This is an electronic version of an article published in Communications in Statistics - Simulation and Computation, 1532-4141, Volume 29, Issue 4, 2000, Pages 1215 - 1237. The published article is available online at:

http://www.informaworld.com/smpp/content~db=all?content=10.1080/03610910008813661

COVERAGE-ADJUSTED ESTIMATORS FOR MARK-RECAPTURE IN HETEROGENEOUS POPULATIONS

J. Ashbridge and I. B. J. Goudie

School of Mathematics and Statistics,
University of St Andrews, St Andrews, Fife, KY16 9SS, Scotland.

Key Words: capture-recapture sampling; jackknife estimator; population size; root mean square error; sample coverage.

ABSTRACT

Consideration of coverage yields a new class of estimators of population size for the standard mark-recapture model which permits heterogeneity of capture probabilities. Real data and simulation studies are used to assess these coverage-adjusted estimators. The simulations highlight the need for estimators that perform well for a wide range of values of the mean and coefficient of variation of the capture probabilities. When judged for this type of robustness, the simulations provide good grounds for preferring the new estimators to earlier ones for this model, except when the number of sampling occasions is large. A bootstrapping approach is used to estimate the standard errors of the new estimators, and to obtain confidence intervals for the population size.

1. INTRODUCTION

We consider the standard model M_h applied to a target population of unknown size N. Independently of other animals and independently of its previous capture history, animal i $(i = 1, 2, \ldots, N)$ is caught in sample j $(j = 1, \ldots, t)$ with probability p_i. It is thus assumed that, for each animal, this probability remains constant from one sample to the next. When an animal is caught it receives a distinguishing mark so that it can be recognised if it is recaptured in a later sample. In order to estimate N, or equivalently the number of animals that remain unseen, some assumption must be made about the capture probabilities. We adopt the standard practice of assuming that $p_1, p_2, \ldots, p_N$ constitute a random sample from some distribution on (0, 1).

The model M_h is one of the class of closed population models presented by Otis *et al*. (1978) and Pollock *et al*. (1990). Estimators for this model have been proposed by many authors, including Pollock and Otto (1983), Smith and van Belle (1984), Chao (1989) and Norris and Pollock (1996), but the ones most commonly favoured are probably still the jackknife estimators of Burnham and Overton (1978). Lee and Chao (1994), however, assert that the coverage-based estimators of Chao, Lee and Jeng (1992) are to be preferred, except when the heterogeneity is very mild. In the latter case they recommend the maximum likelihood estimator for the model M_0 under which the capture probability is the same for all animals and remains constant from one sample to the next.

In the standard form of model M_h which we address, it is assumed that no information in addition to the capture histories is available. Huggins (1989) provides inferential methods for the related problems in which individual or environmental covariates have been recorded.

In §2 below we introduce a new class of estimators for model M_h which are related to the estimator of Overton (1969), but in which the estimates of the capture probabilities are adjusted according to the estimated coverage. A bootstrap method for estimating the standard error of these "coverage-adjusted" estimators is described in §4. In §5 we consider four data sets which have been previously discussed in the literature, and compare the performance of the new estimators with that of a selection of other estimators for this model, which are described in §3. Interval estimation, and the performance of the standard error estimators, are considered in §7. The simulation study presented in §6 provides a more detailed evaluation of the new estimators, and confirms that the performance of all the population-size estimators under consideration depends on the mean and coefficient of variation of the capture probabilities. This emphasises the need for estimators which perform well for a wide range of values of these parameters, and in §8 we therefore use simulation to examine whether the various estimators are robust in this sense. We summarise in §9 which estimators can be recommended to the practitioner on the basis of this analysis.

2. THE NEW COVERAGE-ADJUSTED ESTIMATORS

Let S_i denote the total number of samples in which animal i is caught and let Δ be the set of animals which are detected at least once. Given the values of the parameters $p_1, p_2, \ldots, p_N$, the random variables $S_1, S_2, \ldots, S_N$ are clearly independent binomial random variables. In particular, the probability that animal i is caught at least once is

$$P(i \; \varepsilon \; \Delta) = P(S_i > 0) = 1 - \left(1 - p_i\right)^t \qquad i = 1, 2, \ldots, N. \qquad (1)$$

If the probabilities $p_1, p_2, \ldots, p_N$ were known, there would be a Horvitz-Thompson estimator of N given by

$$\tilde{N} = \sum_{i \in \Delta} \left\{1 - \left(1 - p_i\right)^t\right\}^{-1} = \sum_{i=1}^{N} \phi(i) \left\{1 - \left(1 - p_i\right)^t\right\}^{-1}, \qquad (2)$$

where $\phi(i)$ is one or zero depending on whether or not animal $i \in \Delta$. Taking an expectation of the second sum and using (1), it is easy to see that, when the probabilities $p_1, p_2, \ldots, p_N$ are known, $\tilde{N}$ is an unbiased estimator of N.

In practice the probabilities $p_1, p_2, \ldots, p_N$ are, of course, not known. To derive a viable estimator from (2), Overton (1969) noted that the maximum likelihood estimate of the binomial parameter p_i $(i = 1, \ldots, N)$ is S_i / t. Inserting these estimates in (2) yielded an estimator

$$Q = \sum_{i \in \Delta} \left\{1 - \left(1 - \frac{S_i}{t}\right)^t\right\}^{-1} = \sum_{j=1}^{t} f_j \left\{1 - \left(1 - \frac{j}{t}\right)^t\right\}^{-1},$$

where f_j $(j = 1, 2, \ldots, t)$ denotes the number of animals that are caught exactly j times. Note that, for greater clarity when comparing estimators in the later sections of this paper, we will simply use a single upper case letter to denote each of the estimators of interest.

Pollock and Otto (1983) speculated that it might be possible to improve Overton's estimator Q by finding better estimates of the capture probabilities.

This paper seeks to substantiate this suggestion. To achieve this we use estimates of the coverage C of the observed set Δ, which is defined by the equation

$$C\sum_{1}^{N} p_i = \sum_{i\in\Delta} p_i \,. \tag{3}$$

The most commonly used estimate of C, due to Good (1953), is

$$\hat{C}_1 = 1 - \frac{f_1}{Z}, \tag{4}$$

where $Z = S_1 + S_2 + \ldots + S_N$ is the total number of sightings. Assuming $t \geq 3$, we also make use of two bias-corrected versions of the estimator $\hat{C}_1$. These are given by

$$\hat{C}_2 = \min\left\{1,\ 1 - \frac{f_1}{Z} + \frac{2}{(t-1)}\frac{f_2}{Z}\right\}, \tag{5}$$

$$\hat{C}_3 = \min\left\{1,\ 1 - \frac{f_1}{Z} + \frac{2}{(t-1)}\frac{f_2}{Z} - \frac{6}{(t-1)(t-2)}\frac{f_3}{Z}\right\}. \tag{6}$$

The estimators $\hat{C}_2$ and $\hat{C}_3$ differ from the corresponding coverage estimators of Chao *et al*. (1992) only in being bounded above by unity. The simulation study reported in §6 showed that the original versions of these estimators could exceed unity, albeit only rarely.

