical reviews of accounting ratios, trends in such ratios and resulting investigations of unusual fluctuations and
(d) information about the environment within which the organization works—accounting errors are less prevalent in regulated industries, such as banking, as shown in Maletta and Wright (1996).

Hence, the auditors’ judgment of whether there is a material error by value is not solely based on the point estimate of total error but also on this relatively large amount of other information. This need to combine sparse data and genuine prior information suggests that a Bayesian analysis would be beneficial. Currently, the uncertainty about the amount of error is implicitly but informally taken into consideration when the auditor judges whether the sample and non-sample information together indicate that no material error is present. A Bayesian method would formalize this process. Prior (other) information can be combined with sample information to produce not just a point estimate of the total error but also information about the true uncertainty (not just sampling variability) about the total error. Published Bayesian approaches in auditing go back to Cox and Snell (1979). Steele (1992) provided a guide to using Bayesian statistics in an auditing context, and Johnstone (1995) and Edwards (1995) provided compelling arguments that inference in auditing is more efficient and reliable when carried out in an evidential Bayesian framework.

1.2. A multilocation audit

The UK’s National Audit Office (NAO) is responsible to the UK Parliament for reporting on whether public money has been spent as Parliament intended. Therefore, they audit the vast majority of Government-funded activities and organizations. They commissioned the research that is reported here to tackle the particular problem of a multilocation audit. Large organizations typically have accounts that are spread over many locations and auditors can, in practice, only visit a few locations. Therefore they take a two-stage sample, i.e. a sample of locations is visited, and a sample of transactions is examined in each sampled location. Hence the sample data will be made up of the number and size of errors in the expenditure that is associated
with the transactions sampled in each location that is visited. In addition to the sample data on
erors, the auditor has access to the books of the company, and hence the whole set of recorded
expenditures (termed book values) for every transaction. Since errors are not independent of
book values, in that more expenditure usually results in larger error, it will be necessary to take
account of this known covariate.

An example data set in the context of a Government audit follows. For client confidentiality
we do not describe any details of the body being audited and, to avoid identification, the
numerical audit information has been altered. Nevertheless, all the relevant features of the
data for our analysis have been retained.

In the audited body there are 30 locations or offices, which we have named A, B, . . . , Z, AA,
BB, CC and DD. In our data set, every location has 5850 recorded transactions with a mean book
value of £47. The variance of the book values in selected locations is given in Table 1 in Section 6.
During the audit, the audit team visited locations I, P, Q and T. In each case they examined a sam-
ple of 25 transactions, with book values totalling £2000, £1300, £1300 and £13000 respectively.
Only one error was found, in location T. This transaction had a book value of £100 and the audi-
tors believed that the value should have been £110. Therefore the error was an understatement of
– £10.

Notice how little information has been gained from the sample. The auditors will have learned
a reasonable amount about the likely error rate having observed only one error in a total sam-
ple of 100. However, there is very little information about the size of errors. Hence, it is clear
that if other substantial information about the likelihood and size of errors is available then,
in this example, the auditor’s inferences should and would be strongly influenced by this other
information. This is consistent with the usual role of prior information in an audit. Sample infor-
mation is typically used as confirmation of what is believed on the basis of relevant, non-sample
information. We return to this data set in Section 6.

A potentially important part of the auditor’s prior knowledge relates to knowledge of indi-
vidual circumstances of certain locations. Some locations may have been visited in a previous
audit; others may be believed to be more or less prone to errors for reasons such as a recent
change of management or computer system. It is therefore often inappropriate to assume either
equal or exchangeable error rates across locations.

Therefore the challenge that is faced in this paper is to provide a method of inference for
the error in an account balance that allows for the realistic expression of auditors’ prior know-
ledge, and the integration of sample data with that knowledge, in the context of a multilocation
audit.

1.3. Organization of the paper
In Section 2 we set out the auditing problem, discuss other methodology that the NAO might use
and previous work on analysing single-location and multilocation audit data. Sections 3 and 4
present our model, and the simulation technique for computing posterior inference is developed
in Section 5. Our method is compared with a purely data-based method in the context of the
above audit data in Section 6, with a particular description of the procedure for eliciting the
auditor’s prior knowledge. A final Section 7 briefly discusses the advantages of our approach,
the wider applicability of this work and directions for future research.

2. Auditing methodology
2.1. Set-up and notation
Auditors refer to the recorded expenditure of a transaction as the book value and the ‘true’
expenditure of a transaction, as would be determined by an auditor, as the **audit value**. We denote by $E_{ik}$ the book and audit values of the $k$th transaction in location $i$. The error, denoted by $E_{ik}$, is therefore $E_{ik} = x_{ik} - Y_{ik}$. Without loss of generality, our notation supposes that the auditors visit locations $\{1, \ldots, l\}$ from a total of $L$ locations and examine transactions $\{1, \ldots, n_i\}$ from a total of $N_i$ in location $i$. Book values of all transactions are known but the audit values $Y_{ik} = y_{ik}$, and hence errors $e_{ik} = x_{ik} - y_{ik}$, are only known for the sample. Depending on circumstances, the NAO would like to infer about some or all of the total overstatement error, understatement error, net error and absolute error. For the purposes of this paper we focus on the total net error

$$E = \sum_{i=1}^{L} \sum_{k=1}^{N_i} E_{ik}$$

and note that the method can be easily adjusted to make inferences about the others.

### 2.2. Possible non-Bayesian audit approaches

NAO auditors have always taken into account all the audit evidence that is available to them, both in terms of planning an audit and in arriving at audit conclusions. Particularly in the case of a multilocation audit, however, it was felt that Bayesian methods might be capable of providing auditors with a useful alternative structure for formally combining the different types of audit information that is available.

In the NAO’s earlier methodology, available sample-based evidence is extrapolated to give a point estimate of the total error. Two possible ratio estimators are

$\hat{E}_1 = \frac{x}{\sum_{i=1}^{l} x_i} \sum_{i=1}^{l} n_i \sum_{k=1}^{N_i} E_{ik}$,

$\hat{E}_2 = x \left(1 - \frac{\hat{Y}}{\bar{X}}\right)$,

where $\bar{X} = (L/l) \sum_{i=1}^{l} N_i \bar{X}_i$, $\bar{X}_i$ is the sample mean of book values from location $i$ and $\hat{Y}$ is defined analogously. In the above, the known total book value of transactions at location $i$ is denoted by $x_i = \sum_{k=1}^{N_i} x_{ik}$ and the total book value of all transactions in the account is denoted by $x = \sum_{i=1}^{L} x_i$. Some other estimators that might be used are contained in Panel on Nonstandard Mixtures of Distributions (1989). Note that locations could be stratified if appreciably different error distributions were expected in different locations.

