ESTIMATING POPULATION CHANGE FROM COUNT DATA: APPLICATION TO THE NORTH AMERICAN BREEDING BIRD SURVEY

WILLIAM A. LINK AND JOHN R. SAUER

Patuxent Wildlife Research Center, Laurel, Maryland 20708 USA

Abstract. For birds and many other animal taxa, surveys that collect count data form a primary source of information on population change. Because counts are only indices to population size, care must be taken in using them in analyses of population change. Temporal or geographic differences in the proportion of animals counted can be misinterpreted as differences in population size. Therefore, temporally or geographically varying factors that influence the proportion of animals counted must be incorporated as covariables in the analysis of population parameters from count data. We describe the North American Breeding Bird Survey (BBS) for illustration. The BBS is a major, landscape-level survey of birds in North America; it is typical of many count surveys, in that the same sample units (survey routes) are sampled each year, and change is modeled on these routes over time. We identify covariables related to observer ability, the omission of which can bias estimation of population change from BBS data. Controlling for observer effects or other potential sources of confounding requires the specification of models relating counts to population size. We begin with a partial model specification relating expected counts to population sizes; we describe estimators currently in use in relation to this partial specification. Additional assumptions lead to a class of overdispersed multinomial models, for which we describe estimators of population change and procedures for parsimonious model selection. We illustrate the use of overdispersed multinomial models by an application to data for Carolina Wren (*Thryothorus ludovicianus*).

Key words: counts; generalized linear models; monitoring; North American Breeding Bird Survey; surveys; trends.

INTRODUCTION

Over 650 species of birds breed in North America (Robbins et al. 1983). For the vast majority of these, there is no feasible way to estimate population sizes. Sophisticated mark–resight, capture–recapture, and transect methods can be applied to local populations, but are generally too labor intensive to implement in large-scale surveys. Exceptions tend to be limited to hunted species, which are monitored through aerial surveys with ground counts for visibility adjustments (Lancia et al. 1994). These methodologies, however, are not well suited to the monitoring of most avian populations.

For most species, a less ambitious goal may, nevertheless, be pursued. Counts provided under standardized conditions by skilled observers can be regarded as indices to population size and used to estimate patterns of relative population abundance; the type of patterns studied may be spatial (geographic range) or temporal (population increases or declines through time). However, in any application of count

data, it must be remembered that counts are neither censuses nor density estimates. Their use for estimation of relative abundances requires careful statement of assumptions and rigorous evaluation of the validity of these assumptions (Geissler and Noon 1981, Barker and Sauer 1992). We expand on these points in this paper, using the North American Breeding Bird Survey (BBS) for illustration.

The North American Breeding Bird Survey was initiated in 1966 to monitor bird population change. It is unique in its geographic extent and the large number of species it surveys. Presently, the BBS monitors >400 species of birds, using data obtained on an annual basis at >3500 roadside survey routes in the continental United States, southern Canada, and northern Mexico. Highly competent observers count the numbers of birds seen or heard in a 3-min period at each of 50 stops along each 24.5-mile (39.4-km) survey route, and report the total counts by species. Details of the history of the BBS, data collection protocols, and analytical procedures previously used are given by Robbins et al. (1986), Peterjohn (1994), and Peterjohn et al. (1994).

The design of the BBS is quite typical of a large number of surveys for birds and other animal taxa, in

which counts are conducted at the same sites over a series of years, and change is estimated within and among sites. Because the BBS has been well studied, many factors have been identified that, at best, complicate analysis of its data and, at worst, limit its usefulness as a source of information on population change (e.g., Robbins et al. 1986, Sauer et al. 1994, Peterjohn et al. 1995). We believe that many of the lessons learned in analysis of BBS data are appropriate for other similar surveys; thus, although our emphasis in this presentation is on the BBS, we believe that the conceptual and methodological issues discussed have a much wider relevance and applicability.

FUNDAMENTAL ISSUES RELATED TO THE ANALYSIS OF COUNT DATA

Any analysis of count data makes assumptions about variation in the proportion of animals counted, but such assumptions frequently are not stated explicitly. Temporal patterns of relative abundance in count data are only directly descriptive of corresponding patterns in populations if the proportion counted is constant through time. The same can be said with regard to spatial patterns in count data.

Although it is unlikely that the proportion counted is constant through space or time, variation in this proportion might not bias estimation of relative abundance, provided that there is no spatial or temporal pattern to the variation. If no pattern exists, this variation is simply a source of inefficiency in estimation of relative abundance. If a patterns exists, however, bias is a possibility: changes in the proportion counted could easily be misinterpreted as population change. To illustrate, Fig. 1 presents data for the House Sparrow (*Passer domesticus*) on BBS Route 1 in Alabama. The impression given by plotting the counts against years is of a sharp decline in the population in the 1970s, followed by a steady increase (in the early 1980s) to a plateau in the late 1980s (top panel, Fig. 1; the curve is a LOESS smooth, with tension parameter $f = 0.5$). The bottom panel, however, reveals that these "dynamics" are actually artifacts of observer changes in 1973 and 1981. An analysis of covariance regression shows that, having controlled for the difference in observer abilities, the counts are fairly stable with respect to time (lower panel, Fig. 1).

Clearly, then, it is important to identify factors that can bring about changes in the proportion counted, and to see whether there is evidence of temporal or spatial differences in counts related to these factors, but not to population changes.

