Sampling risk evaluations in a tax fraud case: Some modelling issues

BY Jostein Lillestøl

DISCUSSION PAPER

NHH

Institutt for foretaksøkonomi
Department of Business and Management Science

FOR 05/2020

ISSN: 1500-4066

May 2020

Sampling risk evaluations in a tax fraud case: Some modelling issues

by[1]

Jostein Lillestøl

Department of Business and Management Science
Norwegian School of Economics (NHH), Bergen, Norway

May. 12, 2020

Summary

This work is a follow-up to Lillestøl (2019). The context is the use of sample data to support claims of tax fraud at eateries, where the possibilities of embezzlement are overreporting of take-away sales and underreporting of cash payments. Ratios of sales amounts of opposing types are computed from the sample and used as estimates for the true yearly ratios. Decisions are made by comparison with the reported ratios in the taxpayer's yearly income statement, allowing for sampling risk. To this end, a "risk distribution" is established and its quantiles used as decision limits. There are different ways of doing the calculation and to establish the accompanying risk distribution, among them models based on Gamma-assumptions, as detailed in Lillestøl (2019). They may lead to different results, more or less favorable to the taxpayer. The chosen method therefore must be fair and defensible. In this connection, some relevant issues have surfaced, mainly related to independence and conditioning. The objective of this paper is to explore these issues and provide some recommendations on the choice of method.

[1] This work is supported by Norwegian Centre for Taxation (NoCeT) at NHH.

1. The context

This paper investigates some theoretical issues concerning risk evaluations for a class of sample audit estimates applicable to the tax review of take-away eateries, as detailed in Lillestøl (2019). The context is as follows: A key quantity for the tax authority reported in the yearly income statements is the percentage of yearly take-away sales (15% VAT) as opposed to the percentage of sales consumed at the premises (25% VAT). The latter may be underreported. Another quantity of interest for the tax authority is the percentage of sales paid by card as opposed to paid by cash, as the latter offers opportunity to underreport. Following the suspicion of underreporting, the tax authority has sampled some opening dates and observed the sales on these days, i.e. sales amounts and type of sales. From this, they extrapolate to see if the reported numbers for the whole year are justified or not.[2]

The findings from inspection visits at a pizzeria on three samples dates are given in Table 1, where the opposing typing of sales are named ForHere vs ToGo and Cash vs Card.

Table 1 Sales amount (NOK) and number of sales (#) at inspection on three sampled dates

Date	ForHere NOK (#)	ToGo NOK (#)	Cash NOK (#)	Card NOK (#)	Total NOK (#)
14.05.2014	838 (5)	1664 (5)	969 (2)	1533 (8)	2502 (10)
25.05.2014	969 (6)	6307 (28)	1944 (12)	5332 (22)	7276 (34)
18.06.2014	2274 (11)	2283 (12)	699 (5)	3858 (18)	4557 (23)
Total	4081 (22)	10254 (45)	3612 (19)	10723 (48)	14335 (67)

We assume that there is a true yearly percentage of sales of each opposing type, named type 1 and type 2, which may differ from the reported percentages. The aim of the sampling is to estimate the true percentage accompanied with an uncertainty judgment, say by providing a limit on the true percentage with an attached probability guarantee accounting for sampling errors.

There are several ways to go about this. The approach used by the tax authority, in this case, was to calculate the fractional amounts paid for each opposing type for each day separately. These numbers were then averaged over the three days to provide estimates of the true yearly percentage amounts of each type. Crude risk calculations are possible, although a basis of essentially three observations is weak, and with conditions required for normal approximation hard to justify.[3]

An alternative approach advocated in Lillestøl (2019) is to aggregate all sales amounts of each type and divide by the total sales amount. This is taken as estimates of the true yearly percentages, and risk calculations are possible, based on distributional assumptions on the individual sales. Note that the first approach involves the number of observations each day, in a sense irrelevant, while the second approach goes free of this.[4] Take ForHere to be type 1 sale and ToGo to be type 2 sale. Then the aggregated sales amounts are S_1=4225 and S_2=10110, and so the ratio of ForHere sales is R=4081/(4081+10254)=0.285. In case the reported ForHere% is 20%, the sample appears somewhat unfavorable to the taxpayer. However, if the lower limit guaranteed with high confidence (say 95%) turns out to be less than 20%, there is no clear evidence in disfavor of the taxpayer. So, what is a

[2] Comparisons could be made with percentages known from the eatery business at large, as well as to prior year reports and the one for the current year, first available at the end of the year.
[3] In this case, formal risk calculations were not performed by the tax authority.
[4] In fact, the first day visit to the eatery covered only about half of the opening hours.

reasonable lower limit on ForHere% accounting for sampling error, more than the reported 20% or less than 20%? And, what if the reported ratio was 15% or as low as 10%. This may be crucial for any follow-up actions by the tax authority.

The sample turned out the number of sales and the amount ratios for the two pairs of opposing types of sales displayed in Table 2, together with the reported ratios.

Table 2. Amount ratios: Sampled and reported

Type of sale	ForHere	ToGo	Cash	Card
No. of sales	22	45	19	48
Ratio (R)	28.5%	71.5%	25.2%	74.8%
Reported	10.3%	89.7%	12.7%	87.3%

The paper will examine some questions related to the use of a sampled ratio R as estimate of the long run ratio ρ. With a sample over a time span of three opening days out of about 360 openings days for a year, it is in effect a finite population problem. However, analytically it is more convenient to treat the problem within an infinite horizon context. This Is justified, since sample is small relative to the population. Let S_1 and S_2 be the sums of sampled sales of each opposing type, type 1 and type 2. Focusing on the of type 1 sales we then take

$$R = \frac{S_1}{S_1 + S_2} \quad \text{as estimate of } \rho = \frac{E(S_1)}{E(S_1) + E(S_2)}$$

We take ρ as the true ratio, and when the sampled ratio R conflicts with the one reported, suspicion is raised. However, benefit of doubt should be given to the taxpayer by some calculation accounting for sampling error.[5] The calculation will be performed by establishing a risk distribution of some kindand use an appropriate quantile as decision limit.[6] We will enumerate the opposing types so that the lower quantiles are the relevant ones. The risk distribution can be established several ways, same distribution-free and some based on distributional assumptions. Both frequentist and Bayesian approaches are available. Some also offer estimates of the long run ratio ρ other than R. The proposed models in Lillestøl (2019) all assume independent Gamma-distributed sales amounts, and beyond that, different levels of complexity. In this paper we examine theoretical issues raised in connection with some of these approaches. In case knowledge of the Gamma-distribution and its relation to the Beta-distribution is lacking, some basics are given in an Appendix.

2. The modelling issues

For the moment, we assume that the number of sales of each type regarded as given, say n_1 and n_2. Assume that the sales amounts are independent and follow a Gamma-distribution with common scale parameters, but possibly different shape parameters for the two groups. It then follows by theory (see Appendix) that R will have a Beta-distribution with parameters determined by the two involved shape parameters. The parameters are easily estimated from data by the maximum likelihood method. The fitted distributions are given in Figure 1

[5] Some may argue that E(R) should be taken as the true ratio. Our choice is the more convenient theoretically. In practice, the choice does not matter much, and the users are not likely not think in these terms anyway.
[6] The name risk distribution may be ill conceived. However, it aims to cover different conceptions for establishing a distribution to support claims and decisions taken.

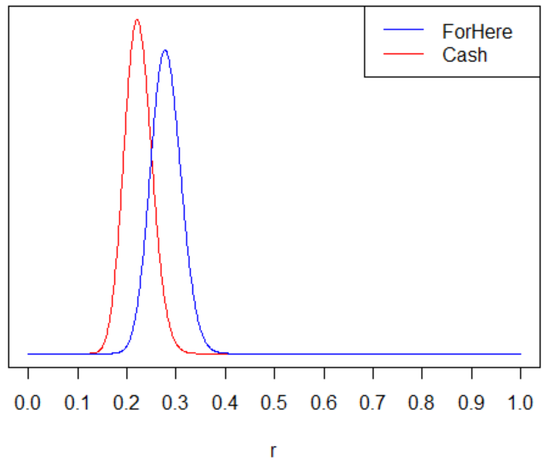

Figure 1. Fitted Beta distribution of R from Gamma-assumptions

This leads to simple evaluations based on the quantiles of the fitted Beta-distributions, easily obtained by most statistical software as given in Table 3.

Table 3. Quantiles of R obtained from fitted Beta-distributions

	ForHere				Cash			
Quantile	1%	5%	10%	20%	1%	5%	10%	20%
Beta-quantile	0.215	0.232	0.240	0.254	0.165	0.180	0.189	0.200

Question: To what extent is this simple calculation justified and fair to the taxpayer?

