



*Appl. Statist.* (2017)

# Estimation of the population size by using the one-inflated positive Poisson model

Ryan T. Godwin

*University of Manitoba, Winnipeg, Canada*

and Dankmar Böhning

*University of Southampton, UK*

[Received October 2014. Final revision August 2016]

**Summary.** In population size estimation, many capture–recapture-type data exhibit a preponderance of ‘1’-counts. This excess of 1s can arise as subjects gain information from the initial capture that provides a desire and ability to avoid subsequent captures. Existing population size estimators that purport to deal with heterogeneity can be much too large in the presence of 1-inflation, which is a specific form of heterogeneity. To deal with the phenomena of excess 1s, we propose the one-inflated positive Poisson model for use as the truncated count distribution in Horvitz–Thompson estimation of the population size.

**Keywords:** Capture–recapture; Count inflation; Horvitz–Thompson estimator; Maximum likelihood; Poisson model

## 1. Introduction

This paper is concerned with estimating the size of a population in a capture–recapture setting. The number of unobserved members in a population,  $n_0$  (the unobserved ‘0’-counts), must be estimated to obtain an estimate for the entire population:  $N = n_0 + n$ , where  $n$  is the observed sample size. Early efforts to estimate population size by using capture–recapture methods are from John Graunt in the early 1600s in England (Hald, 2003) and Pierre Laplace in 1802 in France (Cochran, 1978). A popular estimation strategy, from at least as early as McKendrick (1926), is to use a zero-truncated distribution to explain the observed frequencies of captures, to obtain the number of missing 0s implied by the counterpart untruncated distribution. Once the parameters of the zero-truncated distribution have been estimated, a maximum likelihood estimator (MLE) for  $N$  can be obtained through the Horvitz–Thompson-like estimator (Horvitz and Thompson, 1952):  $\hat{N} = n\{1 - f(0, \hat{\theta})\}$ , where  $f(0, \hat{\theta})$  is the estimated probability that a 0 occurs under the *untruncated* distribution, and where  $\hat{\theta}$  has been estimated under the *zero-truncated* distribution (Overton and Davis, 1969).  $\hat{N}$  is not a true Horvitz–Thompson estimator, since the probability that a 0 occurs under the untruncated distribution is unknown and must be estimated. The Horvitz–Thompson-like estimator has been widely applied, e.g. in van der Heijden *et al.* (2003a, b), Böhning *et al.* (2004), Cruyff and van der Heijden (2008), Vilas and Böhning (2008) and Niwitpong *et al.* (2013), to name a few.

*Address for correspondence:* Ryan T. Godwin, Department of Economics, University of Manitoba, Winnipeg, Manitoba, R3T 5V5, Canada.  
E-mail: ryan.godwin@umanitoba.ca

As in many problems involving count data, a Poisson process is a natural starting point, and indeed the positive Poisson (PP) distribution is the benchmark distribution for the Horvitz–Thompson method. Often the Poisson assumptions are too restrictive, however, and a more flexible distribution is needed because of phenomena such as heterogeneity or contamination, for example. Although several well-established alternative estimators are purportedly capable of dealing with these phenomena, such as the Chao estimator for heterogeneity (Chao, 1987) and the Zelterman estimator for contamination (Zelterman, 1988), they can be extremely positively biased in the presence of a particular form of the phenomena: 1-inflation.

The one-inflated positive Poisson (OIPP) distribution is proposed to deal with 1-inflation for zero-truncated data. The OIPP model is akin to the well-known zero-inflated Poisson model. Mullahy (1986), and Johnson *et al.* (1992), discussed a modification to the Poisson distribution which may handle extra 0s. An extra proportion  $\omega$  is added to the probability of observing a 0, while appropriately decreasing the proportion of remaining counts by  $1 - \omega$ . Lambert (1992) used a zero-inflated Poisson model in the context of manufacturing equipment, which may be in a perfect state producing zero defects, or in an imperfect state, which produces defects according to a Poisson distribution. The zero-inflated Poisson model has since been used in many applications, e.g. in recreation demand (Gurmu and Trivedi, 1996), in patent applications (Crépon and Duguet, 1997) and in dental epidemiology (Böhning *et al.*, 1999).

The main practical usefulness of the OIPP distribution is that it provides the only modelling option for 1-inflation and provides a population size estimate that is much smaller than other estimators, one of which is thought to provide a lower bound in the presence of heterogeneity (Chao, 1987). The Horvitz–Thompson estimator via the OIPP distribution typically estimates between a half and a quarter as many missing 0s as alternative estimators, in cases where 1-inflation is plausible. The OIPP distribution is obtained through modifying the PP process, by adding an extra proportion  $\omega$  to the probability of observing a ‘1’, while decreasing the proportion of remaining counts by  $1 - \omega$ . In justifying the one-inflated process, it is useful to consider a justification for the zero-truncated process. The prevalence of zero-truncated count distributions has been attributed to the tendency of the observational apparatus to become active only when at least one event occurs (Johnson and Kotz, 1969). The OIPP distribution goes one step further in allowing the activation of the observational apparatus to alter the PP process for some proportion  $\omega$  of the population, such that only 1-counts are generated. The act of observation alters the process being observed, but only for a portion of the population.

Specifically, 1-inflation is justified in two ways.

- (a) The subject discovers that the observational experience is more unpleasant than expected, and decides to put forth avoidance effort towards being observed again.
- (b) Additionally, the subject may learn how to avoid being observed subsequently, which we term avoidance ability.

Both (a) and (b) are necessary for the subject to avoid subsequent detections; the subject must have both a desire and an ability to avoid. At least one of (a) or (b) must be learned through the observational experience for 1-inflation to arise but, depending on the situation, avoidance effort or ability may already be present. As may be the case in law enforcement for example, if the observational experience is known to be unequivocally unpleasant, or if membership in the population is unequivocally undesirable, then the subject already has avoidance effort (a). Alternatively, avoidance ability (b) is granted in voluntary treatment programmes where the subject has complete choice regarding whether or not he or she is observed. In the following section the case-studies are introduced, and 1-inflation is justified in the specific context of each, as it arises because of (a) and/or (b).

Avoidance effort and ability may be learned by a portion of the population because of the information that is revealed to the subject during the observational experience. If each subsequent observational experience is similar to the first, then the subject gleans no additional information from subsequent experiences, and perturbation will occur only around the initial experience, giving rise to 1-inflation.

Avoidance effort that is learned on initial capture may be considered as a behavioural response. In the model structure that was introduced by Otis *et al.* (1978), 1-inflation, as it arises due to learned avoidance effort, is similar to models of class  $M_b$ , in which there is a behavioural response to capture. Behavioural response models are common in ecological studies, where subjects may become ‘trap happy’ (and desire to return to the trap), or ‘trap shy’ (for example, see McCrea and Morgan (2014), page 37, and Borchers *et al.* (2002), page 119). 1-inflation due to learned avoidance effort is a limiting form of ‘trap shyness’, since the probability of subsequent captures becomes 0, following the initial capture, for some portion of the population.

## 2. Case-studies

In the field of biology, population size estimation methods have been used for estimating animal abundance, and in epidemiology they have been used for estimating disease incidence. Horvitz–Thompson estimators have been used in these fields, where the PP distribution is the benchmark. Recently, population size estimation has extended to criminology, where apprehensions–re-apprehensions are akin to captures–recaptures. Whether or not the same assumptions under the PP distribution that are realistic for animals are also realistic for humans has been discussed by van der Heijden (2003a). The key assumption is that prior apprehensions do not affect a change in the individual Poisson parameter. Again, the OIPP distribution can model for a simple form of heterogeneity by allowing the first apprehension to alter the behaviour of a proportion  $\omega$  of the population, such that further apprehensions do not occur. In these cases, the observational apparatus is the encounter with law enforcement, for example.

In this section, data from various fields are described, and for each a justification for 1-inflation is provided. In Section 7, several estimators that are popular in the literature are applied to each case-study and compared with the Horvitz–Thompson estimator via the OIPP distribution.

The programs that were used to analyse the data and the data for the Dutch illegal immigrants example can be obtained from

<http://wileyonlinelibrary.com/journal/rss-datasets>

### 2.1. Street prostitutes in Vancouver

The data are on counts of prostitution arrests over a 23-month period and are from Rossmo and Routledge (1990). The data are  $f_1 = 541$ ,  $f_2 = 169$ ,  $f_3 = 95$ ,  $f_4 = 37$ ,  $f_5 = 21$  and  $f_6 = 23$ , where  $f_1$  denotes the frequency of prostitutes arrested only once, etc. The goal is to obtain an estimate for the total number of prostitutes in Vancouver.

Prostitution arrests had very little effect in terms of deterrence, and hence on recaptures, and a Poisson process is plausible. However, Rossmo and Routledge (1990) wrote that

‘prudent ones seem to learn from the experience, not to avoid street prostitution, but to avoid being arrested.... This would lead to a relatively large number of single arrests.’

As in many law enforcement situations, it is likely that subjects desire to avoid detection and avoidance effort (a) may already be present, but avoidance ability (b) is learned by a portion of ‘prudent ones’. It is reasonable that, if avoidance is not learned after the initial capture, it will not be learned subsequently, and 1-inflation will occur.

### 2.2. H5N1 flu outbreaks

In data similar to McKendrick's (1926) cholera data, outbreaks of an infectious disease in individual entities may follow a count distribution, yet observations on the number of infected entities that do not produce outbreaks may be unreliable or impossible to obtain. In this case, the capture–recapture approach may be used to obtain an estimate for the entire population of infected entities.

The data are on reported H5N1 flu outbreaks at a subdistrict level in Thailand, from Vergne *et al.* (2014). Districts with 0-outbreaks are reported; however, we suppose that the reported 0-counts do not represent the number of districts which were infected but did not produce an outbreak. The data are  $f_1 = 410$ ,  $f_2 = 161$ ,  $f_3 = 87$ ,  $f_4 = 46$ ,  $f_5 = 26$ ,  $f_6 = 21$ ,  $f_7 = 8$ ,  $f_8 = 4$ ,  $f_9 = 6$  and  $f_{10} = 10$ . 1-inflation may arise if an  $\omega$ -portion of districts decide to take preventive measures following the information that was learned from reporting the initial outbreak, and justification (b), avoidance ability, is again appropriate. If preventive measures are available but are not taken following the initial outbreak, and subsequent outbreaks provide no additional information or incentive to prevent, 1-inflation will occur.

### 2.3. Rotterdam opiate users

Data on the number of applications for a methadone treatment programme made by opiate users in Rotterdam are from Cruyff and van der Heijden (2008) and are  $f_1 = 1206$ ,  $f_2 = 474$ ,  $f_3 = 198$ ,  $f_4 = 95$ ,  $f_5 = 29$ ,  $f_6 = 19$ ,  $f_7 = 5$ ,  $f_8 = 2$  and  $f_{10} = 1$ . If a participant is absent for 3 days, they are only readmitted following a new application; hence multiple applications by the same individual are possible. Here, justification (a) is plausible. The applicant may discover that the programme itself is more unpleasant than expected, so subsequent treatment attempts are sought elsewhere or somehow otherwise. It may also be that applicants perceive more scrutiny of their illegal activities than expected and decide not to return. Avoidance ability is granted since it is a voluntary programme.

### 2.4. Dutch illegal immigrants

Data on apprehensions of illegal immigrants, who were unable to be effectively expelled, are from van der Heijden *et al.* (2003a), and are  $f_1 = 1645$ ,  $f_2 = 183$ ,  $f_3 = 37$ ,  $f_4 = 13$ ,  $f_5 = 1$  and  $f_6 = 1$ . The homogeneity assumption of the Poisson process assumes that an apprehension does not change the probability of reappréhension. 1-inflation entertains heterogeneity by allowing for a proportion of immigrants to avoid reappréhension following the initial apprehension. Although the illegal immigrants cannot be effectively expelled, the apprehension may nonetheless be unpleasant, and the subject may learn to avoid detection, so both justifications (a) and (b) are applicable.