Since for $i \notin \Delta$ the maximum likelihood estimate of p_i is zero, the maximum likelihood estimate of either of the two sums in (3) is Z/t, implying an unrealistic estimate of the coverage as unity. Consideration of its mean endorses the use of the statistic Z/t as an estimator of $\sum_{1}^{N} p_i$. Adopting one of the more realistic estimators $\hat{C}_u\ (u = 1, 2, 3)$ of C, equation (3) then suggests estimating $\sum_{i\in\Delta} p_i$ by $\hat{C}_u Z/t = \left(\hat{C}_u/t\right)\sum_{i\in\Delta} S_i$. Deflating the maximum likelihood estimate of each capture probability by the same factor $\hat{C}_u$ gives $\hat{C}_u S_i/t$ as the corresponding estimator of p_i for $i \in \Delta$. Insertion of these revised estimators of p_i in (2) yields new estimators of N. The formulae for these coverage-adjusted estimators, which we denote by W, X and Y, are shown in Table I.

3. ALTERNATIVE ESTIMATORS FOR THIS MODEL

Also listed in Table I are a selection of other estimators that are used for Model M_h. Included in this list are L, the maximum likelihood estimator under model M_0, and the estimator R proposed by Darroch and Ratcliff (1980) for mark-recapture experiments in which the samples are of size one. The expression given in Table I for the first-order jackknife U of Overton's estimator Q corrects a typographical error in the source paper. The estimator V, described as a bootstrap estimator by Smith and van Belle (1984), uses a notional resampling procedure to attempt to correct the bias that would result from estimating N by the total number D of distinct animals captured.

The results that we present below indicate that, even within a class of estimators of similar type, it is often not obvious which estimator is to be preferred. We therefore include all three of the estimators for Model M_h considered by Chao *et al*. (1992). As shown in Table I, these three estimators, which we denote by E, F and G, are defined in terms of the number D of distinct captures and the estimated coefficient of variation $\hat{\gamma}_u$ of the capture probabilities given by

$$\hat{\gamma}_u^2 = \max\left\{ 0,\ -1 + \frac{Dt\sum j(j-1)f_j}{\hat{C}_u(t-1)Z^2} \right\} \qquad u = 1, 2, 3, \tag{7}$$

where $\hat{C}_1$, $\hat{C}_2$ and $\hat{C}_3$ are given by (4), (5) and (6) respectively. The estimators E, F and G may be viewed as refinements of R. Note also that each of the coverage-adjusted estimators corresponds to one of the estimators E, F and G in the sense that it depends on the same coverage estimator.

Burnham and Overton (1978) gave the formulae for the first-order jackknife K of the statistic D, and for the corresponding jackknives of second to fifth order. They also gave a selection procedure for determining which to adopt. Rosenberg, Overton and Anthony (1995) suggest that some caution is needed with the jackknife selection procedure and commend the use of first or second order jackknives when population sizes are low and capture probabilities are low and heterogeneous. Another minor shortcoming of this selection procedure, as described in Burnham and Overton (1979), is that it cannot always be automated, since, if the fourth order is rejected, their advice is to select the jackknife of whichever of the first three orders appears most appropriate. The version J of the jackknife estimator which we use avoids this subjectivity by always selecting the first order jackknife *K* when the fourth order is rejected.

We denote by *I* the interpolated jackknife, the original version of which is again due to Burnham and Overton (1978). When the selection procedure for the jackknife *J* chooses the first-order jackknife *K*, then *I* is simply equal to *K*. When the procedure chooses a jackknife of order υ, for $\upsilon > 1$, then *I* is a weighted average of the jackknives of orders (υ - 1) and υ.

We also tried a variant J' of the jackknife estimator which always used the fifth order jackknife when the fourth order was rejected. Rejection of the fourth

TABLE I. Estimators for Model M_h.

Estimator	Reference	Description or formula
E, *F*, *G*	Chao *et al.* (1992)	$\left(D + f_1\hat{\gamma}_u^2\right)/\hat{C}_u$ (u = 1, 2, 3). See (4), (5), (6), (7) for $\hat{C}_1, \hat{C}_2, \hat{C}_3$ and $\hat{\gamma}_u^2$ (u = 1, 2, 3) respectively.
H, I	Burnham and Overton (1978)	Two versions of the interpolated jackknife
J, *K*	Burnham and Overton (1978)	Jackknife, First-order jackknife
L	Otis *et al.* (1978)	MLE under model M_0
Q	Overton (1969)	$\sum_{j=1}^{t} f_j \left\{1 - \left(1 - \frac{j}{t}\right)^t\right\}^{-1}$
R	Darroch and Ratcliff (1980)	$D/\hat{C}_1$
U	Pollock and Otto (1983)	$t\sum_{j=1}^{t} a_{t,j} f_j - \frac{t-1}{t}\sum_{j=1}^{t-1} a_{t-1,j}\left\{(t-j)f_j + (j+1)f_{j+1}\right\}$ where $a_{t,j} = \left\{1 - (1 - j/t)^t\right\}^{-1}$.
V	Smith and van Belle (1984)	$D + \sum_{j=1}^{t} f_j \left\{1 - (j/t)\right\}^t$
W, X, Y	Coverage-adjusted estimators introduced in this paper	$\sum_{j=1}^{t} f_j \left\{1 - \left(1 - \frac{j\hat{C}_u}{t}\right)^t\right\}^{-1}$ (u = 1, 2, 3). See (4), (5), (6) for $\hat{C}_1$, $\hat{C}_2$ and $\hat{C}_3$ respectively.

order jackknife occurs infrequently for $t > 5$, and never for $t \leq 5$, and thus it was unsurprising that J and J' gave identical estimates for each of the four real data sets discussed below. We do not display J' in FIGs. 1 and 2, since in all of the cases illustrated, whenever the two estimators differ, J' has both a larger absolute bias and a larger standard deviation than J.

Following the same interpolation procedure used to obtain I from the selection procedure for J, we can, however, define an interpolated estimator H, based on the procedure defining J'. This estimator H can thus employ weighted averages of the jackknives of fourth and fifth order. In the simulation study in §6, we display H in the plots for $t = 10$, this being the one case in that study in which H has a smaller absolute bias than I.

In this paper we are thus including a wide range of alternative estimators for model M_h, but the number now available in the literature is such that it is impractical to attempt to be exhaustive. We believe, however, that the validity of our conclusions is unlikely to be affected by the estimators that we have omitted. For instance, although Norris and Pollock (1996) showed that their nonparametric maximum likelihood estimator displayed very small bias when applied to model M_h, its standard deviation was usually very large, when compared to that of the estimators of Chao *et al.* (1992).