In case common scale is not realistic, the model may be widened to account for this, leading to an R with a distribution related to the so-called Beta-prime distribution. Although a more complicated distribution, the risk calculation is numerically simple and, for the user, the added complexity is hidden in the software. These two models consider, in a sense, separate incoming streams of customers for each type, say the staying customers and the take-away customers. Alternatively, one could imagine one stream of customers, and add a parameter representing the probability that an incoming customer is a take-away customer. Now the ratio R will have a complicated distribution, but still possible to handle analytically or by simulations. We have here assumed calculations within the classical (frequentist) statistical paradigm, where the model is estimated and then used for risk calculations. In case the Bayesian paradigm is favored, one may seek the posterior distribution of R , based on prior distributions on the underlying parameters. The posterior distribution thus obtained is not analytically tractable and may be obtained by simulation techniques. The implementation will be the same for the simple and widened models.

The assumption of Gamma-distributed sales may be tested on the sampled (although limited) data. Then the required distribution of R follows from the independence assumption by theory. This leads to relatively simple risk calculations. However, two interesting practical and theoretical issues have surfaced in connection with this approach:

1. The taking of sample size as given
2. The forced independence of S_1 and S_2

Taken as objections, a brief refutation may be as follows: (1) The sample size bears no relation to the parameter of interest. It is natural to follow the common practice in observational studies, where statisticians with few exceptions condition on the sample size.[7] Moreover, conditioning makes long run interpretations, like confidence limits, more relevant to the individual case under study.[8]
(2) Independence between customers purchase amounts is the natural assumptions. Clearly, positive dependence between of S_1 and S_2 will arise when calculated on daily basis, but this is due to the number of sales of each type is dependent on a common factor, namely the total number of customers on a specific day. However, our estimate by R is not dependent on this grouping. In fact, the sales amounts occur symmetrically, and may be aggregated in any order. What matters here is the total number of observations and not when they are taken.[9] We then imagine that the size of the study represented by total number of sales n to observe is prechosen and then then observe n_1 and n_2 and take them as fixed in the risk calculation. This is the basis for the simple risk calculation based on the Beta-distribution in Lillestøl (2019). It may be argued that this is unreasonable, and that the variation in the number of sales is relevant and should be taken into account in the risk calculation.

In the following, we will look further into the conditioning arguments for by studying models where n_1 and n_2 are not fixed, but random. In a sense, the two issues (1) and (2) are tied together, as correlation between S_1 and S_2 may be explained within a model where the number of customers of each type, say N_1 and N_2, are correlated random variables, for instance having some kind of bivariate Poisson distribution. Natural models with random number of sales will have the following density factorization

(1) $$f(x,n\,\varphi\,,\psi\,)=f_1(x|n;\varphi\,)\cdot f_2(n;\;\psi)$$

where x represents a vector of observables, including the sales amounts, and n represents the number of sales. Here φ and ψ are parameter vectors, φ of primary interest and ψ of minor interest (nuisance parameters). From this follows that inference about φ should be made from f_1 by conditioning and inference about ψ should be made from f_2 by marginalization.[10] [11]

Two classes of models may be imagined, either modelling one stream of customers or separate streams for the two opposing types of sales. In the former case, the type of sale may be represented by a probability parameter included in the parameter vector of interest φ. The issue is whether a given model fulfill (1) with $n=n_1+n_2$ or with $n=(n_1,n_2)$, that is, condition with respect to the total number of observations or possibly with respect to the individual numbers itself. In the probabilistic one-stream model mentioned above, equation (1) provides justification for conditioning on $n=n_1+n_2$ but not on $n=(n_1,n_2)$.

The structure of extended models allowing random number of observations are given in Figure 2 for one-stream and Figure 3 for two-stream. Here the square boxes contain observable random variables where X_i is the vector of N_i sales mounts of type i (i=1,2). The unboxed quantities are (vector)

[7] Statisticians frequently go beyond that, by taking the observed values of explanatory variables (the regressors) to be given as well (conditional inference).
[8] See the overview article in Statistical Science on the roles of conditioning in inference by N. Reid (1995), in particular the pages 138 and 154.
[9] However, sampling dates for visits reduces the risk collecting data from days with some special customer behavioral pattern. Three days are then a compromise between getting enough representative data and available resources.
[10] See formula (3.1) and accompanying text in Reid (1995).
[11] A more thorough discussion may involve conceptions of sufficiency and ancillarity.

parameters, either of importance or nuisance. The parameters of importance are those which determine the theoretical ratio ρ to be estimated by R (circled), which in turn is determined by the observable variables. The parameters of importance and the definition of ratio ρ may depend on the choice of model, and possibly on the position taken, i.e. whether one have decided to condition on the N's at the outset or not. The sales amounts are all assumed conditionally independent, given the prior quantities in the graph, and having common distribution with expectation μ_i expressed in terms of θ_i for each type of sale (i=1,2). In Figure 3 the dotted line indicates that correlation between N_1 and N_2 is induced by an additional parameter to the model.

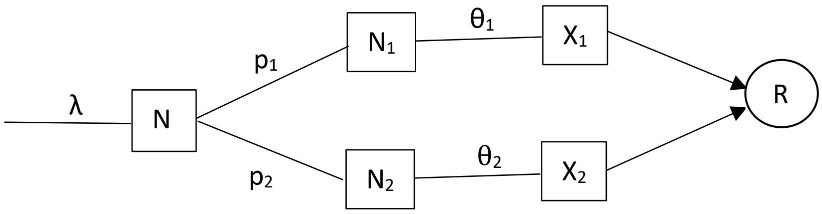

Figure 2. One-stream model

For the one-stream model in Figure 2, the parameters of interest will be $\varphi = (p_1, p_2, \theta_1, \theta_2)$ and the nuisance parameter $\psi = (\lambda)$. Here $\rho = \frac{p_1 \mu_1}{p_1 \mu_1 + p_2 \mu_2}$ is to be estimated. For this model, the joint density of the observables fulfills the factorization of formula (1) above, and the inference and risk calculations can be made conditional on N. This weakens the arguments for taking both n_1 and n_2 fixed as in the non-random case.

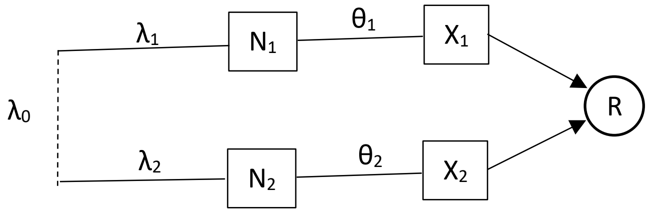

Figure 3. Two-stream model with correlation

For the two-stream model in Figure 3, the parameters of interest will be $\varphi = (\lambda_1, \lambda_2, \theta_1, \theta_2)$ and the nuisance parameter $\psi = (\lambda_0)$. Here $\rho = \frac{\lambda_1 \mu_1}{\lambda_1 \mu_1 + \lambda_2 \mu_2}$ is to be estimated. For this model, the joint density of the observables does not fulfill the factorization of formula (1) above, and inference conditional on the number of sales is unjustified.

In principle, one may adopt a model of the above kind, with distributional assumptions, more or less justified by data. We then have the opportunity to estimate the parameters of interest, and then ρ from the appropriate formula by plug-in, instead of using the simple ratio R of aggregated sales. We have essentially three levels of sophistication:

1. Model-based estimation and risk calculations
2. Estimate by R and do risk calculation in view of a model
3. Estimate by R and do simplistic/schematic risk calculation

In practice, the tax-authority analyst may prefer to stay clear of disputable models and use the empirical ratio R anyway. Some kind of schematic risk judgment may also be preferred. That is, the latter may be preferred in practice. The use of R with (n_1, n_2) fixed and look-up quantiles of the Beta-distribution is of this kind. Extended models trying to explain more, and in some cases irrelevant features of the data, will incur higher risks. On the other hand, neglecting sources of variation, say by unjustified conditioning, may lead to underestimation of the risks. One could hope that the risk calculated on different assumptions are not much different. This may settle any disputes on the use of simple models. However, if they are different, our investigations may indicate the size of added necessary safety margins.

This study will explore some of these issues, partly by theory and partly by simulations. The next section includes some basic statistical theory with specific distributional assumptions.

3. Some statistical theory

Initially we assume a fixed number of sales of each type, coming from Gamma-distributions, later to be used as conditional building blocks for the random models. For further details see Appendix.

Let S_i for i=1,2 be sums of n_i sales each with distribution $Gamma(\alpha_i, \beta)$, i.e. Gamma-distributed with common scale parameter β but possibly different shape parameters α_i. Then S_1 and S_2 are independent with S_i distributed $Gamma(n_i \alpha_i, \beta), i = 1{,}2$. Moreover, the distribution of $R = \frac{S_1}{S_1 + S_2}$ will be $Beta(n_1\alpha_1, n_2\alpha_2)$, not dependent on β. This provides a convenient distribution for calculating decision limits. The only thing we need is estimation of the two shape parameters.