The value of BBS data is enhanced by the standardization of data collection protocols and the careful selection of observers capable of identifying (visually and aurally) all species likely to be encountered on their routes. Nevertheless, considerable differences exist in the abilities of individual observers, and the ability of
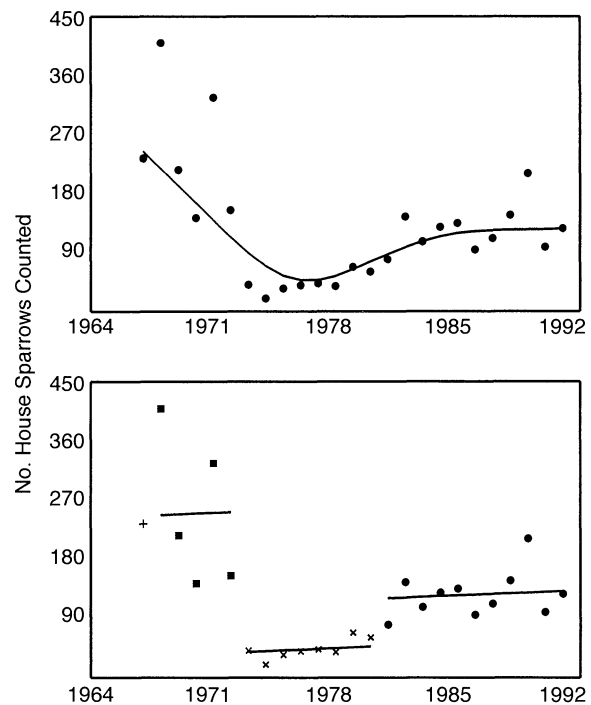

FIG. 1. Count data for the House Sparrow on BBS Route 1, Alabama. The top panel includes LOESS smooth (tension parameter $f = 0.5$) without controlling for observer differences. The bottom panel has distinct symbols for each observer and fitted values from an ANCOVA model.

the average observer has changed through time (Sauer et al. 1994). Simply put, new recruits to the BBS tend to be more skilled observers than the observers they replace. They tend to count more birds, particularly of species that are difficult to count; it has been suggested that these more qualified observers might tend to ignore more common and abundant species. They also identify more species of birds.

Patterns of change in counts thus reflect not only changes in population sizes, but also changes in the pool of observers; population change can be confounded with observer change. It might be anticipated that trend estimates for "difficult" species would be positively biased, as would trend estimates for species richness. On the other hand, it is conceivable that, for some more common species, the effect of increasing observer ability would be to introduce a negative bias to trend estimates.

Trends in observer ability can be documented in large-scale analyses of BBS data. One way of doing so is to estimate interval-specific population change (trend), first having controlled for observer differences, and then without controlling for observer differences. Consistent differences in these pairs of estimates point to the need for such controls; if not needed, such controls are likely to cause inefficiency, rather than bias. In particular, evidence for an increasing trend in ob-
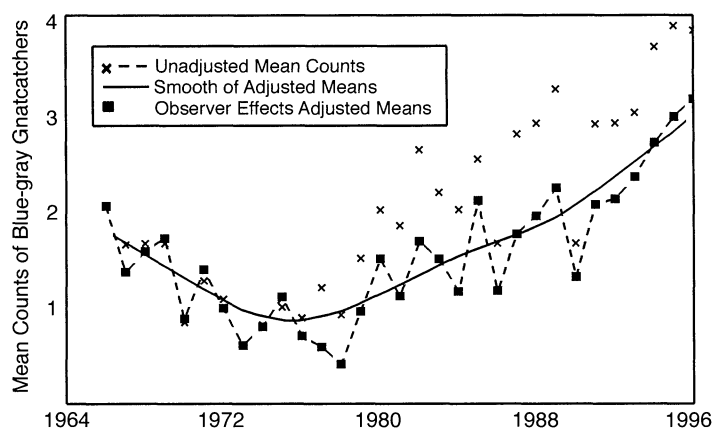
FIG. 2. Mean counts of the Blue-gray Gnatcatcher from BBS in the Upper Coastal Plain physiographic stratum in Maryland. Means adjusted for observer effects (■) are presented with a LOESS smooth (Hastie and Tibshirani 1990) with tension parameter $f = 0.5$. Unadjusted mean counts (×) overstate the pattern of population change. Data consist of 794 observations on 33 BBS routes by 149 observers.

server ability exists if consistently larger (more positive) trend estimates occur when differences are not controlled for. Sauer et al. (1994) analyzed BBS data for 409 species, using regression procedures, and found that 72.6% of the species considered had larger trend estimates when observer covariables were omitted. Sauer et al. (1994) also used several other analyses to document trend in observer ability in BBS data. James et al. (1996) provided results of LOESS-based trend analyses carried out both including and omitting controls for observer effects. We have noted (Link and Sauer 1997b) that 20 of 26 species in their analyses had more positive trend estimates when observers were not controlled for; this result provides further evidence of trend in BBS observer ability.

Further effects relating to observers include a tendency for smaller than expected counts in the first year of an observer's service on a route (perhaps as a consequence of inexperience with the logistics of conducting the survey), and declines in counts associated with hearing loss as observers age (Kendall et al. 1996, Link and Sauer 1997a). If these effects are not controlled for in analysis, the former is likely to result in a positive bias in estimates of population change; the latter is likely to cause a negative bias.

