Some Bayesian methods for two auditing problems

Glen Meeden*
School of Statistics
University of Minnesota
Minneapolis, MN 55455
glen@stat.umn.edu

Dawn Sargent
Minnesota Department of Revenue
600 North Robert St.
St. Paul, MN 55146

September 2005

*Research supported in part by NSF Grant DMS 0406169

Some Bayesian methods for two auditing problems

SUMMARY

Two problems of interest to auditors are i) finding an upper bound for the total amount of overstatement of assets in a set of accounts; and ii) estimating the amount of sales tax owed on a collection of transactions. For the first problem the Stringer bound has often been used despite the fact that in many cases it is known to be much too large. Here we will introduce a family of stepwise Bayes models that yields bounds that are closely related to the Stringer bound but are less conservative. Then we will show how this approach can also be used for solving the second problem. This will allow practitioners with modest amounts of prior information to select inference procedures with good frequentist properties.

AMS 1991 subject classifications Primary 62P05 and 62D05; secondary 62C10.

Key Words and phrases: Accounting, auditing, Stringer bound, stepwise Bayes models and dollar unit sampling.

1 Introduction

Consider a population of N accounts where the book values $y_1, \ldots, y_N$ are known. If each account in the population were audited then we would learn the actual or true values $x_1, \ldots, x_N$. Let Y denote the total of the book values and X the total of the audit values then

$$D = Y - X$$

denotes the total error in the book values. An auditor, who checks only a subset of the accounts, wishes to give an upper bound for D based on information contained in the sample. In the important special case where for each account $y_i - x_i \geq 0$ an upper confidence bound proposed in Stringer (1963) has been much discussed in the literature. It is easy to use but in many cases it is overly conservative. Moreover the 1989 *National Research Council's panel report on Statistical Models and Analysis in Auditing* stated that "... the formulation of the Stringer bound has never been satisfactorily explained." This report is an excellent survey of this and related problems.

Fienberg, Neter and Leitch (1977) present confidence bounds based on the multinomial distribution with known confidence levels for all sample sizes. Their bounds however are quite difficult to compute as the number of errors increase. Two other instances where a multinomial model has been assumed are the approaches of Tsui, Matsumura and Tsui (1985) and McCray (1984). In both these cases one must specify a prior distribution to implement the analysis.

Another important auditing problem is estimating the sales tax liability of a large corporation. Because of the large number of transactions only a sample of them can be checked. As with the Stringer problem most of the records in the sample will contain no errors. Here we will be interested in a giving a point estimate of the additional tax owned by the corporation. Let v denote the tax rate that applies to each transaction. If m is size of a transaction (in dollars) then $x = m(1+v)$ is the true total dollar amount associated with the transaction. For each transaction the corporation announces the book value y which should be x but in practice is sometimes smaller because not all of the necessary tax was paid. We will assume that $m \leq y \leq x$. Let Y be the total of all the book values and X be the total of the true values. Given a sample we wish to estimate $X - Y$.

Here we introduce a family of stepwise Bayes models for these auditing problems. Stepwise Bayes methods have the advantage of producing a posterior on which one can base their inferences but the required prior information is less than what is needed for a fully Bayesian analysis. We show that one member of this family is closely related to the Stringer bound. This gives a new way to think about the Stringer bound. In addition one can use prior information about the population to select a member of the family which is less conservative than the Stringer bound. We then show that essentially the same approach can work for the second problem.

In section 2 we briefly discuss the Stringer bound and argue that it can be given a Bayes like interpretation. In section 3 we review the stepwise Bayes approach to multinomial problems and discuss a family of models that give new bounds for this auditing problem. In section 4 we give some simulation results comparing the bounds. We find that the stepwise Bayes models can have reasonable frequentist coverage probabilities and are less conservative than the Stringer bound. In section 5 we apply our methods to the second problem. We will see that our methods are easy to use and can yield better answers than standard methods. In section 6 we conclude with a brief discussion.

2 A new view of the Stringer bound

2.1 The Stringer bound

We will begin by assuming for each unit that $x_i \geq 0$ and the difference $d_i = y_i - x_i \geq 0$. Then for $y_i > 0$

$$t_i = d_i / y_i$$

is called the tainting or simply the taint of the ith item. Then the error of the book balance of the account is

$$D = Y - X = \sum_{i=1}^{N} d_i = \sum_{i=1}^{N} t_i y_i$$

An important feature of this problem is that a large proportion of the items of the population will have $d_i = 0$. In such cases it is not unusual for the sample to to contain only a few accounts with $d_i > 0$. In fact if either this proportion is large enough or the sample size is small enough then there can be a nontrivial probability of the sample only containing units with $d_i = 0$.