4. THE STANDARD ERROR OF THE NEW ESTIMATORS

The estimators I, J, K, Q, U, V are all linear functions of the frequencies f_j $(j = 1, 2, \ldots, t)$, and thus estimators of their variances are readily obtained (see, for example, Pollock and Otto, 1983). Explicit expressions are also available for the asymptotic variances of the maximum likelihood estimator L and Darroch and Ratcliff's R. For the estimators E, F and G, Chao *et al.* (1992) indicate how, by assuming that the capture probabilities form a random sample from some distribution, estimates of the asymptotic variance can be obtained. Note that, for the estimators of Chao *et al.*, these asymptotic variances are not functions of the sufficient statistics alone.

After trying a number of approaches to estimating the standard error of the coverage adjusted estimators, W, X and Y, we concluded that the most satisfactory was that based on the parametric bootstrap. If the estimate of population size is

$\hat{N}$, realisations are simulated from a population of size $\hat{N}$. Each realisation comprises t samples, in which the capture probability of the i^{th} animal is given by $\hat{C}_u S_i / t$ for $i \in \Delta$.

It is also necessary to assign capture probabilities to the $\varphi = \hat{N} - D$ animals that are estimated to have eluded detection in the original sampling process. The discussion in §2 suggests estimating $\sum_{i \notin \Delta} p_i = \sum_{1}^{N} p_i - \sum_{i \in \Delta} p_i$ by $\left(1 - \hat{C}_u\right) Z / t$, where u =1 for the estimator W, 2 for X and 3 for Y. It is less clear how this probability should be subdivided between the φ animals concerned, and, as with other methods of estimating the standard error, a number of the variants we tried led to a strong negative bias. The most satisfactory way we found to assign probabilities to these animals was to draw a random sample of size φ from a uniform distribution, and then, to rescale these readings so that they summed to $\left(1 - \hat{C}_u\right) Z / t$. Such an assignment is clearly arbitrary, but this is inevitable when all that is known about these animals is that they were not seen. For each simulated realisation of the sampling process, the value of the particular coverage-adjusted estimator is calculated, discarding any resamples for which this value is less than the number of distinct animals observed. The standard deviation of the estimates from the resamples then provides an estimated standard error for the coverage-adjusted estimator. It would appear to be advisable to use at least 1000 resamples for this bootstrap estimator.

5. REAL DATA EXAMPLES

Table II summarises four data sets which have been discussed previously in the literature, and displays the estimates of the coverage given by equations (4), (5) and (6). Otis *et al.* (1978) consider both the trapping data on snowshoe hares, collected but not published by Burnham and Cushwa, and the taxicab data of Carothers (1973), obtained under his sampling scheme *A*. The model selection procedure in program CAPTURE, provided by Otis *et al.* (1978), chooses model M_h for both these data sets and also indicates that this is one of the two most plausible models for the meadow vole data presented in Pollock *et al.* (1990). The heterogeneity model M_h has also been judged appropriate by Norris and Pollock (1996) for the eastern chipmunk data, which is due to Mares, Streilein and Willig (1981).

TABLE II Four standard data sets.

Data	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_i (i>10)
Hares	25	22	13	5	1	2	0	0	0	0	0
Taxicabs	142	81	49	7	3	1	0	0	0	0	0
Voles	29	15	15	16	27	0	0	0	0	0	0
Chipmunks	14	13	18	12	7	5	1	1	0	1	0

Summary statistics	d	z	t	$\hat{C}_1$	$\hat{C}_2$	$\hat{C}_3$
Hares	68	145	6	0.83	0.89	0.86
Taxicabs	283	500	10	0.72	0.75	0.74
Voles	102	303	5	0.90	0.93	0.90
Chipmunks	72	232	13	0.94	0.95	0.95

In Table III the various estimators are evaluated on these four data sets. Both the chipmunk and the taxicab populations are usually treated as being of known sizes, 82 and 420 respectively, although the latter figure is probably more accurately viewed as a close approximation to the true value. For the taxicabs,

TABLE III Values of the estimators, together with their estimated standard errors, on the four data sets.

Estimator	Hares		Taxicabs		Voles		Chipmunks	
	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.	Estimate	s.e.
L	75	3.4	368	14.6	103	1.1	73	1.5
Q	83	4.6	370	11.4	118	4.7	82	3.8
U	100	11.0	504	28.2	144	11.6	87	8.8
R	82	5.5	395	18.4	113	3.5	77	2.3
V	79	3.6	343	8.8	113	3.7	79	3.0
K	89	6.2	411	15.6	125	6.5	85	5.0
J	89	6.2	495	36.4	142	12.2	85	5.0
I	89	6.2	469	25.9	138	10.6	85	5.0
E	89	8.9	416	25.9	123	7.8	79	4.1
F	81	7.4	386	23.2	118	7.1	78	3.9
G	84	8.2	393	23.9	123	7.8	79	4.0
W	90	7.0	439	21.0	121	5.1	83	4.3
X	87	6.2	427	19.9	120	5.1	83	4.1
Y	88	6.6	429	19.6	121	5.2	83	4.3
True size, if known	-		420 approx		-		82	

the closest estimate is provided by *E*, due to Chao *et al.*, with the coverage adjusted estimators *X* and *Y* and the first order jackknife *K* also performing well. For the chipmunk data Overton's estimator *Q* equals the true value, with the next closest estimates being given by the three coverage-adjusted estimators *W, X* and *Y*. Burnham and Overton's jackknife estimators *I, J* and *K* slightly over-estimate, whilst *E, F* and *G* all underestimate. Although the danger of drawing firm conclusions on the basis of a small number of data sets is clear, these data sets do also illustrate the well-known tendency of the maximum likelihood estimator *L* and Darroch and Ratcliff's estimator *R* to underestimate in the presence of heterogeneity. The results suggest that Smith and van Belle's estimator *V* may also be negatively biased, and that Pollock and Otto's *U*, the jackknifed version of *Q*, may over-estimate.

The behaviour of many of the estimators is similar on the snowshoe hare and meadow vole data sets. The largest estimates are again provided by *U*, with those supplied by the Burnham and Overton jackknives also tending to be amongst the largest. Estimators *L*, *R*, *V* and *F* again provide comparatively low population estimates, and on these two data sets the same is true of *Q*. The remaining five estimators *E*, *G*, *W*, *X* and *Y* are not consistently extreme in either direction. For three of the four data sets shown in Table III the three coverage-adjusted estimators in this group of five have smaller standard errors than the estimators *E* and *G*. It should, however, be noted that such comparisons depend on the accuracy of the various estimators of standard error. Further evidence on the standard deviations of the estimators appears in sections 6 and 8 and on the performance of the standard error estimator for *Y* in §7.