Noting that $E(S_i) = n_i \cdot \left(\frac{\alpha_i}{\beta}\right), i = 1{,}2$, it follows that

$$E(R) = \frac{n_1\,\alpha_1}{n_1\,\alpha_1 + n_2\,\alpha_2} = \frac{E(S_1)}{E(S_1) + E(S_2)} = \rho,$$

In the case of common scale, the scale parameters cancels and R becomes an unbiased estimate of ρ. In a sense, we can interpret α_1 and α_2 as weighing factors applied to n_1 and n_2 in the natural estimate of the two types of sales, $\frac{n_i}{n_1 + n_2}$, i=1,2, when all sales amounts have the same expectation.

The above theory reflects a two-stream model with $n_1$and n_2 given. It may cause some concern that the expectation $E(R)$ depends on $n_1$and n_2, inasmuch R is intended to be an estimate of an expected yearly ratio. This may be understood in the following context: In effect, our target will be $\rho = \frac{p_1\,\alpha_1}{p_1\alpha_1 + p_2\alpha_2}$, where p_i is the probability that a sale is of type i (i=1,2). Then R is obtained by replacing p_1 and $p_2$by their observed estimates $\hat{p}_i = \frac{n_i}{n_1 + n_2}$, i=1,2. In this sense, the estimate R is asymptotically unbiased. In the case $S_i \sim Gamma(n_i\alpha_i, \beta_i)$, $i = 1{,}2$ with differing scale parameter, the ratio $R = \frac{S_1}{S_1 + S_2}$ no longer follows the standard a Beta-distribution. We now have

$$E(R) \neq \rho = \frac{E(S_1)}{E(S_1) + E(S_2)} = \frac{n_1\,(\alpha_1/\beta_1)}{n_1\,(\alpha_1/\beta_1) + n_2\,(\alpha_2/\beta_2)},$$

but a similar asymptotic interpretation may be given. Extension of the theory above may provide a suitable distribution for calculating decision limits for the case of not common scale as well (see Appendix).

The above considerations indicate some problems with the two-stream model. We will therefore examine a one-stream model in some generality. At the outset, we do not assume that the sales are Gamma-distributed. We take the total number of observations n as given, in order to contrast the case above when the number of both types are given.

Imagine one stream of sales and for each sale probabilities $p_1 = p$ and $p_2 = 1 - p$ of being of type 1 or type 2. The aggregated sales amounts may be represented by

$$S_1 = \sum_{i=1}^{n} I_i X_i^{(1)} \qquad S_2 = \sum_{i=1}^{n} (1 - I_i) X_i^{(2)}$$

Here $I_i = 1$ if the ith sale is of type 1 and $I_i = 0$ if the ith sale is of type 2. Assume that $I_i,\ i = 1,2, \ldots, n$ are independent with $E(I_i) = P(I_i = 1) = p,\ i = 1,2, \ldots, n.$ Furthermore assume $X_i^{(j)}, i = 1,,2, \ldots, n$ are independent identically distributed with expectation $\mu_j = E\left(X_i^{(j)}\right), j = 1,2.$ Moreover, assume that $(I_i,\ X_i^{(1)},\ X_i^{(2)})$ are mutually independent. Note that the expressions above are just formal representations where sales not belonging to the stream are excluded by zeros. This representation is useful for theory development and for simulation, as will be demonstrated in later sections. It follows that

$$R = \frac{S_1}{S_1 + S_2} \to \rho = \frac{p_1\, \mu_1}{p_1\, \mu_1 + p_2\, \mu_2} \text{ in probability}$$

In the case of Gamma-distributed sales, this is expressed by the Gamma-parameters as $\rho = \frac{p_1\, (\alpha_1/\beta_1)}{p_1(\alpha_1/\beta_1) + p_2(\alpha_2/\beta_2)}$, which is reduced to $\rho = \frac{p_1\, \alpha_1}{p_1\, \alpha_1 + p_2\, \alpha_2}$, in the case of common scale.

For this model we do not readily have a distribution for calculation of decision limits like the Beta distribution above. This can be overcome by various generally applicable numerical techniques, like numerical integration and simulation, as demonstrated in later sections.

Let us end this section by preparation for the extension to models where the number of sales is taken to be random. This is dealt with for both one-stream and two-stream models in separate sections below, where issues of dependence are examined. With one stream of customers arriving randomly, we may assume that the number of sales N in the sampled period is Poisson distributed with parameter λ, in short $N \sim Poisson(\lambda)$. We then have

$$P(N = n) = \frac{\lambda^n}{n!} e^{-\lambda}; n = 0, 1, 2, 3, \ldots$$

Then $E(N) = V(N) = \lambda$, i.e. expectation and variance coincide. If we, as above, assume the type of sale are determined randomly with probabilities p_1 and p_2, then the number of sales of each type in the sampled period will respectively be $N_1 \sim Poisson(\lambda\, p_1)$ and $N_2 \sim Poisson(\lambda\, p_2)$ and independent. The parameter of interest is now

$$\rho = \frac{\lambda p_1\, \mu_1}{\lambda p_1\, \mu_1 + \lambda p_2\, \mu_2} = \frac{p_1\, \mu_1}{p_1\, \mu_1 + p_2\, \mu_2}$$

i.e. the same as in the fixed n case, as λ cancels out.

In the case of separate streams of customers for the opposing types of sales, the Poisson distribution may serve a building block for constructing a bivariate Poisson distribution for the number of sales of each type N_1 and N_2 with correlation, as will be seen in section 7.

4. Risk calculation by resampling

Risk calculations are based on assumptions involving the elements: conditioning, independence and distribution. In Lillestøl (2019) the risk calculations for R are, in the common frequentist setting (section 5), based on the Gamma assumption. Conditional on the number of observations of each type this leads to a Beta-distribution for R in case of common scale and its extension for different scale. If the model is extended, taking the total number as given, and adding a parameter for type 1 (say), the unconditional distribution of R is no longer Beta. In this case, measures of the uncertainties may be derived by numerical integration or approximate reasoning. However, both are heavily dependent on the model chosen and its technicalities. A more widely applicable approach would be to judge the uncertainties by resampling. Different ways of doing this exist, as well as different modes of the calculation from each resample. We have mainly the following:

Modes of resampling:

A. Resampling directly from the set of observations with replacement
B. Resampling from the estimated model, say of Gamma-type

Modes of calculation:

1. Direct computation of R
2. Compute parameter estimates of model and estimate ρ by plug-in.

The risk calculation in Lillestøl (2019) in the frequentist setting are of type B1, but examples of B2 are given in the Bayesian setting. Note that the combination A1 offers the opportunity to do risk calculations without distributional assumptions, and example will be given below. The plug-in estimates may seem impractical but may be used for checks of internal consistency. Within a Bayesian context, where parameters typically are computed by MCMC-methods, the plug-in may seem more natural. We then have samples from the joint posterior of the parameters, and plug-in then gives a sample from the posterior of ρ as well.

We will now examine some variants of the models and their risk calculation, both within the frequentist and the Bayesian framework. Among others, we will contrast the case when the number of observations are taken as given and the case when the number of observations is taken as random determined by a Poisson-model.

First, we consider resampling with replacement from the observed data and calculation of R (case A1). We consider three situations: (i) fixed with (n_1, n_2)=(22, 45) in the case of ForHere/ToGo and with (n_1, n_2)= (19, 48) in the case of Cash/Card, (ii) fixed n= n_1+n_2 =67 and (iii) random N. In the cases (ii) and (iii) we estimate the probability of a type 1 sale as follows: $p = \frac{22}{67}$ in case of ForHere, and $p = \frac{18}{67}$ in the case of Cash. In case (iii) we use λ =67 for the expected number of sales in a period of length as given.

A resampling with 10 000 repeats was made to obtain an empirical distribution of R. Histograms for case (ii) of fixed total n are given in Figur 4 (bin size 0.05), where the curves (in red) are the smoothed versions. Histograms for case (iii) are similar with seemingly the same dispersion, but for cases (i) they are less dispersed. This is seen in Figure 5 where the smoothed histograms for case (i) and (ii) are displayed together, and where case (iii) will overlap the red curve for case (ii) and is not displayed.