In most audit situations the NAO would not explicitly consider an estimated standard error for the estimate of the total error and would therefore rarely compute confidence limits. The reason is that such statistics do not reflect the true uncertainty about the total error, which should take into account the large volume of other information that is available to the auditor. Accuracy of the error estimate is considered implicitly in the design of the audit sample and when using the total error estimate to facilitate decisions. To determine whether to qualify, the NAO would bring together sample-based information on individual transactions, such as is summarized by the error estimate, with other information, and the auditor would use his judgment to assess the relative importance of the different pieces of available evidence. These judgments must be documented and are subject to professional review. The Bayesian approach may simplify and facilitate such a review by offering a more formal and transparent use of the auditor’s judgment.
Other solely data-based methods for analysing multilocation audit data are examined in Kim et al. (1987). They simulated multilocation audit populations and examined the behaviour of standard sample survey methods based on normal approximations and a ‘global’ version of the Stringer bound. The Stringer bound is an upper bound for the total error that is often used in the single-location case when transactions are sampled with probability proportional to size. Kim et al. (1987) found that standard methods based on asymptotic normality assumptions result in upper confidence limits that are, on average, too low—which would lead to account balances being prematurely accepted as accurate—and the global Stringer bound results in limits that are, on average, too high—which would lead to unnecessary extra work to ensure that the account balance is correct. These findings are consistent with similar work in the single-location case; see Neter and Loebbecke (1977) and Reneau (1978).

2.3. Taints
In tackling the problem of inference for audit populations, we must account for the known covariate, the book value of individual transactions. In general, large errors will tend to be associated with larger book values. In common with previous researchers, e.g. Cox and Snell (1979), we define the fractional error or taint of a transaction as

\[ Z_{ik} = \frac{E_{ik}}{x_{ik}}. \]

Auditors will now typically assert that the probability that an individual taint \( Z_{ik} \) lies in any given range is the same as for any other taint \( Z_{ik'} \) in the same location, regardless of their respective book values \( x_{ik} \) and \( x_{ik'} \). From our Bayesian perspective, this means that taints are exchangeable within each location. Our statistical modelling is therefore formulated in terms of taints, with the initial objective of deriving posterior inference about the population of taints from the sample of observed taints. This may then be converted into inference about errors by using the relationship \( E_{ik} = Z_{ik} x_{ik} \).

A key feature of the audit context is the non-standard nature of the taint distribution. First, there will be very many zero taints, corresponding to transactions that are not in error. The taint distribution therefore has a large point mass at zero. There may be other point masses arising from particular types of error. For instance, a transaction that appears in the accounts but should not be there at all (termed a bogus transaction) results in a taint of 1, and this may have a non-zero point mass. Similarly, failure to deduct value-added tax (currently 17.5% of the net value in the UK) results in a taint of 0.149. The distribution for negative taints (where the book value understates the true audit value) may be quite different from that of positive taints (where the book values are overstatements). Neter et al. (1985) gave examples of taint distributions for a few audit populations.

2.4. Single location—previous work
Many models for single-location audit populations have been proposed over the past 30 or more years. Panel on Nonstandard Mixtures of Distributions (1989) gives a comprehensive review; more recent work includes that of Rohrbach (1993) and Helmers (2000).

Tamura (1988) supposed that the audit population is infinite and so approximated inference for the total error \( E \) by inference for \( E(E) = \mathbb{E}(Z) \), where \( \mathbb{E}(Z) \) is the mean taint. Tamura argued that it is difficult to justify the use of a parametric model for the distribution of taints and so used a Bayesian nonparametric model based on the Dirichlet process; Ferguson (1973). The Dirichlet process allows the auditor to specify the expected distribution of taints (in a parametric way if desired); updating is via the empirical distribution function of the observed taints.
The Laws and O’Hagan (2000) model for the taint distribution is also based on the Dirichlet process but an extra stage of modelling is added by introducing a classification of errors according to taint. Each of these error classes corresponds to either a continuous open interval on the real line or a point mass in the taint distribution. In Tamura (1988), inference for $E(E)$ is made via numerical inversion of a characteristic function and numerical integration. Laws and O’Hagan (2000) made inferences for $E$, rather than $E(E)$, by simulation. The unseen population of errors is repeatedly simulated from the model, the total error for each simulation is calculated and an empirical distribution of the total error $E$ is thus generated. Inferences for $E$ are then calculated from this empirical distribution. The simulation method also allows a widening of the scope of inferences. Any feature of the posterior distribution of $E$ can be estimated by the equivalent feature of the simulated distribution. Inferences can also be made about other features of the population that are simulated from the model, such as the number of errors, the rate of errors and the likely size of errors.

2.5. Our approach

2.5.1. Hierarchical prior

The challenge is to develop a sufficiently flexible model for the taint populations in different locations, to allow for the realistic expression of auditors’ prior knowledge. We develop a hierarchical structure for locations based on

(a) beta prior distributions for rates of error at each location,
(b) a beta hyperprior distribution for overall error rates among a class of ‘typical’ locations and
(c) a flexible parameterization to express prior belief about the ways in which ‘untypical’ locations differ from the typical.

2.5.2. Elicitation

This model is combined with the modelling for the single location described in Laws and O’Hagan (2000), allowing the auditor to express prior knowledge about overall error rates, about the division of errors into different classes and about the magnitudes of errors within classes. A correspondingly detailed procedure for eliciting prior information to specify the hyperparameters of the prior structure has been developed in association with the NAO. The elicitation of substantial prior information is an important topic in Bayesian methodology, yet references concerning elicitation for complex applications are sparse. See Kadane and Wolfson (1998) and O’Hagan (1998) for some recent work. We hope that the technique described here will be a useful contribution to this topic.

2.5.3. Computation

We compute posterior inferences by simulation—our model is organized to facilitate direct Monte Carlo sampling from the posterior distribution. The extremely high dimensionality of the posterior (since we have a parameter for every transaction in the account) is easily accommodated by this approach. The heart of the process is a method for drawing from

(a) the univariate, but intractable, distribution of underlying error rates for typical locations and
(b) the low dimensional distribution of rates at which errors fall into error classes.

Conditionally on these draws, the remainder of the simulation is based on drawing from standard distributions.
3. The taint distribution

3.1. Introduction

As in Laws and O’Hagan (2000) we divide the taint distribution into several classes or categories. One class holds those transactions that are not in error, whereas the other classes divide the range of error taints into point masses and continuous intervals on the real line. So we divide the model into three parts but introduce hierarchical priors to describe dependence between locations. Here the three parts of the model are:

(a) a hierarchical model that has a prior structure based on the beta distribution, for the rate at which errors occur, given in Section 3.2,
(b) a hierarchical model, with a Dirichlet distribution as a basis for the prior, describing the probability with which errors fall into the error classes, described in Section 3.3, and
(c) a model with a Dirichlet process prior for the taints within those error classes that are intervals on the real line, presented in Section 3.4.

Because the prior information that is provided by the auditor will typically have a major influence on the final inference statement, it is very important to model the auditor’s beliefs correctly. The auditor typically has different levels of uncertainty about error rates, error class probabilities and the distribution of taints within error classes. Our model is designed to allow these different levels of uncertainty to be fully expressed.

