The Library of the Department of Statistics
North Carolina State University

Modeling visibility bias using Mark-Recapture and Line Transect sampling methodology: An application to Wildlife Bioassessment of species with prominent nesting structures

Russell Alpízar-Jara, Kenneth H. Pollock, and Dawn E. Haines
Biomathematics Graduate Program
Department of Statistics
North Carolina State University
Box 8203, Raleigh, NC 27695-8203, U.S.A

Institute of Statistics
Mimeograph Series No. 2509
North Carolina State University
Raleigh, North Carolina

June, 1998

Abstract

The combined mark-recapture and line transect sampling methodology proposed by Alpizar-Jara and Pollock (1997) is used to illustrate estimation of populations with prominent nesting structures (e.g. bald eagles). In the context of a bald eagle population, the number of nests in a list frame corresponds to a pre-marked sample of nests, and an area frame corresponds to a set of transect strips that could be regularly monitored. An advantage of this method is that by using line transect sampling we allow for visibility bias in the area frame. Unlike for the previous method based on dual frame methodology (Haines and Pollock 1997), we no longer need to assume that the area frame is complete (i.e. we do not need to see all nests in the sampling sites; e.g. transect strips). Also, by combining information from both, list and area frames, we obtain more efficient estimators than those obtained by using only one data frame-type. In addition, we may use line transect methods to allow us to model the drop in detectability as a function of distance and other covariates. We derive an estimator for visibility bias, and generalize the screening estimator to allow for visibility bias. A simulation study is carried out to compare the performance of the Lincoln-Petersen type estimators to the screening estimator.

KEY WORDS: Wildlife density estimation, multiple frame sampling, screening estimator, Lincoln-Petersen, line transect, visibility bias.

1 Introduction

We continue our current line of research to improve traditional sampling designs based on dual frame sampling methodology. Traditional sampling designs do not use existing information that could greatly enhance the estimation of population parameters. Haines and Pollock (1997) propose a methodology that combines incomplete list frames with an area frame which is assumed to be complete. Their work has been based on original multiple frame sampling schemes (list plus area sampling) which have been widely used in agricultural surveys (Hartley 1962, 1974, Kott and Vogel 1995). Haines and Pollock (1997) propose using the Hartley's screening estimator to combine sample information from independent list and area frame sampling schemes to estimate the total number of active eagle nests and the number of successful nests in a predefined region. They suggest the method applies to any species with prominent nesting structures. For a good discussion of the advantages and disadvantages of both list and area frames we refer to Nealon (1984). For a good discussion on design considerations when monitoring bald eagles and their habitats see Bartish (1994).

The dual frame sampling scheme performs well when serious problems of visibility bias are not present collecting data for the area sample. In this paper we consider extensions of the dual frame sampling scheme which allow for modeling the visibility bias in the area sample. The Lincoln-Petersen and the capture-recapture line transect combination methodology (Alpizar and Pollock 1997) can be used when considering the nests already located on the list as the sample of marked nests, and the area sample (i.e. line transect survey) as the recapture sample with both nests previously on the list (marked) and nests not on the list (unmarked). Here we evaluate the performance of these estimators by relaxing the assumption of visibility bias at several levels and by modeling the key parameter, the probability of detecting a nest in the area sample, based on different model assumptions. We will also show how the screening estimator proposed by Haines and Pollock (1997) can be viewed as a special case of the more general approach proposed in this paper.

In section 2, we briefly introduce notation and description of the sampling schemes. Issues related to visibility bias are discussed in section 3. Capture-recapture methods as dual frame estimators are presented in section 4. The Screening estimator is presented in section 5. A generalized screening estimator is presented in section 6. An illustration of our estimation methods is given in section 7. We include some simulation results in section 8. Finally, we conclude with a discussion and future research directions in section 9.

2 Notation

Following the capture-recapture terminology as defined by Alpizar and Pollock (1997) we now define some quantities of interest.

Known fixed quantities

n_1 is the number nests on the list frame.

f_L is a fraction of nests from the list frame to be sampled.

$n_1^* = n_1 f_L$ is the number of nests from list frame to be sampled.

f_A is a fraction of sampling sites (quadrants, circular plots, strip transects, etc.) from the area frame to be sampled.

Statistics and random variables

n_{10} is the number nests on the list frame but not found in the area sample.

n_{01a} is the number nests found in the area sample but not on the list frame.

n_{11a} is the number nests on the list frame also found in the area sample.

n_{1a} is the number of active nests found on the list frame sample.

$n_{2a} = n_{01a} + n_{11a}$ is the number of active nests found in the area sample (all nests found in the area sample are assumed to be active nests).

Parameters of interest

N is the total number of nests in the geographic area of interest.

N_a is the total number of active nests in the geographic area of interest.

n_a is the number of active nests on the list frame.

p_1 is the probability of a nest being on the list frame.

p_{1a} is the probability of an active nest being on the list frame.

$p_{2a} = f_A \beta(\underline{\theta})$ is the probability of a nest being in the area sample.

$p = p_1 + p_{1a}p_{2a} - p_1 p_{1a} p_{2a}$ is the probability of detecting a nest either on the list frame or in the area sample.

$\beta(\underline{\theta})$ is the visibility bias parameter.

$\underline{\theta}$ is a vector parameter describing the detection function. The detection function could be a function of several variables such as distance, number of chicks in the nest, etc.

2.1 An independent estimate of the number of active nests

To estimate the population of active nests N_a we assume that a list frame of eagle nests is available from a previous year. At present, the actual number of active nests on the list is unknown (n_a), and it needs to be estimated. Let n_1 be the number of nests in our list frame. Let p_{1a} be the proportion of active nests on the list. An estimate of p_{1a} can be obtained by randomly drawing a sample of size n_1^* from the list frame (this sample represents a known fraction of nests from the list to be inspected, f_L). Assume that n_{1a} out of n_1^* nests were found to be active. An estimator of p_{1a} is then given by $\hat{p}_{1a} = \frac{n_{1a}}{n_1^*}$, and an estimator of the number of active nests on the list frame is given by $\hat{n}_a = n_1 \hat{p}_{1a}$. Note that $n_{1a}|n_1^* \sim binomial(n_1^*, p_{1a})$, and $\hat{n}_a|n_1 \sim binomial(n_1, p_{1a})$.

2.2 Estimating number of active nests from the area frame

An area sample from the total nest population can be taken by randomly placing sample sites (plots, strip transects, etc) in the geographic area of interest. A known fraction, (f_A), of sampling sites is to be drawn. A sample of size n_{2a} active nests is collected. We denote by n_{11a} the number of active nests that were already on the last year list frame. The area sample could also be used to update last year's list frame. Some new nests not on the last year's list frame will be found, i.e. $n_{01\bar{a}}$. Also some nests on last year's list frame may be destroyed and need to be taken out of this year's list.