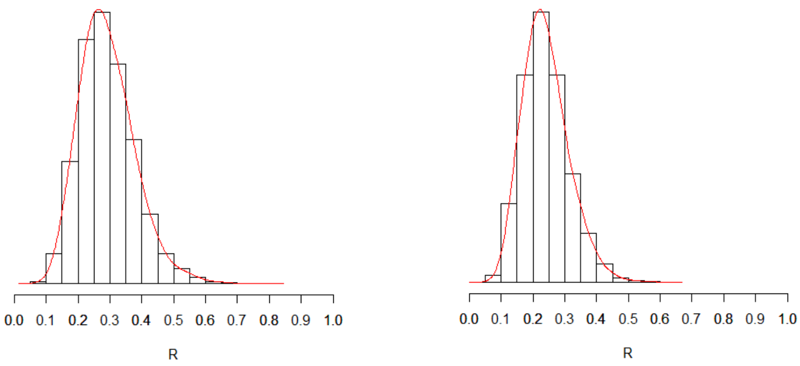

Figure 4. Histograms of R by resampling from actual observations for fixed n: (Left: ForHere/ToGo, Right: Cash/Card)

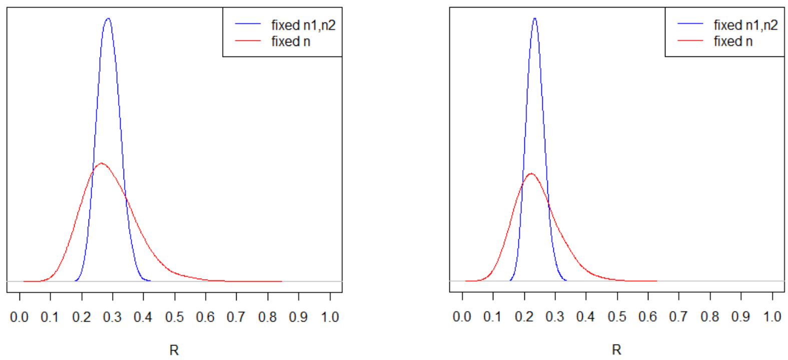

Figure 5. Distributions of R by resampling from actual observations: (Left: ForHere/ToGo, Right: Cash/Card)

We see clearly the difference between conditioning on (n_1, n_2) and on just $n=n_1+n_2$. It may come as a surprise that the distribution in case of random N is not noticeable different from fixed n.[12] For making decisions we are interested in the lower quantiles of the distributions. They are given as in Table 4.[13]

[12] The graphs indicate that the distribution is slightly shifted upward in the case of conditioning on both n_1 and n_2. This may affect the tail probabilities, but our investigations go directly to these, since they are the main basis for decisions.

[13] The boldfaced line in this and later tables indicate the methods that stand out from the discussion at the end.

Table 4. Quantiles of R: Resampling with replacement from actual observations (A1)

	ForHere				Cash			
Quantile	1%	5%	10%	20%	1%	5%	10%	20%
(i) fixed n_1 n_2	0.222	0.238	0.248	0.260	0.179	0.193	0.201	0.213
(ii) **fixed n**	**0.139**	**0.175**	**0.194**	**0.222**	**0.103**	**0.135**	**0.153**	**0.178**
(iii) random N	0.137	0.172	0.193	0.220	0.103	0.136	0.156	0.181

First, we note that case (i) gives quantiles on the level of the Beta-quantiles given in Table 3. In fact, they are slightly above and therefore less favorable to the taxpayer. Next, we see a striking difference between case (i) and (ii) i.e. fixing the number of sales of each type and fixing the total number of sales. Quantiles for the latter case are far below the corresponding ones for the former, reflecting the added uncertainty with one more parameter. On the other hand, we see that the quantiles for the random case are not much different from the fixed total case, despite the adding of one more parameter. We have argued above in favor of conditioning on the total number of sales. In practice, this gives the taxpayer a large benefit of doubt, compared to conditioning on the number of sales of each type.

A commentary in Lillestøl (2019) may leave the impression that the fixed assumption does not matter in the numerical sense. We see here that this is not the case. In fact, choosing one conception over the other may lead to different conclusions. In the current case, where the ratios reported by the taxpayer are 10.3% ForHere and 12.7% Cash, the choice would have no implications when using common 5% risk level. However, at the 1% risk level, the low Cash% quantile comes to rescue for the taxpayer.

A demonstration that the choice of model does not matter for the conclusion of guilt may of course strengthen the case for the tax authority. However, one should be prepared to handle the case when different approaches lead to different conclusions. What if taxpayer had reported ratios of about 20%? We are then back to the validity of the conditioning argument given above, stating that the statistically sound practice is to report conditioned on the total number of observations.

We have seen that adding a risk element for type of sale have led to wider distribution, and lower quantiles (in favor of the taxpayer). However, adding another risk element, the total number of sales (keeping its expectation fixed), did not alter the risks noticeably. It is of some interest to see what happens if we go in the other direction and remove a risk element. This will be the case if the amounts are neglected altogether, and the fraction $n_1/(n_1 + n_2)$ is used as estimate. This may be justified if we knew that all sales amounts are (approximately) equal. We then have a model with just one parameter (p) in the fixed case, and two parameters (p and λ) in the random case. We then get the quantiles given in Table 5.

Table 5. Quantiles: Resampling with probabilities as observed fraction (neglecting amounts)

	ForHere				Cash			
Quantile	1%	5%	10%	20%	1%	5%	10%	20%
Fixed n	0.209	0.239	0.254	0.284	0.164	0.194	0.209	0.239
Random N	0.200	0.235	0.254	0.280	0.160	0.194	0.203	0.237

Now the quantiles are just below the ones in Table 4, the case of fixing the number of sales of both opposing types, but taking advantage of the amounts This illustrates (i) that the variation in the sales amounts is a major component of the risk and (ii) the danger of assuming unjustified simplicity and end up with illusionary low risks. On the other hand, the number of sales of the opposing types are the main determinants of the level of the ratio.

In the following sections we will examine separately the role of conditioning and independence within random arrival models, first one stream model and then the two-stream model. For each model, we first examine some of its properties, in general and then adding the assumption of Poisson arrivals and Gamma sales amounts. Then follows actual quantile calculations with comparisons for different modes of calculation. Both frequentist and Bayesian calculations are examined.

5. One stream random arrival model: Properties

In this section we will examine some consequences of taking the total number of sales N random. At the outset, we do not specify the distribution of N or the distribution of the sales amounts but, for simplicity, we assume that all sales amounts (type 1 or type 2) comes from a common distribution.

Let the observations be $(X_{i,}I_{i,})$; $i = 1,2 \ldots., N$, where $X_{i,}$ is the ith sales amount and $I_{i,}$ is the indicator for type of sale, i.e. equal to 1 for type 1 sale and equal to 0 for type 2 sale. The total number of sales N is assumed random. The aggregated sales amounts of each type may now be written

$$S_1 = \sum_{i=1}^{N} I_i X_i \qquad S_2 = \sum_{i=1}^{N} (1 - I_i) X_i$$

Assume that the sales amounts are independent from a common distribution with expectation $E(X)$ and variance $V(X)$, and that the type of sales (e.g. cash or card) are independent and independent of its amount and determined by a common probability p of type 1 sale. Assume further that the total number of sales is determined independent of the other variables by a distribution with expectation $E(N)$ and variance $V(N)$. In particular, it is of interest to examine the case when N is Poisson-distributed.

The expected sums of sales amounts are

$$E(S_1) = p \cdot E(N) \cdot E(X) \qquad E(S_2) = (1 - p) \cdot E(N) \cdot E(X)$$

and their variances are

$$V(S_1) = p \cdot E(N) \cdot V(X) + (p \cdot (1 - p) \cdot E(N) + p^2 \cdot V(N)) \cdot \big(E(X)\big)^2$$

$$V(S_2) = (1 - p) \cdot E(N) \cdot V(X) + (p \cdot (1 - p) \cdot E(N) + (1 - p)^2 \cdot V(N)) \cdot \big(E(X)\big)^2$$

The covariance becomes

$$C(S_1, S_2) = p \cdot (1 - p) \cdot (V(N) - E(N)) \cdot (E(X))^2$$

In the case of N Poisson-distributed with parameter λ we have $E(N) = V(N) = \lambda$, and get

$$V(S_1) = \lambda \cdot p \cdot (V(X) + \big(E(X)\big)^2)$$

$$V(S_2) = \lambda \cdot (1 - p) \cdot (V(X) + \big(E(X)\big)^2)$$

We see that the covariance is zero in the Poisson case. For distributions with $V(N) > E(N)$, so-called overdispersion, the covariance is positive. A prominent example of this, is the negative binomial distribution, which offers a good alternative to the Poisson distribution, in case data indicates overdispersion. For $V(N) < E(N)$, the covariance becomes negative, which in particular is so for $V(N) = 0$. However, this is the unconditional distribution. After conditioning on N we have independence. With the Poisson assumption we have unconditional independence as well.