In such a case the standard methods of finite population sampling are of little use.

Before presenting the Stringer bound we need some more notation. Given an observation from a binomial(n, p) random variable which results in m successes let $\hat{p}_u(m; 1-\alpha)$ denote an $(1-\alpha)\%$ upper confidence bound for p.

Suppose we have a sample of n items where k of them have a positive error and the remaining $n-k$ have an error of zero. For simplicity we will assume that the k taints of the items in error are all unique and strictly between zero and one. Let these k sampled taints be ordered so that $1 > t_1 > t_2 > \cdots > t_k > 0$. Then the $(1-\alpha)\%$ Stringer bound is defined by

$$\hat{D}_{u,st} = Y\big\{\hat{p}_u(0; 1-\alpha) + \sum_{j=1}^{k} [\,\hat{p}_u(j; 1-\alpha) - \hat{p}_u(j-1; 1-\alpha)\,]\, t_j\big\} \qquad (1)$$

To help understand why the Stringer bound works let p be the proportion of the items with error in the population and assume that the number of items with error in the sample follows a binomial(n, p) distribution. We first consider the case where $k = 0$, that is all the items in the sample have no error or all the taints are zero. Now $\hat{p}_u(0; 1-\alpha)$ is a sensible upper bound for p and if we assume for each item with a positive error that $x_i = 0$, that is for each item in error the actual error is as large as possible, then $Y\hat{p}_u(0; 1-\alpha)$, the Stringer bound in this case, is a sensible upper bound for D. Next we consider the case when $k = 1$. Now assume that for each item in the population their taint is either one or t_1. Then $\hat{p}_u(1; 1-\alpha) - \hat{p}_u(0; 1-\alpha)$ is a sensible upper bound for the proportion of items in the population whose taint equals t_1. So $Y[\hat{p}_u(1; 1-\alpha) - \hat{p}_u(0; 1-\alpha)]t_1$ is a sensible upper bound for the total error associated with such items. Now if we add this to our previous upper bound for all the items in the population with a taint of one we obtain the Stringer bound for this case. Continuing on in this way for samples which contain more than one error we see that the Stringer method adjusts in a marginal fashion bounds which assume that the errors in the population are as large as possible consistent with the observed errors.

2.2 A Bayesian interpretation

Here we will discuss how the logic underlying the Stringer bound can be given a Bayesian slant. The $(1-\alpha)\%$ Stringer bound given in equation 1 uses the $(1-\alpha)\%$ upper confidence bounds for the probability of success based on a binomial sample. Remembering that for the auditing problem a success is getting an item with a positive error it follows that

$$
\begin{aligned}
\alpha &\doteq \sum_{j=0}^{m} \binom{n}{j} [\hat{p}_u(m; 1-\alpha)]^j [1-\hat{p}_u(m; 1-\alpha)]^{n-j} \\
&= 1 - \sum_{j=m+1}^{n} \binom{n}{j} [\hat{p}_u(m; 1-\alpha)]^j [1-\hat{p}_u(m; 1-\alpha)]^{n-j} \\
&= 1 - \int_0^{\hat{p}_u(m;1-\alpha)} \frac{n!}{m!(n-m-1)!} u^{m+1-1}(1-u)^{n-m-1}\, du
\end{aligned}
$$

where the first equation is just a restatement of the fact that $\hat{p}_u(m; 1-\alpha)$ is a $(1-\alpha)\%$ upper confidence bound and the third uses the well known relationship between the cumulative distribution function of a binomial random variable and the incomplete beta function. Hence $\hat{p}_u(m; 1-\alpha)$ is the $(1-\alpha)$th quantile of a beta$(m+1, n-m)$ distribution. This is the crucial observation for what follows.

Consider now a sample where all n items had no error. This is not an uncommon occurrence in auditing problems. In such cases it is reasonable to assume that the population contains items in error even items with a taint equal to one. Let p_0 be the proportion of items in the population with taint equal to one. If we believe that there is at least one member of the population with such a taint then a sensible posterior given a sample with no items in error is

$$\propto p_0^{1-1}(1-p_0)^{n-1} \tag{2}$$

In this case we are assuming that all the items in the population are of two types; their taints are either zero or one. Now under this posterior the expected value of the total error, D, is $Y_{\bar{s}}/(n+1)$ where $Y_{\bar{s}}$ is the total book value of the items not in the sample. Why is this? We can imagining selecting a value u from a beta$(1, n)$ distribution and then selecting a random sample of size $u * (N-n)$ from the items not observed in the first sample. Then assuming that each of these $u*(N-n)$ items had a taint equal to one the sum of their book values would be a simulated value for D under this posterior. If we did this repeatedly for many choices of u then the mean of the set of simulated values for D will converge to $Y_{\bar{s}}/(n+1)$. On the other hand if we wanted a sensible $(1-\alpha)$-upper bound for D we could use the $(1-\alpha)$th quantile of our simulated values. Recall for this case the Stringer bound is just $Y\hat{p}_u(0; 1-\alpha) = Y_{\bar{s}}\hat{p}_u(0; 1-\alpha)$, since none of the sampled units had any error, and where $\hat{p}_u(0; 1-\alpha)$ is the $(1-\alpha)$th quantile of the beta$(1, n)$ distribution.