As a general indication of the importance of these observer effects, we include Fig. 2, which provides an estimate of the pattern of change in the population of Blue-gray Gnatcatchers (*Polioptila caerulea*) in the Upper Coastal Plain of Maryland. This estimate was obtained by controlling for observer effects in a manner to be described in later sections. In addition to having controlled for change in observer ability, we included in the model a significant "first year of count" effect and an "observer senescence effect" (for observers counting at a time >20 years after their first year); these effects were estimated to be 23% and 43% diminutions of counts, respectively. Superimposed on the graphic are mean counts, which are not corrected for observer effects. The pattern of change in the Blue-gray Gnatcatcher population is very poorly represented by the

unadjusted mean counts. Thus, a crucial component of the analysis of BBS data is an appropriate accounting for the effects of observers on counts.

To this point (and in most of this paper), our discussion has focused on estimation of temporal patterns of population change from count data. BBS data are also used to address many hypotheses about spatial patterns of population size and change. Observer effects are only one of many factors that could influence the proportion of birds detected on different BBS routes (e.g., differences in habitats; Sauer et al. 1995). Controlling for differences in observer effects among routes is much more difficult than controlling for observer effects within routes. Many of the published analyses of spatial patterns in abundance attempt no sort of adjustments for observer effects (e.g., Maurer 1994). Flather and Sauer (1996) controlled for within-route observer effects, but did not attempt to incorporate among-route adjustments. Changes in the ability of the pool of observers can clearly cause bias in temporal analyses of population change; the same biases can exist in spatial analyses of change, if the pattern of observer change is spatially heterogeneous. Thus, we believe that evaluation of these (and other) between-route effects should be a high priority for future spatial analyses of BBS data.

MODELS RELATING COUNTS TO POPULATION SIZES

BBS data cannot be used to directly estimate population sizes, because of the nature of count data and the possible bias of roadside route locations. However, the pattern of changes in counts for individual observers can be expected to generally follow the pattern of change in population size, which we refer to as the population trajectory. This limited assumption about the associations of change in counts and population sizes allows us to describe models relating counts to population trajectories and methods for selecting and fitting models of population trajectories to BBS data.

Let N_{ij} denote a local population size in some region including route i in the jth year of counts, t_j. The notions

of a local population size, and the region associated with it, are of necessity vague, as BBS sampling cannot guarantee either a census or a known fixed area of sampling. A referee has suggested that the following model description could benefit by substituting "density" for local population size, but, in our view, this simply introduces an alternative set of ambiguities. For population monitoring, the important consideration is not the actual population size or density, but the pattern of change through time associated with it. Thus, we suppose that the vector of population sizes is a single realization of a stochastic process with

$$\log\{E(N_{ij})\} = \theta_i + h_i(t_j) \tag{1}$$

where $h_i(t_1) = 0$, so that θ_1 is the logarithm of the expected population size in year t_1, and $E(N_{ij})$ denotes the expected value of the population size. We refer to $f_i(t) = \exp\{h_i(t)\}$ as the population trajectory on route i, and to $h_i(t)$ as the log-trajectory. Let μ_{ij} denote the expected count on route i in the jth year of observation. Denoting the observer who provides the jth count on route i by index $j(i)$, we assume that

$$\log\{\mu_{ij}\} = \gamma_{j(i)} + \delta_i(t_j) + h_i(t_j) \tag{2}$$

where the parameters $\gamma_{j(i)}$ represent observer effects, the function $\delta_i(t)$ describes time-dependent nuisance factors affecting counts but not related to the population size, and $h_i(t)$, as in Eq. 1, describes changes in the population. An application of this model to data collected on Carolina Wren (*Thryothorus ludovicianus*) is presented later in the paper.

Model statements in Eqs. 1 and 2 describe a relation between the pattern of change in counts and the pattern of change in population size. For instance, a simple special case is when $\delta_i(t)$ is identically equal to zero; then, the ratio of expected values for counts made by the same observer in two different years is the same as the ratio of expected values for population sizes in the same two years, and the population trajectory is reflected in the pattern of counts.

The population trajectory can be thought of as a scaled, and possibly smoothed, version of the vector of population sizes. The requirement that $h_i(t_1) = 0$ sets the trajectory value, $f_i(t)$, equal to 1 at time $t = t_1$. Values of the trajectory at other time points can be thought of as describing the total proportional change in the population over the time period from t_1 to t_2. Thus, for example, if $f_i(t_2) = 0.80$, then the population declined ~20% over the time interval $(t_1, t_2]$. If $t_2 - t_1 = 5$ years, then the 20% total rate of decline can be expressed as an annual rate of $0.80^{(1/5)} - 1 = -4.36\%$, which is the geometric mean rate of change over the interval.

We suggest that the term "trend" thus be defined as the geometric mean rate of change over a specified interval. The geometric mean rate of change over time

period $(t_a, t_b]$ on route i can be written in terms of the log-trajectory as $\beta_i(t_a, t_b) - 1$, where

$$\beta_i(t_a, t_b) = \exp\left\{\frac{h_i(t_b) - h_i(t_a)}{t_b - t_a}\right\}. \tag{3}$$

This is in keeping with the vernacular use of the term "trend" as a description of an overall tendency, without regard to fluctuations in the trajectory. Alternative definitions of "trend" are possible; our purpose here is to suggest that the term be reserved for some simple summary of the more complex pattern in the trajectory.

We note that if the trajectory is not linear on the log scale, the trend must be understood as an interval-specific quantity; the trend may well be positive over some period of time and negative over another. In discussion of population change, we view it as critical that the terms "trajectory" and "trend" be clearly distinguished. The distinction between the terms has important consequences for analysis and interpretation of survey data. Trend describes the change in a population over a specific interval; trajectory describes the manner in which the change occurred. Trajectory estimation provides greater detail, at greater cost in terms of data resources.