In this paper we slightly modified the original combined approach (Alpizar and Pollock 1997) by assuming that an independent sample from the list is taken to estimate the number of active nests. Eventually we would like to focus our attention to active nests only, or nests with some characteristic of interest such as number of successful nests as an example. There are at least three possible estimators for the active nest population, the combined capture-recapture line transect estimator, the Lincoln-Petersen estimator , and the screening estimator. The advantages and disadvantages of these methods will be evaluated in this paper.

3 Modeling visibility bias

Pollock and Kendall (1987) review several estimation procedures to account for visibility bias in aerial surveys. They recommend to utilized the Chapman's (1951) modification of the Lincoln-Petersen estimator under several circumstances. For instance, when complete ground counts cannot be assured, the completeness assumption can be relaxed if some nests from the area sample can be identified as nests previously on the list. More recently a series of papers have emerged in the literature about combination of capture-recapture and line transect methods to correct for visibility bias in aerial surveys (Laake *et al.*, 1997; Borchers *et al.*, 1996; Manly *et al.*, 1996; Alpizar and Pollock, 1996; to mention some references).

The dual frame sampling scheme is excellent if any area sampling done does not suffer from visibility bias. With very prominent nesting structures visibility bias may not be a serious problem, however, there may be some situations where nests are missed in the area sample. The key aspect under our framework is how we model the parameter p_{2a}, probability of a nest being included (detected) in the area sample. p_{2a} is a function of the fraction of sampling sites included in the area sample (f_A), and the visibility bias parameter (β). Several assumptions can be made. If we assume that detectability is a function of covariates such as distance, then the visibility parameter (β) is a function of the parameter set describing the detectability function ($\underline{\theta}$), as in the case of line transect methodology (Buckland *et al.*, 1993). In this case we must use the full likelihood function (1) as suggested in section (4.1.1) of this paper. When the use of distance information is impractical, we could use the modified Lincoln-Petersen estimator and assume that $\beta(\underline{\theta}) = \beta$ is constant. We could also use the Lincoln-Petersen method if plots are searched rather than using line transects. Sometimes a combined model using grouped distances categories such as the stratified Lincoln-Petersen and line transect model suggested by Alpizar and Pollock (1996) can be used as an alternative approach. They propose this model to be used in a context of multiple observers for aerial surveys where grouped distance information could be collected . If $\beta < 1$, which is usually the case, even if the entire sampling plots are searched, then we use the Lincoln-Petersen model. In this paper we compare the performance of this model ($\beta < 1$) vs the model in which ($\beta = 1$) which corresponds to the screening estimator suggested by Haines and Pollock (1997).

4 Capture-recapture methods

4.1 Combined model as a dual frame estimator

The most general model combines capture-recapture and line transect data to obtain population estimates. If we have a list frame, during the line transect monitoring study an observer follows the center line measuring perpendicular distances for both, nests already on the list and newly found nests. Some nests already on the list might not be seen and new nests will appear. Alpizar and Pollock (1997) proposed a population estimator for a mark-resight study where in the resighting stage perpendicular distance data is recorded. Under their approach one could analyze the collected data as a capture-recapture study, as a line transect study or develop a combined model to allow estimation of the probability of seeing an active nest (or an object) on the transect line (i.e. $g(0)$) and to test whether or not that probability is one ($g(0) = 1$). Their simulation study shows a significant gain in precision of population estimates when using this combined approach. Alpizar and Pollock (1997) focus on comparing the combined estimator to the Lincoln-Petersen and the line transect estimators. We refer to their paper for these comparisons. In this paper we shall focus on the comparisons between the Lincoln-Petersen and the Screening estimator.

4.1.1 Combined approach: full likelihood function, $p_{2a} = f_A\beta(\underline{\theta})$

$$L(\cdot) = L_1 \cdot L_2 \cdot L_3 = f_1(n_{1a}|n_1^*; p_{1a}) \cdot f_2(n_{10}, n_{01a}, n_{11a}|N; p_1, p_{1a}, p_{2a}) \cdot f_3(x_1, \ldots, x_{n_{2a}}|n_{2a}; \underline{\theta}) \tag{1}$$

with

$$L_1(\cdot) = f_1(n_{1a}|n_1^*; p_{1a}) = \binom{n_1^*}{n_{1a}} [p_{1a}]^{n_{1a}} [(1 - p_{1a})]^{n_1^* - n_{1a}} \tag{2}$$

$$L_2(\cdot) = f_2(n_{10}, n_{01a}, n_{11a}|N; p_{1a}, p_{2a}) = \binom{N}{n_{10}, n_{01a}, n_{11a}} [p_1(1 - p_{1a}p_{2a})]^{n_{10}} [(1 - p_1)p_{1a}p_{2a}]^{n_{01a}} \times [p_1 p_{1a} p_{2a}]^{n_{11a}} [(1 - p_1)(1 - p_{1a}p_{2a})]^{N - n_{10} - n_{01a} - n_{11a}} \tag{3}$$

$$L_3(\cdot) = f_3(x_1, \ldots, x_{n_{2a}}|n_{2a}; \underline{\theta}) = \prod_{i=1}^{n_{2a}} f(x_i|\underline{\theta}) = \prod_{i=1}^{n_{2a}} \frac{g(x_i)}{\int_0^w g(x)dx} \tag{4}$$

where

$$p_{2a} = f_A\beta(\underline{\theta}) = \frac{\int_0^w g(x|\underline{\theta})dx}{w}, \quad hence, \quad \beta(\underline{\theta}) = (f_A w)^{-1} \int_0^w g(x|\underline{\theta})dx \tag{5}$$

There are two sampling situations where a Lincoln-Petersen type estimator arises. Alpizar and Pollock (1997) have shown that under the assumption that detectability on the transect line is less than one (i.e. $g(0) < 1$), the combined estimator reduces to a Lincoln-Petersen type estimator which takes into account distance information. In this situation, the sampling plot (i.e. strip transect) does not need to be searched completely and visibility bias, $\beta(\underline{\theta})$, is modeled by assuming that sighting probability decreases as a function of the perpendicular distances from the center line. The other situation occurs when a complete search of the sampling sites (i.e. quadrants, circular plots, strip transects) is required. Under some circumstances, collecting distance information might be difficult an the entire sampling plot need to be searched. A Lincoln-Petersen type estimator can also be used in this setting, but here we assume that the visibility parameter is a constant function of distance, and hence $\beta(\underline{\theta}) = \beta$. In this paper we compare the performance of the Lincoln-Petersen and the Screening estimators.