Note both bounds assume that any actual taint must be one. In addition they can both be interpreted as arising from the same posterior for the number of possible taints in the population The difference is that Stringer uses the product of the $(1-\alpha)$th quantile of this distribution and $Y_{\bar{s}}$ to get his bound while the stepwise Bayes approach uses the $(1-\alpha)$th quantile of the distribution of D induced from the posterior.

Next consider a sample where $n-1$ of the items had no error and one had an error with a taint $t_1 \in (0, 1)$. If we let p_0 be the proportion of the population with taint equal to one, p_1 be the proportion of the population with taint equal to t_1 and p_2 the proportion of the population with no error then arguing as before a sensible posterior for (p_0, p_1, p_2) is Dirichlet$(1, 1, n-1)$. From this we see that the marginal posterior of $p_0 + p_1$ given the sample is beta$(2, n-1)$. Recall that Stringer's bound involves the $(1-\alpha)$th quantile of this distribution and of the beta$(1, n)$ distribution and the value t_1.

Now we will give another bound based directly on the Dirichlet$(1, 1, n-1)$ distribution. We begin by generating an observation $p' = (p'_0, p'_1, p'_2)$ from this distribution. Then we randomly divide the unsampled items into three groups whose sizes are proportional to p'. In the first group we assume all the taints are one, in the second all the taints are t_1 and in the third all the taints are zero. Using these assigned taint values for all the unsampled items we find the corresponding value of D by adding all these simulated errors to the one observed error. This gives one possible simulated value for D from its posterior distribution induced by the Dirichlet$(1, 1, n-1)$ distribution. Repeating this for many choices of p' and finding the $(1-\alpha)$th quantile of the resulting set of simulated values for D we would have a sensible upper bound for D.

More generally assume we have observed k distinct taints $1 > t_1 > t_2 > \cdots > t_k > 0$ and $n-k$ taints equal to zero in the sample. Let p_o be the proportion of the population with a taint of one and p_{k+1} the proportion with a taint of zero. For $i = 1, \ldots, k$ let p_i be the proportion with taint equal to t_i. In this case our posterior will be Dirichlet$(1, 1, \ldots, 1, n-k)$ where there are $k+1$ parameters equal to one in this distribution. Now just as in the simpler cases this posterior induces a distribution for D. It is easy to generate simulated values of D and take the $(1-\alpha)$th quantile of a large set of such simulated values as an upper bound for D. One would expect that this bound would be somewhat smaller than the Stringer bound and in fact we will see that this is so.

3 A family of stepwise bounds

3.1 The multinomial model

Consider a random sample of size n from a multinomial distribution with $k+1$ categories labeled $0, 1, \ldots, k$. Let p_j denote the probability of observing category j and W_j be the number of times category j appears in the sample. Then $W = (W_0, \ldots, W_k)$ has a multinomial distribution with parameters n and $p = (p_0, \ldots, p_k)$. Now W/n is the maximum likelihood estimator (mle) of p and is admissible when the loss function is the sum of the individual squared error losses and the parameter space is the k-dimensional simplex

$$\Lambda = \{p : p_i \geq 0 \text{ for } i = 0, 1, \ldots, k \text{ and } \sum_{j=0}^{k} p_j = 1\}$$

For many problems both Bayesian and non Bayesian statisticians would be pleased to be presented with an "objective" posterior distribution which yielded inferences with good frequentist properties. Then give the data they could generate sensible answers without actually having to worry about selecting a prior distribution. For the multinomial model one such posterior distribution is well known. Given the data $w = (w_0, \ldots, w_k)$ let I_w be the set of categories for which $w_i > 0$. We assume I_w contains at least two elements and let

$$\Lambda_w = \{p : \sum_{i \in I_w} p_i = 1 \text{ and } p_i > 0 \text{ for } i \in I_w\}$$

Then we take as our "objective posterior"

$$g(p|w) \propto \prod_{i \in I_w} p_i^{w_i - 1} \qquad \text{for } p \in \Lambda_w \tag{3}$$

Note the posterior expectation of the above is just the mle. This fact can be used to prove the admissibility of the mle. Although not formally Bayesian against a single prior these posteriors can lead to a non-informative or objective Bayesian analysis for a variety of problems. It is the stepwise Bayes nature of these posteriors that explains their somewhat paradoxical properties. Given a sample each behaves just like a proper Bayesian posterior but one never had to explicitly specify a prior distribution. These posteriors are essentially the Bayesian bootstrap of Rubin (1981). See also Lo (1988).