Eqs. 1 and 2 provide a partial specification of a general model for the relation between count data and corresponding population sizes. The specification is incomplete in several ways, most notably in that it does not describe the probability mechanism generating counts. Furthermore, without appropriate constraints on the functions $\delta_i(t)$ in Eq. 2, it is impossible to estimate the trajectory from count data. Nevertheless, Eqs. 1 and 2 provide a basic framework for estimating population change from BBS count data, allowing for the control of observer effects and time-related nuisance effects that are unrelated to the status of the population under consideration. For simplicity, we will refer to Eqs. 1 and 2 as the "means model" for count data. In subsequent sections, we discuss specific applications of the means model; that is, we describe the selection of functions $f_i(t)$ and $\delta_i(t)$ relating counts to population size.

The means model described is indexed by subscripts "i" corresponding to routes. This allows for the possibility of spatial variation in trajectory and trends. In practice, it may be found that BBS data are not adequate for the estimation of complex trajectories at the route level, so that it may be necessary to assume equal trajectories for geographically proximate routes. Trend estimates are then obtained at whatever scale on which it is reasonable to estimate a common trajectory (the "common trajectory region"). Summarizing trend estimates over larger geographic regions requires that individual estimates be combined by an abundance-weighted average. Ideally, these abundance weights would be based on population sizes; these are, of course, unknown. The usual expedient is to use the size

of the common trajectory region and a mean count for observers on routes in that region. The appropriateness of such abundance weights depends on a number of assumptions, the validity of which require further investigation.

ANALYTIC METHODS USED FOR BBS AND OTHER COUNT DATA

We now describe several analytic procedures that have been applied to count data in terms of the means model we have presented.

Approximate ANCOVA methods.—We use the term "approximate ANCOVA methods" to describe techniques that were used for over a decade to analyze BBS data (Geissler and Sauer 1990). These techniques were implemented assuming a simple form of the means model (Eq. 2), in which $\delta_i(t)$ is identically zero and the log-trajectory is linear, viz.,

$$\log\{\mu_{ij}\} = \gamma_{j(i)} + \beta_i(t_j - t_1).\qquad(4)$$

Analysis was carried out by least squares estimation of Eq. 2, with $E\{\log(Y_{ij} + C)\}$ used in place of $\log(\mu_{ij})$; here, Y_{ij} denotes the count on route i in year j, and C denotes a small positive constant (usually 0.5) added to all counts to avoid domain errors in the application of the logarithm. No specific error structure was used in evaluating the estimates of trend β_i. Variances were only estimated for composite trend estimates, through bootstrapping.

Applications of the approximate ANCOVA method implicitly acknowledged the possibility of a nonlinear log-trajectory. The possibility was accommodated by estimating period-specific trends, using only the data over the range of years for which an estimate of the geometric mean rate of change was desired. Thus, for example, Robbins et al. (1989) fit "early" and "late" trends corresponding to the periods 1966–1978 and 1978–1987.

Were it desirable, ANCOVA methods could be adapted to other specifications of the means model, thus allowing nonzero $\delta_i(t)$, and the modeling of nonlinear log-trajectories. However, the ANCOVA methods can only be thought of as approximating Eq. 2 of the means model, because $E\{\log(Y_{ij} + C)\}$ is not the same as $\log\{E(Y_{ij})\}$. For common species, the approximation works reasonably well, but for rarer species, the approximation can lead to biased estimation of the trend and its variance (Link and Sauer 1994, Sauer et al. 1996).

Estimating equation methods.—An improvement over the ANCOVA methods is provided by analyses using estimating equations that correspond to the score functions of a Poisson regression model with means governed by Eq. 2. Link and Sauer (1994) described trend estimators of this sort, assuming that $\delta_i(t)$ is identically zero, and that the trajectory is log-linear (as in Eq. 4). Again, these overly restrictive assumptions are

not essential to the application of the method. The assumption of Poisson distributions used in obtaining the estimating equations is not essential to their use, especially if model-based estimates of variances for trend estimates are not used. In the 1992–1993 BBS summary (Peterjohn et al. 1994), estimating equation estimates of trend were used with bootstrap estimates of variance.

LOESS-based methods.—The foregoing methods were developed under the requirement that a specific parametric model for the log-trajectory be chosen. Alternative analyses of BBS data have used nonparametric and semiparametric procedures to estimate the log-trajectory under the minimal assumption that it is a smooth function. James et al. (1996) described the use of LOESS (locally weighted least squares regression; Hastie and Tibshirani 1990) to analyze BBS data. This procedure is similar to a running average: fitted values are obtained by local smoothing of nearby data values. The degree of smoothing in LOESS-based analyses is governed by a "tension parameter," f, taking values between zero and one. When f is small, the fitted trajectory essentially "connects the dots" for counts by each observer. When f is close to 1, the fitted trajectory becomes closer to a straight line.

James et al. (1996) present LOESS analyses of BBS data, with and without controls for observer effects. We note that the model of James et al. differs slightly from the means model (Eqs. 1 and 2), in that it assumes observer effects to be additive on the square-root scale. Although we caution against the use of models that do not incorporate observer effects (Link and Sauer 1997b), we regard semiparametric LOESS with controls for observer effects as a useful tool for preliminary examination of population trajectories. Methods for optimal selection of the tension parameter have been developed, but, to our knowledge, have not been applied to the analysis of BBS data. We caution against the uncritical examination of nonlinearities in LOESS estimates of trajectories, especially when the tension parameter has been arbitrarily selected.