3.2. Error rate

Let $\theta_i$ be the probability that a transaction in location $i$ is in error. In a sample of $n_i$ transactions taken from location $i$, the number that are in error, denoted by $r_i$, is modelled by a binomial distribution with parameters $n_i$ and $\theta_i$, conditional on the usual exchangeability assumptions.

We give $\theta_i$ a conjugate beta prior with parameters $\omega_i g_i$ and $\theta_0 / (1 - g_i)$, where $\theta_0$ is some population parameter and $\theta_i$ and $\theta_j, i \neq j$, are assumed independent given $\theta_0$:

$$E(\theta_i|\theta_0) = g_i(\theta_0)$$

and

$$\text{var}(\theta_i|\theta_0) = E(\theta_i|\theta_0)\{1 - E(\theta_i|\theta_0)\}/(\omega_i + 1),$$

and so $\omega_i$ determines the variability of $\theta_i$ about $E(\theta_i|\theta_0)$. Throughout what follows, the parameter $\omega_i$ retains its interpretation as a relative precision parameter and we concern ourselves with the choice of $g_i(\theta_0)$.

Suppose that prior knowledge is such that the error rates at the different locations are exchangeable. We might then choose $E(\theta_i|\theta_0) = \theta_0$ for $i = 1, \ldots, L$, with $\omega_1 = \ldots = \omega_L = \omega_0$. Then, $\theta_0$ is the mean of the $\theta_i$s and $\omega_i$ determines the variability of the $\theta_i$s about $\theta_0$. Under this specification, the parameters of the beta prior for $\theta_i$ are $\omega_0 \theta_0$ and $\omega_0 (1 - \theta_0)$.

However, auditors usually have different knowledge, opinions and judgment concerning the error rates at different locations. To allow for this, we provide a parameterization that allows a shift of the mean of $\theta_i$ away from the average or ‘typical’ value $\theta_0$. Hence $\theta_0$ need no longer be the mean of each $\theta_i$ but still provides a mechanism for linking locations. We introduce a location parameter $\delta_i$ so that $E(\theta_i|\theta_0)$ is given by

$$g_i(\theta_0) = \frac{\theta_0 \delta_i}{1 + \theta_0 (\delta_i - 1)}.$$ 

A value of $\delta_i > 1$ results in a shift of $E(\theta_i|\theta_0)$ to above $\theta_0$, and a value of $\delta_i$ such that $0 < \delta_i < 1$
brings $\mathbb{E}(\theta_i|\theta_0)$ below $\theta_0$. Choosing $\delta_i = 1$ gives $\mathbb{E}(\theta_i|\theta_0) = \theta_0$. Therefore, those locations judged by auditors to have higher error rates will be given larger values for $\delta_i$. The variability of $\theta_i$ about its mean is still governed by $\omega_i$. It then follows that the beta prior for $\theta_i$ has parameters

$$
g_{1i}(\theta_0) = \frac{\omega_i\theta_0\delta_i}{\Gamma(\theta_0(\delta_i - 1))},
g_{2i}(\theta_0) = \frac{\omega_i(1 - \theta_0)}{\Gamma(\theta_0(\delta_i - 1))}. \tag{1}
$$

We describe those locations with $\delta_i = 1$ and hence $\mathbb{E}(\theta_i|\theta_0) = \theta_0$ as typical or base locations. Then the population parameter $\theta_0$ determines the mean error rate in typical locations, with values of $\delta_i \neq 1$ describing departures of $\mathbb{E}(\theta_i|\theta_0)$ from the typical average for atypical locations. To allow locations to learn about each other through $\theta_0$, the parameter $\theta_0$ is given a prior or ‘hyperprior’ distribution. We choose a beta distribution, with parameters denoted by $a_1$ and $a_2$, as the prior for $\theta_0$.

The auditors take a sample of transactions from each of a sample of $l$ locations. The vector of sample sizes taken is $n = (n_1, n_2, \ldots, n_l)^T$ and we denote the numbers of errors found in the samples by $r = (r_1, r_2, \ldots, r_l)^T$. Under the above model the marginal posterior density function of $\theta_0$ is easily shown to be

$$
p(\theta_0|n, r) \propto \theta_0^{a_1 - 1}(1 - \theta_0)^{a_2 - 1} \prod_{i=1}^l \frac{B\{g_{1i}(\theta_0) + r_i, g_{2i}(\theta_0) + n_i - r_i\}}{B\{g_{1i}(\theta_0), g_{2i}(\theta_0)\}}. \tag{2}
$$

where $B(x, y) = \Gamma(x) \Gamma(y)/\Gamma(x + y)$ and $\Gamma(\cdot)$ is the gamma function. Conditionally on $\theta_0$, the posterior of $\theta_i$ is a beta distribution with parameters $g_{1i}(\theta_0) + r_i$ and $g_{2i}(\theta_0) + n_i - r_i$, where $n_i = r_i = 0$ for those locations that are not visited in the audit. For those locations that were not part of the sample, the posterior distribution of $\theta_i|\theta_0$ is not directly updated by the data, but updating does occur, indirectly, through the influence of the data on the distribution of $\theta_0$.

Ultimately, we are interested in the errors in the unseen population. Conditionally on $\theta_i$, the number of errors $R_i$ among the unseen transactions at location $i$ has a binomial distribution with parameters $N_i - n_i$ and $\theta_i$.

### 3.3. Error classes

Section 3.2 models the point mass in the distribution of taints at zero by modelling the rate at which transactions are in error. This section and the next section model the taint distribution of erroneous transactions. An inspection of audit data and discussions with auditors have led us to split the taint distribution into classes. In populations of taints there are usually identifiable point masses at values other than zero. For example, in Section 6, we model a point mass at taint 1. Therefore, some of the classes correspond to point masses in the distribution and the others signify open intervals between the point masses. This section describes a hierarchical model for the probabilities that an erroneous transaction falls into each of the error classes.

Suppose that there are $p$ error classes, and, for location $i$, $\psi_{ij}$ is the probability that an error is of class $j, i = 1, 2, \ldots, L$ and $j = 1, 2, \ldots, p$, where $\psi_{ip} = 1 - \Sigma_{j=1}^{p-1} \psi_{ij}$. If the auditor discovers $r_i$ errors in the sample from location $i$ then, under the usual assumptions, $r_i = (r_{i1}, r_{i2}, \ldots, r_{ip})^T$, the numbers of errors in the $p$ error categories, has a multinomial distribution with parameters $r_i$ and $\psi_i = (\psi_{i1}, \psi_{i2}, \ldots, \psi_{ip})^T$.