Carothers (1973) also considered estimation based on 20 different subsets of his complete taxicab data set, obtained under his sampling scheme A. On 16 of these subsets, the estimate provided by *W* is closer to the true value of 420 than that given by *E*, the corresponding estimator in the class due to Chao *et al.* (1992). Similarly on 17 of these subsets, the estimators *X* and *Y* improve on *F* and *G* respectively. It should, however, be noted that the 20 estimates are not independent since many of the subsets have observations in common.

6. A SIMULATION STUDY

The relative merits of the various estimators for model M_h were further investigated by simulation. Some previous authors have chosen to divide their simulated target populations into distinct subsets, assigning to each animal in a particular subset a common detection probability. For the simulations in this paper, however, the capture probabilities are random deviates, drawn from beta distributions, the family of distributions used for the majority of the simulations given in Burnham and Overton (1979). Lee and Chao (1994) stress that, in choosing an appropriate estimator due attention must be paid to the coefficient of variation of the distribution from which the capture probabilities are drawn. They indicate that the appreciation of the importance of this quantity dates back to Cormack (1966). We therefore chose to select systematically each beta distribution used, in order to investigate possible dependence on its mean μ and coefficient of variation γ.

We took a grid of points in the (μ, γ) plane, noting that the parameters α and β of the beta distribution with density $\left[\Gamma(\alpha+\beta)/\{\Gamma(\alpha)\Gamma(\beta)\}\right]p^{\alpha-1}(1-p)^{\beta-1}$ for $0 < p < 1$ satisfy the equations

$$\alpha = -\mu + \left\{(1-\mu)/\gamma^2\right\}, \quad \beta = \alpha(1-\mu)/\mu .$$

In selecting a grid, we first judged that the values of (μ, γ) of interest lay within the rectangle

$$\Omega \equiv \left\{(\mu, \gamma) : 0.04 \le \mu \le 0.2,\ 0.3 \le \gamma \le 0.8\right\}.$$

The range of mean detection probabilities within Ω is similar to that adopted in previous studies. We chose 0.3 as the minimum value for γ, since model M_h is rarely adopted for small values of γ. Menkens and Anderson (1988) note that, particularly when the population size is small or the data are sparse, mild departures from the model M_0 are unlikely to be detected by the model selection procedure of Otis *et al.* (1978). Lee and Chao (1994) tentatively suggest 0.4 as the value of γ below which it is appropriate simply to use *L,* the maximum likelihood estimator under model M_0. The evidence of the above taxicab and chipmunk data sets, for which $\hat{\gamma}_1$ takes the values 0.327 and 0.331 respectively, led us to use the slightly lower level. Some studies on population size estimation have entertained values of γ that are much larger than our maximum level of 0.8, but such values appear hard to justify empirically. Our choice of 0.8 follows that of Chao *et al.* (1992).

Noting that many of the points on the perimeter of Ω correspond to relatively extreme scenarios, we chose from the interior of the rectangle the grid of nine points given by

$$\{(\mu, \gamma) \ : \ \mu = 0.067, 0.120, 0.173; \gamma = 0.383, 0.550, 0.717\}.$$

If Ω is partitioned into nine smaller rectangles of equal size in the obvious manner, the chosen points thus lie at the centres of these smaller rectangles. For each value of t, we considered each point on the grid in turn and drew independent random samples, of size N, from the beta distribution corresponding to that point. Each random sample thus represented a set of capture probabilities for the N animals. From each such sample, a realisation of the trapping process was then simulated, by generating, for each animal, t independent deviates from the Bernoulli distribution with the appropriate capture probability, to represent whether or not it was caught on each of t sampling occasions. For each of the data sets thus generated, the value of each of the different estimators was determined. As many of the estimators are only finite if at least one recapture occurs, any data set not meeting this condition was discarded. The simulation procedure continued until 10^4 data sets for which the condition did hold had been generated for each point on the grid.

Based on its values in the 10^4 independent simulations, the mean m and standard deviation s of each estimator were then determined, with these statistics thus being conditional on at least one recapture occurring. FIG. 1 shows the results of this simulation for each of the nine points on the grid in the case $t = 5$. Note that, as each of the 10^4 data sets was based on a different set of capture probabilities, the values of the standard deviation s incorporate not only the sampling variation that would be experienced in practice when sampling from a particular population with fixed detection probabilities, but also the variability due to sampling from the particular beta distribution. We have, however, conducted further simulations which indicate that the component of the variance due to the latter source is very small. Indeed, were it to be excluded, the consequent difference to the plots in FIG. 1 would be hard to detect visually.

An initial inspection of these plots suggests that the performance of some of the estimators may be similar across much of the grid. In all nine plots negative