### 2.5. Illegal ownership of firearms

The data are from van der Heijden *et al.* (2003b) and are  $f_1 = 2561$ ,  $f_2 = 72$  and  $f_3 = 5$ . van der Heijden *et al.* (2003b) constructed the frequency of individuals captured from police records and argued that contagion resulting from police behaviour cannot be large, since 'many offenders are apprehended as the result of their own behavior (such as displaying a gun in a public place)'. 1-inflation may arise due to reason (b): a portion of the population simply learn not to display their gun in a public place, following the initial capture.

### 2.6. Drunk-driving

The data are again from van der Heijden *et al.* (2003b) and are  $f_1 = 8877$ ,  $f_2 = 481$ ,  $f_3 = 52$ ,

$f_4 = 8$  and  $f_5 = 1$ . Contagion due to police behaviour is ruled out since the apprehensions mostly occur because the police stop and check all passing drivers. Owing to the stigma that is attached to drunk-driving, however, there may be a strong behavioural response following the initial capture, so a portion of individuals exit the population of drunk-drivers, and 1-inflation is justified by reason (b).

2.7. *Snowshoe hares*

Keith and Meslow (1968) presented data on the number of times that individual snowshoe hares were captured–recaptured, from live trapping at six different study areas, during 1962–1967. Keith and Meslow (1968) used the PP distribution to estimate the population size of the hares via the Horvitz–Thompson estimator and used the  $\chi^2$  goodness of fit to inform trap responsiveness, where it is found that the data generally indicate trap response homogeneity. Unlike the benchmark wild cottontail rabbits data set of Edwards and Eberhardt (1967), there is no information on the frequencies at capture occasions; hence time dependence models cannot be applied ( $M_t$  and  $M_{th}$ ), and these models are not used in this paper. Since we shall entertain the notion of heterogeneity as it arises through 1-inflation, however, we shall compare our model with several others which are capable of dealing with heterogeneity.

In the capture–recapture of animals, the observational apparatus is the trap, and the justification for 1-inflation is quite intuitive. The inflation parameter  $\omega$  is the proportion of hares whose behaviour is modified by the trapping experience, such that the animal will never be captured again. Both of the justifications for 1-inflation apply here:

- (a) the trapping experience is learned to be unpleasant and the animal chooses to avoid subsequent trappings, and/or
- (b) the trapping experience teaches the animal how to avoid subsequent trappings.

If no information is learned by the second capture that has not already been learned at the initial capture, then the behavioural response will arise as 1-inflation.

3. **Accommodating excess 1s in the positive Poisson model**

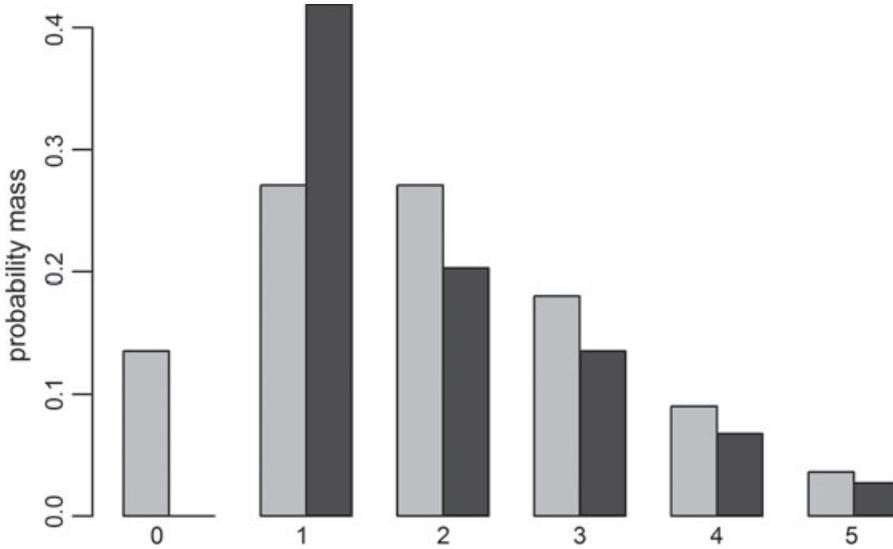
There are two ways to introduce 1-inflation into a zero-truncated Poisson distribution, depending on whether the distribution is first truncated and then 1 inflated, or vice versa. The former will be called the OIPP distribution, and the latter the positive one-inflated Poisson (POIP) distribution. In this section, it will be shown that the MLEs from one of the distributions may be trivially obtained from the other; however, in a later section it will be shown that each distribution gives rise to a different Horvitz–Thompson estimator.

3.1. *One-inflated Poisson distribution*

Let  $f_y(\lambda)$  be the mass function for the Poisson distribution. Then, a one-inflated Poisson distribution is

$$\omega \left\{ \begin{array}{ll} f_y(\lambda), & y=0, \\ 1 - f_0(\lambda) + (1 - \omega) f_y(\lambda), & y=1, \\ (1 - \omega) f_y(\lambda), & y > 1, \end{array} \right\} \tag{1}$$

where  $f_0(\lambda)$  is the probability that a 0 occurs under a Poisson distribution. In distribution (1), the excess probability of a 1 is drawn from the positive counts only, i.e. the inflation process may alter non-zero counts only. This data-generating process is in keeping with the idea that the



**Fig. 1.** A Poisson process (□) is the true underlying data-generating process; however, the observational apparatus only becomes active once a positive count has been generated (hence the missing 0s); further, the observational apparatus perturbs a proportion  $\omega$  of the population, such that only 1-counts are generated, which shifts probability mass from  $y > 1$  to  $y = 1$ ; probabilities from the OIIPP distribution (■) have been normalized by  $1 - f_0^{\text{Poisson}}(\lambda = 2; \omega = \frac{1}{4})$

activation of the observational apparatus alters the Poisson process, and that the observational apparatus is activated only once a positive count has been generated. Unobserved members of the population are not altered in expression (1). Zero-truncating expression (1) results in the following OIIPP distribution:

$$f_1^{\text{OIIPP}} = \omega + (1 - \omega) \frac{\lambda}{\exp(\lambda) - 1}, \quad (2)$$

$$f_y^{\text{OIIPP}} = (1 - \omega) \frac{\lambda^y}{\{\exp(\lambda) - 1\}y!} \quad y = 2, 3, \dots,$$

where  $\lambda > 0$  is the shape parameter for the OIIPP distribution. Note that, in expression (2), it is as if the Poisson process has first been truncated, and then 1 inflated; hence 0-counts are left unaltered. Fig. 1 illustrates both of the features that differentiate the OIIPP process from a Poisson process: 0-truncation and 1-inflation.

For expression (2) to be a proper mass function,  $\omega < 1$ . If 1-deflation is ruled out *a priori* then the logistic link  $\omega = \{1 + \exp(-\vartheta)\}^{-1}$  may be used so that  $\vartheta$  may be estimated without constraint while maintaining  $0 < \omega < 1$ . 1-deflation might be ruled out if the observational experience is unequivocally unpleasant. In the snowshoe hares example, however, it is possible that receiving the bait might induce an animal to return for recapture, thus giving rise to 1-deflation. Note that using the logistic link causes the null hypothesis of no 1-inflation ( $\omega = 0$ ) to be on the boundary of the parameter space, and the likelihood ratio (LR) test statistic to follow a  $0.5\chi^2$ - (or  $\chi^2$ -) distribution (Chernoff, 1954). The score test statistic, however, follows a standard  $\chi^2$ -distribution (Moran, 1971; Chant, 1974), regardless of how  $\omega$  is reparameterized. This is because the calculation of the score test statistic relies on only the sample data and the MLEs from the PP distribution.

To allow for 1-deflation as well as 1-inflation, the link  $\omega = \theta/(1 + \theta)$  may be used. This link also ensures that the null hypothesis of no 1-inflation does not lie on the boundary of the parameter space; hence the associated LR test statistic follows a  $\chi^2$ -distribution. This link also ensures that  $-\lambda\{\exp(\lambda) - \lambda - 1\}^{-1} < \omega < 1$ , where the left-hand side of the inequality is required for the probability that a 1-count remains positive. In Section 6, Monte Carlo evidence confirms the above conjectures on the null distributions of score and LR test statistics under the two different reparameterizations of  $\omega$ .

In the OIPP model,  $\lambda$  may vary through covariates. We use the convention  $\lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$ , where  $\mathbf{X}$  is an  $n \times p$  matrix consisting of  $p$  regressors, and  $\boldsymbol{\beta}$  is a  $p \times 1$  matrix of coefficients. This link ensures that  $\lambda_i > 0$ , although unlike the Poisson model this link provides for no convenient interpretation of  $\boldsymbol{\beta}$ , since the mean of the OIPP distribution is  $E[y] = \lambda \exp(\lambda) \{\exp(\lambda) - 1\}^{-1}$ . Using  $\omega = \theta/(1 + \theta)$  and  $\lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$ , expression (2) becomes

$$f_1^{\text{OIPP}} = \frac{\theta}{1 + \theta} + \frac{1}{1 + \theta} \frac{\lambda_i}{\exp(\lambda_i) - 1},$$

$$f_{y_i}^{\text{OIPP}} = \frac{1}{1 + \theta} \frac{\lambda_i^{y_i}}{\{\exp(\lambda_i) - 1\}_{y_i!}} \quad y_i = 2, 3, \dots$$

When referring to an element of the  $\mathbf{X}$ - or  $\boldsymbol{\beta}$ -matrices, the following notation is used.  $X_{ir}$  is the  $r$ th (of  $p$ ) covariate for the  $i$ th (of  $n$ ) observation, and  $\beta_r$  is the  $r$ th (of  $p$ ) parameter. Under independent sampling from the above mass function, the log-likelihood is

$$l = \sum_{i=1}^n \left( -\ln(1 + \theta) + I_1 \ln \left\{ \theta + \frac{\lambda_i}{\exp(\lambda_i) - 1} \right\} + (1 - I_1)[y_i \ln(\lambda_i) - \ln\{\exp(\lambda_i) - 1\}] \right),$$

where  $I_1$  is an indicator function taking the value 1 when  $y_i = 1$ , and 0 otherwise. The first derivatives of the log-likelihood are given by

$$\frac{\partial l}{\partial \beta_r} = \sum_{i=1}^n \left[ I_1 \frac{1}{\theta \exp(\lambda_i) - \theta + \lambda_i} \left\{ \lambda_i X_{ir} - \frac{\exp(\lambda_i) \lambda_i^2 X_{ir}}{\exp(\lambda_i) - 1} \right\} + (1 - I_1) \left\{ y_i X_{ir} - \frac{\exp(\lambda_i) \lambda_i X_{ir}}{\exp(\lambda_i) - 1} \right\} \right] \quad r = 1, \dots, p,$$

$$\frac{\partial l}{\partial \theta} = \sum_{i=1}^n \left[ -\frac{1}{1 + \theta} + I_1 \left\{ \frac{\exp(\lambda_i) - 1}{\theta \exp(\lambda_i) - \theta + \lambda_i} \right\} \right].$$

The second derivatives of the log-likelihood are provided in the Web appendix A, and R code for maximum likelihood estimation of the OIPP model via numerical methods may be found in the Web appendix C and at <http://wileyonlinelibrary.com/journal/rss-datasets>.