We use a hierarchical prior structure for $\psi_i$ so that locations are allowed to have different $\psi_i$s but so that the $\psi_i$s are not independent. Auditors cannot usually identify different priors for
ψ_i at different locations. Hence we suppose that the ψ_i's are exchangeable and have a Dirichlet distribution with parameter vector γψ_0, where ψ_0 = (ψ_01, ψ_02, ..., ψ_0p)^T is the vector of mean probabilities and γ determines the prior variability of the ψ_i's about ψ_0. The vector of means ψ_0 is then given a prior distribution, a Dirichlet prior with parameters b_1, b_2, ..., b_p. Although exchangeability of the ψ_i's is assumed here and is often appropriate, the approach of introducing parameters to describe departures from exchangeability for the θ_i's in Section 3.2 can also be used.

Given observed numbers r = (r_1T, r_2T, ..., r_pT)T in the error classes in the l visited locations, the marginal posterior density of the population parameters ψ_0 is found to be

$$p(ψ_0|r) ∝ \prod_{j=1}^{P} b_j^{−1} \prod_{i=1}^{l} \frac{D(γψ_01 + r_{i1}, γψ_02 + r_{i2}, ..., γψ_0p + r_{ip})}{D(γψ_01, γψ_02, ..., γψ_0p)}$$

(3)

where D(x_1, x_2, ..., x_p) = \Pi_{j=1}^{P} Γ(x_j)/Γ(Σ_{j=1}^{P} x_j).

The posterior distribution of ψ_i conditional on ψ_0 is Dirichlet with parameters γψ_01 + r_{i1}, γψ_02 + r_{i2}, ..., γψ_0p + r_{ip}, where r_{i1} = r_{i2} = ... = r_{ip} = 0 for those locations that are not visited in the audit.

We are interested in the classification of the errors in the unseen population. Suppose that there are R_i errors among the unseen transactions at location i; then, conditionally on ψ_i, the distribution of the errors among the p classes, specified by R_i = (R_{i1}, R_{i2}, ..., R_{ip})T, has a multinomial distribution with parameters R_i and ψ_i.

### 3.4. Taints within error classes

In Section 3.3, we have provided a model for the number, and the probabilities, of transactions falling into a set of error classes. To complete our model of the taint distribution it remains to model the taints within the error classes that denote open intervals on the real line. For a single-location audit, Laws and O'Hagan (2000) used a Bayesian nonparametric model based on the Dirichlet process. We now deal with the problem of carrying over this modelling method to a multilocation scenario. One possibility would be to have a hierarchical structure that links locations in a similar way to the models of Sections 3.2 and 3.3, with location-specific functions serving the role of the δ_i-parameter in Section 3.2 in describing departures from the ‘typical’. However, introducing such a structure would unnecessarily complicate the model, as auditors are unlikely to be able to provide reliable prior elicitation concerning any differences in the taint distributions within an error class for different locations that carry out much the same transactions. In addition, since so few errors are observed in the sample, the data would not be adequate to provide any meaningful learning about such differences and inference would thus be dependent on rather arbitrary prior information. To permit pooling of information about the taint distribution in error classes, our method allows locations to be grouped by the auditors according to the likely taint distribution within the error classes. Locations that carry out very similar transactions and so have similar book values will also encounter similar taints and taints. We group such locations together. Locations within the same taint group are assumed to have the same distribution of taints within each error class. Groups are then assumed to be independent. It is simple to extend this principle so that there are different groups of locations for different error classes, but this is unlikely to be necessary.

Suppose that there are M taint groups. The worked example in Section 6 takes the special case of M = 1, where all locations are assumed to have the same distribution of taints within an error class. Once locations have been grouped, the model for the taints within taint group m,
$m = 1, 2, \ldots, M$, of error class $j$ is as in Laws and O’Hagan (2000). A brief outline is repeated here for clarity and convenience.

The prior distribution for the taints in taint group $m$ of error class $j$ is in two stages. Conditionally on their common cumulative distribution function (CDF) $F_{jm}$, the taints are independent and identically distributed with CDF $F_{jm}$. The CDF $F_{jm}$ has a Dirichlet process prior distribution with parameter $\pi_{jm} F_0jm$, where $F_0jm$ is a CDF which represents the prior expectation of $F_{jm}$, and where $\pi_{jm} > 0$ is a prior precision parameter. Suppose that $r_{jm}$ taints in group $m$ of error class $j$ are observed in the sample, and $F_{\delta jm}$ denotes the empirical CDF of the $r_{jm}$ observations. The posterior distribution of $F_{jm}$ is a Dirichlet process with parameter $\pi_{jm} F_0jm + r_{jm} F_{\delta jm}$. Since the taints in group $m$ of error class $j$ are independent given $F_{jm}$, the CDF of the posterior predictive distribution of an unseen taint in group $m$ of error class $j$ is

$$P(T \leq t | \text{data}) = E\{F_{jm}(t) | \text{data}\} = \frac{\pi_{jm} F_{0jm}(t) + r_{jm} F_{\delta jm}(t)}{\pi_{jm} + r_{jm}}.$$

4. The book value distribution

A book value is the expenditure involved in a transaction as recorded in the accounts of an organization. Owing to the increased use of computerized accounting packages, usually one can obtain computer files containing all the book values in the organization. Therefore, there is no need to devise a statistical model as the population is known and in the simulation method of the next section we can draw book values from the population of any given location. However, in many situations there are practical difficulties (such as problems of incompatibility between systems) in putting together a full record of the book values from every location. In addition, there are situations in which a full record of book values is not available but instead the auditor has only some simple summary statistics of the book value distribution in each location. So, as an alternative to sampling from the full population, we could sample from an approximating continuous distribution that acts as a proxy for the known distribution of the book values. We suggest approximating the known book value distribution in each location by a log-normal distribution with the same mean and variance. The log-normal distribution should reflect well any skewness in the book value distribution, and, as we require inference about the distribution of the sum of errors, inferences should be reasonably insensitive to other features of the book value distribution—e.g. the presence of multiple modes.

5. Simulation-based inference

The hierarchical model set-up in Section 3 is strongly suggestive of a simulation approach in which unconditional inferences can be calculated by repeatedly simulating the whole hierarchy, starting from the top, i.e. with the population hyperparameters. Our simulation method proceeds by simulating the errors in the whole population of transactions from the model, and then summing to give the total error $E$. Hence, we can draw a large sample $E^{(1)}, E^{(2)}, \ldots, E^{(\text{large})}$ of values of $E$ from its posterior distribution. Inferences are then calculated by using the simulated distribution to estimate features of the distribution of $E$. The algorithm for generating one observation $E^{(s)}$ of $E$ is as follows.

**Step 1:** for the numbers of errors,

(a) generate $\theta_0^{(s)}$ from the posterior distribution of the mean error rate in typical locations $\theta_0$ given in expression (2),
(b) generate $\theta_i^{(s)}$ from the beta posterior distribution of the error rate at a location given the mean error rate in typical locations, i.e. $\theta_i|\theta_0 = \theta_i^{(s)}$, for $i = 1, 2, \ldots, L$, and 
(c) generate $R_i^{(s)}$ from the distribution of the number of errors in the unseen population given the error rate. This has a binomial distribution with parameters $N_i - n_i$ and $\theta_i = \theta_i^{(s)}$, for $i = 1, 2, \ldots, L$.