LOESS-based trend analyses can be carried out using bootstrap estimates of variance; James et al. (1996) fit trajectories on a route-by-route basis, and assess the variability of their combined estimate of trend by the sample variance of the abundance-weighted trend estimates. As is the case with ANCOVA and estimating-equations methods of trend estimation, LOESS-based methods do not specify the within-route variance structure of counts.

NEW PROCEDURES BASED ON OVERDISPERSED MULTINOMIAL MODELS

Recently developed methods for estimation of trajectory and trend from count data control for differences among observers by treating each observer's vector of counts as an overdispersed multinomial random variable. We have described these methods in formal

terms elsewhere (Link and Sauer 1997a); here, we describe their motivation and application to BBS data. The overdispersed multinomial models can be regarded as extensions of the estimating-equations estimators, developed with the intention of specifying a within-route variance structure of counts. These procedures allow for parsimonious estimation of population trajectories in the context of the means model (Eqs. 1 and 2).

Omission of observer effects in analyses can bias trend estimates; on the other hand, modeling observer differences requires the inclusion of a large number of parameters in the model. The median number of years of counts by BBS observers is estimated to be 4.8; typically, a new observer effect must be modeled for each five years of data on a route. To give a concrete example, routes with 18 years of data have up to 10 observers, with an average of 4 observers. Parametric modeling of a quadratic log-trajectory for the data from such a route would involve two parameters of interest (linear and quadratic terms) and four nuisance parameters (observer effects $\gamma_{j(i)}$). If a common quadratic log-trajectory were applied to a collection of 10 such routes, roughly 42 parameters would be needed, of which only two are of interest. This proliferation of nuisance parameters can have serious consequences in the implementation of numerical optimization procedures. Newton-Raphson optimization includes the inversion of numerous $p \times p$ matrices, where p is the number of parameters; thus, modeling observer effects can be computationally burdensome.

A solution to this difficulty is to carry out parametric analyses, having conditioned on sufficient statistics for nuisance parameters (McCullagh and Nelder 1989). The statistical procedure of conditioning on sufficient statistics replaces a full parametric likelihood with a conditional likelihood that is essentially equivalent in terms of the parameters of interest, but that does not involve the nuisance parameters. The practical consequence of this is immense: in the foregoing example, maximum likelihood estimation is reduced from a 42-dimensional optimization to a two-dimensional optimization.

In the Poisson regression model used to develop the estimating-equations estimates described in the previous section, the process of conditioning on sufficient statistics for the observer parameters $\gamma_{j(i)}$ has appealing and intuitive consequences. Conditioning on the total counts by each observer results in replacing the full likelihood by a product of multinomial likelihoods, one for each observer. The multinomial random variables corresponding to each observer have as many cells as years in which the observer counted; the index of the multinomial is the total count by the observer, and the cell probabilities are proportional to the observer's expected counts. The advantage of this new, conditional likelihood is that it does not depend on the observer

effects, $\gamma_{j(i)}$. All of the parameters of the full likelihood are retained, with most or all of the information about them, but the nuisance parameters are omitted.

Suppose, for a simple example, that the population were increasing at a constant rate of 10% per year, so that expected values of counts would follow Eq. 4, with $\beta = 1.10$. Then, the counts for an individual observer that counted in years 1, 2, 3, 4, and 7 would be expected to be in the proportions

$$\beta^1 :: \beta^2 :: \beta^3 :: \beta^4 :: \beta^7.$$

Because the cell probabilities must sum to 1, it is easy to verify that the cell probabilities are 15.6%, 17.2%, 18.9%, 20.8%, and 27.6%. Given this model, and knowing that an observer had counted a total of 100 birds, we would expect that ~16 of the birds were observed in year 1, ~17 in year 2, etc.

In practice, we do not know the value of β, but must work back from the data to the parameters. Suppose that, of 100 birds counted by an individual, 15 were in year 1, 21 in year 2, 28 in year 3, 10 in year 4, and 26 in year 7. The multinomial likelihood to be maximized is proportional to:

$$\left(\frac{\beta^1}{\beta^1 + \beta^2 + \beta^3 + \beta^4 + \beta^7}\right)^{15} \left(\frac{\beta^2}{\beta^1 + \beta^2 + \beta^3 + \beta^4 + \beta^7}\right)^{21}$$

$$\times \left(\frac{\beta^3}{\beta^1 + \beta^2 + \beta^3 + \beta^4 + \beta^7}\right)^{28} \left(\frac{\beta^4}{\beta^1 + \beta^2 + \beta^3 + \beta^4 + \beta^7}\right)^{10}$$

$$\times \left(\frac{\beta^7}{\beta^1 + \beta^2 + \beta^3 + \beta^4 + \beta^7}\right)^{26}.$$

Although no closed form exists for the value β that maximizes this expression, the maximizer is easily obtained using Newton-Raphson or other numerical procedures. The maximum likelihood estimator of β for this likelihood is 1.054; the data indicate a 5.4% per year rate of increase in the population. Under the multinomial model, the standard error of this estimate is 4.9%, based on the information matrix.

In this example, data for a single observer have been treated as a single multinomial random variable. BBS routes typically have several observers; in this case, the likelihood that is maximized is the product of multinomial likelihoods, one for each observer, but governed by the same set of parameters. In the general case of the means model (see Eq. 2), the cell probability corresponding to count Y_{ij} is

$$\pi_{ij} = \frac{\mu_{ij}}{\sum_{k(i)=j(i)} \mu_{ik}} = \frac{\exp\{\delta_i(t_j) + h_i(t_j)\}}{\sum_{k(i)=j(i)} \exp\{\delta_i(t_k) + h_i(t_k)\}}. \quad (5)$$

Note that the nuisance parameters $\gamma_{j(i)}$ (from Eq. 2) have factored out of the numerator and denominator in the right-hand side of Eq. 5. In our subsequent discussion, we will denote the collection of π's corresponding to the jth observer on route i by a vector π_j and suppress

dependence of j on i in our notation. The total number of birds counted by this observer will be denoted by T_j, and the observer's vector of counts denoted by $\mathbf{Y}_j$.