The expectation and variance of R are not easily derived, but first order approximations are given by

$$E(R) \approx p \qquad V(R) \approx \frac{p(1-p)}{E(N)}\left(1 + \frac{V(X)}{(E(X))^2}\right)$$

Example 1: Let $E(N) = 60$, p= 0.3, $E(X) = 200$ and $V(X) = 100^2$, i.e. the standard deviation is half the expected sales amount. We then get $V(R) \approx \frac{0.3\cdot 0.7}{60}\left(1 + \frac{100^2}{200^2}\right) = 0.004375$, so that the standard deviation of the ratio is approximately is 0.065. When sales are Gamma distributed, we may simulate the distribution of R by taking the shape parameter $\alpha = 4$ and scale parameter $\beta = 0.02$. In addition to the case with N being Poisson with expectation 60, we simulate the case with fixed n at 60 and the case with both types fixed at 18 of type 1 and 42 of type 2, at fraction 0.3. Simulation with 10 000 repetitions gave the smoothed histograms in Figure 5 representing the probability density of R.

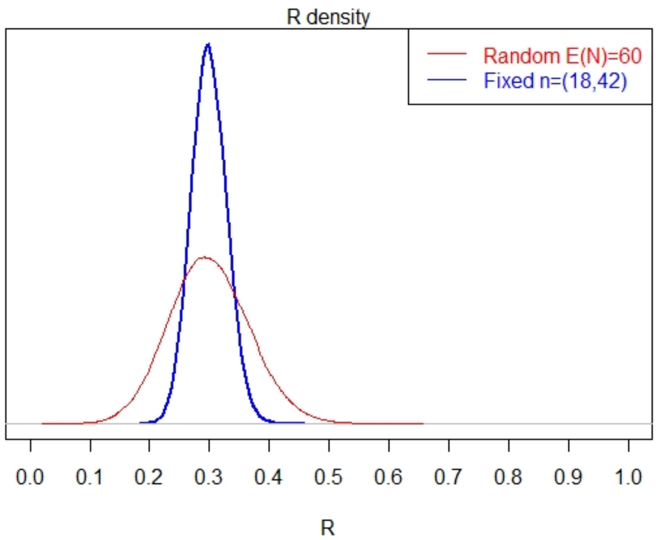

Figure 5. Simulated distributions of R as smoothed histograms for Example 1

We see the narrow distribution in the fixed n=(18,42) case in comparison with the wider distribution in the random E(N)=60 case. It turns out that the fixed total n=60 case is hardly distinguishable from the random case and is not shown. Lower quantiles for the three cases are given in Table 6.

Table 6. Quantiles of simulated distribution of R for Example 1

Quantile	1%	5%	10%	20%	50%
Fixed n=(18,42)	0.234	0.251	0.261	0.275	0.299
Fixed n=60	0.156	0.194	0.216	0.242	0.300
Random E(N)=60	0.151	0.191	0.215	0.243	0.299

We see that the consequence of assuming the total number of sales fixed, which may be justified as above, is only microscopic different from allowing the total to be random. On the other hand, the consequence of assuming the number of sales of each type fixed, is dramatic in disfavor of the taxpayer.

6. One stream random arrival model: Risk calculations from the data

We will now examine whether the conclusions of the preceding sections remain the same for estimation and risk calculations of various types for our data, assuming Gamma-distributed sales. We will examine this within both the frequentist and Bayesian framework.

Frequentist framework

For both situations, fixed and random, we assume the individual sales are Gamma-distributed. A common scale was justified for the ForHere/ToGo typology, while different scales were required for the Cash/Card typology. Parameter estimates obtained by maximum likelihood are given in Table 7.

Table 7. Estimates of Gamma-parameters: Shape and Scale

Type of sale	ForHere	ToGo	Cash	Card
Shape	2.72	3.43	1.86	4.72
Scale	0.0150	0.0150	0.0097	0.0210

Remark. As check we compute the Cash-ratio estimate $\rho = \frac{\frac{18}{67}\cdot\frac{1.86}{0.0097}}{\frac{18}{67}\cdot\frac{1.86}{0.0097}+\frac{49}{66}\cdot\frac{4.72}{0.0210}} = 0.252$, which is right on the direct calculation of R in Table 2. For the For-Here ratio, using the constant scale formula, we get $\rho = \frac{\frac{22}{67}\cdot 2.72}{\frac{22}{67}\cdot 2.72+\frac{45}{66}\cdot 3..43} = 0.279$, which is somewhat lower than the direct calculation. This illustrates that it may be worthwhile to adopt the non-restricted model to have internal consistency.

The resampling is performed as follows: Type of sale is generated using p equal to the fraction observed and amounts according to a Gamma-distribution with parameters as estimated in Table 7. The total number of observations are the actual n=67 (fixed case) or generated according to a Poisson distribution with expectation 67 (random case). With this data, we either calculate R directly (case B1) or estimate Gamma-parameters and use the formula for ρ (case B2). That is

$$\rho = \frac{p_1\cdot\frac{\alpha_1}{\beta_1}}{p_1\cdot\frac{\alpha_1}{\beta_1}+p_2\cdot\frac{\alpha_2}{\beta_2}} \text{ or } \rho = \frac{\lambda_1\cdot\frac{\alpha_1}{\beta_1}}{\lambda_1\cdot\frac{\alpha_1}{\beta_1}+\lambda_2\cdot\frac{\alpha_2}{\beta_2}} \text{, whichever appropriate}$$

A resampling with 10 000 repeats was made to obtain an empirical distribution, from which we obtain the lower quantiles as given in Table 8.

Table 8. Quantiles: Four frequentist modes of calculation

	ForHere				Cash			
Quantile	1%	5%	10%	20%	1%	5%	10%	20%
B1 Fixed n	**0.149**	**0.185**	**0.204**	**0.228**	**0.120**	**0.155**	**0.175**	**0.200**
B1 Random N	0.147	0.184	0.204	0.229	0.121	0.154	0.173	0.199
B2 Fixed n	0.152	0.186	0.204	0.228	0.123	0.157	0.175	0.199
B2 Random N	0.149	0.185	0.204	0.229	0.120	0.156	0.175	0.200

We see, not too surprisingly, that the two modes of calculation (B1 and B2) give almost identical quantiles. Again, it seems that taking the total number of sales random does not noticeably affect the calculated risks.

Remark. The One-stream Gamma model with type of sale taken to be random allows analytic probability calculations by conditioning on the number of Type 1 sales n_1 for given total number of sampled sales n. This opportunity is briefly addressed and dismissed in the Appendix.

The Bayesian framework

Consider the Bayesian formulation with sales independent Gamma-distributed with different shape and scale-parameters for each type of sale, and with type (e.g. Cash/Card) determined by independent Bernoulli-variables with constant probabilities (p, 1-p). This gives a model with five parameters, for which we assume independent non-informative priors.[14] The posterior distributions of the parameters are obtained by MCMC simulations, using the R2OpenBUGS implementation of the WinBUGS package, see Sturz et.al. (2005). A sample of 10 000 vectors from the joint posterior distribution was obtained. The means of the marginal posteriors (Bayes estimates) are given in Table 9.

Table 9. Means of marginal posteriors of Type 1 probability and the Gamma-parameters

Type of sale	ForHere	ToGo	Cash	Card
Prob	0.338	0.662	0.283	0.717
Shape	2.48	3.54	1.79	4.65
Scale	0.0136	0.0156	0.0094	0.0207

Each sample from the joint posterior distribution of the model parameters may be used to provide samples from the posterior av the quantity of interest. We have essentially two different methods: The plug-in method is repeated calculations of ρ from the formula $\rho = \frac{p_1\,(\alpha_1/\beta_{)})}{p_1(\alpha_1/\beta_{1)} + p_2(\alpha_2/\beta_{2)}}$ and the simulation method is repeated calculation of R from simulated sales amounts for each sample from the joint posterior of the parameters. While the former method aims at a posterior parameter distribution, the latter aims at the posterior predictive distribution, which is slightly different. From

[14] This is implemented by assuming a Gamma(0.001, 0.001) prior for each of the four Gamma-parameters, that is with expectation 1 and variance 1000, and a uniform prior over the [0,1]-interval for the Bernoulli-parameter.

the 10 000 derived observations of the quantity of interest, we obtained the quantiles given in the first two rows of Table 10. [15]

Table 10. Quantiles: Three Bayesian modes of calculation

Variable	ForHere				Cash			
Quantile	1%	5%	10%	20%	1%	5%	10%	20%
Plug-In	**16.6**	**19.7**	**21.7**	**24.1**	**13.3**	**16.2**	**17.8**	**20.0**
Predictive	15.1	18.6	20.6	23.1	11.3	14.4	16.3	18.8
Random Pred	11.2	15.9	18.5	21.9	8.4	12.3	14.7	17.8

For both methods, plug-in and predictive, the total number of sales is taken as fixed at n= 67. We see that the quantiles, obtained by conditioning, are slightly different from the corresponding frequentist ones in Table 8. They were not expected to be identical, since the ways of generating repeats are different. In the Bayesian case, the repeats come from the MCMC algorithm, generating samples from the joint posterior. It is likely that this makes more effective use of the correlation of the parameters. This have turned out differently for the two methods. For the plug-in method, quantiles are raised for both ForHere and Cash. For the predictive method, the quantiles are lowered for Cash, but not appreciably for ForHere. For illustration we have added calculations for the predictive case, taking the number of sales not fixed, coming from a Poisson distribution with expectation 67. Contrary to the frequentist calculations above, this lowers the quantiles appreciably. However, since we have justified conditioning on the total, this causes no worry.