4.2 The Lincoln-Petersen estimator

The Lincoln-Petersen method is the simplest of the capture-recapture models. Next we briefly discuss the main assumptions of this model.

(1) The nest population is constant in size. This assumption may be violated since nests from the list are becoming inactive. This is some sort of nest mortality, and then the Lincoln-Petersen estimator is an estimator of the number of nests in the population at the time of the first sample (when list frame was last updated, hopefully not to long ago from when the area sample is taken).

(2) Nests are equally likely to be seen within each sampling occasion (i.e. every nest on the list have the same probability of been sampled, and every nest in the area frame have the same probability of been sighted).

(3) There is perfect matching between the nests that are on the list and nests found during the area sampling. This assumption is violated if is impossible to identify which nests found during the area sampling were already on the list.

(4) The sightings of different nests within and between sampling occasions occur independently.

The Lincoln-Petersen estimator is given by

$$\hat{N}_{LP} = \frac{n_1 n_{2a}}{n_{11a}}. \tag{6}$$

Note that (6) is undefined when $n_{11a} = 0$, hence the moments of this estimator do no exist. However, when the probability of $n_{11a} = 0$ is negligible, as it is often the case in many experimental situations, this may not be a strong limitation. If probability of $n_{11a} = 0$ is zero, the approximate variance of this estimator is given by

$$\widehat{var}(\hat{N}_{LP}) = \frac{n_1 n_{2a}(n_1 - n_{11a})(n_{2a} - n_{11a})}{n_{11a}^3}. \tag{7}$$

Chapman (1951) shows that (6) tends to be positively biased for a large range of values of n_1 and n_{2a}. He proposes a modified version which is less biased, and which always has finite moments. The Chapman estimator is given by

$$\hat{N}_{CH} = \frac{(n_1+1)(n_{2a}+1)}{(n_{11a}+1)} - 1 \tag{8}$$

with approximate variance

$$\widehat{var}(\hat{N}_{CH}) = \frac{(n_1+1)(n_{2a}+1)(n_1-n_{11a})(n_{2a}-n_{11a})}{(n_{11a}+1)^2(n_{11a}+2)}. \tag{9}$$

4.2.1 Number of active nests

The number of active nests under the Lincoln-Petersen model is given by

$$\hat{N}_{LP_a} = \hat{N}_{LP}\hat{p}_{1a} = \frac{n_1 n_{2a}}{n_{11a}}\frac{n_{1a}}{n_1^*} = \frac{n_{1a}n_{2a}}{n_{11a}}\frac{1}{f_L} \tag{10}$$

with approximate variance

$$\widehat{var}(\hat{N}_{LP_a}) = \frac{1}{f_L^2}\frac{n_{1a}n_{2a}(n_{1a}-n_{11a})(n_{2a}-n_{11a})}{n_{11a}^3} \tag{11}$$

where f_L is the known fraction of nests from the list to be sampled. In the case of the Chapman estimator, the number of active nest can be estimated by

$$\hat{N}_{CH_a} = \frac{(n_1\hat{p}_{1a}+1)(n_{2a}+1)}{(n_{11a}+1)} - 1$$

$$\hat{N}_{CH_a} = \frac{(\frac{n_{1a}}{f_L}+1)(n_{2a}+1)}{(n_{11a}+1)} - 1 \tag{12}$$

with approximate variance

$$\widehat{var}(\hat{N}_{CH_a}) = \frac{(\frac{n_{1a}}{f_L}+1)(n_{2a}+1)(\frac{n_{1a}}{f_L}-n_{11a})(n_{2a}-n_{11a})}{(n_{11a}+1)^2(n_{11a}+2)}. \tag{13}$$

Variances (7), (9), (11), and (13) are merely approximations, and can be derived using conditional arguments on n_1, n_{1a}, and n_{2a} (Chapman 1951, Seber 1970). A Chapman estimator of the number of active nests could also be obtained simply by multiplying equation (8) by $\hat{p}_{1a}$.

4.3 Estimation and properties of the visibility bias parameter

From the full likelihood approach it can be shown that the MLE of the product

$$p_{1a}p_{2a} = p_{1a}f_A\beta$$

is given by

$$\hat{p}_{1a}f_A\hat{\beta} = \frac{n_{11a}}{n_1},$$

hence β can be estimated by

$$\hat{\beta} = \frac{n_{11a}}{n_{1a}}\frac{f_L}{f_A} \tag{14}$$

and its properties can be studied since it is a ratio of two independent binomial random variables, $n_{11a} \sim binomial(N, p_1p_{1a}f_A\beta)$ and $n_{1a} \sim binomial(N, p_1p_{1a}f_L)$.

Using a conditional argument it can be shown that $E(\hat{\beta}) \approx \beta$, and an approximate variance for (14) is given by

$$\widehat{var}(\hat{\beta}) = \hat{\beta}^2\left(\frac{1}{n_{11a}} + \frac{1}{n_{1a}} - \frac{2}{\hat{N}_{LP}}\right) \tag{15}$$

Later we shall show the relationship between the estimators of population size and the visibility bias parameter. It can be shown that if $p_{1a} < 1$ and $\beta < 1$ then the number of active nests using the Lincoln-Petersen and the screening estimators are equivalent provided the visibility parameter is included in the likelihood of the screening estimators. This result becomes obvious as we have shown that the two estimators are derived from the same general likelihood approach.

5 The screening estimator of population size, $p_{1a} = 1$, $\beta = 1$, $p_{2a} = f_A$

Haines and Pollock (1997) estimate the total number of active nests in the population using the sample means of active nests in the nonoverlaping domain of the area frame and number of active nests on the list frame. Haines (1997) shows that the screening estimator naturally arises as the maximum likelihood estimator of a capture-recapture model that assumes that the list frame size

(n_1) and the proportion of sample sites to be surveyed (f_A) are known. Under her approach and our notation the screening estimator is given by

$$\hat{N}_S = n_1 + \frac{n_{01a}}{f_A}, \quad where \quad p_{2a} = f_A \tag{16}$$

with approximate variance

$$\widehat{var}(\hat{N}_S) = \frac{n_{01a}(1 - f_A)}{f_A^2}. \tag{17}$$

n_{01a} is what Haines and Pollock (1997) named as the number of nests in the nonoverlaping domain.

5.1 Number of active nests

Now we derive an estimator of the number of active nests using our likelihood approach (1) and assuming that $p_{1a} < 1$. The number of active nests for the screening estimator is given by

$$\hat{N}_{S_a} = n_1\hat{p}_{1a} + \frac{n_{01a}}{f_A} = \frac{n_{1a}}{f_L} + \frac{n_{01a}}{f_A} \tag{18}$$

with approximate variance

$$\widehat{var}(\hat{N}_{S_a}) = \frac{n_{1a}(1 - \hat{p}_1\hat{p}_{1a}f_L)}{f_L^2} + \frac{n_{01a}(1 - \hat{p}_{1a}f_A)}{f_A^2}. \tag{19}$$

The main advantage of the screening estimator is that if the list and area frame are complete (no visibility bias) then it will be more efficient that the Lincoln-Petersen estimator. However, as we shall show in our simulation study, the screening estimator can be severely negatively biased due to the incompleteness of the area sample.