Step 2: for the numbers of errors in each error class,

(a) generate $\psi_0^{(s)}$ from the posterior distribution of the population mean probabilities of falling into each error class $\psi_0$ given in expression (3),
(b) generate $\psi_i^{(s)}$ from the conditional Dirichlet posterior distribution of the probabilities of falling into each error class in a location given the population mean, i.e. $\psi_i|\psi_0 = \psi_0^{(s)}$, for $i = 1, 2, \ldots, L$,
(c) generate $R_i^{(s)}$ from the distribution of the number of errors falling into each error class at a location given the number of errors and the probabilities of falling into the classes. Under our model $R_i$ has a multinomial distribution with parameters $R_i^{(s)}$ and $\psi_i = \psi_i^{(s)}$, for $i = 1, 2, \ldots, L$.

Step 3: for taints within error classes, generate $R_{ij}^{(s)}$ taints from error class $j$ in location $i$, for all $j$ and all $i$. If the error class is a point mass then the value of the taint from that class is determined by the location of the point mass. For those error classes that correspond to intervals on the real line, taints are generated from the appropriate Dirichlet process model for the taint group to which the location belongs.

Step 4: for book values, sample $R_i$ book values without replacement from the population of unseen book values for location $i$ or, if they are not available, the approximating log-normal distribution of book values for location $i$, allocating them to the $R_i$ taints generated for that location. Repeat for all $i$.

Step 5: for errors,

(a) calculate all the simulated errors in the population via error = book value $\times$ taint,
(b) sum the value of the errors generated from each location and the observed sample errors to give the total error in that location, $E_i^{(s)}$ say, and
(c) compute $E^{(s)} = \sum_{i=1}^{L} E_i^{(s)}$.

This algorithm is repeated many times to produce simulated empirical posterior distributions of the $E_i$s and $E$. Note that our simulation method allows inferences to be made for other targets of inference, since the method also generates simulated values from the unconditional posterior distributions of $\theta_0$, the $\theta_i$s, $\psi_0$, the $\psi_i$s, the $R_i$s and the $R_i$s, as well as the distribution of taints and errors.

Methods for generating draws from the well-known parametric distributions that are used in our model are well documented (Ripley, 1987). A method for generating a series of draws from a Dirichlet process model is described in Laws and O’Hagan (2000). To sample from the posterior distributions of $\theta_0$ and $\psi_0$ we use the rejection sampling algorithm (Ripley (1987), page 60). It is usually straightforward to devise a version of this algorithm for sampling random vectors with a small number of components. Here we need to sample from $\theta_0$ which is a scalar and $\psi_0$ which has $p$ components, where $p = 3$ in the example of Section 6. We sample from the distribution of $\eta_0 = \log(\theta_0/(1 - \theta_0))$ and $\zeta_0 = (\zeta_{01}, \zeta_{02}, \ldots, \zeta_{0(p-1)})^T$, where $\zeta_{0j} = \log(\psi_{0j}/\psi_{0p})$, for $i = 1, 2, \ldots, p - 1$, using as the rejection envelope a (multivariate) Cauchy density function with the same mode and modal dispersion (matrix), both of which are determined numerically. We then obtain $\theta_0$ and $\psi_0$ by the appropriate inverse transformation. The transformations are a
familiar device to achieve approximate (multivariate) normality, and the Cauchy and heavy-tailed t-distributions are also commonly used in importance sampling algorithms.

6. Analysis of example data set

In this section we describe the use of our model and compare it with simple data-based estimators that the NAO have used in the past, in the context of the Government audit described in Section 1.2.

In the model for the taint distribution that is proposed in Section 3, we must determine

(a) the classification of errors, as proposed in Section 3.3 and
(b) the grouping of locations according to the taints that are likely to occur (Section 3.4).

After consultation with auditors and inspection of past audit data the following classification of errors was found to be appropriate to the audit in question. There are three classes \( p = 3 \), namely

(a) understatements, where \( Z_{ik} < 0 \), and so the book value is less than the audit value (class 1),
(b) overstatements, where \( Z_{ik} \in (0, 1) \), and so the book value is greater than the audit value, and the audit value is greater than 0 (class 2), and
(c) bogus transactions, defined in Section 2.3, where \( Z_{ik} = 1 \), and so the audit value is equal to 0.

A taint greater than 1 relates to a transaction that is written as a credit \( x_{ij} > 0 \) but should have been a debit \( y_{ik} < 0 \). This very rarely occurs in practice.

For the distributions of taints within error classes 1 and 2 we treat the locations as one homogeneous group and therefore choose \( M = 1 \) in Section 3.4. The motivation for this is that, since the locations provide very similar services, the taints are similar. Also, as errors are rare, little information is available in the sample about the distributions of taints within the error classes and the best use of the sample information is to combine it to update a global taint distribution for each error class.

Of the 30 locations, only locations O and Q were believed to be atypical as regards error rates. Both were expected to have fewer errors than the other locations.

6.1. Elicitation of prior parameters

In view of the major role that non-sample information plays in the audit process, the elicitation of the auditor’s expert prior opinion is extremely important. Here we use the context of this example to describe our elicitation procedure, in which the auditor is asked a series of questions about the various aspects of the model. We stress that we had many discussions with auditors to ensure that the questions asked are appropriate in that

(a) they match up with the way in which an auditor approaches an audit and
(b) the auditor is capable of answering them.

The auditor is asked to provide an upper bound for some unknown quantities and to specify at the outset the probability \( \alpha \) with which each upper bound is exceeded, with the default value of \( \alpha \) set at 0.05.

6.1.1. Error rate parameters

6.1.1.1. Average. Recall that uncertainty about the average error rate in typical locations, \( \theta_0 \), is described by a beta distribution with parameters \( a_1 \) and \( a_2 \). To elicit \( a_1 \) and \( a_2 \), the auditor is
asked to provide his or her ‘best’ estimate of the average error rate in typical locations, which
we shall denote by \( \hat{\theta}_0 \), and an upper bound on the average error rate, denoted by \( \theta_0 \). ‘Best’ is
interpreted as the ‘most likely value’ for \( \theta_0 \) and so we set the mode of the beta distribution equal
to the auditor’s estimate, i.e.

\[
\frac{a_1 - 1}{a_1 + a_2 - 2} = \hat{\theta}_0.
\]  

To make use of the upper bound we note that \( P(\theta_0 > \hat{\theta}_0) = P(\logit(\theta_0) > \logit(\hat{\theta}_0)) \), where
\( \logit(\theta) = \ln(\theta/(1 - \theta)) \). O’Hagan (1994), page 281, showed that when \( \theta_0 \) has a beta\((a_1, a_2)\) distribution then, under certain conditions, the distribution of \( \logit(\theta) \) can be well approximated
by a normal distribution with mean \( \ln\{h(a_1)/h(a_2)\} \) and variance \( h(a_1)^{-1} + h(a_2)^{-1} \), where
\( h(a) = a - 0.5 - 1/24a \). Therefore, we set

\[
\ln\{h(a_1)/h(a_2)\} + \Phi^{-1}(1 - \alpha)\{h(a_1)^{-1} + h(a_2)^{-1}\}^{1/2} = \logit(\hat{\theta}_0), \tag{5}
\]

where \( \Phi^{-1}(\cdot) \) is the inverse of the standard normal CDF. The parameters of the beta distribution
for \( \theta_0 \) are then found by solving for \( a_1 \) and \( a_2 \) in equations (4) and (5) by using a numerical
method such as Newton–Raphson iteration.