The tax authority will typically have prior information of the sales amounts. This may be quantified in terms of informative prior distributions. The obtained posteriors may be interesting, at least for throwing light on their suspicion, but may be dismissed as evidence in preparing the case against the taxpayer.

Well justified calculations are those in Table 8 and 10 marked in boldface. Suppose the decision limit is the common 5% quantile, and recalling the reported ForHere 10.3% and Cash 12.7%, we see that this is lower whatever justifiable calculation chosen, and therefore nails the taxpayer. If the more conservative 1% quantile is taken decision limit, it remains so for ForHere, but for Cash the decision will depend on which of the justifiable calculations chosen.

7. Two-stream random arrival model: Properties

In this section and the subsequent one we will assume that the sales of the opposing two types comes from separate streams. This may not be reasonable in the context of the case at hand but may be so in other contexts. As indicated above, this leads to a narrower risk distribution, less favorable to the taxpayer in the case of fixed n_1 and n_2. We have argued that this is unreasonable, and that the number of sales should be taken as random and now denoted N_1 and N_2 (see Figure 2). The total sales amounts of each type are written as

$$S_1 = \sum_{j=1}^{N_1} X_j^{(1)} \qquad S_2 = \sum_{j=1}^{N_2} X_j^{(2)}$$

[15] In Lillestøl (2019), the plug-in case (first row) was erroneously reported as the predictive case (second row, not reported). However, the difference is minor and does not alter the conclusions.

Assume that the sales amounts $\{X_j^{(i)}; j = 1,,2,..,N_i; i = 1,2\}$ are all independent and with common distribution for each stream independent of N_1 and N_2, which are possibly positively correlated. Let for i=1,2 $E(N_i) = \lambda_i$ and $E\left(X_j^{(i)}\right) = \mu_i$. It follows that the expectation, variance and covariance of the sum of sales are

$$E(S_i) = E(N_i) \cdot E\left(X_j^{(i)}\right)$$

$$V(S_i) = E(N_i) \cdot V\left(X_j^{(i)}\right) + V(N_i) \cdot (EX_j^{(i)})^2$$

$$C(S_1, S_2) = EX_j^{(1)} \cdot EX_j^{(2)} \cdot C(N_1, N_2)$$

With N_1 and N_2 independent, S_1 and S_2 are still independent, but correlation between N_1 and N_2 implies correlation between S_1 and S_2 as well. We now add distributional assumptions as follows:

Assume for i=1,2 $X_j^{(i)} \sim Gamma(\alpha_i, \beta_i); j = 1,2, \dots, N_i$ where $N_i \sim Poisson(\lambda_i)$. Positive correlation is achieved by taking $N_i = N_{00} + N_{0i}$; $i = 1,2$, with independent Poisson variables $N_{0i} \sim Poisson(\lambda_{0i}); i = 0,1,2$, so that $\lambda_i = \lambda_{00} + \lambda_{0i}; i = 1,2.$ [16] With the assumptions above, the covariance becomes $C(N_1, N_2) = \lambda_{00}$ and the correlation is therefore given by:

$$\rho(N_1, N_2) = \frac{\lambda_{00}}{\sqrt{\lambda_1} \cdot \sqrt{\lambda_2}}$$

It follows, with shorthand notation $\mu_i = EX_j^{(i)} = \alpha_i/\beta_i$, that the expectations, variances, covariance and correlation for the total amounts of each type is given by

$$E(S_i) = \lambda_i \cdot \mu_i \qquad V(S_i) = \lambda_i(1 + \alpha_i) \cdot \frac{\alpha_i}{\beta_i^2}$$

$$C(S_1, S_2) = \lambda_{00} \cdot \mu_1 \cdot \mu_2$$

$$\rho(S_1, S_2) = \sqrt{\frac{\alpha_1 \alpha_2}{(1 + \alpha_1)(1 + \alpha_2)}} \cdot \rho(N_1, N_2)$$

The parameter of interest is now[17]

$$\rho = \frac{\lambda_1 \cdot \mu_1}{\lambda_1 \cdot \mu_1 + \lambda_2 \cdot \mu_2} = \frac{\lambda_1 \cdot \frac{\alpha_1}{\beta_1}}{\lambda_1 \cdot \frac{\alpha_1}{\beta_1} + \lambda_2 \cdot \frac{\alpha_2}{\beta_2}}$$

The expectation of R is not readily available, but first order Taylor approximation shows that $E(R) \approx \rho$. Second order approximation indicates that there will be a bias proportional to $\lambda_{00}(\mu_1 - \mu_2)$, i.e. likely underestimation in the case of $\mu_1 < \mu_2$, and positive correlation.[18] The approximation also indicates that the variance of R will decrease with $C(N_1, N_2)$.

---

[16] For application and software for this distribution see Karlis and Ntzoufras (2005)

[17] Do not confuse the two different uses of the Greek letter ρ.

[18] Note that the above provides approximations for the unconditional finite sample distribution and does not provide any asymptotic results.

The question is now how much this will affect conclusions and which of the parties will be favored. Relevant here is not so much the mean, but the (estimated) tail probabilities.

We will now study the effect of correlation on the yearly (population) level and in the context of a sample of days and projections made for the year. The study is based on simulations, which go as follows:

1. Simulate N_{0i} for i=0,1,2 and compute N_1 and N_2 .
2. Simulate $X_j^{(i)}$ for j=1,2, …,N_i i=1,2 and compute S_1 and S_2
3. Repeat 1-2 for the number of periods and add to S_1 and S_2
4. Compute ratio $R = S_1/(S_1 + S_2)$

We will present a numerical example with specifications close to the actual case in Lillestøl (2019). Here the sample was three days from a population of 360 days. With day as period, we may repeat and then accumulate the results. However, due to the additive property of both the Poisson distribution and the Gamma-distribution, this is not necessary, as long as the distribution parameters for the days are the same. For convenience, we may just as well take three days as the period length and a year of 120 periods.

Example 2: Assume $\alpha_1 = \alpha_2 = 4$ and $\beta_1 = \beta_2 = 0.02$, so that sales are Gamma distributed with expectation 200 and standard deviation 100, regardless of type. Assume that $\lambda_1 = 21$ and $\lambda_2 = 39$, so that the total expected number of sales is 60.[19] For chosen λ_{00} we have the correlation

$$\rho(N_1, N_2) = \frac{\lambda_{00}}{\sqrt{21}\sqrt{39}} = 0.0349 \cdot \lambda_{00}$$

Now $\lambda_{00} = 6$ implies $\lambda_{01} = 15$ and $\lambda_{02} = 33$ and $\rho(N_1, N_2) = 0.209$, while $\lambda_{00} = 9$ implies $\lambda_{01} = 12$ and $\lambda_{02} = 30$ and $\rho(N_1, N_2) = 0.314$. In any case, the expected totals are $\lambda_1 \cdot \mu_1 = 4200$ and, $\lambda_2 \cdot \mu_2 = 7800$ with standard deviations $\sigma_1 = 1024.7$ and $\sigma_2 = 1396.4$, respectively, so that

$$\rho = \frac{4200}{4200 + 7800} = 0.35$$

This case will be compared with the case of $\lambda_1 = 21$ and $\lambda_2 = 39$ with no correlation, I.e. $\lambda_{00} = 0$.

Simulations of 10 000 repeats for the given specifications and calculation of R turned out means slightly below 0.35 for all specifications of correlation, in fact 0.3495 in the case of no correlation, and close to that in the cases of correlation as well. The standard deviations were equal to 0.0693 in the case of no correlation, and 0.0637 and 0.0605 respectively for the cases of increasing correlation. Thus, the mean of the distribution of R is seemingly not affected by the correlation. On the other hand, the spread of the distribution is clearly decreasing by increasing positive correlation. However, the effect is minor. This is illustrated in Figure 6 by smoothed histograms of the simulated R, to represent the true probability density of R for the case of covariance 0 and 9, i.e. correlation 0.314.[20]

[19] Splitting into daily expectations this is 7+7+7 and 13+13+13 respectively.