6 A generalized screening estimator, $p_{1a} < 1, \beta < 1, \quad p_{2a} = \beta f_A$

A generalized screening estimator can be obtained when incorporating the bias parameter (14) in equations (16) and (18). Also, it can be shown that this generalized screening estimator is equivalent to the Lincoln-Petersen estimators of equations (6) and (10) respectively. For the total population size the estimator is given by

$$\hat{N}_{GS} = n_1 + \frac{n_{01a}}{\hat{p}_{1a}\hat{\beta} f_A}, \tag{20}$$

and for the number of active nest the estimator is given by

$$\hat{N}_{GS_a} = n_1\hat{p}_{1a} + \frac{n_{01a}}{\hat{\beta} f_A} = \frac{n_{1a}}{f_L} + \frac{n_{01a}}{\hat{\beta} f_A}. \tag{21}$$

7 Illustrative example

We consider a hypothetical situation in which an eagle nests population is to be estimated using the methods presented in the previous sections. Following an example with similar numbers to those presented by Haines and Pollock (1997) but ignoring stratification we have

$n_1 = 700$ nests on the list frame.

$f_L = 0.8$ is the fraction of nests from the list frame to be sampled.

$n_1^* = 700(0.8) = 560$ sample size of nests from list frame.

$f_A = 0.05$ is a fraction of sampling sites (quadrants, plots, strip transects, etc.) from the area frame to be sampled.

$n_{1a} = 507$ active nests are found on the list frame sample.

$n_{01a} = 16$ nests are found in the area sample but not on the list frame.

$n_{11a} = 21$ nests are found on the list frame and in the area sample.

$n_{2a} = n_{01a} + n_{11a} = 16 + 21 = 37$ is the number of active nests found in the area sample.

We could further assume that the area sample were strip transects of width $w = 20$ units and distances were collected. In this case the parameter $\beta(\underline{\theta})$ needs to be estimated using distance information. We refer to Alpizar and Pollock (1997) for estimation using this approach.

7.1 Estimation of parameters of interest

Using the above model equations, maximum likelihood estimators and their standard errors given in parenthesis are as follow: exists in is close form and are given by

$$\hat{N}_{LP} = \frac{(700)(37)}{21} = 1234, \quad (SE = 174.3)$$

$$\hat{N}_{CH} = \frac{(700+1)(37+1)}{(21+1)} - 1 = 1210, \quad (SE = 161.2)$$

$$\hat{N}_S = 700 + \frac{16}{0.05} = 700 + 320 = 1020, \quad (SE = 78.0).$$

Now we estimate the number of active nests and corresponding standard errors,

$$\hat{N}_{LP_a} = \hat{N}_{LP}\hat{p}_{1a} = (1234)\frac{507}{560} = 1118, \quad (SE = 156.9)$$

$$\hat{N}_{CH_a} = \hat{N}_{CH}\hat{p}_{1a} = (1210)\frac{507}{560} = 1096, \quad (SE = 145.7)$$

$$\hat{N}_{S_a} = (700)\frac{507}{560} + \frac{16}{0.05} = 634 + 320 = 954, \quad (SE = 81.1).$$

The estimate of the visibility bias parameter is

$$\hat{\beta} = \left(\frac{21}{507}\right)\left(\frac{0.80}{0.05}\right) = 0.663, \quad (SE = 0.145).$$

8 Simulation

A simulation study was carried out considering combinations of the following parametric scenarios:

$N = \{500, 5000\}$, $f_L = \{0.7, 1.0\}$, $p_1 = \{0.7, 0.9\}$,

$p_{1a} = \{0.7, 1.0\}$, $f_A = \{0.05, 0.2\}$, $\beta = \{0.7, 0.8, 0.9, 1.0\}$

A total of $2 \times 2 \times 2 \times 2 \times 2 \times 4 = 128$ possible combinations. We ran 1000 Monte Carlo simulation repetitions, and for each case computed the root mean square error (RMSE), the relative bias (%BIAS), and the standard error (SE) of three estimators: The Screening estimator ($\hat{N}_S$), the Chapman estimator ($\hat{N}_{CH}$), and the Lincoln-Petersen estimator ($\hat{N}_{LP}$). We have done these analyses considering both, the estimators of the number of active nests in the population (10, 12, and 18) and the estimators of the total population size. We report only on results from the estimates of the number of active nests since this is the most important parameter.

The above scenarios allow the study of the following special cases:

Searching all nests from the list, $f_L = 1.0$ If it is possible to search all nests from the list to examine which of them are still active then $n_1^* = n_1 f_L = n_1$. In our simulations we also allow for sampling only a fraction of the nests from the lists ($f_L = 0.7$). We assume that nests from the list are relatively easy and inexpensive to verify whether they are active or not, and often a large proportion of them can be sampled.

All nests on the list are active, $p_{1a} = 1.0$ Under this assumption $L_1(\cdot) = 1$, and our model reduces to the simple Lincoln-Petersen model. We emphasize that we slightly modified the original combined approach given by Alpizar and Pollock (1997). We introduce the likelihood $L_1(\cdot)$ in order to estimate the proportion of active nests on the list. The underlying assumption is that the list frame is out of date and not all nests on the lists are active. The same model can be used if we are interested in any other characteristic of the nest population (i.e. proportion of successful nests). In our simulations we also allow for this proportion to be as low as 70%, assuming a rapidly changing population of nests with 30% inactive nests.

Note also that we allow for two levels of population size (high and low abundance, $N = \{500, 5000\}$), and two levels of probability of nests on the list frame, $f_L = \{0.7, 1.0\}$. The probability of a nest being on the list is usually high. We assume that lists are kept fairly updated, and that nests on the lists come from inventories of all the nests that have been reported over several years. Thus, nests on list represents a large proportion of the nests in the total population. On the contrary, we assume two low levels for the fraction of sampling plots to be drawn, $f_A = \{0.05, 0.2\}$. Usually it is very expensive or impractical to sample from an area frame. Several levels of visibility bias are considered $\beta = \{0.7, 0.8, 0.9, 1.0\}$, including the case in which there is no visibility bias ($\beta = 1$).