In this example the auditor chose \( \hat{\theta}_0 = 0.01 \) and \( \tilde{\theta}_0 = 0.04 \). This results in \( a_1 = 2.579 \) and
\( a_2 = 157.3 \).

6.1.1.2. Typical locations. The error rate in an individual location is given by \( \theta_i \). Recall that
the distribution of \( \theta_i|\theta_0 \) is a beta distribution, with, in typical locations, parameters \( \omega_i\theta_0 \) and
\( \omega_i(1 - \theta_0) \). The parameter \( \omega_i \) governs the variability of the error rate in a location about the
average \( \mathbb{E}(\theta_i|\theta_0) \). To elicit \( \omega_i \) for typical locations we suppose that \( \theta_0 \) is fixed at its most likely
value, i.e. \( \theta_0 = \hat{\theta}_0 \), and ask for an upper bound, \( \tilde{\theta}_i \), say, on the error rate in a typical location given
that the average in typical locations is \( \hat{\theta}_0 \). Using a normal approximation for the distribution of
\( \logit(\theta_i) \) given \( \theta_0 = \hat{\theta}_0 \) we again use equation (5), in this case with \( a_1 = \omega_i\hat{\theta}_0 \), \( a_2 = \omega_i(1 - \hat{\theta}_0) \)
and \( \tilde{\theta}_0 = \tilde{\theta}_i \). Again, a numerical solution for \( \omega_i \) can be obtained.

In our example, the auditor chose \( \tilde{\theta}_i = 0.04 \), resulting in \( \omega_i = 122.4 \) for typical locations.

6.1.1.3. A typical locations. For the atypical locations O and Q, we require values for both \( \delta_i \)
and \( \omega_i \). These are elicited by asking for an estimate (the mode) of the error rate in the location,
denoted by \( \tilde{\theta}_i \), and an upper bound on that error rate, \( \tilde{\theta}_i \), given that the mean error rate in typical
locations is \( \hat{\theta}_0 \). The parameters of the beta distribution of \( \theta_i|\theta_0 = \hat{\theta}_0 \) are given by \( g_{1i}(\hat{\theta}_0) \) and
\( g_{2i}(\hat{\theta}_0) \), where \( g_{1i}(\cdot) \) and \( g_{2i}(\cdot) \) are as defined in equations (1). Values for these parameters can be
derived from the mode and upper bound as described above, using equations (4) and (5), but
replacing \( a_k \) by \( g_{ik} \), \( k = 1, 2 \), \( \hat{\theta}_i \) by \( \tilde{\theta}_i \) and \( \tilde{\theta}_0 \) by \( \hat{\theta}_0 \). Values for \( \delta_i \) and \( \omega_i \) are then calculated from the
values of \( g_{1i}(\hat{\theta}_0) \) and \( g_{2i}(\hat{\theta}_0) \) by rearranging equations (1).

In both of the atypical locations O and Q the auditor specified \( \hat{\theta}_i = 0.005 \) and \( \tilde{\theta}_i = 0.02 \),
resulting in \( \delta_i = 0.8039 \) and \( \omega_i = 324.1 \). These values indicate a lower error rate than in typical
locations, and less uncertainty about the error rates in these two locations.

6.1.2. Error class parameters
The prior uncertainty about the population mean probability of falling into each of the error
classes is described by a Dirichlet\((b_1, b_2, b_3)\) prior distribution for \( \psi_0 = (\psi_{01}, \psi_{02}, \psi_{03})^T \). To obtain values for \( b_1, b_2 \) and \( b_3 \) we ask the auditors for an estimate of how the transactions split
between the different error classes, denoted by \( \hat{\psi}_0 = (\hat{\psi}_{01}, \hat{\psi}_{02}, \hat{\psi}_{03})^T \), and an upper bound on
the average probability of an understatement error, denoted by \( \hat{\psi}_{i1} \). We require an upper bound for only one error class because this uncertainty statement will, under our model, determine the uncertainty about the other classes. Clearly we can report the upper bounds for the other two classes that result from these statements so that the auditors can verify that they are consistent with their beliefs. The estimate of the split between the error classes is interpreted as the joint mode of \( \psi_0 \) and so gives

\[
\frac{b_1 - 1}{b_1 + b_2 + b_3 - 3} = \hat{\psi}_{01},
\]

\[
\frac{b_2 - 1}{b_1 + b_2 + b_3 - 3} = \hat{\psi}_{02}.
\]

The marginal distribution of \( \psi_{01} \), the probability of an understatement, is a beta distribution with parameters \( b_1 \) and \( b_2 + b_3 \). Here we relate this upper bound to the parameters of the model by using equation (5) but with \( a_1 = b_1, a_2 = b_2 + b_3 \) and \( \theta_0 = \hat{\psi}_{01} \). The parameters of the prior of \( \psi_0 \) are then found by solving numerically for \( b_1, b_2 \) and \( b_3 \) in equation (6) and the altered equation (5).

In this application, the auditor specified a ‘most likely’ split of errors given by \((\hat{\psi}_{01}, \hat{\psi}_{02}, \hat{\psi}_{03})^T = (0.4, 0.5, 0.1)^T \). The upper bound on the average rate of errors in the understatement category was given as \( \hat{\psi}_{01} = 0.7 \). This resulted in a Dirichlet prior for \( \psi_0 \) with parameters \( b = (2.351, 2.688, 1.338)^T \).

The amount by which the error class probabilities in locations might vary about the average is governed by the parameter \( \gamma \). First we assume that the average class probabilities take their most likely value, i.e. \( \psi_0 = \hat{\psi}_0 \). Then, under our model, \( \psi_{i1} \), which is the probability that an error falls in an understatement category in a particular location, has a beta distribution with parameters \( \gamma \hat{\psi}_{01} \) and \( \gamma(\hat{\psi}_{02} + \hat{\psi}_{03}) \). We then ask the auditor to give an upper bound on \( \psi_{i1} \), denoted by \( \hat{\psi}_{i1} \) say. The parameter \( \gamma \) is then obtained by numerically solving for \( \gamma \) in equation (5), where here we replace \( a_1 \) by \( \gamma \hat{\psi}_{01} \), \( a_2 \) by \( \gamma(\hat{\psi}_{02} + \hat{\psi}_{03}) \) and \( \theta_0 \) by \( \hat{\psi}_{i1} \).