[20] The corresponding calculation of R based on simulation of 1000 repeats for one year of daily data averaged 0.3499 with standard deviation 0.0058. Taking the average as an estimate of the theoretical ρ, its standard error is about 0.0006. Noting that our parameter specification leads to $\rho = 0.35$, this is taken as an empirical verification of the approximate expectation formula.

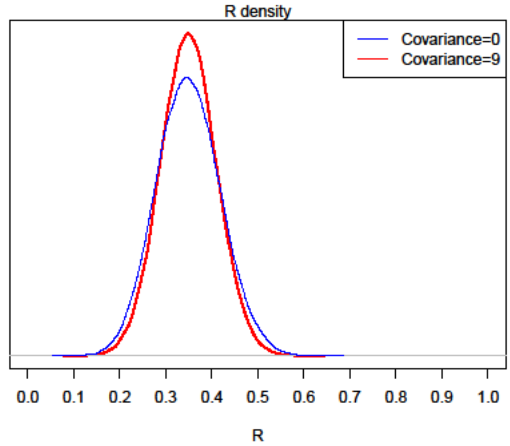

Figure 6. Distribution of R for zero correlation versus moderate correlation

Lower quantiles of the distribution are for the cases $\lambda_{00} = 0, 6, 9$ in Table 11.

Table 11. Quantiles: Three different correlation

Quantile	1%	5%	10%	20%	30%	40%	50%
$\lambda_{00} = 0$	0.1950	0.2385	0.2615	0.2910	0.3129	0.3312	0.3490
$\lambda_{00} = 6$	0.2038	0.2455	0.2680	0.2955	0.3153	0.3324	0.3490
$\lambda_{00} = 9$	0.2090	0.2503	0.2725	0.2985	0.3176	0.3338	0.3490

We see that the introduction of positive correlation in the data, only reduces the variability of R to a minor degree.

8. Two-stream random arrival model: Risk calculations based on the data

In this section we will examine to what extent the introduction of random number of sales in the two-stream model will widen the risk distribution and make it more favorable to the taxpayer. Jn order to calculate the risk distribution for the two-stream Poisson-Gamma model, we need estimates of the three additional parameters of the bivariate Poisson distribution. In principle, they may be obtained by the EM-algorithm, e.g using the algorithm of Karlis (2003). However, in our data we only have three days of observed number of sales of the opposing types, which is the minimum for carrying this any further.[21] It turns out that values close to those of Table 12 give a good fit to the data, simplified for illustrative purposes. The table shows no correlation between the number of ForHere and ToGo customer. The correlation between the number of Cash and Card customers is $19/\sqrt{19 \cdot 48} = 0.63$.

[21] In technical terms we are left with zero degrees of freedom, with no opportunity to judge the uncertainty.

Table 12. Well-fitted Lambda-parameters

Type of sale	ForHere/ToGo	Cash/Card
λ_{00}	0	19
λ_{01}	22	0
λ_{02}	45	29

The risk distribution is established by simulating the model with Gamma-assumptions (using the parameter estimates given above in Table 6 followed by direct computation of R. First (n_1, n_2) is taken as fixed and then with (N_1, N_2) simulated as dependent Poisson variables from lambdas as given in Table 12. The distributions are displayed in Figure 7 for ForHere/ToGo (left and Cash/Card (right). Quantiles of the distribution are given in Table 13.

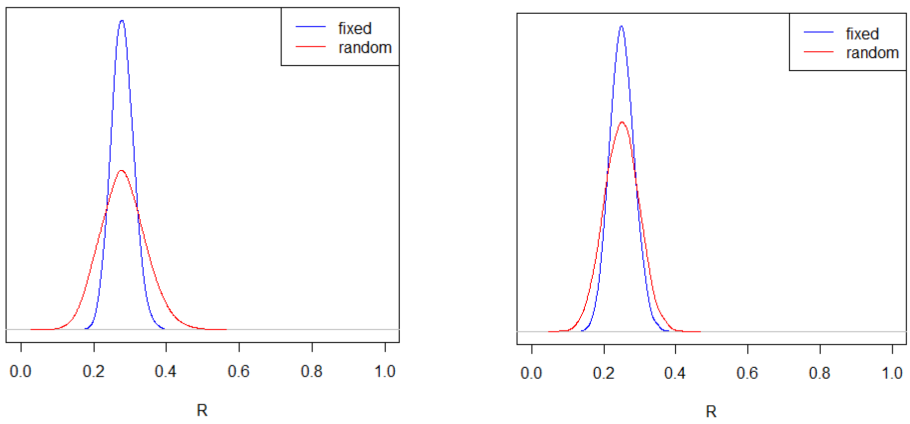

Figure 7. Densities of simulated R for ForHere/ToGo (left) Cash/Card (right)

Table 13. Quantiles for two-stream frequentist model

	ForHere				Cash			
Quantile	1%	5%	10%	20%	1%	5%	10%	20%
Fixed	0.212	0.229	0.240	0.253	0.180	0.197	0.208	0.223
Random	0.148	0.183	0.203	0.230	0.139	0.172	0.189	0.211

We see that fixed computation gives relatively high lower quantiles to the disfavor of the taxpayer. The introduction of randomness has lowered the quantiles considerably, more so for ForHere (independent arrivals) than for Cash (dependent arrivals), in line with theory indicating that the variance decreased with increased correlation. Note that the effect of random arrivals for the two-stream model is different from the one-stream model, where random N and fixed n gave approximately the same quantiles. For the one-stream model, the ForHere 1% quantile was about 0.15 and for Cash about 0.12. The above shows again that it would be unfair to the taxpayer to condition on both numbers. Here the adding of randomness obtains about the same fairness as conditioning on n in the one-stream model, where we had no need to add randomness to the total number of customers.

Bayesian approach

For the Bayesian approach two calculations are presented with the Poisson-Gamma distributional assumption, one with (n_1,n_2) fixed and one the full model with random (N_1,N_2). Again, non-informative priors are used as input to an MCMC-algorithm.

The posterior distribution of ρ is obtained by the plug-in method and portrayed in Figure 8. The corresponding lower quantiles are given in Table 14.

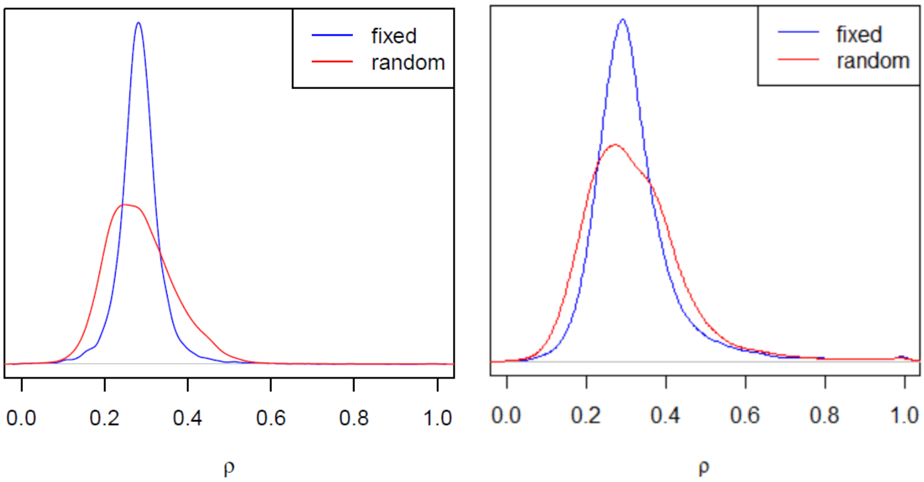

Figure 8. Posterior densities of Rho for ForHere/ToGo (left) Cash/Card (right)

Table 14. Quantiles for Bayesian two-stream model (plug-in approach)

	ForHere				Cash			
Quantile	1%	5%	10%	20%	1%	5%	10%	20%
Fixed	0.141	0.206	0.229	0.251	0.133	0.190	0.217	0.246
Random	0.125	0.170	0.192	0.216	0.110	0.158	0.187	0.222

We see that the low quantiles for the two cases are lower than the corresponding frequentist case. This may reflect that the Bayesian approach with a non-informative prior is better to pick up the true joint variability involving the lambda-parameters, where the frequentist estimation is based on data with zero degrees of freedom. In particular, note the high quantiles in the frequentist fixed case, which are clearly deceptive.