8.1 Simulation results

A series of figures 1, 2, 3 (a,b) summarize the main simulation results. Figures 1, 2 and 3 show plots of the $RMSE$, $\%BIAS$ and SE respectively. These quantities are plotted against the different levels of the visibility bias parameter in the x axes, and for all the other combinations of parameters as indicated in the plots. Figures $\#a$ corresponds to a population of 500 nests (objects), and figures $\#b$ corresponds to a population of 5000 nests. Although we also examined the performance of the Lincoln-Petersen estimator, in our simulation results we decided to show only the Chapman estimator since in practice this estimator is used more often.

Note that in terms of RMSE, as a measure of assessing the combined effect of bias and precision of the estimators, for a population of size $N = 500$ the RMSE of the $\hat{N}_{S_a}$ is generally smaller than that of the $\hat{N}_{CH_a}$. However, when all the nests on the lists are active ($p_{1a} = 1.0$), and the proportion of sampled sites is 20% ($f_A = 0.2$), the RMSE of the $\hat{N}_{S_a}$ is not always smaller than that of the $\hat{N}_{CH_a}$. For some level of visibility bias (β_0), and usually when visibility bias is severe ($\beta < 0.85$) the $\hat{N}_{CH_a}$ performs better than the $\hat{N}_{S_a}$ (Figure 1a). We also note that the RMSE of the $\hat{N}_{CH_a}$ is always smaller than that of the $\hat{N}_{LP_a}$, but larger than that of the $\hat{N}_{S_a}$ (especially for low values of p_1 and f_A, i.e. $p_1 = 0.7$ and $f_A = 0.05$).

For a large population ($N = 5000$), it is very clear that for almost all cases there is some level of visibility bias, (β_0), at which the $\hat{N}_{CH_a}$ does better than the $\hat{N}_{S_a}$ as visibility bias increases (Figure 1b). We found not difference between the RMSE of the $\hat{N}_{LP_a}$ and $\hat{N}_{CH_a}$.

These results indicate that there is an obvious trade off between bias and precision of the estimators. Therefore, there are specific situations in which a particular estimator is more appropriate to be used. It is convenient now to examine the performance of the estimators in terms of relative bias and standard errors.

In terms of relative bias, the $\hat{N}_{S_a}$ shows severe negative bias for all scenarios except when no visibility bias is present ($\beta = 1$). The $\hat{N}_{CH_a}$ is basically unbiased (Figures 2a,b). In general, the $\hat{N}_{CH_a}$ show less bias than the $\hat{N}_{LP_a}$. For a large population the last two estimators are certainly unbiased. In terms of standard errors the $\hat{N}_{S_a}$ is always more precise than the $\hat{N}_{CH_a}$ estimator (Figures 3a,b). Note also that as visibility bias decreases the SE of $\hat{N}_{CH_a}$ also decreases; however, the SE of $\hat{N}_{S_a}$ slightly increases.

9 Discussion and future work

In this paper we have proposed a general framework to estimate population parameters under the multiple frame scheme. Our methods allow for modeling the visibility bias in the area sample, and we propose alternative approaches to deal with incompleteness of the list frame. Combining information from a list frame and an area sample we can either use the combined mark-recapture and line transect sampling design, the two sample Lincoln-Petersen model, or the Screening estimator to obtain population estimates. The screening estimator is ideal for the study of wildlife species with prominent nest sites where visibility bias may not be a serious problem.

We have shown that the Screening estimator can be viewed as a special case of a more general approach proposed in this paper. We have also shown that the Screening estimator performs better than the Lincoln-Petersen estimator if any sampling done does not suffer from visibility bias, and the list frame is complete. However, in the cases in which substantial visibility bias exists, the Screening estimator could be severely negatively biased, and often has a larger $RMSE$ than the Chapman or the Lincoln-Petersen estimators.

Testing model assumptions is possible using the likelihood ratio framework. For example we could test the null hypothesis that there is no visibility bias ($H_0 : \beta = 1$). It is also possible to tests whether the proportion of active nests on the list equals one ($H_0 : p_{1a} = 1$).

Acknowledgments

We thank Christine Bunk, Project Officer, and Tim Barish, BEST coordinator, for their strong support of this project. This work has been funded by the National Biological Service, Department of the Interior.

10 References

Alpízar-Jara and Pollock (1996) A combination line transect and capture-recapture sampling model for multiple observers in aerial surveys. *Journal of Environmental and Ecological Statistics*, 3(4), 311-327.

Alpízar-Jara and Pollock, K. H. (1997). Population estimation using a combination of line transect and capture-recapture for mark-resighting experiments. *Submitted to Biometrics.*

Bartish, T. (1994). Design considerations for trust species monitoring: Bald eagles and their habitat in the Great Lakes Region. Biomonitoring of Environmental Status and Trends (BEST) program internal report. December. National Biological Service, U.S. Department of Interior.

Borchers, D. L., S.T. Buckland, E. D. Clarke, and S.L. Cumberworth. (1996). Horvitz-Thompson estimators for double-platform line transect surveys. *Submitted to Biometrics.*

Buckland, S.T., D.R. Anderson, K.P. Burnham, and J.L. Laake. (1993) Distance sampling: Estimating abundance of biological populations. Chapman and Hall, London. 446pp.

Chapman, D.G. (1951). Some properties of the hypergeometric distribution with applications to zoological censuses. University of California. *Publ. Stat.* **1**, 131-160.

Haines, D.E. (1997) Unpublished Ph.D. Dissertation. Estimating population parameters using multiple frame and capture-recapture methodology.

Haines, D.E. and Pollock, K.H. (1997) Estimating bald eagle population size using dual frame sampling techniques. *Submitted to Journal of Environmental and Ecological Statistics.*

Hartley, H. O. (1962). Multiple Frame Surveys. Proceedings of the Social Statistics Section. *American Statistical Association.* 203-206.

Hartley, H. O. (1974). Multiple frame methodology and selected applications, *Sankhyá.* **36.**

Kott, P. S. and Vogel, F. A. (1995). Multiple-Frame Business Surveys, Surveys Methods for Business, Farms, and Institutions (B. G. Cox, ed.), John Wiley and Sons. 185-203.

Laake, J. L., Calambokidis, J., Osmek, S. D., and Rugh, D. J. (1997) Probability of detecting Harbor Porpoise from aerial surveys: Estimating $g(0)$. *Journal of Wildlife Management*, **61(1)**, 63-75.

Manly, B. F., L. L. McDonald, and G. W. Garner. (1996). The double-count method with two independent observers. *Journal of Agriculture, Biological, and Environmental Statistics*, **1(2)**, 170-189.

Nealon, J. P. (1984). Review of the multiple and area frame estimators. Staff Report 80. U.S. Department of Agriculture. Statistical Reporting Service. Washington, D.C.