Cohen and Kuo (1985) showed that admissibility results for the multinomial model can be extend to similar results for finite population sampling. For more details on these matters see Ghosh and Meeden (1997). When the sample size is small compared to the population size the finite population setup is asymptotically equivalent to assuming a multinomial model. Since the later is conceptually a bit simpler, in what follows we will restrict attention to the multinomial setup. We next discuss a slight modification of the stepwise Bayes posteriors which gives the mle. This new family of posteriors will then be applied to the auditing problem.

3.2 Some new stepwise Bayes models

The posteriors of the previous section were selected because they lead to the mle. Here we present a slight modification of them to get posteriors which will be sensible for our auditing problem.

To help motive the argument imagine one believes that $p_0 > 0$ but there is a possibility that it could be quite small. In such a case it can be quite likely that the category zero would not appear in a sample. But even when $w_0 = 0$ one would still want the posterior probability of the event that $p_0 > 0$ to be positive. Note this would not be true for the posteriors of the previous sections.

Keeping in mind that category zero will play a special role one extra bit of notation will be helpful. If w is any sample for which $w_i > 0$ for at least one $i \neq 0$ then we let $I_{0,w}$ be category zero along with the set of categories for which $w_i > 0$ and

$$\Lambda_{0,w} = \{p : \sum_{i \in I_{0,w}} p_i = 1 \text{ and } p_i > 0 \text{ for } i \in I_{0,w}\}$$

For such a w our posterior has the form

$$g(p|w) \propto p_0^{w_0+1-1} \prod_{i \in I_{0,w} \text{ and } i \neq 0} p_i^{w_i - 1} \qquad \text{for } p \in \Lambda_{0,w} \tag{4}$$

It useful to compare equations 3 and 4. The posterior in equation 3 gives positive probability only to categories that were observed in the sample and essential treats them the same. That is the posterior depends only on the observed sample frequencies. This is why it can be thought of as an objective or non-informative posterior. On the other hand the posterior in equation 4 always assigns positive probability to the event $p_0 > 0$ whether or not the zero category has appeared in the sample. But the observed categories are treated similarly.

We should also point out that in the first posterior the actual values of the categories play no role in the argument. The categories could be given any set of labels and the same argument applies. This is true for the second posterior as well with the exception that one must be able to identify the exceptional category, zero, before the sample is observed. With no loss of generality the actual values of the other categories can depend on the sample.

The family of posteriors in equation 4 is exactly the one described in section 2.2 when giving the Bayesian interpretation of the Stringer bound. There the special category corresponding to p_0 was units with a taint equal to one. We will denote this case as the "objective" stepwise Bayes bound. Note it is based on exactly the same assumptions that underlie the Stringer bound.

In principle there is no reason to assume that the unobserved category has taint equal to one. One could choose any value of $t \in (0, 1]$ for the unobserved taint. Selecting a $t < 1$ will give a smaller bound than the "objective" choice of $t = 1$. Moreover it allows one to make use of their prior information about the population in a simple and sensible way. The theory describe in section 3.1 for the family of posteriors given in equation 3 extends in a straightforward way to the family of posteriors given in equation 4. Again one can check Ghosh and Meeden (1997) for details.

4 Some simulation results

In this section we present some simulation results which compare our proposed bounds to the Stringer bound. We also computed a bound proposed by Pap and Zuijlen (1996). Bickel (1992) studied the asymptotic behavior of the Stringer bound and they extended his work demonstrating the asymptotic conservatism of the bound. They proposed a modified Stringer bound which asymptotically has the right confidence level. Their bound is easy to compute and will be include in the simulation studies. We will call it the adjusted Stringer bound. You should check their work for further details.

Although not stated explicitly in the previous section the new bounds are based on the assumption that the size of an item's taint is independent of its book value. This should be clear when one recalls how the simulated values for D are generated. We should not expect the new method to work well for populations where this assumption is violated.