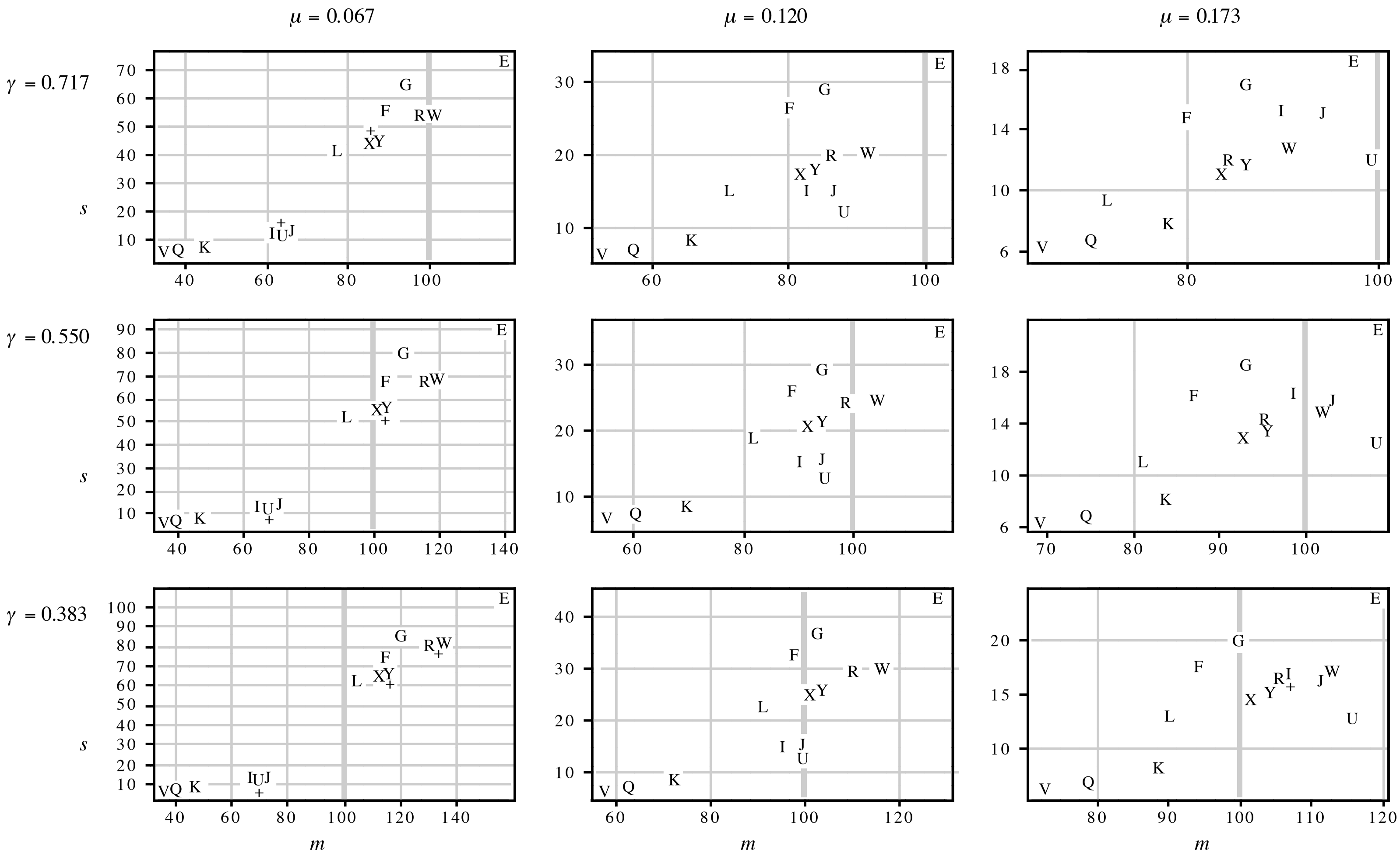

FIG.1. Plots of the standard deviation *(s)* against the mean *(m)* for each estimator in the case $N = 100$, $t = 5$ at each location on the nine-point grid. The positions of estimators adjacent to + signs have been slightly perturbed for greater clarity.

bias is exhibited by Smith and van Belle's estimator *V*, Overton's estimator *Q* and the first order jackknife *K*, and in some cases this bias is severe. The maximum likelihood estimator *L* also tends to suffer from negative bias for the larger values of γ. In all nine cases the standard deviations of *W*, *X* and *Y* are less than those of *E*, *F* and *G* respectively. Estimator *X* is the coverage-adjusted estimator with the smallest standard deviation, but the standard deviation of *Y* is usually only marginally larger.

The pattern in the relative performances of the estimators across the nine plots should also be noticed. For each fixed value of μ, the relative positions of the estimators in the plots change little as γ increases, though there is a general tendency for the means of the estimators to decrease. The plots are, however, less stable under changes in the value of μ. Observe, in particular, that three of the jackknife estimators, namely *I*, *J* and *U*, have relatively small means for $\mu = 0.067$, but amongst the largest means for $\mu = 0.173$.

When choosing an estimator, however, the differences between the plots are of more significance than the similarities. The jackknife estimators *J* and *I*, widely accepted as the most useful estimators for this model, perform very well at some points on the grid, such as $(\mu, \gamma) = (0.120, 0.383)$, but underestimate badly for $\mu = 0.067$. The estimator *E* has the smallest bias for $(\mu, \gamma) = (0.120, 0.717)$, but has a strong positive bias when $\gamma = 0.383$. Pollock and Otto's *U* is arguably the best estimator for $\mu = 0.120$, but does less well when μ is small. FIG. 1 suggests that the estimators *F* and *G* of Chao *et al.* and the coverage-adjusted estimators *X* and *Y* avoid the severe biases sometimes shown by some of the other estimators, but the relative performance of these estimators in terms of bias varies from one plot to another.

It should be noted that the behaviour of the estimators in these examples, all of which are in the case $t = 5$, is not necessarily typical of their behaviour for other values of *t*. In particular it will be seen below that the estimators *Q* and *V* are more useful for larger values of *t*. The same is true of the first order jackknife *K*, which appears poor in comparison to *J* and *I* when $t = 5$.

Many authors have found it convenient to emphasise an ordering of the available estimators based on some one-dimensional statistic. Usually this statistic has been either the root mean square error (RMSE) or the relative mean square error, which, for a fixed value of *N*, is equivalent. For a different population size model, Goudie, Pollock and Ashbridge (1998) emphasised the danger that criteria based on mean square error tend to reward negative bias, often leading one to favour precise, but seriously biased, estimators. Many of the examples that we have examined have led us to believe that the same danger is often present for model M_h.

Inspection of the plots in FIG. 1 shows that the estimators with the largest means tend also to have the largest standard deviations, and this is a standard phenomenon in population size models of this type. RMSE, and related criteria, highlight the estimators with small standard deviation and therefore typically lead the user to select a negatively biased estimator. For instance, in the case where $N = 100$, $t = 5$ and $(\mu, \gamma) = (0.067, 0.717)$, the simulations suggest that the best estimator under RMSE is the jackknife estimator *J*, with a mean of 66.0 and a standard deviation of 13.1. Arguably preferable alternatives are Darroch and Ratcliff's *R* with a mean of 98.2 and a standard deviation of 54.0 or the coverage-adjusted estimator *W* with a mean of 101.9 and a standard deviation of 54.2. Despite their large variances, it seems more sensible to use one of these alternatives than the estimator *J* which one knows will seriously underestimate with high probability.

In our view, examples of this type suggest that much previous advice for this model has been too reliant on the RMSE criterion, and that this in part accounts for the prominence that has been afforded by users of the model to the jackknife estimator *J*. FIG. 1 clearly demonstrates that the jackknife estimators *J* and *I* should not be seen as invariably the appropriate estimators to adopt, and that the relative performances of all the estimators are dependent on both μ and γ. The latter comment raises the interesting possibility of seeking better identification of the range of values of (μ, γ) for which any particular estimator is to be preferred and basing the choice of estimator of *N* on preliminary estimates of μ and γ. This is not a route we judge to be currently viable since estimation of the coefficient of variation γ would appear to be at least as hard as estimation of *N*.

It would, however, appear ill-advised to dismiss lightly the evident variability in the relative performance of the estimators from one beta distribution to another. Many users will wish to adopt estimators which can be relied on to perform well across a range of situations. In §8 we will therefore examine the robustness of the various estimators.

7. CONFIDENCE INTERVALS AND EVALUATION OF THE STANDARD ERROR ESTIMATORS

A parametric bootstrap procedure for estimating the standard errors of the coverage-adjusted estimators was described in §4. The performance of these standard error estimators was also assessed by simulation at each of the nine points on the grid used in the last section. In the manner previously described, 10^3 independent data sets, each including at least one recapture, were generated. For each data set, the standard error of a particular coverage-adjusted estimator was then estimated based on a further 10^3 independent resampled data sets.