In this example, the upper bound on \( \psi_{i1} \), given that \( \psi_0 = \hat{\psi}_0 \), was given by the auditor as \( \psi_{i1} = 0.7 \). This results in a value of \( \gamma = 7.591 \).

### 6.1.3. Taints within error classes

Error classes 1 and 2 above correspond to the distribution of taints in an open interval on the real line. For each of these classes the auditor is required to specify the expected distribution of taints in the class.

In error class 1, understatements, auditors felt that taints close to 0 are most likely, and that an exponential-type distribution would best describe the expected distribution of taints, i.e. a distribution with CDF \( F_{01}(t) = \exp(\lambda_1 t) \) for \( -\infty < t < 0 \). The parameter \( \lambda_1 \) is elicited by asking the auditor for the median of the understatement taint distribution, \( m_1 \) say. Then \( \lambda_1 = -\ln(2)/m_1 \). Here the auditor gave \( m_1 \) as \(-0.17\) and so \( \lambda_1 = 4.077 \). The prior weight parameter \( \pi_1 \) was chosen as 10 to reflect auditors’ beliefs that prior knowledge about this taint distribution is comparable in strength with observing a fairly small number of actual taints.

For error class 2 the auditor needs to describe the expected taint distribution for the open interval \((0, 1)\). Again the auditors thought that an exponential form would be most appropriate with taints close to 0 being most likely, and so they chose to express their expected distribution through a truncated exponential distribution with CDF \( F_{02}(t) = \{1 - \exp(-\lambda_2 t)\}/\{1 - \exp(-\lambda_2)\} \) for \( 0 < t < 1 \). To determine \( \lambda_2 \), a median overstatement taint is specified by the
Multilocation Auditing

445

auditor, \( m_2 \) say, and \( \lambda_2 \) is found by numerical solution of the equation \( \{1 - \exp(-m_2\lambda_2)\}/\{1 - \exp(-\lambda_2)\} = 0.5 \). In this example, the auditor gave \( m_2 = 0.17 \), resulting in \( \lambda_2 = 3.967 \). The prior weight parameter \( \pi_2 \) was again chosen to be 10.

6.1.4. Implications of prior parameter choices

One of the advantages of our simulation method is the range of posterior inferences that can be made about quantities of interest, such as the total error and error rates (both across and within locations). With only minor changes, this simulation method can also be used to draw prior inferences about such quantities. We provide such information to the auditor to aid elicitation. The elicitation process so far has been expressed solely in terms of the taint distribution. We provide the auditor with the resulting prior inferences for, among others, the total error distributions, so that the auditor can verify that these prior inferences match up with his or her beliefs.

6.2. Simulation-based inferences

The NAO’s main objectives are inferences for the total error \( E \) and the location-specific errors, the \( E_i \)'s. We report the results of the use of the algorithm given in Section 5 for simulating observations on the total population error from the model described in Section 3. For comparisons we also simulate the population based on the prior information only and compare our results with inferences from non-Bayesian methods.

Throughout what follows we restrict ourselves to discussing inferences for the organization as a whole and for locations A (a typical location that is not visited in the audit), I (a typical location that is visited but at which no errors were found), O (an atypical location that is not visited), Q (an atypical location that is visited but at which no errors were found) and T (the typical location, visited in the audit and at which one error was found). Calculations are based on 10000 simulations of the population.

6.2.1. Error totals

Fig. 1 plots estimates of the prior and posterior density functions, calculated from the simulated observations of the total error by using a normal kernel (Silverman, 1986). Table 1 gives the prior and posterior means and standard deviations of the generated observations on organization error and location error for locations A, I, O, Q and T. Table 1 also gives the values of the two simple point estimates \( \hat{E}_1 \) and \( \hat{E}_2 \) that are defined in Section 2.2 and the estimated

<table>
<thead>
<tr>
<th></th>
<th>Book value</th>
<th>Prior</th>
<th>Posterior</th>
<th>Data-based values</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>variance</td>
<td>Mean</td>
<td>Standard deviation</td>
<td>Mean</td>
</tr>
<tr>
<td>Organization</td>
<td>—</td>
<td>27870</td>
<td>33070</td>
<td>18460</td>
</tr>
<tr>
<td>Location A</td>
<td>16000</td>
<td>941</td>
<td>1842</td>
<td>628</td>
</tr>
<tr>
<td>Location I</td>
<td>17000</td>
<td>932</td>
<td>1792</td>
<td>530</td>
</tr>
<tr>
<td>Location O</td>
<td>15000</td>
<td>765</td>
<td>1360</td>
<td>516</td>
</tr>
<tr>
<td>Location Q</td>
<td>22000</td>
<td>742</td>
<td>1355</td>
<td>476</td>
</tr>
<tr>
<td>Location T</td>
<td>450000</td>
<td>943</td>
<td>3012</td>
<td>543</td>
</tr>
</tbody>
</table>
The posterior beliefs about the total error indicate a lower total error than the prior and, except for location T, lower standard deviations. This is consistent with the sample result of fewer errors than expected. The observed error was also an understatement and, as the auditors believe that overstatements are more common, the shift towards a belief in more
understatements will further reduce the net error. Note also that the data observed in locations I and Q result in lower means and standard deviations than for the equivalent unvisited locations A and O. The lower means result from the fact that we have observed zero errors in these locations, and the lower standard deviations also result from the increase in information concerning these particular locations. The two frequentist point estimates of the organization error total differ markedly, as well as being clearly inadequate because the only observed error is an understatement, whereas auditors expect overstatements to dominate in the population as a whole. The frequentist methods also fail to deliver realistic point estimates and standard errors for locations I and Q as well as for those locations that are not visited in the audit.

6.2.2. Other inferences
By using our method the auditors have available the whole distribution of the total error in each location and in the organization as a whole. It is of particular interest to estimate the total error as a percentage of the total expenditure, as this is a mechanism for comparisons across organizations. This is achieved simply by dividing the simulated total error values by the total expenditure. It can also be very useful to view the posterior densities, as given in Fig. 1, to calculate moments, and to estimate quantiles and posterior probabilities of interest. The last two are easy to obtain. For example, a 95th percentile of the total error distribution is estimated by the 95th percentile of the simulated distribution. The prior and posterior 95th percentiles of the total error for the above example are £91,070 (1.104% of expenditure) and £64,230 (0.7787% of expenditure). Posterior probabilities of exceeding a specific value $M$ are estimated by the proportion of simulated values exceeding $M$. In the above example, the prior and posterior probabilities of exceeding 1% of expenditure, which is £82,485, are 0.0656 and 0.0225 respectively. Frequentist methods either fail completely to allow such inferences or else produce unrealistic answers because they are entirely sample based.

Yet further inferences are easily available. We could compare locations. The auditor may wish to rank locations by the likely size of error, or else to calculate probabilities that the error in one location exceeds that in another. Again one just uses the simulated values as a surrogate for the true joint distribution. Other outputs of the procedure can be used to make relevant inferences—e.g. the error rate at location I has estimated prior mean 0.01595 and posterior mean 0.0119, a reduction that fits intuition as no errors were found at that location. This highlights the flexibility of our approach, which offers auditors a depth of analysis that is not available to them through existing frequentist methods.