Pollock, K. H. and W. L. Kendall. (1987) Visibility bias in aerial surveys: a review of estimation procedures. *Journal of Wildlife Management*, **51(2)**, 502-510.

Seber. G. A. F. (1970). The effects of trap response on tag-recapture estimates. *Biometrics*, **26**, 13-22.

Figure 1a: Root Mean Square Error vs. Visibility Bias, N=500

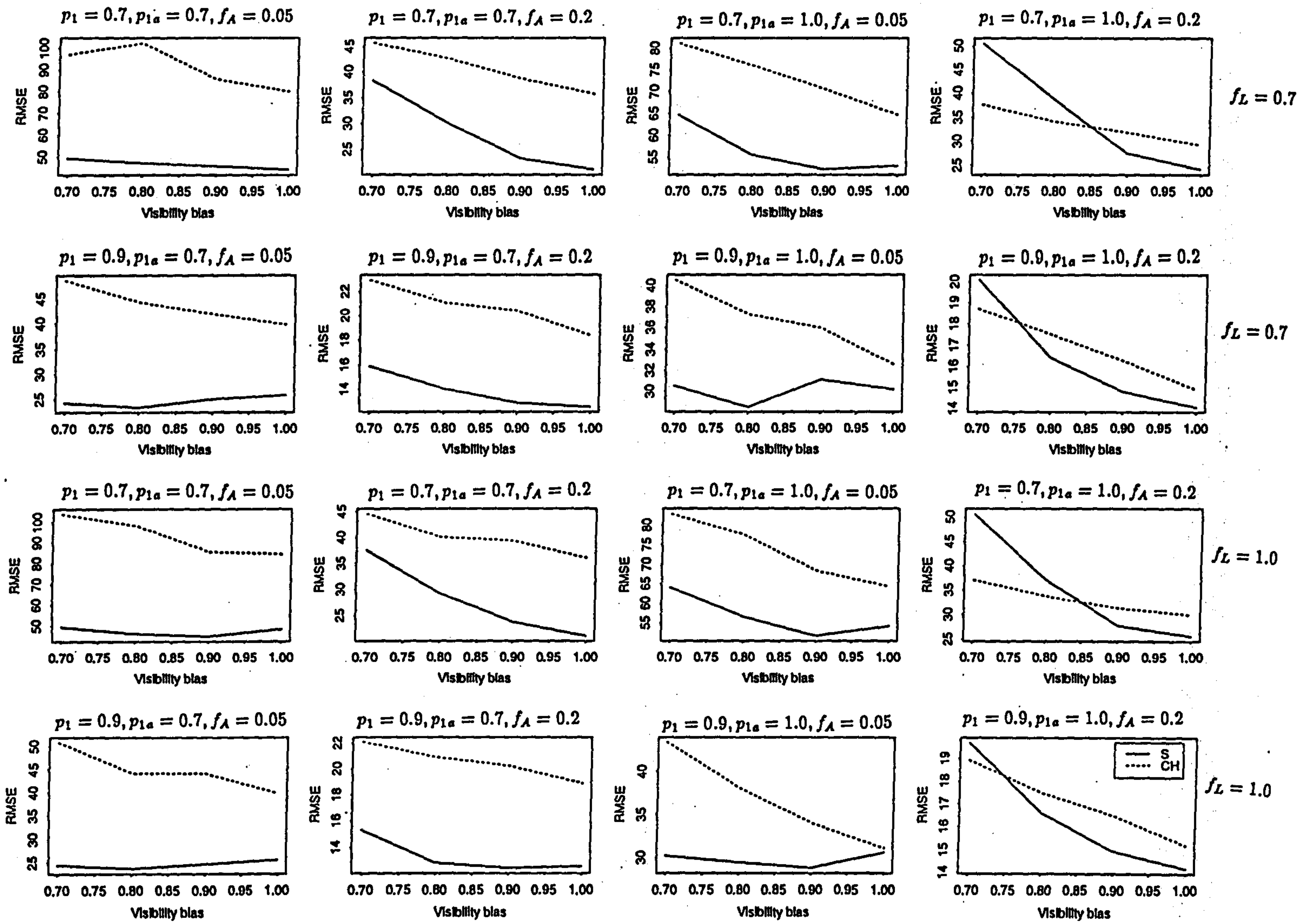

Figure 1b: Root Mean Square Error vs. Visibility Bias, N=5000

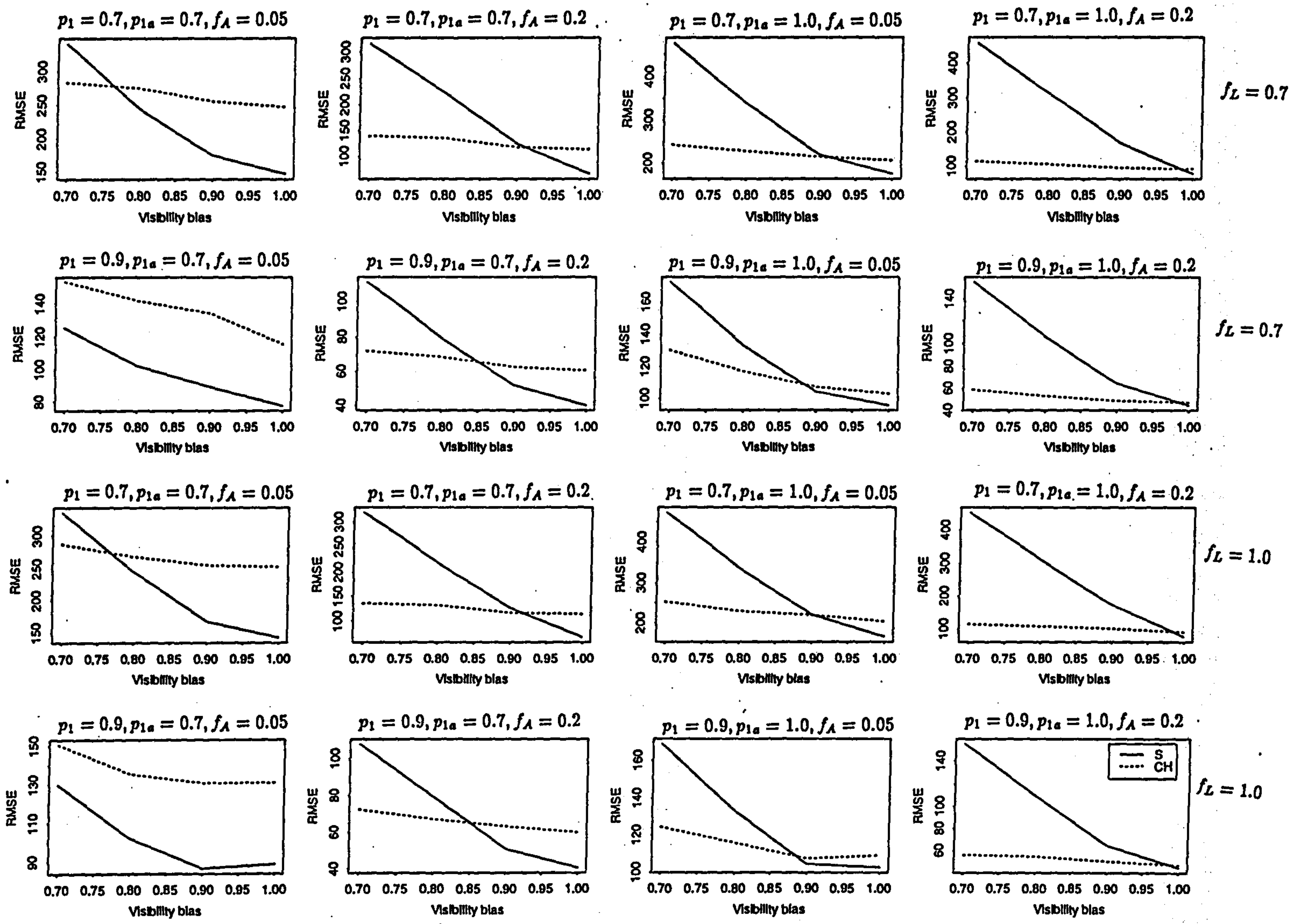

Figure 2a: Relative Bias vs. Visibility Bias, N=500