For $t = 5$, 10, 15 and 20 and for each of the nine points on the grid, Table IV shows, for the case $N = 100$, the observed standard deviation s of the estimator Y over the original 10^3 data sets and the mean $\hat{s}$ of the 10^3 bootstrap estimates of this quantity. These results suggest that the overall behaviour of the

TABLE IV The mean $\hat{s}$ of 10^3 bootstrap estimates of the standard deviation of the coverage-adjusted estimator Y, in the case $N = 100$, compared to the sample standard deviation s of Y. Also shown is the corresponding confidence interval coverage (CIC) for central intervals nominally at the 95% level.

		$\mu = 0.067$			$\mu = 0.12$			$\mu = 0.173$		
t	γ	s	$\hat{s}$	CIC	s	$\hat{s}$	CIC	s	$\hat{s}$	CIC
5	0.383	64.4	58.0	96	27.9	26.0	94	14.7	14.5	96
	0.550	56.2	50.5	93	19.9	20.7	89	13.3	12.7	91
	0.717	50.3	43.6	88	17.7	17.1	75	11.4	10.6	67
10	0.383	20.6	20.6	96	10.3	10.4	98	6.8	7.5	94
	0.550	18.4	17.6	88	9.8	9.5	93	7.0	6.9	96
	0.717	15.2	14.4	74	8.9	8.4	67	6.8	6.4	75
15	0.383	12.3	12.6	97	6.7	7.5	93	4.5	5.6	81
	0.550	11.8	11.3	92	6.8	7.0	97	5.1	5.4	96
	0.717	11.1	9.9	72	7.1	6.5	80	5.4	5.2	87
20	0.383	9.2	9.8	97	5.1	6.1	84	3.4	4.4	76
	0.550	9.0	9.0	96	5.6	5.9	96	3.9	4.5	96
	0.717	8.8	8.1	79	6.0	5.6	89	4.9	4.5	92

standard error estimators is satisfactory, though with a small negative bias for small t and a slight positive one for large t.

Confidence intervals for the population size N can also be found using the bootstrap estimates obtained by using estimator Y after resampling. If an estimate of N is obtained from each of η resamples, and $\hat{N}_i$ is the i^{th} $(i = 1, 2, \ldots, \eta)$ such estimate when they are ordered in increasing size, an approximate central 100(1 - 2α)% confidence interval for N is given by $\left(\hat{N}_j, \hat{N}_k\right)$, where $j = (\eta + 1)\alpha$ and $k = (\eta + 1)(1 - \alpha)$. If j or k is not an integer, the value may be rounded or else linear interpolation used. (See Buckland, 1984.)

In Table IV, as well as the standard error estimates, for each of the situations covered, we show the actual confidence interval coverage (CIC) for central intervals that are nominally at the 95% level. Earlier authors (Otis *et al.*, 1978; Chao, 1989) have reported problems of low confidence interval coverage when using jackknife estimators with this model. Rosenberg, Overton and Anthony (1995) showed that the coverage based on the first order jackknife K could at times be very poor, and sought to improve it using an adjusted variance estimator. Table IV shows that the same problem of low coverage is also sometimes present here. For $\gamma = 0.550$, the CIC of the nominally 95% confidence intervals is usually good, and in none of the cases falls below 0.88, but for γ = 0.717, and, to a lesser extent γ = 0.383, the coverage can be poor, albeit not as low as some of the levels reported for the estimator K. The coverage resulting from the use of the bias-corrected percentile method of Efron (1982) shows no overall improvement. Computations indicate that such an improvement can be achieved by the ad hoc modification of using the $100(1 - \alpha_1 - \alpha_2)$ % non-central interval $\left(\hat{N}_j, \hat{N}_k\right)$, given by $j = (\eta + 1)\alpha_1$ and $k = (\eta + 1)(1 - \alpha_2)$, with $\alpha_1 > \alpha_2$, but these non-central intervals are usually somewhat longer.

8. ROBUSTNESS

In view of the results in §6, the practitioner may seek reassurance that any proposed estimator will give reliable results over most, and preferably all, of the rectangle Ω. To examine the robustness of the estimators over the rectangle we modified our simulation procedure as follows. Instead of drawing random samples from the specific beta distribution corresponding to a single point in this rectangle, a random sample of points was selected according to a bivariate uniform distribution on Ω. At each point thus selected, a random sample of size *N* was drawn from the beta distribution corresponding to that point. As before, this sample of size *N* gave the capture probabilities for each of the *N* animals. To obtain a realisation of the trapping process, *t* independent deviates from the appropriate Bernoulli distribution were generated for each animal to indicate capture or non-capture on each of the *t* sampling occasions.

Independent simulations were conducted for each of the 18 values of (t, N) obtained by considering t = 5, 10, 15, 20, 25 and 30 for each of the parameter values N = 50, 100 and 200. In each of the cases, random selection of points in Ω concluded when 10^4 data sets had been obtained in which at least one recapture occurred, and the different estimators were evaluated on each of the data sets. Based on these 10^4 readings, the mean *m*, the absolute bias |*b*| and the standard deviation *s* of each estimator were then calculated, with all these measures thus being conditional on at least one recapture occurring. In FIG. 2, for each of the 18 cases, we plot $s/\sqrt{m}$ against $|b/N|$. The use of $s/\sqrt{m}$, the square root of the dispersion factor, yields approximate stability of the variability for different population sizes *N*. Adoption of the relative absolute bias $|b/N|$ as the horizontal variable is also designed to assist the experimenter to compare the relative performances of the estimators over an appropriate range of values of *N*. As well as the number *t* of sampling occasions, each row of FIG. 2 is labelled by the coverage *C*. The value shown is strictly the average coverage of the simulations in the case N = 100, but this average coverage varies little with *N*, always differing by less than 0.01 for the other two cases.

Estimators that perform relatively poorly for any particular value of *t* are not shown in the plots corresponding to that value in FIG.2. All the estimators listed in the 'large bias' column for a given value of *t* have relative biases which place them outside the plotted region. It is of interest to note which these are. The negative bias of the maximum likelihood estimator *L* is such that it falls outside each of the 18 plots. Other estimators which do not appear due to their large negative biases are Smith and van Belle's estimator *V* when $t \leq 25$, Overton's *Q* for $t \leq 15$ and Darroch and Ratcliff's *R* when $t \geq 15$. Similarly large positive biases exclude the jackknife estimator *J* when $t \geq 25$ and Pollock and