6.3. Sensitivity to prior choices
In using our elicitation method, auditors have found most difficulty with the concept of specifying quantities conditionally, especially when giving an upper bound on the error rate in a typical location conditionally on the average error rate taking its most likely value. Also, our practical experience indicates that experts have difficulty specifying accurately their beliefs about quantiles (upper bounds), and that rounding or digit bias often occurs. This leads us to question the sensitivity of our procedure to prior specifications. As a consequence we have carried out two small sensitivity studies based on the example data set. For brevity we focused our studies on the sensitivity of our estimate of the mean and standard deviation of the total error distribution.

As a consequence of the auditors’ comments our first study looked at the sensitivity of inference to changes to $\bar{\theta}_i$ in Section 6.1.1, the conditional upper bound on the error rate in typical
locations. The other prior quantities in Section 6.1 were held fixed at the values specified there. Note that, in Section 6.1.1, $\bar{\theta}_i$ also denotes the upper bounds in the two atypical locations but as there were only two such locations these were kept at the value given in Section 6.1.1. We varied $\tilde{\theta}_i$ from 0.035 to 0.045 where the upper end of this range is almost as large as the estimate of the unconditional 95th percentile of $\tilde{\theta}_i$ in the example in Section 6.1. Hence, we can find out what happens when the auditor is thinking unconditionally, when he or she should be thinking conditionally. Estimates of the mean of the total error varied from 17950 to 19090, whereas the standard deviation varied from 24300 to 25350.

In the second sensitivity study we generalized our approach to take in all uncertainty judgments. We investigated scenarios at two extremes. The first scenario was the case of an auditor who specifies less uncertainty about the error rates and error class probabilities than in Section 6.1, with $\bar{\theta}_0 = 0.03$, $\bar{\theta}_i = 0.03$ in typical locations, $\bar{\theta}_i = 0.015$ in locations O and Q, $\bar{\psi}_01 = 0.6$ and $\bar{\psi}_i1 = 0.6$. The second scenario was the case of an auditor specifying more uncertainty than in Section 6.1, with $\bar{\theta}_0 = 0.05$, $\bar{\theta}_i = 0.05$ in typical locations, $\bar{\theta}_i = 0.025$ in locations O and Q, $\bar{\psi}_01 = 0.8$ and $\bar{\psi}_i1 = 0.8$. The estimates of the posterior means were 14820 and 22850 for the first and second scenarios respectively. The standard deviations were 15850 and 34940.

It is clear that the first study exhibits reasonable robustness to changes in one uncertainty judgment. In the second study the results for the two extreme scenarios match with intuition but show large changes in both the mean and the variance. Does this show too much sensitivity? The two scenarios are quite extreme; for example $\bar{\theta}_0$ and $\bar{\theta}_i$ change by 25% from their values in Section 6.1. Further practical experience will show how variable auditors will be in specifying their judgments on similar accounts, and how sensitive the auditors require our routines to be to differences between auditor inputs.

7. Discussion

7.1. Advantages and limitations of the method

The major strength of our method is that it provides a way of making inference statements based on both auditors’ judgment and sample information for a multilocation audit. As in the example data set, often in auditing the sample information is severely limited and so the auditors’ judgment has a large influence on the resultant inference. This is consistent with audit practice. As the data are sparse the auditors’ information about the population is crucial. Such information also ensures that the method is robust to small changes in the data.

The model itself may seem complex but it reflects both the practical situation and the structure of the beliefs of the auditors. The auditor can express different beliefs about the rate of errors in different sites and the posterior inference will be properly sensitive to these different beliefs. Also, Bayesian modelling allows inference for unvisited locations.

Our model for taints has a general structure that allows the identification of likely point masses depending on the application at hand. The set-up described in Section 6, where the taint distribution is divided into understatements, overstatements and bogus transactions, will be appropriate for many audits. The use of the Dirichlet process results in a resampling effect which reflects the belief that observed taints may represent unknown point masses in the taint distribution.

Many statistical methods in auditing depend on the assumption that monetary unit sampling has been employed, e.g. Fienberg et al. (1977), Dworin and Grimlund (1984) and McCray (1984), but our method does not require this assumption, although it is allowed. Also, the form of inference is flexible. Auditors at the NAO are primarily interested in point estimates and
posterior probabilities of the total error, but since the entire posterior distribution is simulated any feature of that distribution can be estimated. For example, a full probabilistic comparison of locations could be made.

One of the major limitations of the audit scenario is the sparseness of the data and hence model checking is difficult. Here we justify our choice of model by promoting its generality and the fact that it is based on the auditors’ knowledge of the population and previous studies of audit data such as that in Neter et al. (1985).

7.2. Further developments
In audit practice small errors that are found in the sample are often ignored when making inferences about the population (Elder and Allen, 1998). Clearly this is at best an approximation, but we could consider treating small errors (or perhaps, more correctly, errors in small transactions) differently from errors in large transactions if there is a belief that the error rate is different for different sizes of transaction. We might expect that internal audit procedures are more rigorous for large expenditure items and consequently errors in small transactions occur more frequently than errors in large transactions. Kim et al. (1987) dealt with this by stratifying the population according to the book value. A similar approach might be considered here.

Audit sampling is part of a much larger audit process. In this paper we have largely considered inference from sample data as a distinct process, with the connection between our model and the remainder of the audit process being made via the prior parameters of the model. It may be more sensible, if possible, to look more carefully at the audit as a whole and to build our method into some sort of coherent expert system, as proposed by Dutta and Srivastava (1993) and Boritz and Wensley (1990). The NAO have built the elicitation procedure into an EXCEL work sheet to help to integrate our method into the audit process but there is much still to do. Bonner et al. (1996) provided motivation for the development of decision aids in making conditional probability judgments such as those that are needed to elicit the parameters of our model.

Finally, the task of audit planning—choosing which sites to visit and how large a sample to take—is the subject of some on-going research.

7.3. Wider applicability
The model described in this paper has been designed to deal with a population structure that arises in auditing. However, it is clear to us that many of the ideas will be useful across the breadth of statistical applications. The whole model will be useful in situations in which errors are rare and where it is reasonable to assume that the fractional error, i.e. the taint, is independent of the known covariate or unit size, the book value in the case of audit.

Some aspects of our model will be usable in applications which do not deal with the modelling of rare errors, especially the hierarchical modelling of binary and categorical data, and the way in which real prior opinions about the different levels of a factor (the locations) are input into the model.

Acknowledgements
This research was funded by the UK NAO and published with their kind permission. The authors would like to thank the staff at the NAO, especially John Haworth for his patient advice, Alison Eve and Robin Swan, and the NAO’s statistical advisors, Professor Vic Barnett and Professor Fred Smith.
References