### 3.2. The positive one-inflated Poisson distribution

The OIPP distribution supposes that the additional probability that a 1 occurs is drawn only from the positive counts, i.e. the inflation process changes a portion of the positive counts into 1-counts. An alternative but related data-generating process supposes that the inflation process may also draw from the 0-counts. The mass function for such a distribution is

$$\begin{aligned} w + (1 - w) f_y(\lambda), & \quad y = 1, \\ (1 - w) f_y(\lambda), & \quad y \neq 1, \end{aligned} \tag{3}$$

where  $f_y(\lambda)$  is again the mass function for the Poisson distribution. Zero-truncating expression (3) results in the following POIP distribution:

$$f_1^{\text{POIP}}(\lambda, w) = \frac{\exp(\lambda)w + (1-w)\lambda}{\exp(\lambda) - (1-w)},$$

$$f_y^{\text{POIP}}(\lambda, w) = \frac{(1-w)\lambda^y}{\{\exp(\lambda) - (1-w)\}y!}, \quad y=2, 3, \dots$$

In this mass function, it is as if the Poisson process is first inflated, thereby drawing probability from all possible counts (including 0), and then truncated. The parameter  $\lambda$  has the same interpretation in both the OIPP and the POIP distributions. The inflation parameters ( $\omega$  from the OIPP and  $w$  from the POIP) do not have the same interpretation. However, by reparameterizing,

$$w = 1 - \frac{(1-\omega)\exp(\lambda)}{1-\omega+\exp(\lambda)-1},$$

the POIP distribution becomes identical to the OIPP distribution. Hence, MLEs for one distribution can be easily obtained from the other.

#### 4. Estimating the size of an unknown population via the one-inflated positive Poisson model

The Horvitz–Thompson estimator, using the OIPP model, is  $\hat{N}^{\text{OIPP}} = n/\{1 - f(0, \hat{\lambda})\}$ , where  $f(0, \hat{\lambda})$  is the estimated probability that a 0 occurs under the Poisson distribution, and where  $\hat{\lambda}$  has been estimated under the OIPP distribution. The MLEs of the OIPP distribution ( $\hat{\lambda}$  and  $\hat{\omega}$ ) enjoy the usual properties of consistency, asymptotic normality, asymptotic efficiency and invariance, provided that the following regularity conditions hold (Greene (2012), page 515, for example):

- (a) the first three derivatives of the log-likelihood are continuous and finite;
- (b) the expectations of the first two derivatives may be obtained;
- (c) the third derivatives of the log-likelihood are bounded by an integrable function.

These conditions are almost always met, and the present case is no exception. Owing to the property of invariance,  $\hat{N}^{\text{OIPP}}$  is also consistent, asymptotically normal and asymptotically efficient. The consistency property is verified in the Monte Carlo experiments, since the bias and root-mean-square error (RMSE) of  $\hat{N}^{\text{OIPP}}$  go to 0 as  $N$  increases.

##### 4.1. Alternative Horvitz–Thompson estimators

A very important point arises in the difference between the two one-inflated data-generating processes. In the OIPP distribution, where 1-inflation draws only from the positive counts, the Horvitz–Thompson estimator for the population size is  $n/\{1 - f_0(\hat{\lambda})\}$ . In the POIP distribution, where 1-inflation draws from all counts including 0s, the Horvitz–Thompson estimator for the population size is  $n/\{1 - (1 - \hat{w})f_0(\hat{\lambda})\}$ . In both cases,  $f_0$  is the mass function for the Poisson distribution,  $\hat{\lambda}$  is the estimated shape parameter from either the OIPP or POIP distribution and  $\hat{w}$  is the estimated inflation parameter from the POIP distribution. Note that the Horvitz–Thompson estimator under the OIPP distribution is always at least as large as the Horvitz–Thompson estimator under the POIP distribution, since in the POIP process some 1-counts are observed that would otherwise have been 0s if not for the activation of the observational apparatus. Hence, using the POIP version can provide a lower bound for the population size estimate in the presence of 1-inflation.

There is no way to tell from the sample data which of the two Horvitz–Thompson estimators is appropriate. In terms of estimation, choice of the OIPP or POIP distribution is arbitrary since estimates for one distribution may be trivially obtained from the other; however, the appropriate Horvitz–Thompson estimator critically depends on speculating on one of the two data-generating processes. The choice can be important. For example, if  $f_0(\hat{\lambda}) = \hat{w} = \frac{1}{3}$  (which is characteristic of some of the illustrative applications considered), then the population size estimate will be  $\frac{1}{3}$  larger under the OIPP distribution. We shall focus on the OIPP distribution, since the case-studies considered seem characteristic of processes that do not perturb the unobserved counts.

#### 4.2. An EM algorithm for estimating $\hat{N}^{\text{OIPP}}$

As an alternative to estimating the parameters of the OIPP distribution numerically, an EM algorithm is proposed in this section. Monte Carlo evidence (which is not reported) indicates that maximum likelihood estimation via an EM algorithm results in slightly less bias and mean-squared error than does maximum likelihood estimation via packaged numerical methods.

##### 4.2.1. Unobserved complete likelihood (*M-step*)

Let  $z_i$  denote a binary variable indicating whether the sample value 1 is arising from the extrapolation part or not. Clearly,  $z_i$  is not observed. Then, we have the following complete, unobserved likelihood function:

$$L = \prod_{y_i=0} \exp(-\lambda) \prod_{y_i=1} [\omega \{1 - f_0(\lambda)\}]^{z_i} \{(1 - \omega)\lambda \exp(-\lambda)\}^{1-z_i} \prod_{y_i>1} \{(1 - \omega)\lambda^{y_i} \exp(-\lambda)/y_i!\}.$$

The log-likelihood may be written as

$$\begin{aligned} l = & (\Sigma_1 - N)\lambda + (n - \Sigma_1) \log(1 - \omega) + \Sigma_1 \log\{1 - f_0(\lambda)\} + \Sigma_1 \log(\omega) \\ & + \sum_{i=1}^N y_i \log(\lambda) - \Sigma_1 \log(\lambda) - \log(y_i!), \end{aligned}$$

where  $\Sigma_1$  is the unobservable number of extra-Poisson 1s,  $N$  is the unobservable full sample size (including the missing 0s) and  $n$  is the observable sample size (which excludes the 0s). Note that the log-likelihood is orthogonal with respect to the parameters  $\lambda$  and  $\omega$ . Solving the score equations leads to the maximum likelihood estimators

$$\begin{aligned} \hat{\omega} &= \Sigma_1/n, \\ \hat{\lambda} &= \frac{\sum_{i=1}^N y_i - \Sigma_1}{N - \Sigma_1 + \Sigma_1 f_0(\lambda) \{1 - f_0(\lambda)\}^{-1}}. \end{aligned}$$

##### 4.2.2. *E-step*

$f_0(\lambda)$  and  $\Sigma_1$  (the number of extrapolation 1s due to the inflation process) are unobservable and must be replaced by their expected values:

$$\begin{aligned} E[f_0(\lambda)] &= N_0/N, \\ E[\Sigma_1] &= n_1 P(z_i = 1 | y_i = 1; \lambda, \omega), \end{aligned}$$

where  $n_1$  is the observed number of 1s in the sample. By Bayes's theorem:

$$E[\Sigma_1] = n_1 \frac{P(y_i = 1 | z_i = 1; \lambda, \omega) P(z_i = 1 | \lambda, \omega)}{P(y_i = 1 | \lambda, \omega)} = \frac{n_1 \omega \{1 - \exp(-\lambda)\}}{\omega \{1 - \exp(-\lambda)\} + (1 - \omega) \lambda \exp(-\lambda)}.$$

These values have to be used in the M-step.

#### 4.2.3. *The EM algorithm*

A possible way to obtain an estimate for  $\hat{N}^{\text{OIPP}}$  is by the following algorithm.

*Step 1:* choose an initial value for  $\hat{N}^{\text{OIPP}}$  and  $\hat{\Sigma}_1$ .

*Step 2:* estimate  $\hat{\omega}$  by  $\hat{\omega} = \hat{\Sigma}_1/n$ .

*Step 3:* estimate  $\hat{\lambda}$  by

$$\hat{\lambda} = \left( \sum_{i=1}^N y_i - \hat{\Sigma}_1 \right) / \{ \hat{N}^{\text{OIPP}} (1 - \hat{\omega}) \}.$$

*Step 4:* estimate  $\hat{\Sigma}_1$  by

$$\hat{\Sigma}_1 = \frac{n_1 \hat{\omega} \{1 - \exp(-\hat{\lambda})\}}{\hat{\omega} \{1 - \exp(-\hat{\lambda})\} + (1 - \hat{\omega}) \hat{\lambda} \exp(-\hat{\lambda})}.$$

*Step 5:* repeat steps 2–4 until the estimate  $\hat{\Sigma}_1$  converges (or any of the other estimates converge).

*Step 6:* estimate  $\hat{N}^{\text{OIPP}}$  by

$$\hat{N}^{\text{OIPP}} = \frac{n}{1 - \exp(-\hat{\lambda})}.$$

*Step 7:* repeat steps 2–6 until the estimate  $\hat{N}^{\text{OIPP}}$  converges.

R code for implementing this algorithm is provided in the Web appendix C and at <http://wileyonlinelibrary.com/journal/rss-datasets>.

## 5. Comparing $\hat{N}^{\text{OIPP}}$ with other estimators

The Horvitz–Thompson population size estimator via the OIPP distribution is compared with several other estimators that are popular in the literature: the Horvitz–Thompson via PP distribution, the Horvitz–Thompson via finite Poisson mixture model, the Chao estimator (Chao, 1987) and the Zelterman estimator (Zelterman, 1988). In turn, each of these estimators will be defined and discussed in terms of their percentage bias, %bias, and percentage RMSE, %RMSE, under an OIPP data-generating process. For the Chao and Zelterman estimators, analytic expressions for %bias are obtained. For the remaining estimators, closed form solutions are difficult or impossible to obtain since the MLEs themselves do not admit a closed form and must be solved numerically. Therefore, %bias and %RMSE are simulated for all estimators (Table 1).

### 5.1. *The one-inflated positive Poisson distribution as the data-generating process*

The OIPP distribution is the assumed data-generating process behind the simulation results that are presented in Table 1. The algorithm which generates OIPP variates is now outlined, serving to elucidate the OIPP process.

*Step 1:* generate  $N$  counts according to the Poisson distribution.

*Step 2:* remove the 0-counts from the sample, giving a new sample of size  $n$ . The number of 0s removed,  $n_0$ , is to be estimated, so that an estimate for  $N = n_0 + n$  may be obtained.