To see how the new bounds can work we constructed nine populations of 500 items. In each case the book values are the same, a random sample of 500 from a gamma distribution with shape parameter ten and scale parameter one. The sum of the book values for this population is 4942.3. The nine populations are divided into three groups, each with three populations. Within each group there was a population with 2%, 5% and 10% of the items in error. To construct the true values for the items we did the following. In the first group we selected a random sample of 2%, 5% and 10% of the items and set their x value equal to the product of their y value and a uniform(0,1) random variable. All the uniform(0,1) random variables were independent. For the remaining items their x values were set equal to their y values. The second group was constructed in exactly the same manner except that a uniform(0.01,0.1) distribution was used to generated the items in error. Note in both these groups the size of the taint should be independent of the book value. Perhaps it should be more difficult in the second group to find an upper bound for D because the taints will be much larger on the average and so D will be larger. The third group was constructed to violate the assumption that the size of the taint is independent of the book value. Here rather than selecting the items in error at random we selected the items with the largest 2%, 5% and 10% book values and then multiplied each of those in the set by a uniform(0.0,0.1) random variable.

To compare the performance of the bounds over the nine populations we took 500 samples of sizes 40 and 80 where the items were selected at random without replacement with probability proportional to the book values. In auditing problems this is commonly referred to as *Dollar Unit Sampling*. For each sample we computed the 95% Stringer bound and the adjusted bound proposed by Pap and Zuijlen (1996). Finally we computed the objective stepwise Bayes bound where we assume the unobserved taint $t = 1.0$. The new bound was found approximately for each sample by simulating 500 values of D from its posterior. We then use the 0.95 quantile of these simulated values as our bound.

The results are given in Table 1 when the sample size was 40. (The results when the sample size was 80 were very similar.) The first thing to note is that Stringer bound is indeed quite generous (column six of the table) and exceeds its nominal level. The objective stepwise bounds are indeed shorter than the Stringer bounds but in most cases still quite generous. The frequency of coverage of these bounds never falls below the nominal 0.95 level, a quite acceptable performance. The adjusted Stringer bound is clearly the best in the first seven populations but significantly under covers in the last two populations. Recall that in these populations the size of the taint does depend on the book value and were constructed as extreme cases. Perhaps it is not surprising that for these small sample sizes of 40 and 80 an asymptotic argument cannot improve on Stringer for all possible populations,

Let t denote the value of the unobserved taint that our model assumes is in the population even when it is not observed in the sample. Instead of assigning it the value one we could let it represent our prior guess for the average value of all the positive taints in the population. This will yield shorter bounds and allows one to incorporate prior information into the problem. To see how this could work we ran simulations for the nine populations with $t = 0.5$. This is in fact approximately the average value of the taints in the first three populations but much to small for for the last six. The results are given in table two. We see that reasonable prior information can result in significantly smaller bounds without sacrificing approximate correct coverage probability.

When all the items in the sample have no error the frequency of coverage depends entirely on the model underlying the procedure. This is true to a lesser extent when the sample contains just one item with a positive taint. This is true even for the Stringer bound, For example for the third population in the third group where the proportion of items in error is 10% we took 500 samples of size 18 and compared the Stringer and objective stepwise Bayes bound. The frequency of coverage for the Stringer bound was 0.970 and for the stepwise Bayes bound 0.976. The ratio of the average length of the stepwise bound and the average length of Stringer's bound was 0.923. This occurred because in the 15 cases where all the sampled items had no error the Stringer bound never covered while the stepwise Bayes bound covered in three samples. The sample size was selected so that The Stringer bound would be too small when the sample contained no items in error.

The stepwise Bayes bounds proposed here essentially assume a multinomial model where the number of categories used depends on the observed sampled. Two other instances where a multinomial model has been assumed in a nonparametric Bayesian setting are the approaches of Tsui, Matsumura and Tsui (1985) and McCray (1984). In these cases however one must specify a full prior distribution in contrast to just the one parameter needed here. Finally, one can extend our method to populations which contains items with both positive and negative taints. In such problems one needs to have a good guess for a lower bound for the taints,

5 Estimating Sales Tax Liability

5.1 Formulating the problem

In this section we consider the problem of estimating the Sales Tax Liability of a large corporation based on an audit of a sample of their transactions. A corporations records are usually so detailed, complex, and voluminous that an audit of all the records would be impractical. Therefore, sampling methods are used to determine whether the correct amount of sales tax was paid for a given period. As with the Stringer problem most of the records in the sample will contain no errors. Here we will be interested in a point estimate and not an upper bound. Even though this problem is slightly different than Stringer's problem our stepwise Bayes approach can be applied here as well.