9. Conclusion

In this paper we have examined the consequences of several modes of calculating decision limits, based on different models and different conditioning, in the frequentist as well as in the Bayesian context. Theoretical analysis as well as computations have led to the conclusion in favor of conditioning with respect to the total number of sales and this gives similar results across different modes of computation, by resampling from actual observations or based on the one-stream model with Gamma-distributional assumption, direct simulation or plug-in, frequentist or Bayesian. This have turned out justifiable decision limits, fair to the taxpayer. It is demonstrated that conditioning on both types will typically lead to decision limits unfair to the taxpayer, as the case with the two-stream model with Gamma-Beta theory. Recommended methods are the ones boldfaced in Tables 4, 8 and 10. The choice may depend on the context and to what extent the parties and the decision maker (e.g. a court) will understand and accept evidence based on statistical assumptions. With risk of low acceptance, the resampling method of Table 4 may be a wise choice, but with thrust that more advanced methods are accepted, the methods displayed boldface in Table 8 or Table 10 may be defensible, in accordance with the tentative conclusions in Lillestøl (2019). Note however that their quantiles do not coincide completely. Of course, the best situation would be that the conclusion drawn from the sample evidence is the same irrespective of choice between the three recommended modes of computation (or other reasonable modes as well). Then, of course, the method simplest to communicate is chosen, which can be supported by another reasonable one, if challenged. For the current case the reported 10.3% ForHere sale is below the 1% quantile for all three recommended methods and the conclusion of underreporting is well supported. The reported 12.7% Cash-sales is below the 5% quantile for any of the three methods, and is close to the 1% quantiles for the two methods based on the distributional assumption, but is above for the resampling method. Thus, a definite conclusion may be dependent on which strength of evidence is requested.

Instead of relating to tabled quantiles one could simply report the calculated (estimated) probability of a result as extreme as the one reported or beyond, with our choice of Type 1 sales, a low tail probability. Of course, this may be calculated by any of the methods examined above. More direct analytic computations are also possible, as demonstrated in the appendix with R-code for the one-stream comma scale Gamma-model with given total number of observations. However, they are not recommended for reasons given in the appendix..

10. References

Karlis, D. (2003) An EM algorithm for multivariate Poisson distribution and related models, *Journal of Applied Statistics*, volume 30, Issue 1.

Karlis, D. and Ntzoufras J. (2005). Bivariate Poisson and diagonal inflated bivariate Poisson regression models in R. *Journal of statistical Software*, volume **14**, issue 10.

Krishnamoorthy, K. (2015) Handbook of Statistical Distributions with Applications, 2 ed., London: Chapman & Hall.

Lillestøl, J. (2019) Sample statistics as convincing evidence: A tax fraud case, Law, Probability and Risk

Reid, N. (1995) The roles of conditioning in inference. Statistical Science, vol. 10, no. 2, 138-157.

Sturtz, S., Ligges, U. and Gelman, A. (2005) R2WinBUGS: A Package for Running WinBUGS from R, Journal of Statistical Software, 12(3), 1-16.

Appendix: Distribution theory

Let $X \sim Gamma(\alpha, \beta)$ denotes a sale amount X with a Gamma-distribution determined by parameters $(\text{shape} = \alpha, \text{scale} = \beta)$. The probability density is

$$f(x; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot x^{\alpha-1} e^{-\beta x} \ ; \ x \geq 0$$

with expectation $E(X) = \alpha/\beta$ and variance $V(X) = \alpha/\beta^2$. Here $\Gamma(\cdot)$ is the Gamma-function.

Let $X_i, i = 1,2,3, \ldots$ be consecutive sales assumed independent with distribution $X_i \sim Gamma(\alpha_i, \beta)$ i=1,2,3,.. .i.e. with common scale-parameter β, but possibly differing shape-parameters α_i. Then the distribution of the sum $S = X_1 + X_2 + \cdots + X_n \sim Gamma(\alpha_1 + \alpha_2 + \cdots + \alpha_n, \beta)$, i.e. $S \sim Gamma(n\alpha, \beta)$ in the case of common $\alpha_i = \alpha$.

Let S_1 and S_2 be sums of independent sales with distribution $S_i \sim Gamma(\alpha_i, \beta)$ i=1,2. Then the ratio $R = \frac{S_1}{S_1 + S_2}$ has a Beta-distribution with parameters $(\text{shape1} = \alpha_1, \text{shape2} = \alpha_2)$, denoted $R \sim Beta(\alpha_1, \alpha_2)$. The probability density is

$$f(r; \alpha_1, \alpha_2) = \frac{r^{\alpha_1 - 1}(1-r)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)} \ ; 0 \leq r \leq 1$$

where $B(\cdot\ ,\ \cdot)$ is the Beta-function. The expectation foe the Beta-distribution is $E(R) = \frac{\alpha_1}{\alpha_1 + \alpha_2}$.
In the case that S_1 and S_2 are obtained as sums of n_1 and n_2 common scale Gamma-variables $R \sim Beta(n_1\alpha_1, n_2\alpha_2)$, from which it follows that

$$E(R) = \frac{n_1\, \alpha_1}{n_1\, \alpha_1\ + n_2\, \alpha_2}$$

In the case $S_i \sim Gamma(\alpha_i, \beta_i)$, $i = 1,2$ with differing scale parameter, the ratio $R = \frac{S_1}{S_1 + S_2}$ no longer follows the standard a Beta-distribution, but now has probability density is given by

$$f_1(r;\ \alpha_1, \alpha_2, \beta_1, \beta_2) = \frac{r^{\alpha_1 - 1}(1-r)^{\alpha_2 - 1}}{B(\alpha_1, \alpha_2)} \cdot \frac{{\beta_1}^{\alpha_1} {\beta_2}^{\alpha_2}}{(r\beta_1 + (1-r)\beta_2)^{\alpha_1 + \alpha_2}};\ \ 0 \leq r \leq 1$$

Here the first factor is ordinary Beta-density, which obtains in the case of $\beta_1 = \beta_2$, when the second factor is one. Its expectation and variance are complicated formulas, but approximate expressions may be obtained. However, tail probabilities are easily calculated numerically to any degree of accuracy. In the case that each of S_1 and S_2 are obtained as sums of n_1 and n_2 common scale Gamma-variables, possibly different for the two groups, the density becomes $f_1(r;\ n_1\alpha_1, n_2\alpha_2, \beta_1, \beta_2)$.

Take the number of sales of each type (N_1, N_2) to be random, and with constant probabilities (p_1, p_2) for each type of sale. For a given total number of sales n, if follows (assuming independence) that $N_i \sim Binomial(n, p_i)$, $i = 1,2$. By conditioning it follows that the marginal density of R is

$$f_2(r;\ n, p_i, \alpha_1, \alpha_2, \beta_1, \beta_2) = \sum_{k=0}^{n} f_2(r; k\alpha_1, (n-k)\alpha_2, \beta_1, \beta_2)\, bin(k; n, p_i)\ ;\ i = 1,2$$

In particular, this holds for the common scale case $\beta_1 = \beta_2$ and cancel out. Similar formulas hold for cumulative probabilities, replacing the density functions f_i by the cumulative ones $F_i; i = 1,2$.

There is a caveat here, since there is a positive probability of no sales of a specific type, leading to a distribution with positive probability at the end of range at zero and one. However, this probability is negligible in real applications..

The cumulative formula is here used to calculate the (estimated) probability of a sampled result just as extreme or beyond the 10.3% reported for ForHere sales with common scale model.

```
# Probability of ratio as extreme or beyond the one reported
#  computed for given n by conditoning on possible Type 1 outcomes
#  based on estimated One-stream common scale Gamma-model
n=67 # Number of observations
r=0.103 # Observed Type 1 (ForHere) ratio
a1=2.72 # Estimated Shape1 (ForHere)
a2=3.43 # Estimated Shape2 (ToGo)
b=0.0150 # Estimated common scale (not needed)
p=22/67 # Estimated Type 1(ForHere) probability
n1=seq(0,n) # Set possible type 1 outcomes
# Probability as cumulative Beta weighted by Binomial
prob=sum(pbeta(r,n1*a1,(n-n1)*a2)*dbinom(n1,n,p))
prob
> 0.0004268
```

The R-code take advantage of the built-in function *pbeta* returning cumulative Beta-probabilities for vectorized combinations of (shape1, shape2). However, be aware of possible numerical problems causing false results and complete crash. This may be due to how the underlying integral algorithm handles the behavior of the density at zero and/or the evaluation of the Beta-function for large shapes, which again is most critical at zero sales of any type. In the case above with 67 observations, there were no problem, but the problem occurs with a slightly larger dataset. For the above reasons, this analytic approach is not recommended, as good alternatives are found outlined in the paper.

NHH

NORGES HANDELSHØYSKOLE
Norwegian School of Economics

Helleveien 30
NO-5045 Bergen
Norway

T +47 55 95 90 00
E nhh.postmottak@nhh.no
W www.nhh.no