Closer examination of the example data set reveals an inadequacy in the fit of the multinomial random variable to the data; expected cell frequencies under the simple model of a linear log-trajectory are 17.5, 18.5, 19.5, 20.5, and 24.0. There is considerable variation in the counts about these values. The Pearson chi-squared statistic for goodness-of-fit takes the value 10.005 on 3 degrees of freedom ($df = 5$ cells $- 1$ parameter estimated $- 1$); the corresponding P value of 0.0185 indicates an unsatisfactory fit of the model to the data. Two possible explanations can be offered: first, that the specified means model (linear log-trajectory, with $\delta_i(t) = 0$) is incorrect, or second, that the assumption of Poisson distributions for counts is incorrect.

We have found that, even with the specification of fairly complex means models, there is a tendency for the product multinomial likelihood model to yield unacceptable goodness-of-fit statistics for BBS data. This is not surprising; the explanation is that the assumption of Poisson distributions in counts is inadequate. Fortunately, however, this inadequacy does not require that the appealing multinomial approach be jettisoned.

When count data show overdispersion relative to Poisson random variables, a convenient model is the negative binomial, obtained by assuming a gamma mixture. This model specifies that the counts are Poisson random variables, but with random means, sampled from gamma distributions. This model can be thought of as incorporating the effects of population stochasticity on the counts (for details, cf. Link and Sauer 1997a).

Conditioning on totals for negative binomial counts has a remarkably similar consequence to conditioning on totals for Poisson random variables. Conditioning on totals of Poisson random variables yields a multinomial distribution; conditioning on totals for negative binomial random variables (i.e., overdispersed Poisson random variables) yields an overdispersed multinomial distribution. This distribution is known as the Dirichlet compound multinomial (DCM). Under the conditional multinomial model, the mean vector and variance–covariance matrix for $\mathbf{Y}_j$ can be written as $\boldsymbol{\mu}(\boldsymbol{\pi}_j)$ and $\mathbf{V}(\boldsymbol{\pi}_j)$, respectively. Under the DCM model, the mean vector is the same, and the variance–covariance matrix is only slightly modified, to $\phi_j\mathbf{V}(\boldsymbol{\pi}_j)$; the quantity ϕ_j is referred to as the overdispersion parameter. This overdispersion parameter is estimated by the Pearson chi-squared statistic. Thus, in the foregoing example, the standard error of the estimated β under the multinomial model was estimated to be 4.9% this needs to be scaled by a factor of $\sqrt{10.005} = 3.16$, to 15.7%.

Maximum likelihood estimation under the DCM model presents considerable difficulties in implemen-

tation. Newton-Raphson optimization algorithms tend to fail to converge. However, this difficulty can be averted by the use of a quasilikelihood approximation (McCullagh and Nelder 1989) to the DCM likelihood. The quasilikelihood approach begins by treating the counts as Poisson random variables: the multinomial model is fitted and the overdispersion parameter is estimated by a goodness-of-fit statistic. The multinomial model is then refitted, after the observations have been weighted by the overdispersion parameters. The process of fitting the multinomial model and estimating overdispersion parameters can be carried out repeatedly; we suggest that one or two iterations will be sufficient.

The principal benefit of this overdispersed multinomial approach is that it allows for consideration of within-route variation. All of the other procedures described can only provide estimates of precision for composite estimates of trend. The modeling of within-route variation also allows for model selection via likelihood ratio tests, so that the merits of alternative means models can be investigated, in a manner that we shall proceed to describe.

MODEL SELECTION UNDER THE DCM MODEL

The DCM likelihood for $\mathbf{Y}_j$, evaluated at estimates $\hat{\pi}_{ij}$, is approximated by

$$\frac{T_j!}{\prod Y_{ij}!}\left[\frac{\Gamma\left(\sum T_j\hat{\pi}_{ij}/a_j\right)}{\prod \Gamma\left(T_j\hat{\pi}_{ij}/a_j\right)}\right]\frac{\prod \Gamma(T_j\hat{\pi}_{ij}/a_j + Y_{ij})}{\Gamma\left(\sum T_j\hat{\pi}_{ij}/a_j + T_j\right)}. \quad (6)$$

Here, $\Gamma(x)$ is the gamma function, and a_j is a parameter related to the overdispersion by

$$a_j = T_j\left(\frac{\phi_j - 1}{T_j - \phi_j}\right).$$

Given specifications of the means model (i.e., candidate parametric forms for $h_i(t)$ and $\delta_i(t)$ in Eqs. 1 and 2), estimates of the parameters π_{ij} can be obtained using the quasilikelihood procedures previously outlined. Substituting these values in Eq. 6 provides the basis for likelihood ratio tests between nested models.

For non-nested models, we recommend the use of Akaike's information criterion as a means for screening potential models. Akaike's information criterion (AIC) is defined by

AIC $= -2$(maximum log likelihood)
$+ 2$(number of parameters estimated).

Low values of AIC indicate acceptable models. Thus, AIC favors models with large likelihoods and few parameters. Generally speaking, models within one or two AIC points are regarded as equally acceptable.