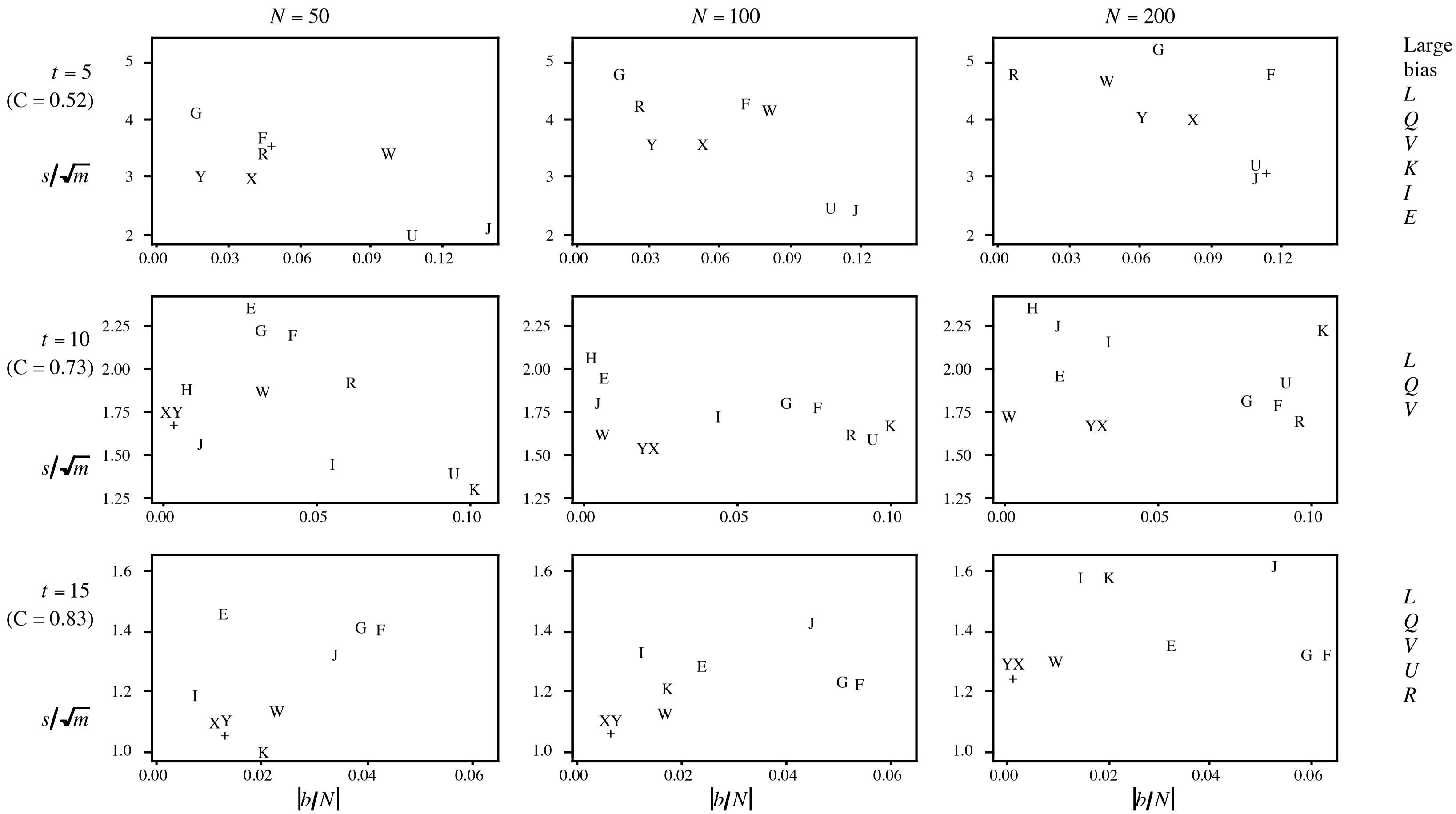

FIG.2. Plots of $s/\sqrt{m}$ against absolute relative bias $|b/N|$, when $t = 5$, 10 and 15, for the estimators that perform well over the rectangle Ω. The positions of estimators adjacent to + signs have been slightly perturbed for greater clarity.

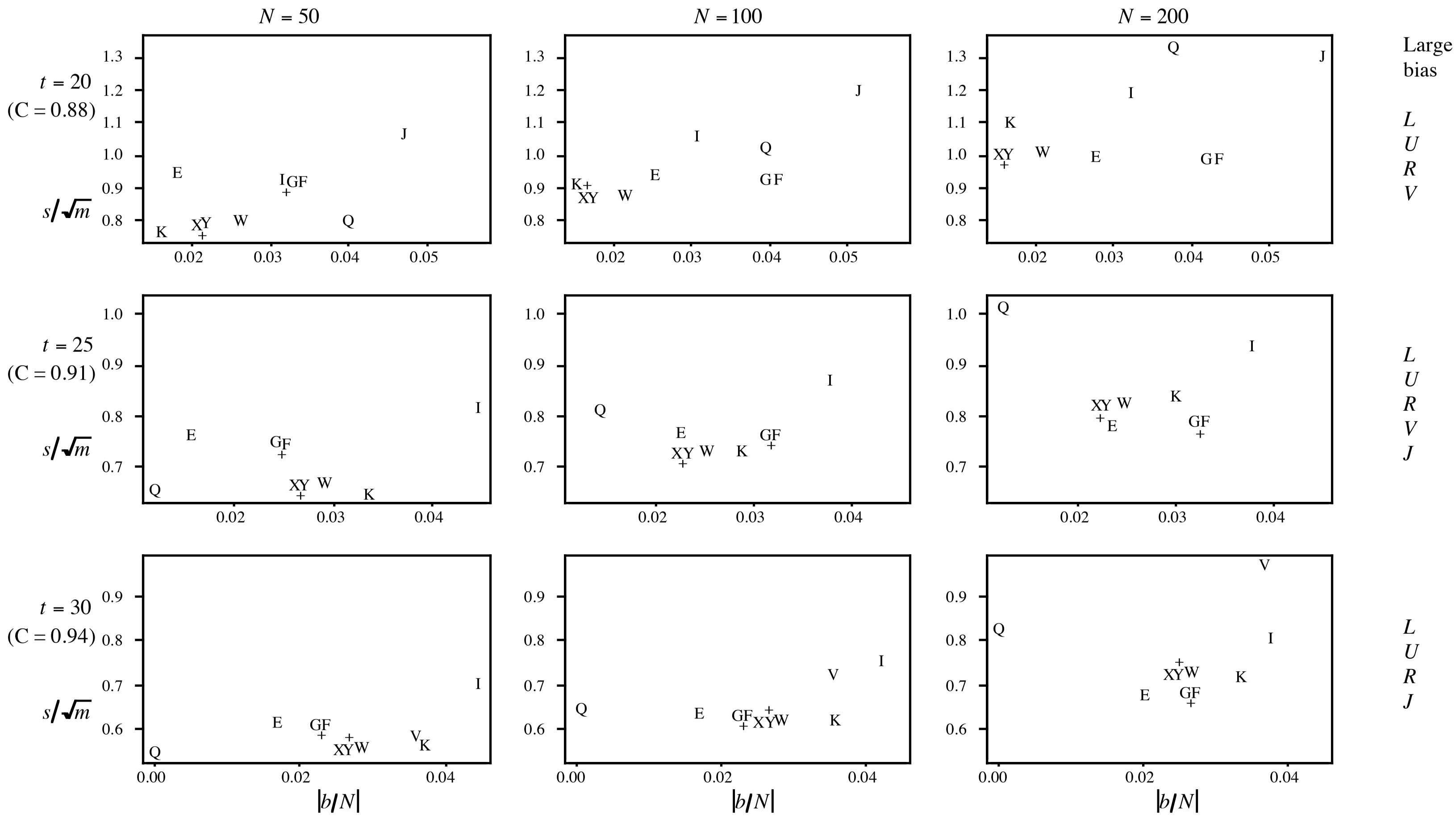

FIG.2. (cont.) Plots of $s/\sqrt{m}$ against absolute relative bias $|b/N|$, when $t = 20$, 25 and 30, for the estimators that perform well over the rectangle Ω. The positions of estimators adjacent to + signs have been slightly perturbed for greater clarity.