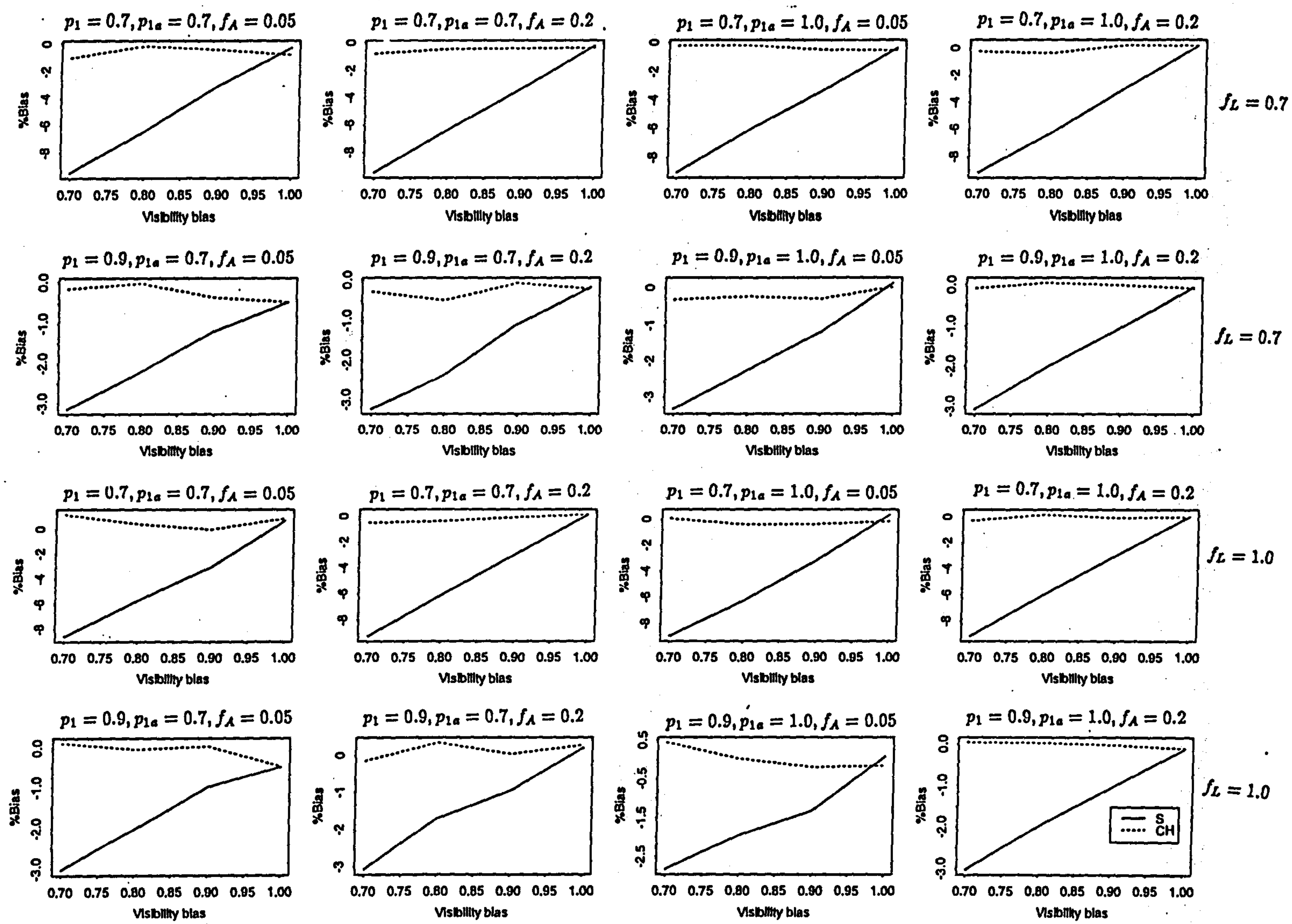

Figure 2b: Relative Bias vs. Visibility Bias, N=5000

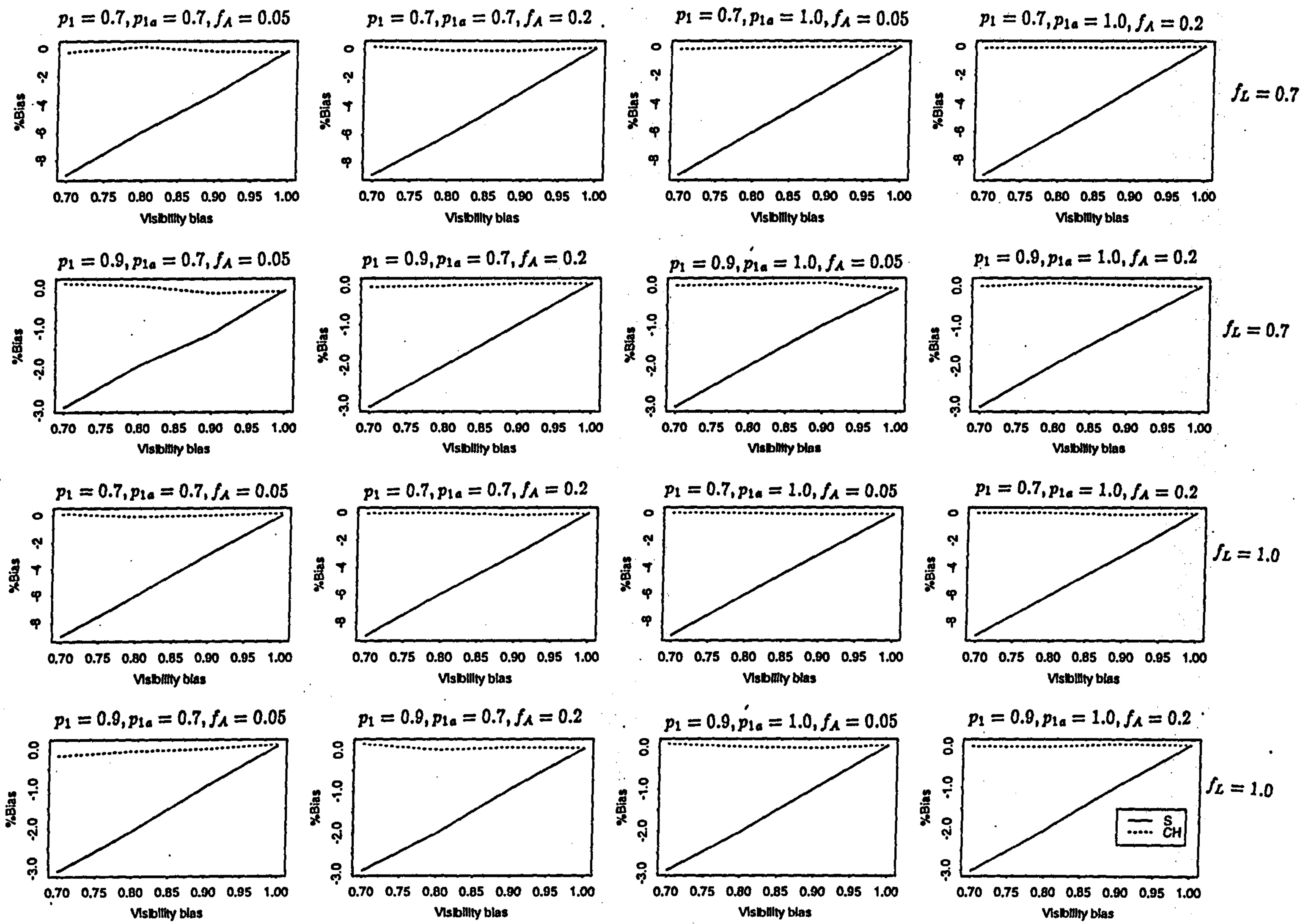

Figure 3a: Standard Error vs. Visibility Bias, N=500

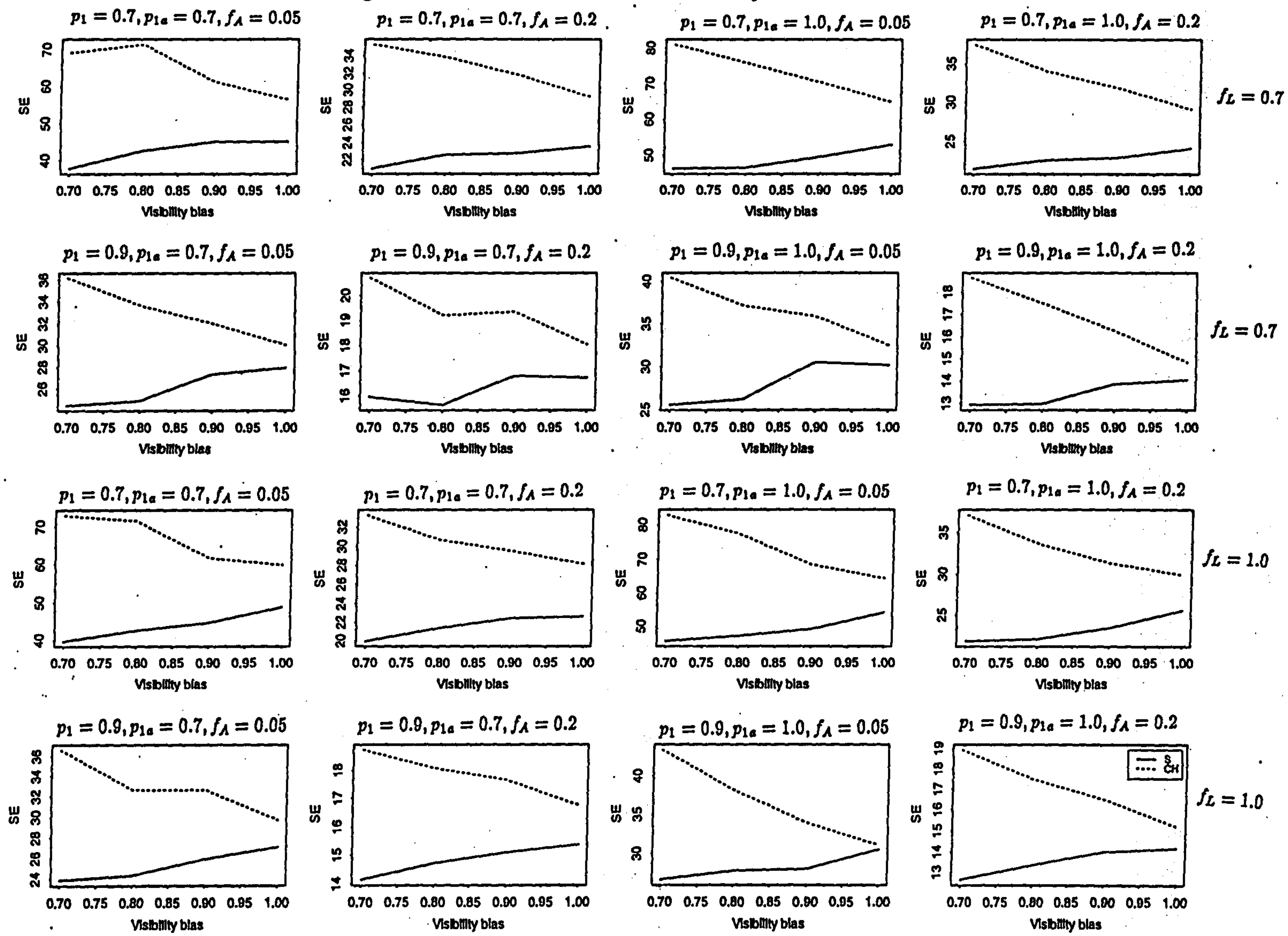

Figure 3b: Standard Error vs. Visibility Bias, N=5000