Let v denote the tax rate that applies to each transaction. If m is size of a transaction (in dollars) then $x = m(1+v)$ is the true total dollar amount associated with the transaction. For each transaction the corporation announces the book value y which should be x but in practice is sometimes smaller because not all of the necessary tax was paid. We will assume that $m \leq y \leq x$. For transaction i we define the taint to be

$$t_i = (x_i - y_i)/y_i$$

Note it is always the case that $0 \leq t_i \leq v$. Please note that although we are using the same notation and terminology as in the Stringer problem book and taint have different meanings here. When a transaction is audited we learn x because we learn the value of m and v is known.

Now in our approach, given the sample, we wish to use the observed taints and the known book values for the unobserved units to simulate completed copies of the possible errors in the population. To simplify matters assume for a moment there is only one stratum. As in the Stringer problem we will assume that there is a special category. Here it will corresponding to a taint with a value $t^* > v$. Even though this is an impossible value which we can never observe it will be useful in the following.

First consider a sample w where all the observed taints equal zero. If p_0 is the probability associated with the taint t^* and $1 - p_0$ is the probability associated with the taint zero then our stepwise Bayes posterior given w will be

$$g(p_0|w) \propto p_0^{1-1}(1-p_0)^{n-1} \qquad \text{for } 0 < p_0 < 1$$

This is just the analogue of equation 4 for this new problem. That is, when the sample contains no errors we are still allowing for the possibility of error in the unsampled transactions.

Next suppose the sample contains at least one transaction with a taint greater than zero. Of course almost all of the observed taints will be zero. For such a sample we will use the "objective posterior" given in equation 3. There is no need to consider a special category for such a sample as we did in the Stringer case. There we were trying to get an upper bound here we are constructing a point estimate. In this case just using the observed taints will be good enough. There is no reason to bias it upwards as we did there.

Whether the sample has only zero taints or contains some positive taints it is clear how we proceed. We use our posterior to randomly partition the unobserved transactions into subsets associated with each taint. Then for every unobserved unit we use its book value along with its assigned taint to generate a completed population. For each simulated population we can then find the total error in the amount of sales tax paid. Then by averaging over many such simulations we have an estimate of the tax owed by the corporation.

Actually when the sample only contains transactions with no error one need not perform the simulation to find the point estimate. Arguing as we did in the paragraph following equation 2 one finds the estimate to be

$$\frac{1}{n+1} t^* \sum_{i \notin s} y_i$$

Note this method always assumes that there were errors made even when the sample contains no errors.

When the population is stratified one approach would be to follow the above program independently within each stratum. We have found however that our method works better if one just ignores the stratification once the sample has been taken. This amounts to assuming that the taints are roughly exchangeable across the entire population not just within each stratum. Often this seems to be a reasonable way to use the information in all of the observed taints. Of course if this assumption is false our method will work poorly and then one just assumes exchangeability with the strata.

5.2 An example

When the Minnesota department of Revenue receives the population of book values from a corporation some of transactions are removed. For example those with negative book values or extremely large book values are examined individually. Even with this pruning most populations are still heavily skewed to the right since there are many more smaller than larger dollar transactions. They then stratify the population using the cumulative square root rule on the book values. Once the total sample size has been chosen they use Neyman allocation to allot the sample sizes within the strata. See Cochran (1977) for details. They then do simple random sampling within each stratum subject to the constraint that the mean of the sampled book values within a stratum must be within 2% of the stratum mean of the book values.

We have a data set which comes from a larger sample of a corporation's tax records. For each transaction we know both the y and x values. This will allow us to do simulations to see how our method compares to standard practice. The population, which we denote by *Corp*, contains 1300 transactions. The book values (in dollars) range from 5 to 25,798. The 25th, 50th and 75th quantiles are 333, 7,110 and 25,798. The total tax liability for this population is 14,210 There are 70 transactions in error. In Figure 1 we have plotted the taints against their book values for this population. The taints do look to be approximately exchangeable.

For this population $v = .065$. For our simulations we set our special value $t^* = .07$. We compared our method to standard practice which just uses the usual stratified estimator. We considered three different sampling plans. The first used no stratification while the next two broke the population into two and three strata as described just above. In each case we then took 1000 samples of size 100 and compared the usual estimator, Stdest, to the stepwise Bayes estimator, SwBest1, which assumes exchangeability across the entire population. We see from the first set of results from Table 3 both estimators are approximately unbiased but SwBest1 is clearly preferred to Stdest.