*Step 3:* randomly alter  $\omega n$  counts such that  $y = 1$ . These altered counts represent members of

**Table 1.** Percentage bias (and percentage RMSE in parentheses) of  $\hat{N}^{OIPP}$ ,  $\hat{N}^{PP}$ ,  $\hat{N}^{Mix}$ ,  $\hat{N}^{Chao}$  and  $\hat{N}^{Zelt}$ , the population size estimated under the Horvitz–Thompson estimator via the PP distribution, the Horvitz–Thompson estimator via the OIPP distribution, the Horvitz–Thompson estimator via a two-component Poisson mixture, the Chao estimator and the Zelterman estimator respectively when the OIPP is the data-generating process†

<i>N</i>	<i>Results for λ=1</i>					<i>Results for λ=2</i>				
	$\hat{N}^{OIPP}$	$\hat{N}^{PP}$	$\hat{N}^{Mix}$	$\hat{N}^{Chao}$	$\hat{N}^{Zelt}$	$\hat{N}^{OIPP}$	$\hat{N}^{PP}$	$\hat{N}^{Mix}$	$\hat{N}^{Chao}$	$\hat{N}^{Zelt}$
<i>ω=0.1</i>										
100	0.4 (15.4)	8.2 (16.7)	27.6 (102.4)	14.7 (27.2)	16.4 (30.9)	0.4 (5.5)	3.4 (6.4)	12.9 (45.6)	10.5 (15.0)	14.8 (21.4)
200	0.1 (11.2)	7.0 (12.2)	23.4 (84.8)	12.4 (19.4)	13.8 (21.8)	0.2 (4.0)	3.1 (4.9)	14.2 (47.3)	9.7 (12.1)	13.5 (17.1)
500	0.1 (7.6)	6.2 (8.7)	20.0 (65.8)	11.1 (14.3)	12.3 (16.0)	0.1 (2.5)	3.0 (3.9)	16.0 (43.4)	9.2 (10.2)	12.6 (14.1)
1000	0.2 (5.6)	5.9 (7.3)	19.4 (59.7)	10.5 (12.3)	11.7 (13.7)	0.1 (1.8)	3.0 (3.4)	23.1 (39.7)	9.1 (9.6)	12.5 (13.2)
<i>ω=0.3</i>										
100	6.0 (23.5)	26.8 (34.6)	78.5 (186.5)	50.0 (65.1)	54.4 (71.4)	1.1 (6.7)	12.9 (15.1)	132.9 (265.5)	44.0 (49.7)	57.7 (65.7)
200	3.5 (16.4)	24.5 (28.4)	86.5 (198.5)	45.1 (51.5)	49.1 (56.3)	0.4 (4.5)	12.2 (13.4)	200.6 (337.3)	41.8 (44.5)	54.7 (58.5)
500	1.7 (10.5)	23.2 (24.9)	111.4 (238.9)	42.6 (45.1)	46.3 (49.1)	0.2 (2.7)	12.0 (12.5)	327.9 (461.1)	40.4 (41.5)	52.8 (54.2)
1000	0.9 (7.2)	22.7 (23.5)	132.7 (270.5)	41.6 (42.9)	45.2 (46.6)	0.1 (1.9)	11.9 (12.1)	433.2 (552.9)	40.1 (40.6)	52.3 (53.0)
<i>ω=0.5</i>										
100	14.1 (40.9)	62.6 (73.9)	150.7 (270.6)	120.7 (147.9)	128.4 (157.9)	1.7 (7.9)	31.4 (34.1)	397.1 (529.0)	118.7 (133.1)	144.2 (162.5)
200	7.0 (24.4)	57.4 (62.6)	165.1 (292.4)	109.1 (119.8)	116.0 (127.6)	0.6 (5.1)	29.8 (31.2)	537.7 (635.2)	111.6 (117.0)	135.6 (142.4)
500	2.6 (13.7)	54.4 (56.5)	194.7 (335.2)	103.0 (106.7)	109.6 (113.6)	0.3 (3.1)	29.3 (29.8)	655.8 (718.1)	107.6 (109.5)	130.5 (132.9)
1000	1.3 (8.6)	53.5 (54.4)	220.6 (366.8)	100.8 (102.6)	107.2 (109.1)	0.1 (2.14)	29.1 (29.34)	724.8 (766.6)	106.6 (107.6)	129.4 (130.6)

†Results are averaged over 10000 replications.

the population who are instead observed only once, because of the activation of the observational apparatus.

This algorithm was repeated 10000 times for each of several parameter configurations and sample sizes. In each repetition the various estimates are obtained, and average %bias and average %RMSE for the various estimators are shown in Table 1. The average percentage bias of  $\hat{N}^{OIPP}$  (for example) is calculated as  $(100/N) \sum_{j=1}^{10000} (\hat{N}_j^{OIPP} - N)/10000$ , and the average percentage RMSE of  $\hat{N}^{OIPP}$  is calculated as  $(100/N) \{ \sum_{j=1}^{10000} (\hat{N}_j^{OIPP} - N)^2 / 10000 \}^{1/2}$ , where *j* indexes the Monte Carlo replication.

Not surprisingly, when the OIPP distribution is the data-generating process,  $\hat{N}^{OIPP}$  itself has very little bias. The largest %bias for  $\hat{N}^{OIPP}$  is 14.1% and occurs when  $\lambda = 1$  and  $\omega = 0.5$ , but in all cases %bias disappears quickly as *N* increases, reflecting the consistency of the MLE.

**5.2. Horvitz–Thompson estimator via positive Poisson ( $\hat{N}^{PP}$ ) model**

This is the benchmark Horvitz–Thompson estimator and is defined as  $\hat{N}^{PP} = n / \{1 - f(0, \hat{\lambda})\}$ ,

**Table 2.** Percentage of boundary problems in  $\hat{N}^{\text{Mix}}$  (Horvitz–Thompson estimator under a two-component Poisson model) occurring in the Monte Carlo experiments (Table 1)

$N$	Results (%) for $\omega=0.1$		Results (%) for $\omega=0.3$		Results (%) for $\omega=0.5$	
	$\lambda=1$	$\lambda=2$	$\lambda=1$	$\lambda=2$	$\lambda=1$	$\lambda=2$
100	1.0	0.2	3.2	6.6	6.7	28.0
200	0.6	0.2	3.6	10.4	7.9	41.1
500	0.3	0.1	5.2	20.4	10.8	53.2
1000	0.2	0	7.0	30.9	11.4	60.9

where  $\hat{\lambda}$  has been estimated via the PP model. From Table 1 it can be seen that  $\hat{N}^{\text{PP}}$  can be quite biased in the presence of 1-inflation, and that the bias does not disappear as  $N$  grows. Under parameter configurations that are compatible with the case-studies, %bias for  $\hat{N}^{\text{PP}}$  can be 20–50%.

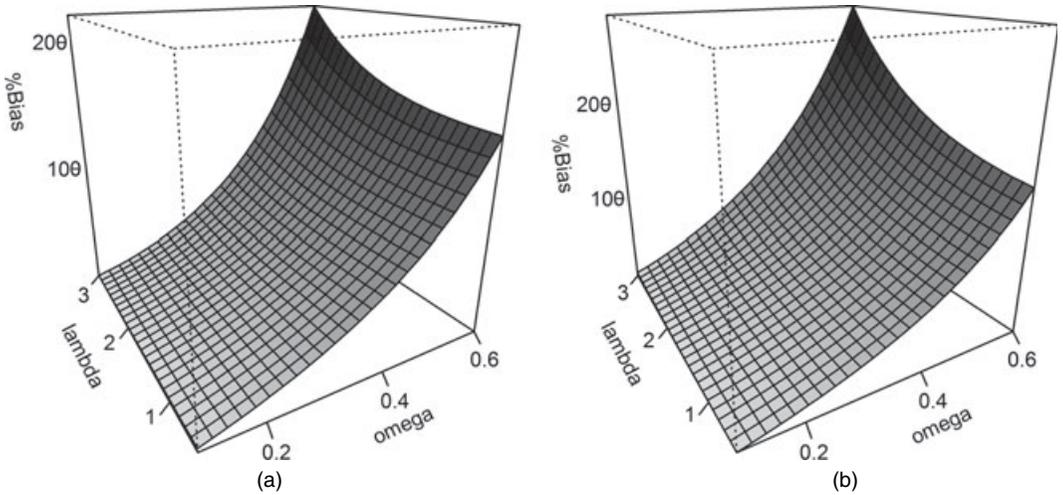
### 5.3. Horvitz–Thompson via Poisson mixture ( $\hat{N}^{\text{Mix}}$ ) model

To deal with unobserved heterogeneity, a Poisson mixture model may be used (Böhning *et al.* (2005), for example). The model consists of  $k$  components, with the weights of the components represented by  $q_j$ , and the Poisson parameters by  $\lambda_j$ , for  $j=1, \dots, k$ . Typically,  $k$  is iteratively increased until two components are estimated to be nearly identical, and then  $k$  is selected via an information criterion. For the purposes of the case-studies, information criteria are not used, and instead  $k$  is allowed to be as large as possible. For the purposes of the simulation experiments, however,  $k=2$ . In simulation experiments (which are not reported here) %bias and %RMSE increased as  $k$  increased, so %bias for  $\hat{N}^{\text{Mix}}$  in Table 1 are lower bounds. Direct maximization of finite Poisson mixture models is difficult or impossible, and the EM algorithm is the method of choice. To estimate  $\hat{N}^{\text{Mix}}$  we use the EM algorithm that was outlined in Böhning *et al.* (2005).

Of all the population size estimators considered, the two-component finite Poisson mixture model fares the worst in terms of %bias and %RMSE. Additionally, in many cases the two-component Poisson mixture model fails to converge; the weight of one of the components approaches 1 whereas the component itself approaches the boundary of zero (i.e.  $q_j \rightarrow 1$  and  $\lambda_j \rightarrow 0$ ), and so  $\hat{N}^{\text{Mix}} \rightarrow \infty$ . In the simulation experiments, the EM algorithm is halted once  $\hat{N}^{\text{Mix}} > 10N$ , so the large %bias of  $\hat{N}^{\text{Mix}}$  in Table 1 is due in part to the failure of the mixture model to converge. For certain values of  $\lambda$  and  $\omega$  the occurrence of this boundary problem increases with  $N$ , suggesting that  $\hat{N}^{\text{Mix}}$  is an inconsistent estimator. The percentage occurrence of the boundary problem in  $\hat{N}^{\text{Mix}}$  from the Monte Carlo experiments of Table 1 is presented in Table 2. Even if the boundary cases are disregarded,  $\hat{N}^{\text{Mix}}$  is the most biased of all the population size estimators, with 100% bias not uncommon.

### 5.4. Chao estimator ( $\hat{N}^{\text{Chao}}$ )

The Chao estimator assumes but does not estimate a Poisson mixture and is supposed to give a lower bound estimate in the case of unobserved heterogeneity. The Chao estimator is not robust to 1-inflation, however, since 1-inflation cannot be expressed as a mixture of Poisson distribu-



**Fig. 2.** %bias of (a)  $\hat{N}^{\text{Chao}}$  and (b)  $\hat{N}^{\text{Zelt}}$ , for various values of  $\omega$  and  $\lambda$

tions. The Chao estimator is  $\hat{N}^{\text{Chao}} = n + f_1^2 / (2f_2)$ , where  $n$  is the observable sample size, and  $f_1$  and  $f_2$  are the observed counts of 1s and 2s respectively. In the presence of 1-inflation, the percentage bias of the Chao estimator is  $\% \text{bias}(\hat{N}^{\text{Chao}}) = 100[\exp(\lambda)\{\lambda^2(1-\omega)/A\}^2 - \exp(-\lambda)]$ , where  $A = \exp(-\lambda)\lambda^2(1-\omega) / \{\exp(-\lambda)\lambda(1-\omega) + 1 - \exp(-\lambda)\}$ . The percentage bias of the Chao estimator is positive for  $\omega > 0$ , is invariant to the sample size (thus  $\hat{N}^{\text{Chao}}$  is inconsistent in the presence of 1-inflation) and is quite severe when considering values of  $\omega$  and  $\lambda$  that are characteristic of the case-studies. For example, when  $\omega = 0.3$  and  $\lambda = 1$ , %bias for  $\hat{N}^{\text{Chao}}$  is 41%. Note that the 41% bias implies that the Chao estimator has estimated the number of missing 0s to be more than twice the actual number. The intuitive reason for the bias of  $\hat{N}^{\text{Chao}}$  (and  $\hat{N}^{\text{Zelt}}$  which is considered next) is its reliance on the frequency of 1-counts, which is troublesome in the presence of 1-inflation. A plot of %bias for  $\hat{N}^{\text{Chao}}$  is shown in Fig. 2(a). The simulation results of Table 1 show that %bias is more severe in finite samples than the analytic expression would indicate.