AIC provides a reasonable criterion for model selection. Nevertheless, the selection of appropriate mod-

els for count data is a difficult task. Care must be taken to ensure that the collection of candidate models includes models describing biologically relevant sources of variation in population change. It is also important to consider candidate models that account for temporal changes in the proportion of birds counted, allowing for changes related to individual observers and to changes in the pool of observers.

AN EXAMPLE: THE CAROLINA WREN IN MARYLAND

We illustrate the process of model selection using data for the Carolina Wren (*Thryothorus ludovicianus*). We use data collected during the period 1966–1992 on the 20 BBS routes of the Northern Piedmont stratum in Maryland. An exceptionally hard winter (1976–1977) is thought to have substantially reduced population numbers for the Carolina Wren.

There were 78 different observers participating in the collection of these data; of these, 26 only provided data for a single year, and another four did not count any Carolina Wrens. Data for these 30 individuals provide no information about population change; their contribution to the likelihood is constant with respect to the parameters of interest, so they can be excluded from the analysis. The remaining 48 observers participated for an average of 8.6 yr (quartiles: 2.3, 6.5, 11.8 yr), and provided a total of 411 counts. The average count was 5.4 birds/yr (quartiles: 2.0, 4.0, and 7.6 birds/yr).

We applied the means model (Eq. 2) to a stratum-level analysis of these data. That is, we assumed a common log-trajectory $h(t)$ for the collection of routes. Candidate log-trajectories were of the form

$$h(t) = \sum_{k=1}^{7} a_k t^k - \xi I(t > 1976).$$

That is, they allowed up to a seventh-degree polynomial, and the inclusion of an intervention effect, ξ, related to the hard winter. Here, and in the sequel, the function $I(x)$ is the indicator function, taking the values one or zero, depending on whether x is true or false.

We controlled for the possibility of initiation effects (lower than expected counts in an observer's first year of service) by setting

$$\delta(t) = \Delta I(\text{Year } t \text{ is observer's first year of participation}).$$

For each of the seven possible degrees of polynomial, we considered four different models, obtained by the decisions whether or not to impose restrictions $\Delta = 0$ or $\xi = 0$. Table 1 reports trend estimates, estimates of their standard errors, and values of AIC for the 28 candidate models. The fit of models with nonzero ξ (i.e., those that include an intervention corresponding to the hard winter of 1976–1977) is clearly superior to that of models in which the restriction $\xi = 0$ is made.

Also, little benefit is seen in modeling a first-year

TABLE 1. Model selection for Carolina Wren data. The first three columns describe the model (degree of polynomial used in trajectory; indicator for observer initiation effects, Δ; indicator for intervention effect related to the hard winter of 1976, ξ). Model selection is based on minimizing the Akaike information criterion (AIC). The final two columns summarize the fitted trajectory: $\hat{\beta}$ is the estimated trend for the period 1966–1992, and $\text{SE}(\hat{\beta})$ is the estimated standard error of trend estimate.

Degree	Δ	ξ	AIC	$\hat{\beta}$	$\text{SE}(\hat{\beta})$
1	0	0	1722.84	5.78	0.76
2	0	0	1691.89	3.39	0.77
3	0	0	1636.35	10.33	1.18
4	0	0	1599.25	18.45	2.27
5	0	0	1586.78	11.89	1.85
6	0	0	1550.28	8.55	1.48
7	0	0	1552.74	8.37	1.53
1	1	0	1725.00	5.71	0.79
2	1	0	1693.18	3.30	0.80
3	1	0	1636.84	10.91	1.23
4	1	0	1596.76	19.17	2.27
5	1	0	1585.42	12.94	1.92
6	1	0	1552.47	9.03	1.57
7	1	0	1555.16	8.86	1.64
1	0	1	1493.55	7.72	0.49
2	0	1	1491.63	8.78	0.71
3	0	1	1475.96	11.48	1.05
4	0	1	1477.92	11.99	1.35
5	0	1	1471.52	9.94	1.34
6	0	1	1469.88	9.47	1.37
7	0	1	1471.72	9.03	1.40
1	1	1	1495.05	7.58	0.50
2	1	1	1493.13	8.65	0.73
3	1	1	1478.42	11.38	1.09
4	1	1	1480.40	11.86	1.40
5	1	1	1473.94	9.76	1.42
6	1	1	1472.28	9.21	1.43
7	1	1	1474.28	8.68	1.56

count effect for observers; this stands in contrast to the Blue-gray Gnatcatcher example (Fig. 2) and to a similar analysis of the Wood Thrush (*Hylocichla mustelina*), in which a first-year effect corresponding to an 8% undercount was noted (Link and Sauer 1997a). Three acceptable models (fifth-, sixth-, and seventh-degree polynomials, with $\Delta = 0$ and $\xi \neq 0$) are essentially equivalent in their fits; of these, the minimum AIC value is for the sixth-degree polynomial. The differences among these trajectory estimates are slight; the effect of the choice is most likely to be consequential if the trajectory is to be used to make projections of population status beyond the range of years in the data set. We would discourage such projections, noting that a major feature of this trajectory is the intervention due to a severe winter, an effect that could not have been predicted from the data for preceding years.