To see how sensitive SwBest1 is to the exchangeability assumption we constructed two extreme cases. First we found the values of the 70 positive taints and ordered them from smallest to largest. We then took the original book values corresponding to these 70 taints again ordered from smallest to largest and gave each of them a new x value. We applied the smallest taint to the transaction with the smallest book value to get its new true value of x. Next we applied the second smallest taint to the second smallest of the 70 book values and so on. We denote this population by $Corp1$. For $Corp1$ the larger taints will be associated with the larger book values. We also construct $Corp2$ by associating in reverse order the 70 taints to the book values. For $Corp2$ the larger taints will be associated with the smaller book values. For these two populations the exchangeability assumption is strongly violated. For each of them we then did a set of simulations similar to what we did for population $Corp$. Not surprisingly SwBest1 becomes quite biased. But even so it still does better than Stdest for population $Corp1$. Even though Stdest is unbiased it has a huge variance which allows Sebest1 to do better in this case.

We also did some simulations which included a second stepwise Bayes estimator, SwBest2, which only assumes exchangeability within each stratum. These results are given in Table 4. It is interesting that SwBest1 still does the best for $Corp$ and $Corp1$ but is beaten by SwBest2 for $Corp2$. These results suggest that SwBest1 is reasonably robust against departures from exchangeability.

6 Discussion

These objective stepwise Bayes models are closely related to the Polya posterior which is closely related to many of the standard procedures in finite population sampling. For more details see Ghosh and Meeden (1997). Hence it is not surprising that the methods presented here should yield procedures with good frequentist properties and that they can be based on rather simple types of prior information.

It seems problematical to us that for small sample sizes one will be able to find bounds which dramatically improve on the Stringer bounds across all possible populations, i.e. bounds which are significantly smaller but still have approximately the nominal coverage probability. Consideration of difficult populations like the third population in group three in section 4 indicates why this is so. For larger sample sizes most samples will have several items with positive taints. For such samples the Stringer bounds are not much larger than those of the objective stepwise Bayes model. Even though there is no known argument proving that the Stringer bound achieves at least its nominal level for all possible populations we believe that the material presented here indicates that the Stringer bound is using the data in a sensible although perhaps a slightly conservative way. This is consistent with the results of Fienberg, Neter and Leitch (1977) .

References

[1] W. C. Cochran. *Sampling Techniques*. John Wiley & Sons, New York, 1977.

[2] M. Cohen and L. Kuo. The admissibility of the empirical distribution function. *Annals of Statistics*, 13:262–271, 1985.

[3] S. E. Fienberg, J. Neter, and R. A. Leitch. Estimating the total overstatement error in accounting populations. *Journal of the American Statistical Association*, 72:295–302, 1977.

[4] M. Ghosh and G. Meeden. *Bayesian Methods for Finite Population Sampling*. Chapman and Hall, London, 1997.

[5] A. Y. Lo. A Bayesian bootstrap for a finite population. *Annals of Statistics*, 16:1684–1695, 1988.

[6] J. H. McCray. A quasi-Bayesian audit risk model for dollar unit sampling. *Accounting Review*, 59:35–51, 1984.

[7] Panel on nonstandard mixtures of distributions. Statistical models and analysis in auditing. *Statistical Science*, 4:2–33, 1989.

[8] Donald B. Rubin. The Bayesian bootstrap. *Annals of Statistics*, 9:130–134, 1981.

[9] K. W. Stringer. Practical aspects of statistical sampling in auditing. In *Proc. Bus. Econ. Statist. Sec.*, pages 405–411. ASA, 1963.

[10] R Development Core Team, editor. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, www.R-project.org, 2005.

[11] K. W. Tsui, E. M. Matsumura, and K. L. Tsui. Multinimial-dirichlet bounds for dollar unit sampling in auditing. *Accounting Rev.*, 60:76–96, 1985.

Appendix

The bounds presented here are easy to compute approximately. The key fact used in the program is that a Dirichlet random variable can be generated from a sum of independent gamma distributions. For the Stringer problem a program has been written in the statistical package R which allows one to simulate values of D once the value t has been specified. A interested reader can use this program in $RWeb$ at one of the author's web site

`http://www.stat.umn.edu/~glen/`

Once there click on the "Rweb functions" link and select "simulateD.html". Then you can construct simple examples to see how the methods presented here work in practice.

Table 1: The frequency of coverage of the 95% Stringer bound, Stg, the 95% adjusted Stringer bound, Adj, and the 0.95 objective stepwise Bayes bound, SwB. The results are based on 500 samples of size 40 for the nine populations. Also included is the value of D and the average values of the stepwise Bayes bound and the adjusted Stringer bound as a percentage of the average value of the Stringer bound.