### 5.5. Zelterman estimator ( $\hat{N}^{\text{Zelt}}$ )

Although 1-inflation can be considered as a form of contamination (since, for an  $\omega$ -portion of the population, the initial capture causes the probability of subsequent captures to be 0), the Zelterman estimator, which is supposed to be robust to contaminations, is also quite biased (and is inconsistent). The Zelterman estimator is  $\hat{N}^{\text{Zelt}} = n / \{1 - \exp(-2f_2/f_1)\}$ , and, under 1-inflation,  $\% \text{bias}(\hat{N}^{\text{Zelt}}) = 100\{1 - \exp(-\lambda)\} / [1 - \exp\{-A \exp(-\lambda)\}]$ , where  $A$  is defined as before. In all instances considered, %bias for  $\hat{N}^{\text{Zelt}}$  is slightly greater than %bias for  $\hat{N}^{\text{Chao}}$ . A plot of %bias for  $\hat{N}^{\text{Zelt}}$  is shown in Fig. 2(b).

### 5.6. Summary of simulation results

In general, for all estimators, %bias is worse the larger the  $\omega$ . The percentage bias, for all estimators and parameter configurations, is always positive. The %bias and %RMSE for the alternative estimators can be quite severe, with %bias for the estimators of  $N$  ranging from 20% to 400%. These percentages understate the failure of the alternative estimators, since the %bias of the unobserved frequency actually being estimated ( $n_0$ ) is much higher. With only one

**Table 3.** Percentage bias (and percentage RMSE in parentheses) of  $\hat{N}^{OIPP}$ ,  $\hat{N}^{PP}$ ,  $\hat{N}^{Mix}$ ,  $\hat{N}^{Chao}$  and  $\hat{N}^{Zelt}$ , when a two-component Poisson distribution is the data-generating process†

<i>N</i>	<i>Results (%) for <math>q_1 = 0.25, \lambda_1 = 0.5</math> and <math>\lambda_2 = 2</math></i>					<i>Results (%) for <math>q_1 = 0.5, \lambda_1 = 0.5</math> and <math>\lambda_2 = 2</math></i>				
	$\hat{N}^{OIPP}$	$\hat{N}^{PP}$	$\hat{N}^{Mix}$	$\hat{N}^{Chao}$	$\hat{N}^{Zelt}$	$\hat{N}^{OIPP}$	$\hat{N}^{PP}$	$\hat{N}^{Mix}$	$\hat{N}^{Chao}$	$\hat{N}^{Zelt}$
100	-12.5 (13.8)	-10.5 (12.0)	1.4 (13.9)	-6.4 (11.8)	-3.9 (14.1)	-24.2 (25.2)	-20.0 (21.3)	2.1 (23.0)	-11.5 (17.3)	-7.6 (18.4)
200	-12.6 (13.3)	-10.7 (11.5)	0.7 (11.3)	-7.1 (9.7)	-5.0 (10.3)	-24.6 (25.1)	-20.3 (21.0)	1.7 (22.3)	-12.3 (15.1)	-8.8 (14.4)
500	-12.7 (12.9)	-10.9 (11.2)	0.4 (9.2)	-7.5 (8.6)	-5.8 (8.0)	-24.9 (25.1)	-20.6 (20.7)	2.8 (23.8)	-13.0 (14.1)	-9.8 (12.0)
1000	-12.7 (12.8)	-10.9 (11.1)	0.3 (8.5)	-7.7 (8.2)	-6.0 (7.1)	-25.0 (25.1)	-20.6 (20.8)	3.9 (24.1)	-13.2 (13.7)	-10.0 (11.2)

†Results are averaged over 10000 replications.

exception in the Monte Carlo experiments, the ranking of the various estimators from most biased to least biased is  $\hat{N}^{Mix}$ ,  $\hat{N}^{Zelt}$ ,  $\hat{N}^{Chao}$ ,  $\hat{N}^{PP}$ ,  $\hat{N}^{OIPP}$ .

**5.7. Finite Poisson mixture model as the data-generating process**

To challenge the OIPP model, %bias and %RMSE of the estimators are again simulated, but where the data-generating process is a two-component Poisson mixture model. Results are displayed in Table 3. Whereas  $\hat{N}^{OIPP}$  is the most biased of all the estimators when the finite mixture model is the data-generating process, the situation is not nearly as severe as when the OIPP is the data-generating process.

**6. Diagnostic tests**

Several tools to aid in the specification of the OIPP model are discussed. In Section 6.1, a score test for no 1-inflation is developed. In Section 6.2, the score test is compared with an LR test via Monte Carlo experiments. In Section 6.3 a ratio plot from Böhning *et al.* (2013) is discussed. In Section 6.4, the  $\chi^2$  goodness-of-fit test is discussed but is cautioned against.

**6.1. A score test for no 1-inflation in a positive Poisson model**

In the case of covariates, the score test statistic is

$$S = \frac{\left\{ \sum_{i=1}^n (I_1[\{\exp(\hat{\lambda}_i) - 1\} / \hat{\lambda}_i] - 1) \right\}^2}{(\hat{v} - \hat{\lambda})'(\mathbf{1} \backslash \hat{\lambda}) - ((\hat{v} - \hat{\varphi}) \backslash \hat{v})' \mathbf{X} [\mathbf{X}' \text{diag}\{\hat{\varphi} \circ (\hat{v} - \hat{\lambda}) \backslash \hat{v}^2\} \mathbf{X}]^{-1} \mathbf{X}' ((\hat{v} - \hat{\varphi}) \backslash \hat{v})}$$

where  $\hat{\lambda} = \exp(\mathbf{X}\hat{\beta})$ ,  $\hat{\beta}$  has been estimated by maximum likelihood,  $\hat{v} = \exp(\hat{\lambda}) - 1$  and  $\hat{\varphi} = \exp(\hat{\lambda}) \circ \hat{\lambda}$ , where ‘ $\circ$ ’ and ‘ $\backslash$ ’ denote Hadamard multiplication and division respectively. Under the null hypothesis,  $S \sim \chi^2_{(1)}$ .

In the case of no covariates, the score statistic becomes much simpler to calculate. Beyond knowing the MLE  $\hat{\lambda}$  estimated under the null hypothesis of no inflation, the only additional information that is required to calculate the score test statistic is  $n_1$ , the number of 1s observed to be in the sample. The score test statistic in the case of no covariates is given by

$$S = \frac{[n_1\{\exp(\hat{\lambda}) - 1\}/\hat{\lambda} - n]^2 \hat{\lambda} \exp(\hat{\lambda}) \{\exp(\hat{\lambda}) - \hat{\lambda} - 1\}}{n[\exp(\hat{\lambda})\{\exp(\hat{\lambda}) - \hat{\lambda} - 1\}^2 - \{\exp(\hat{\lambda}) - \hat{\lambda} - 1\}^2]}.$$

Under the null hypothesis,  $S \sim \chi^2_{(1)}$ . A detailed derivation of the score statistic may be found in the Web Appendix B. R code for calculating the above statistic may be found in the Web appendix C and at <http://wileyonlinelibrary.com/journal/rss-datasets>.

6.2. Comparing the score and likelihood ratio tests

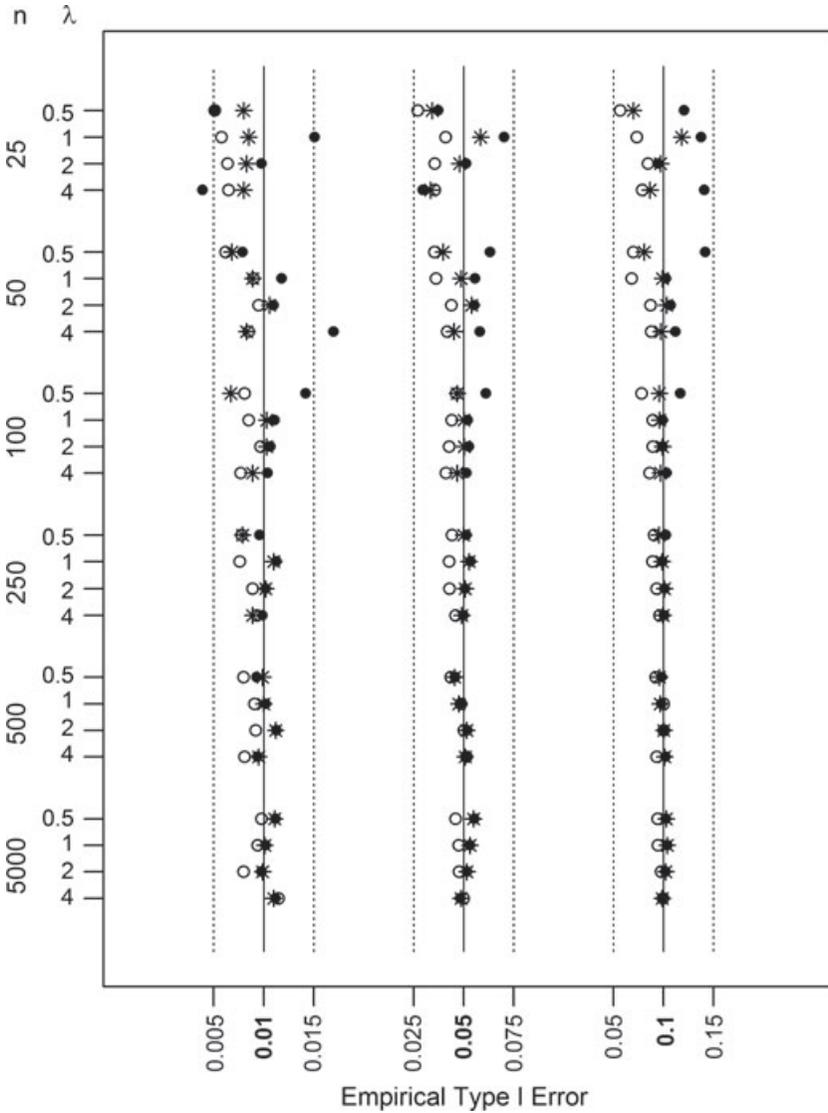
The following section examines the appropriateness of the  $\chi^2$ -distribution as an approximation to the sampling distribution of the above derived test statistic  $S$  in finite samples, for the test of  $H_0$ : PP model versus  $H_A$ : OIPP model. The empirical probability of a type I error and empirical ‘power’ are simulated. The Prob(type I error) of a test is also the size of the test provided that the null hypothesis  $H_0$  is simple, or the test is pivotal, neither of which is true in the present case. The test is likely to be *asymptotically* pivotal, however, as many such tests in similar contexts are (Horowitz and Savin (2000), page 378). Provided that the test is asymptotically pivotal, the critical value that is used to determine the Prob(type I error) of the test may be based on an approximation using first-order asymptotic theory. In view of the Monte Carlo results, it will be argued that the  $\chi^2$ -distribution provides a suitable approximation, so the test indeed appears to be asymptotically pivotal.