Otto's U when $t \geq 15$. When $t = 5$, the estimators K and I are omitted due to large negative bias and the estimator E due to large positive bias.

Recalling that the plots are based on independent simulations, it is noteworthy that the relative positions of some of the estimators remain the same over many of the plots. When $t \geq 10$, E is always the least biased of the estimators of Chao *et al*. (1992), and F the most biased of these three. For $t = 5$, the coverage-adjusted estimator Y is less biased than X. For larger values of t, the performance of these two estimators is very similar, but, for $t \geq 15$, X is usually marginally less biased than Y. The jackknife estimator J is less biased than the interpolated jackknife I in all the plots for $t = 5$ or 10, but the reverse is always true for the larger values of t. In all the plots for $t = 10$, the second version H of the interpolated jackknife is less biased than I. This is particularly so for small N, but the absolute bias of H is only slightly smaller than that of J, which has a smaller standard deviation. The relative performance of the first order jackknife K improves steadily as t rises to 20, but tails off again if t is increased further.

9. RECOMMENDATIONS ON POINT ESTIMATORS

The main motivation for FIG. 2, however, was to determine which of the estimators for model M_h to recommend to the practitioner seeking a robust performance across the rectangle Ω. It is clear from FIG. 2 that no one estimator is uniformly best for all sample sizes t. Nevertheless, if a single estimator had to be chosen for all values of $t \leq 20$, the coverage-adjusted estimator Y appears to be the best choice. Each of the estimators E, I, J, K, L, Q, R, U and V suffers from a large absolute relative bias for at least one value of t in this range, whilst F and G usually have lower precision. FIG. 2 has shown that the estimator X may be marginally preferable to Y in terms of bias for $t = 15$ and 20, but X is less good for small sample sizes. The performance of W, the third of the coverage-adjusted estimators, only matches that of X and Y in the case where $t = 10$.

For $t = 5$, Darroch and Ratcliff's R or Chao *et al.*'s G are viable alternative recommendations, particularly if N is thought likely to be either large or small respectively. Taken as a whole, FIG. 2 offers little endorsement for use of the jackknives. In the case $t = 10$, however, there is some evidence to support the jackknife *J*, particularly if *N* is thought to be small, whilst the first-order jackknife *K* performs strongly when $t = 20$.

For the larger sample sizes, $t = 25$ and 30, the conclusions are rather different. Although the coverage-adjusted estimators *X* and *Y* remain amongst the best few available estimators, our results commend Overton's *Q* or, particularly for $t = 25$, Chao *et al*.'s *E*.

The recommendations given in this section should be regarded as guidelines rather than as an attempt to be prescriptive. Different practitioners will inevitably wish to assign different relative weights to the bias and the standard deviation of estimators, and their preference patterns will differ accordingly. The reader is therefore encouraged to formulate his or her own summary of FIG. 2.

ACKNOWLEDGEMENT

The research work of Jonathan Ashbridge has been supported by the Engineering and Physical Sciences Research Council.

BIBLIOGRAPHY

Buckland, S.T. (1984). "Monte Carlo confidence intervals," *Biometrics* **40**, 811-817.

Burnham, K.P. and Overton, W.S. (1978). "Estimation of the size of a closed population when capture probabilities vary among animals," *Biometrika* **65**, 625-633.

Burnham, K.P. and Overton, W.S. (1979). "Robust estimation of population size when capture probabilities vary among animals," *Ecology* **60**, 927-936.

Carothers, A.D. (1973). "Capture-recapture methods applied to a population with known parameters," *Journal of Animal Ecology* **42**, 125-146.

Chao, A. (1989). "Estimating population size for sparse data in capture-recapture experiments," *Biometrics* **45**, 427-438.

Chao, A., Lee, S.-M. and Jeng, S.-L. (1992). "Estimating population size for capture-recapture data when capture probabilities vary by time and individual animal," *Biometrics* **48**, 201-216.

Cormack, R.M. (1966). "A test for equal catchability," *Biometrics* **22**, 330-342.

Darroch, J.N. and Ratcliff, D. (1980). "A note on capture-recapture estimation," *Biometrics* **36**, 149-153.

Efron, B. (1982). *The jackknife, the bootstrap and other resampling plans*. Philadelphia: Society for Industrial and Applied Mathematics.

Good, I.J. (1953). "The population frequencies of species and the estimation of population parameters," *Biometrika* **40**, 237-264.

Goudie, I.B.J., Pollock, K.H. and Ashbridge, J. (1998). "A plant-capture approach for population size estimation in continuous time," *Commun. Statist. - Theor. Meth.* **27**, 433-451.

Huggins, R.M. (1989). "On the statistical analysis of capture experiments," *Biometrika* **76**, 133-140.

Lee, S.-M. and Chao, A. (1994). "Estimating population size via sample coverage for closed capture-recapture models," *Biometrics* **50**, 88-97.

Mares, M.A., Streilein, K.E. and Willig, M.R. (1981). "Experimental assessment of several population estimation techniques on an introduced population of eastern chipmunks" *Journal of Mammalogy* **62**, 315-328.

Menkens, G.E., Jr. and Anderson, S.H. (1988). "Estimation of small-mammal population size," *Ecology* **69**, 1952-1959.

Norris, J.L. and Pollock, K.H. (1996). "Nonparametric MLE under two closed capture-recapture models with heterogeneity," *Biometrics* **52**, 639-649.

Otis, D.L., Burnham, K.P., White, G.C. and Anderson, D.R. (1978). "Statistical inference from capture data on closed animal populations," *Wildlife Monographs* **62**, 1-135.

Overton, W.S. (1969). "Estimating the numbers of animals in wildlife populations," *Wildlife Management Techniques*, 3rd edition, R.H. Giles, Jr. (ed), 403-455. Washington, D.C. : The Wildlife Society.

Pollock, K.H., Nichols, J.D., Brownie, C. and Hines, J.E. (1990). "Statistical inference for capture-recapture experiments," *Wildlife Monographs* **107**, 1-97.