Group (pop)	D	Freq of Coverage Stg	SwB	Adj	As a % of avStg D	avSwB	avAdj
1(2%)	61.7	1.0	1.0	1.0	13%	89%	86%
1(5%)	150.2	1.0	1.0	1.0	25%	88%	83%
1(10%)	252.0	1.0	1.0	1.0	35%	85%	86%
2(2%)	75.9	1.0	1.0	1.0	16%	91%	82%
2(5%)	239.0	1.0	1.0	1.0	34%	92%	75%
2(10%)	480.6	0.990	0.976	0.934	48%	92%	76%
3(2%)	180.1	1.0	1.0	1.0	29%	93%	74%
3(5%)	415.2	0.996	0.996	0.868	45%	94%	75%
3(10%)	770.7	0.970	0.968	0.812	58%	95%	80%

Table 2: The behavior of the 0.95 stepwise Bayes bound when the unsampled taint $t = 0.5$. The results are based on 500 samples of size 40. The last column gives the ratio of the average value of the $t = 0.5$ bounds to the average value of the $t = 1.0$ bounds.

Group (pop)	D	ave Bnd	Freq of Cov	$t.5/t1$
1(2%)	61.7	269	1.0	0.66
1(5%)	150.2	397	1.0	0.75
1(10%)	252.0	542	0.904	0.85
2(2%)	75.9	301	1.0	0.69
2(5%)	239.0	569	0.884	0.88
2(10%)	480.6	909	0.942	0.98
3(2%)	180.1	490	0.824	0.85
3(5%)	415.2	841	0.914	0.96
3(10%)	770.7	1309	0.96	1.01

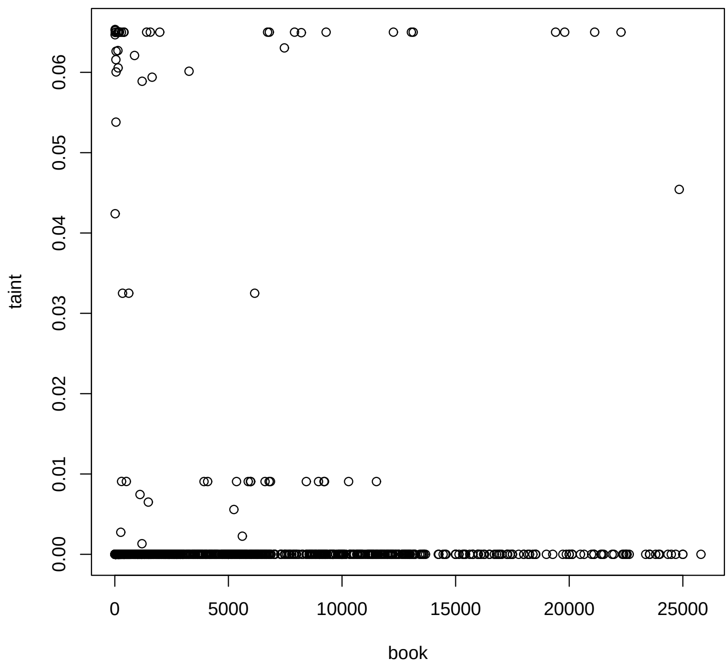

Figure 1: The plot of the taints against the book values for the 1300 transactions.

Table 3: The average value of the estimators and their average absolute errors for Stdest (the standard estimator) and SwBest1 (the Stepwise Bayes estimator which ignores stratification). The results are based on 1000 samples of size 100.

		Stdest		SwBest1	
Population (liability)	Number of Strata	Ave ptest	Ave abserr	Ave ptest	Ave abserr
Corp	1	14,466	9,775	14,132	5,389
(14,210)	2	14,180	7,081	15,077	5,428
	3	14,361	5,987	14,568	5,053
*Corp*1	1	21,757	11,613	14,636	8,584
(21,781)	2	21,080	8,414	17,758	6,951
	3	21,434	8,023	18,317	7,018
*Corp*2	1	5,467	2,535	13,505	8,560
(5,292)	2	5,128	2,747	10,143	5,548
	3	5,196	2,244	9,108	4,754

Table 4: The average absolute errors, based on 1,000 samples of size 100 of Stdest (the Standard estimator) , SwBest1 (the Stepwise Bayes estimator which ignores stratification) and SwBest2 (the Stepwise Bayes estimator which only assumes exchangeability within strata).

Population	Number	Average Absolute Error		
(Liability)	of Strata	Stdest	SwBest1	SwBest2
Corp	1	9,296	5,189	
(14,210)	2	6,637	5,541	6,637
	3	6,481	5,681	6,143
*Corp*1	1	11,204	8,115	
(21,781)	2	8,626	6,945	8,066
	3	7,897	6,918	7,648
*Corp*2	1	2,469	8,469	
(5,292)	2	2,760	5,680	2,806
	3	2,359	5,083	3,207