The Monte Carlo simulations use parameter values of  $\lambda = 0.5, 1, 2, 4$  from the PP distribution (for the empirical Prob(type I error) investigations), and  $\lambda = 0.5, 1, 2, 4$  and  $\omega = 0.1, 0.2, \dots, 0.9$  from the OIPP distribution (for the empirical power investigations). In all experiments, the sample sizes considered were  $n = 25, 50, 100, 250, 500, 5000$ . Each experiment consists of 10000 iterations. PP variates were generated by using the VGAM (Yee, 2011) package, whereas 1-inflation was generated by using VGAM and randomly distributed uniform numbers. Covariates were also considered. The inclusion of covariates allows  $\lambda_i$  to take many different values in a given sample and has been shown to affect the empirical Prob(type I error) and empirical power (Horowitz (1994), page 404). Previous Monte Carlo studies have shown that the particular characteristics of any covariates in the model can affect the empirical Prob(type I error) and power of statistical tests. For example, King and Giles (1984) found that differing degrees of auto-correlation among the covariate matrix of a linear regression model can cause substantial changes in the empirical Prob(type I error) and power of  $t$ -tests used after pre-testing for auto-correlation. Covariates are incorporated through three different designs for the  $\mathbf{X}$ -matrix, which are discussed in the Web appendix D.

The score test is compared with the LR test. The LR test statistic is determined by  $\xi = 2\{l_A\{\tilde{\lambda}, \tilde{\theta}\} - l_A(\hat{\lambda}, 0)\}$ , where  $l_A(\lambda, \theta)$  is the likelihood function for an OIPP model (with  $A$  denoting the model under the alternative hypothesis),  $\tilde{\lambda}$  and  $\tilde{\theta}$  are the MLEs under the OIPP model and  $\hat{\lambda}$  is the MLE under the PP model. Asymptotically, the score test and the LR test are the same (Rao, 1948), and therefore the LR test statistic should also be asymptotically  $\chi^2_{(1)}$ . The inclusion of this test in the Monte Carlo experiments allows the performance of the score test to be directly compared with a competitor.

6.2.1. Empirical probability of type I error

Fig. 3 presents the empirical Prob(type I error) of both the score and the LR test, for various sample sizes and values of  $\lambda$ , in the case of no covariates. Results for the case of covariates may be found in the Web appendix D. The score test seems to dominate the LR test, in terms of Prob(type I error). The various parameter configurations for  $n$  and  $\lambda$  give rise to 24 experiments.



**Fig. 3.** Empirical Prob(type I error) of tests for no 1-inflation by using the score (\*), LR (●) and LR with logistic link (○), in the case of no covariates

At 0.01-significance, the score test is closer to nominal significance than the LR test for 18 of the 24 parameter configurations. At 0.05- and 0.10-significance the score test is closer to nominal significance for 16 of 24 and 15 of 24 of the parameter configurations respectively. The mean absolute percentage of the deviation between the empirical Prob(type I error) relative to the nominal significance is 7.7%, across all of the reported empirical Prob(type I error) for the score test. The same mean absolute percentage deviation is 13.7% for the LR test. For example, the dotted vertical lines in Fig. 3 represent a 50% deviation from the nominal significance level (the full vertical lines). With the inclusion of covariates in the data-generating process, the dominance of the score test over the LR test remains unchanged. Across all experiments, the mean odds that the score test will dominate over the LR test are 3.5:1. No clear pattern among parameter

values or sample sizes can be discerned, in predicting which test will perform better. There is no apparent advantage in the LR test, in terms of probability of type I error distortions.

The suitability of the  $\chi^2_{(1)}$ -distribution as an approximation to the sampling distribution of the score test is now discussed. If the results for  $\lambda = 0.5$  are excluded, the remaining simulated empirical Prob(type I error)s are very close to nominal. In the case of no covariates, the maximum absolute deviation of the empirical Prob(type I error) from nominal is 0.002 at significance 0.01, 0.017 at significance 0.05 and 0.018 at significance 0.1.

Fig. 3 also displays the empirical Prob(type I error) for when the logistic link is used, as opposed to the link  $\omega = \theta/(1 + \theta)$  that is advocated in this paper. When using the logistic link, the null hypothesis is now on the boundary of the parameter space, and the LR test should follow a  $0.5\chi^2_{(1)}$ -distribution, whereas the score test still follows a  $\chi^2_{(1)}$ -distribution. This is verified in the Monte Carlo experiments. In Fig. 3, the empirical type I errors of the score test are identical under the  $\chi^2_{(1)}$ -distribution regardless of link, whereas empirical type I errors for the LR test under a logistic link are obtained from the  $0.5\chi^2_{(1)}$ -distribution.

When covariates are introduced, the maximum deviation increases approximately three times at significance levels 0.01 and 0.1, and increases approximately two times at significance level 0.05, almost regardless of covariate design. At a sample size of 100, the maximum deviation is 0.005 at significance 0.01, across all covariate designs. At significance 0.05 this maximum deviation becomes 0.011, and at significance 0.1 this maximum deviation becomes 0.017. Referring to the simulation results in Horowitz (1994), deviations of the magnitude that was discussed above are unlikely to be considered a problem. It is concluded that the empirical Prob(type I error) results indicate that the sampling distribution of the score test statistic is suitably approximated by the  $\chi^2_{(1)}$ -distribution, for sample sizes 100 or larger.

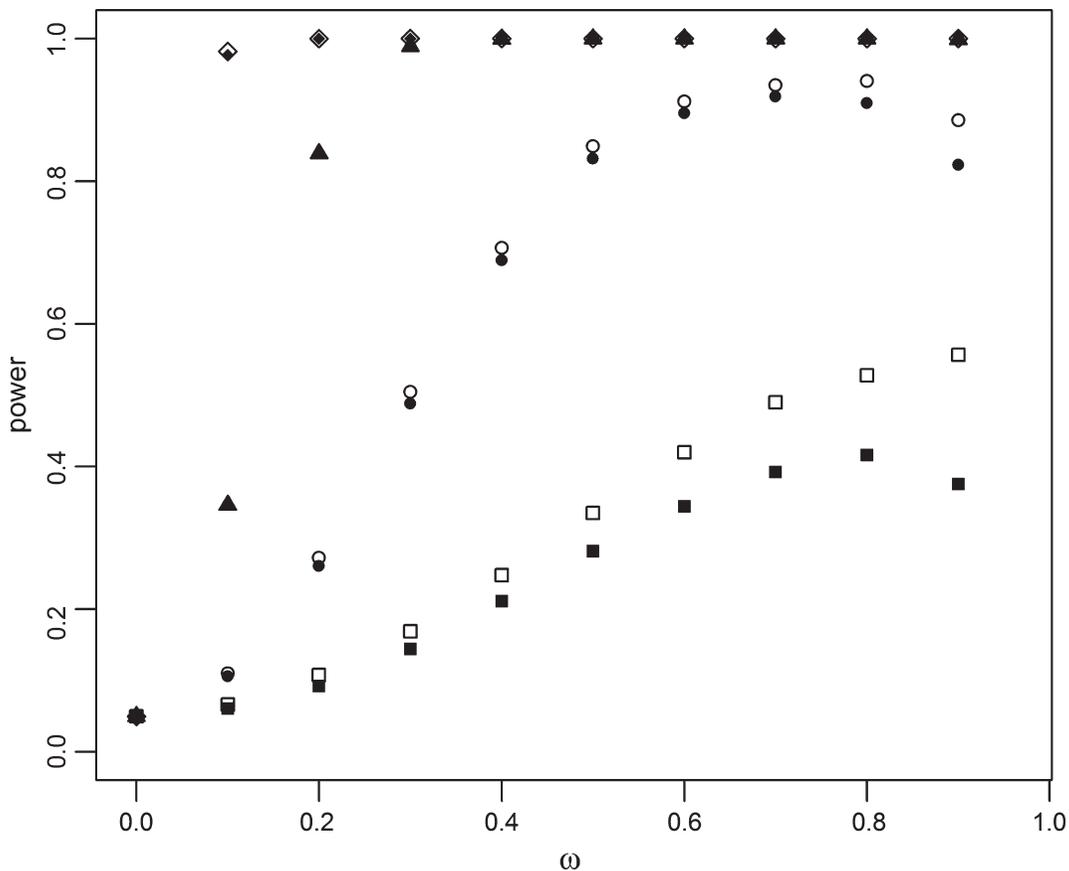
### 6.2.2. Empirical power

Monte Carlo experiments on empirical power serve to compare the score and LR tests further, and to assess further the suitability of the  $\chi^2_{(1)}$ -distribution as an approximation to the finite sample distribution of  $S$ . Fig. 4 displays results for  $n = 250$ ,  $\lambda = 0.5, 1, 2, 4$  and  $\omega = 0.1, 0.2, \dots, 0.9$ , in the case of no covariates. Results for various sample sizes may be found in the Web appendix D.

In the case of no covariates, the score test clearly dominates the LR test in terms of empirical power. At  $n = 25$ , of the 36 experiments that did not result in the empirical powers of the two tests being equal, the score test had higher empirical power 86% of the time. This percentage rises steadily until the score test dominates 100% of the time at  $n = 250$ .

The story is different when covariates are introduced. At  $n = 25$  and  $n = 50$ , the LR test completely dominates the score test. When  $n = 100$ , the score test has higher empirical power than the LR test 44–47% of the time (depending on covariate design). The score test begins to overtake the LR test at  $n = 250$ , performing better 79–83% of the time. At  $n = 500$ , the score test completely dominates. The Web appendix D reports the power results for only one of the designs, as there is little difference in results between designs.

From the empirical Prob(type I error) results in Section 6.2.1, it can be seen that the LR test more frequently rejects the null hypothesis of no 1-inflation when the null hypothesis is true, compared with the score test. This has the effect of shifting the empirical power curve for the LR test upwards. Since the empirical power of the LR test is likely to appear more favourable than the score test at the expense of the small Prob(type I error) distortions of both tests, a more fair comparison between the two tests is accomplished by considering size-adjusted empirical power. When adjusted empirical power is considered, the dominance of the score test becomes



**Fig. 4.** Empirical power of the score ( $\square$ ) and LR ( $\blacksquare$ ) tests of no 1-inflation in the PP model, for various values of  $\lambda$  ( $\square$ ,  $\lambda = 0.5$ ;  $\circ$ ,  $\lambda = 1$ ;  $\triangle$ ,  $\lambda = 2$ ;  $\diamond$ ,  $\lambda = 4$ ) and  $\omega$ , for the case of no covariates, for  $n = 250$ , and with nominal significance  $\alpha = 0.05$ : each point represents the proportion of times that the null hypothesis was rejected, out of 10 000 replications

more pronounced. In the case of no covariates, under the sample sizes that were considered, the score test dominates nearly 100% of the time.

The results change as covariates are introduced. The LR test once again outperforms the score test in small samples, with the score test overtaking the LR test as  $n$  increases. The main difference for the adjusted power results is that the score test begins to dominate at  $n = 100$  instead of  $n = 250$ . Results for size-adjusted power are not reported.

To summarize the Monte Carlo results:

- the  $\chi^2_{(1)}$ -distribution is a suitable approximation for the finite sample distributions of the score and LR test statistics, for  $n \geq 100$ , and
- the score test generally dominates the LR test for  $n \geq 100$ .

This section concludes by addressing an unsettling feature of Fig. 4. The power of both tests initially rises with  $\omega$  but then appears to drop as  $\omega$  reaches 0.7 or 0.8. For  $\omega = 1$  in the OIPP model,  $\lambda \rightarrow 0$  in the PP model and  $S(\beta, 0) \rightarrow 0$ . This property is undesirable but is expected, and is indeed observed for van den Broek's (1995) score test for 0-inflation in a Poisson distribution. A more detailed explanation, along with additional Monte Carlo results, is contained in the

Web appendix E. The power drop tendency, however, is unlikely to be of practical concern for two reasons. First, it is a finite sample problem which becomes less pronounced as  $n$  grows. Second, the phenomenon arises when the sample consists mainly of 1s and such a sample is unlikely to contain enough variation to be of interest.

6.3. Ratio plot

Besides the score test, another tool which may assist in the specification of the OIPP distribution is the log-ratio plot that was proposed by Böhning *et al.* (2013). The log-ratio plot takes advantage of the fact that  $(y + 1)f_{y+1}/f_y = \lambda$ , where  $f_y$  is the mass function for the PP distribution. Hence, the empirical log-ratio will plot as a straight line if the PP distribution is the correct underlying data-generating process. In the case of 1-inflation,  $(y + 1)f_{y+1}/f_y < \lambda$  for  $y = 1$ , and the first point on the log-ratio plot should be lower than the others.

6.4. Goodness-of-fit tests

Goodness-of-fit tests may also aid in the specification of competing count data models. The Kolmogorov–Smirnov test (Kolmogorov, 1933; Smirnov, 1939) and the Cramér–von Mises test (Cramér, 1928; von Mises, 1928) are two potential non-parametric options. For discrete data, the asymptotic distribution of these test statistics is not  $\chi^2$  (Conover, 1972; Gleser, 1985; Choulakian *et al.*, 1994). These tests are not considered in this paper.

Pearson’s  $\chi^2$  goodness-of-fit test is typically employed for count data models. The test statistic follows a mixed  $\chi^2$ -distribution, where the exact form of the mixture may depend on the parameter values of the model (Chernoff and Lehmann, 1954). Ignoring the mixture component in the case of the Poisson model has been shown to induce very little error (Lancaster, 1969); however, in the case of the negative binomial model ignoring the mixture component will ‘seriously inflate the Type I error’ (Pahl (1969), page 145). Given these findings, it may be unwise

**Table 4.**  $\chi^2$  goodness-of-fit test statistic averaged over 10000 replications, for the OIPP distribution, PP distribution and finite Poisson mixture model ( $k = 2$ ), under an OIPP data-generating process

$N$	Results for $\lambda = 1$			Results for $\lambda = 2$		
	OIPP	PP	Mixture	OIPP	PP	Mixture
$\omega = 0.1$						
100	2.6	6.4	2.0	8.4	21.2	3.9
200	3.8	10.9	2.3	8.5	16.9	4.6
500	5.2	13.9	3.0	10.3	25.7	5.7
1000	7.5	15.2	4.9	7.5	6.4	21.2
$\omega = 0.3$						
100	2.3	12.8	2.4	9.5	111.2	4.0
200	3.3	17.0	3.3	7.7	59.7	4.6
500	4.7	31.2	5.2	10.6	153.8	6.6
1000	7.7	61.9	8.6	8.3	124.6	6.5
$\omega = 0.5$						
100	1.9	30.6	3.8	11.9	1878.3	4.2
200	3.0	58.3	6.3	7.0	400.7	4.9
500	4.1	167.0	13.4	11.0	1595.3	6.7
1000	6.7	249.3	24.0	8.3	840.7	6.5

to consider  $p$ -values from the  $\chi^2$ -distribution in Pearson's test in the context of the OIPP model. Additionally, it may be dangerous to rely on the goodness of fit of the observed counts, when interest lies in estimating the unobserved 0-count. This is pertinent with regard to the Poisson mixture model. From the simulation experiments of Table 1, where the data were generated according to the OIPP distribution, the finite Poisson mixture model fits the data extremely well, despite providing an extremely biased estimate of  $N$ , and despite suffering from a boundary problem. Along with the simulation experiments in Table 1, the  $\chi^2$  goodness-of-fit test statistic was recorded for each parametric model. The results were averaged over the 10000 replications and are presented in Table 4, showing that the finite mixture model most often fits the data better than the OIPP distribution, even when the OIPP distribution is the data-generating process.

## 7. Results

In this section, the data are analysed by using the various population size estimators. Table 5 reports some estimation results.  $\hat{N}^{\text{OIPP}}$  is always smaller than the alternative estimators, and in most cases it is much smaller. Similarly to the simulation experiments, the values of the estimators maintain the inequality  $\hat{N}^{\text{Mix}} > \hat{N}^{\text{Zelt}} > \hat{N}^{\text{Chao}} > \hat{N}^{\text{PP}} > \hat{N}^{\text{OIPP}}$ , without exception. Compared with  $\hat{N}^{\text{OIPP}}$ ,  $\hat{N}^{\text{PP}}$  is 13–188% larger,  $\hat{N}^{\text{Chao}}$  is 43–214% larger,  $\hat{N}^{\text{Zelt}}$  is 49–214% larger and  $\hat{N}^{\text{Mix}}$  is 96–552% larger. Given the estimated 1-inflation under the OIPP distribution, these percentages are in keeping with the simulation experiments.

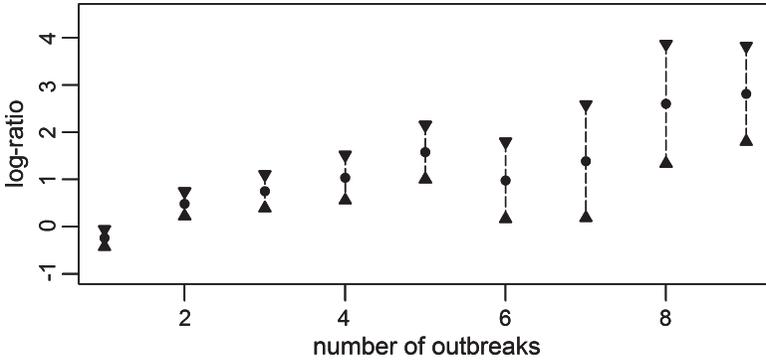
In most cases, there is a discrepancy between the homogeneity-only model  $\hat{N}^{\text{PP}}$  and the alternative models for heterogeneity,  $\hat{N}^{\text{Mix}}$ ,  $\hat{N}^{\text{Chao}}$  and  $\hat{N}^{\text{Zelt}}$ . There is also a discrepancy between  $\hat{N}^{\text{OIPP}}$  and  $\hat{N}^{\text{PP}}$ . Together, these two discrepancies suggest that 1-inflation is not the only source of heterogeneity. The exceptions are the illegal firearms and drunk-driving data, where it is plausible that 1-inflation is the only source of heterogeneity.

1-inflation is estimated between 0.34 and 0.69 across the various case-studies. In the prostitutes data, for example, the interpretation is that 44% of prostitutes are 'prudent' in the sense that they gain an avoidance ability on initial capture, such that they are not recaptured. In all cases the  $p$ -value on the score test statistic is less than 0.01. Despite these results, the OIPP model is not overzealous in its finding of 1-inflation. There are plenty of examples where 1-inflation is not found, such as in the benchmark India cholera epidemic data (McKendrick, 1926).

In addition to the score test supporting the presence of 1-inflation, in general the log-ratio plots do as well. Fig. 5 displays the log-ratio plot for the H5N1 flu data, where the first point appears lower than the rest, indicating that there are too many 1s for a PP process to be appropriate. In this example, whereas the difference between  $\hat{N}^{\text{PP}}$  and  $\hat{N}^{\text{OIPP}}$  is approximately 100, the OIPP

**Table 5.** Population size estimates, estimated 1-inflation  $\hat{\omega}$ , the smallest estimated component from the Poisson mixture model,  $\min(\hat{\lambda}_j)$ , and the number of mixture components,  $k$ , for the various data sets

<i>Data set</i>	$\hat{N}^{\text{OIPP}}$	$\hat{N}^{\text{PP}}$	$\hat{N}^{\text{Mix}}$	$\hat{N}^{\text{Chao}}$	$\hat{N}^{\text{Zelt}}$	$\hat{\omega}$	$\min(\hat{\lambda}_j)$	$k$
Prostitutes in Vancouver	1017	1240	2273	1752	1907	0.44	0.34	2
H5N1 flu outbreaks	829	935	1627	1301	1432	0.42	0.39	3
Rotterdam opiate users	2501	2937	11203	3563	3727	0.34	0.00	3
Dutch illegal immigrants	3455	7080	13374	9274	9425	0.63	0.13	2
Illegal firearms	15334	44201	—	48184	48248	0.68	—	2
Drunk-driving	30479	78710	138313	91333	91710	0.69	0.06	2
Snowshoe hares	143	162	476	211	232	0.34	0.08	3



**Fig. 5.** Ratio plot (on a log-scale) for the H5N1 flu data of Vergne *et al.* (2014) (●) with approximate 95% confidence limits (▼, ▲)

model estimates less than a third the number of infected subdistricts with zero outbreaks as does the PP distribution.

In general,  $\hat{N}^{\text{Mix}}$  increases with each additional component. For example, in the Rotterdam opiate users data, the two-component mixture estimates a population size of 3903, whereas the addition of a third component changes this estimate to 11203. 1-inflation is probably appropriating one of the components, which is estimated to be very close to the boundary of zero (since there are not many 2-counts), which in turn is implying too many missing 0s.

For the Dutch illegal immigrants data, van der Heijden *et al.* (2003b) fitted the PP model by using covariates on nationality, gender and age. In the PP and OIPP model,  $f(0, \hat{\theta}) = \sum_{i=1}^n \{1 - \exp(\lambda_i)\}^{-1}$ , where  $\lambda_i$  varies through the inclusion of covariates. Using covariates, the estimate for  $N$  nearly doubles;  $\hat{N}^{\text{OIPP}}$  is 6272 and  $\hat{N}^{\text{PP}}$  is 12690, and the estimated 1-inflation decreases from 0.63 to 0.53, with a score test statistic of 39.6 ( $p$ -value 0.00). In this case, the covariate information does not account for the excess of 1s, and modelling the observed heterogeneity increases  $\hat{N}^{\text{OIPP}}$  and  $\hat{N}^{\text{PP}}$  in a fashion that maintains their relative sizes.

In the illegal firearms and drunk-driving data,  $\hat{N}^{\text{OIPP}}$  gives an estimate that is approximately a third that of the other estimates. The extremely large estimate for  $\hat{N}^{\text{Mix}}$  is due to the failure of the mixture model to converge. For the illegal firearms data, one of the mixture components continually approached 0 whereas the estimate for  $N$  continually increased, and the EM algorithm had to be halted. Note that the mixture model fares worst for the data that have the highest estimated 1-inflation.

For the snowshoe hares data we also compare our results with the bias-corrected Chao estimator (Chao, 1989),  $\hat{N}^{\text{BCChao}} = n + f_1(f_1 - 1) / \{2(f_2 + 1)\} = 207$ , and the Good–Turing estimator (Good, 1953),  $\hat{N}^{\text{G-T}} = n / (1 - f_1/S) = 174$ , where  $S = 0f_0 + 1f_1 + \dots$ . Once again,  $\hat{N}^{\text{OIPP}}$  provides by far the smallest estimate. For six different data sets from Keith and Meslow (1968), 1-inflation is estimated to be between 0.22 and 0.35, with  $p$ -values from the score tests of 0.30, 0.10, 0.01, 0.02, 0.00 and 0.20. Contrary to Keith and Meslow’s finding of homogeneous behaviour in the hares, these results suggest some heterogeneity in trap response; a portion of animals  $\omega$  respond to the trapping experience by avoiding the trap whereas the behaviour of the remaining  $1 - \omega$  is unaffected, for example.

### 8. Concluding remarks

The OIPP model was proposed to deal with 1-inflation in a zero-truncated Poisson distribution and is used as the truncated count distribution in Horvitz–Thompson estimation of the

population size. The OIPP model allows for a heterogeneous population, where a proportion of the population  $\omega$  responds to the initial capture by effectively avoiding recapture. In general, avoidance of detection depends on

- (a) an avoidance effort put forth by the subject and
- (b) an avoidance ability of the subject.

That is, the subject must both desire and be able to avoid detection. The observational experience reveals information to the subject that may endow avoidance effort and/or ability, for some portion of the population. An avoidance effort occurs when the experience is learned to be less pleasant than expected (e.g. Rotterdam opiate users and snowshoe hares). Avoidance ability is learned straightforwardly by some portion of the population and is plausible in cases involving law enforcement (e.g. Vancouver prostitutes and illegal firearms). Since it is likely that no new information is revealed after the first capture, recapture avoidance manifests as 1-inflation.

A score test for 1-inflation was also derived, and through simulation experiments it was shown that the score test has better properties than the LR test, and that the score test statistic's distribution is reasonably approximated by  $\chi^2_{(1)}$ , for  $n > 100$ . The score test statistic is easy to compute, having the advantage that the data need only to be fitted under the simpler PP model. A ratio plot is also recommended as a diagnostic tool for 1-inflation, but goodness-of-fit tests are cautioned against.

Simulations reveal that, when the OIPP distribution is the data-generating process, alternative estimators that are popular in the population size estimation literature can be extremely positively biased. 50–100% bias for estimators of  $N$  is not uncommon, and these percentages are much higher when considering  $n_0$ . Because of their reliance on the observed frequency of 1-counts, the Chao and Zelterman estimators are quite biased (10–100%). Perhaps surprisingly, the finite Poisson mixture model is either very biased or fails to converge, but it always fits the observed data very well and often fits better than does the OIPP model (which is the data-generating process). Despite these alternative models purporting to deal with unobserved heterogeneity, the Horvitz–Thompson estimator via the homogeneity-only PP model always has less bias than these alternative estimators. The Horvitz–Thompson estimator via the OIPP distribution is shown to have very small bias given an OIPP data-generating process.

The OIPP model is used in seven case-studies, where it provides much smaller population size estimates than the alternative estimators, and where 1-inflation is estimated to be between 0.34 and 0.69. For all the case-studies, the basis for 1-inflation is argued for in the context of the initial observational experience creating an avoidance effort and/or an avoidance ability towards recapture.

The main practical importance of  $\hat{N}^{\text{OIPP}}$  is to provide an asymptotically unbiased population size estimator in the case of 1-inflation, and a lower bound estimator if other forms of heterogeneity exist (van der Heijden *et al.*, 2003a; Böhning *et al.*, 2005). When the OIPP model is not sufficiently comprehensive covariates may be included to reduce the underestimation bias of  $\hat{N}^{\text{OIPP}}$ , and in future work the OIPP model could be extended to handle unobserved heterogeneity in the same ways that the PP model has been extended. However, if 1-inflation is not explicitly modelled as a specific form of heterogeneity, then existing population size estimation may be much too large.

## Acknowledgements

We thank the Associate Editor and referees for thoughtful and helpful comments that much improved the paper.

## References

- Böhning, D., Baksh, M. F., Lerdsuwansri, R. and Gallagher, J. (2013) Use of the ratio plot in capture–recapture estimation. *J. Computat Graph. Statist.*, **22**, 135–155.
- Böhning, D., Dietz, E., Kuhnert, R. and Schön, D. (2005) Mixture models for capture-recapture count data. *Statist. Meth. Appl.*, **14**, 29–43.
- Böhning, D., Dietz, E., Schlattmann, P., Mendonça, L. and Kirchner, U. (1999) The zero-inflated Poisson model and the decayed, missing and filled teeth index in dental epidemiology. *J. R. Statist. Soc. A*, **162**, 195–209.
- Böhning, D., Suppawattanabodee, B., Kusolvisitkul, W. and Viwatwongkasem, C. (2004) Estimating the number of drug users in Bangkok 2001: a capture–recapture approach using repeated entries in one list. *Eur. J. Epidem.*, **19**, 1075–1083.
- Borchers, D. L., Buckland, S. T. and Zucchini, W. (2002) *Estimating Animal Abundance: Closed Populations*. New York: Springer Science and Business Media.
- van den Broek, J. (1995) A score test for zero inflation in a Poisson distribution. *Biometrics*, **51**, 738–743.
- Chant, D. (1974) On asymptotic tests of composite hypotheses in nonstandard conditions. *Biometrika*, **61**, 291–298.
- Chao, A. (1987) Estimating the population size for capture-recapture data with unequal catchability. *Biometrics*, **43**, 783–791.
- Chao, A. (1989) Estimating population size for sparse data in capture-recapture experiments. *Biometrics*, **45**, 427–438.
- Chernoff, H. (1954) On the distribution of the likelihood ratio. *Ann. Math. Statist.*, **25**, 573–578.
- Chernoff, H. and Lehmann, E. L. (1954) The use of maximum likelihood estimates in  $\chi^2$  tests for goodness of fit. *Ann. Math. Statist.*, **25**, 579–586.
- Choulakian, V., Lockhart, R. A. and Stephens, M. A. (1994) Cramér-von Mises statistics for discrete distributions. *Can. J. Statist.*, **22**, 125–137.
- Cochran, G. (1978) Laplace's ratio estimator. In *Contributions to Survey Sampling and Applied Statistics* (ed. H. A. David), pp. 3–10. New York: Academic Press.
- Conover, W. J. (1972) A Kolmogorov goodness-of-fit test for discontinuous distributions. *J. Am. Statist. Ass.*, **67**, 591–596.
- Cramér, H. (1928) On the composition of elementary errors: First paper, Mathematical deductions. *Scand. Act. J.*, **1**, 13–74.
- Crepon, B. and Duguet, E. (1997) Research and development, competition and innovation pseudo-maximum likelihood and simulated maximum likelihood methods applied to count data models with heterogeneity. *J. Econometr.*, **79**, 355–378.
- Cruyff, M. J. L. F. and van der Heijden, P. G. M. (2008) Point and interval estimation of the population size using a zero-truncated negative binomial regression model. *Biometr. J.*, **50**, 1035–1050.
- Edwards, W. R. and Eberhardt, L. (1967) Estimating cottontail abundance from livetrapping data. *J. Wildlif. Mangmnt.*, **31**, 87–96.
- Gleser, L. J. (1985) Exact power of goodness-of-fit tests of Kolmogorov type for discontinuous distributions. *J. Am. Statist. Ass.*, **80**, 954–958.
- Good, I. J. (1953) The population frequencies of species and the estimation of population parameters. *Biometrika*, **40**, 237–264.
- Greene, W. H. (2012) *Econometric Analysis*, 7th edn. Upper Saddle River: Prentice Hall.
- Gurmu, S. and Trivedi, P. K. (1996) Excess zeros in count models for recreational trips. *J. Bus. Econ. Statist.*, **14**, 469–477.
- Hald, A. (2003) *A History of Probability and Statistics and Their Applications before 1750*. Chichester: Wiley.
- van der Heijden, P. G., Bustami, R., Cruyff, M. J., Engbersen, G. and Van Houwelingen, H. C. (2003a) Point and interval estimation of the population size using the truncated Poisson regression model. *Statist. Modllng*, **3**, 305–322.
- van der Heijden, P. G., Cruyff, M. and Van Houwelingen, H. C. (2003b) Estimating the size of a criminal population from police records using the truncated Poisson regression model. *Statist. Neerland.*, **57**, 289–304.
- Horowitz, J. L. (1994) Bootstrap-based critical values for the information matrix test. *J. Econometr.*, **61**, 395–411.
- Horowitz, J. L. and Savin, N. E. (2000) Empirically relevant critical values for hypothesis tests: a bootstrap approach. *J. Econometr.*, **95**, 375–389.
- Horvitz, D. G. and Thompson, D. J. (1952) A generalization of sampling without replacement from a finite universe. *J. Am. Statist. Ass.*, **47**, 663–685.
- Johnson, N. L. and Kotz, S. (1969) *Discrete Distributions*. Boston: Houghton Mifflin.
- Johnson, N. L., Kotz, S. and Kemp, A. W. (1992) *Univariate Discrete Distributions*, 2nd edn. New York: Wiley.
- Keith, L. B. and Meslow, E. C. (1968) Trap response by snowshoe hares. *J. Wildlif. Mangmnt.*, **32**, 795–801.
- King, M. L. and Giles, D. E. A. (1984) Autocorrelation pretesting in the linear model: estimation, testing and prediction. *J. Econometr.*, **25**, 35–48.
- Kolmogorov, A. (1933) Sulla determinazione empirica di una legge di dis-Tribuzione. *G. Ist. Ital. Att.*, **4**, 1–11.
- Lambert, D. (1992) Zero-inflated Poisson regression, with an application to defects in manufacturing. *Technometrics*, **34**, 1–14.

- Lancaster, H. O. (1969) *The Chi-squared Distribution*. New York: Wiley.
- McCrea, R. S. and Morgan, B. J. (2014) *Analysis of Capture–Recapture Data*. Boca Raton: CRC Press.
- McKendrick, A. G. (1926) Applications of mathematics to medical problems. *Proc. Edinb. Math. Soc.*, **44**, 98–130.
- von Mises, R. E. (1928) *Wahrscheinlichkeit, Statistik und Wahrheit*. Vienna: Springer.
- Moran, P. A. P. (1971) Maximum likelihood estimation in non-standard conditions. *Proc. Camb. Philos. Soc.*, **70**, 441–450.
- Mullahy, J. (1986) Specification and testing of some modified count data models. *J. Econometr.*, **33**, 341–365.
- Niwitpong, S. A., Böhning, D., van der Heijden, P. G. and Holling, H. (2013) Capture–recapture estimation based upon the geometric distribution allowing for heterogeneity. *Metrika*, **76**, 495–519.
- Otis, D. L., Burnham, K. P., White, G. C. and Anderson, D. R. (1978) Statistical inference from capture data on closed animal populations. *Wildlif. Monogr.*, **62**, 1–135.
- Overton, W. S. and Davis, D. E. (1969) *Wildlife Management Techniques: Estimating the Numbers of Animals in Wildlife Populations*, pp. 403–455. Washington DC: Wildlife Society.
- Pahl, P. J. (1969) On testing for goodness-of-fit of the negative binomial distribution when expectations are small. *Biometrics*, **25**, 143–151.
- Rao, C. R. (1948) Large sample tests of statistical hypotheses concerning several parameters with applications to problems of estimation. *Math. Proc. Camb. Philos. Soc.*, **44**, 50–57.
- Rossmo, D. K. and Routledge, R. (1990) Estimating the size of criminal populations. *J. Quant. Crim.*, **6**, 293–314.
- Smirnov, H. (1939) Sur les écarts de la courbe de distribution empirique. *Rec. Math.*, **6**, 3–26.
- Vergne, T., Paul, M. C., Chaengprachak, W., Durand, B., Gilbert, M., Dufour, B. and Grosbois, V. (2014) Zero-inflated models for identifying disease risk factors when case detection is imperfect: application to highly pathogenic avian influenza H5N1 in Thailand. *Prev. Veter. Med.*, **114**, 28–36.
- Vilas, V. D. R. and Böhning, D. (2008) Application of one-list capture–recapture models to scrapie surveillance data in Great Britain. *Prev. Veter. Med.*, **85**, 253–266.
- Yee, T. W. (2011) VGAM: vector generalized linear and additive models. *R Package Version 0.8-2*. University of Auckland, Auckland. (Available from <http://CRAN.R-project.org/package=VGAM>.)
- Zelterman, D. (1988) Robust estimation in truncated discrete distributions with application to capture-recapture experiments. *J. Statist. Plannng Inf.*, **18**, 225–237.

#### Supporting information

Additional ‘supporting information’ may be found in the on-line version of this article:

‘Web appendix for “Estimation of the population size using the one-inflated positive Poisson model”’.