The effect due to the severe winter (1976–1977) is estimated as an 83.0% drop in the population size (95% CI: 75.1%, 88.4%). This effect, and the pattern of population change estimated under the sixth-degree polynomial model, are displayed in Fig. 3. Error bars there are for comparison of population sizes with the population in 1966. Despite the intervention, over the en-
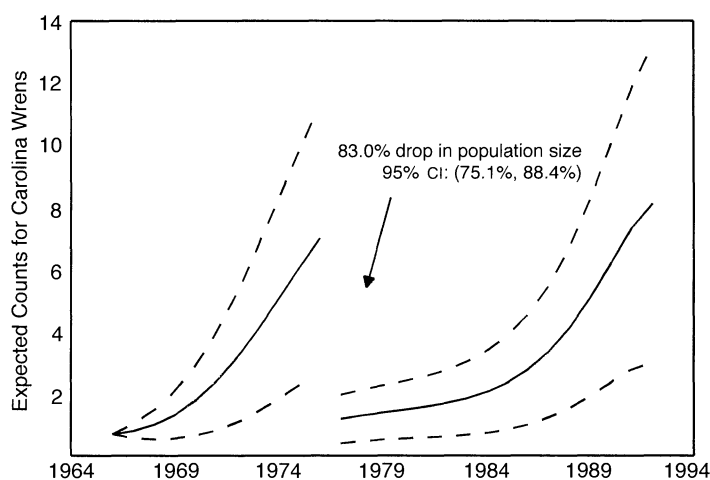
FIG. 3. Estimated population trajectory (with 95% confidence bounds) for the Carolina Wren in the Northern Piedmont physiographic stratum in Maryland. Note the 83% population reduction associated with a hard winter (1976–1977).

tire period (1966–1992), the population of Carolina Wrens is estimated to have grown at an average annual rate of almost 10% per year (9.47 ± 1.37%, mean ± 1 SE). Note (Table 1) that there is considerable variability in the estimates of trend; however, among the acceptable models and the overparameterized models, smaller differences are found. The indication is that the insufficiently parameterized models produce biased estimates of trend.

Fig. 3 highlights the importance of interpreting "trend" as an interval-specific quantity: for example, the trend for the period 1974–1980 is negative, whereas that for the entire period is positive. In the presence of a nonlinear trajectory, trend is simply a convenient summary for a particular interval of time, and should not be used as the sole descriptor of population change.

A final comment on this example is that the trajectory need not be modeled as a polynomial on the log scale. Any other function governed by estimable parameters could be considered. In fact, given adequate replication, a "year-effects" model can be implemented, allowing adjusted yearly estimates of relative population size. The adjusted mean counts in Fig. 2 were obtained in this fashion.

COMPOSITE ANALYSIS FOR LARGER GEOGRAPHIC AREAS

Geographic aspects of the BBS sample can greatly complicate analysis. Bird species differ in abundance over their ranges, and if population change is not consistent in regions of high and low abundance, simple averages of population trends among survey routes or state-stratum regions will not be unbiased estimates of change for the total population. The simple average of trends provides an estimate of the mean population trend, but to estimate total population change, each route-specific trend estimate must be weighted by the population size for the route, which is, of course, unknown. The use of adjusted mean counts as surrogates

for population size (under assumptions that there is little regional variability in sighting proportions) may be the only option if regional heterogeneity in trend exists.

Estimation of regional trends is further complicated by large differences in the precision of trend estimates among individual routes and among state-stratum regions. Simple averaging of estimates that differ greatly in precision can lead to highly inefficient estimates. Area weights are necessary to accommodate differences in the areas covered by each state-stratum area.

Geissler and Sauer (1990) addressed these issues by incorporating several weights into the estimation of composite trends. They weighted trends estimated on routes by (1) an estimate of mean abundance for the route; (2) an index of precision for the route (based on the design matrix of the ANCOVA analysis); and (3) an area weight for the state-stratum in which the route occurred. This rather ad hoc approach had the serendipitous effect of mitigating the influence of routes with incomplete coverage, which tend to positively bias estimates of trend, due to unmodeled start-up effects (Kendall et al. 1996). At the same time, this weighting scheme incorporated the abundance weight to allow estimation of trend for the total population.

If trend and population size are independent, a precision weighted average of local trend estimates provides a reasonable composite estimate of trend, without the need for abundance weighting. For small geographic areas, it may even be reasonable to make the assumption of a common trajectory, and thus avoid the problems of abundance weighting. This is the approach we used in the analysis of Carolina Wren data of the previous section.

The reasonableness of the equal trajectory assumption can be assessed under the overdispersed multinomial model. We applied such a hypothesis-testing procedure to the Carolina Wren data. First, we chose a parsimonious model using AIC, as previously de-

scribed (the selected model had a sixth-degree polynomial log-trajectory and an intervention effect corresponding to the severe winter of 1976–1977, for a seven-parameter model). Fifteen of the 20 BBS routes in the Maryland Northern Piedmont Stratum had adequate data to fit the seven-parameter model. We fit the seven-parameter model to these routes twice, first allowing the parameters to vary among routes, and then under the assumption of a common model. The null hypothesis of a common model was then tested by comparing twice the difference in maximum loglikelihoods to the appropriate chi-squared distribution. The value $\chi^2 = 117.31$ on 98 degrees of freedom corresponds to a P value of 0.089, indicating only weak evidence of heterogeneity in trajectories. Because $\sqrt{117.31/98} = 1.09$, it may be desirable to increase estimated standard errors by 9%.

We suggest that large-scale regional summaries be conducted using empirical Bayes procedures. The individual trend estimates for the component regions are "shrunk" toward the prior mean estimated for the collection, accommodating the differences in precision of the estimates (Link and Sauer 1996). These empirical Bayes estimates can then be averaged with weights of relative abundance and area to provide an estimated trend for the population total.

Spatial comparisons of relative abundance from BBS data must be viewed with some skepticism. The fundamental difference between spatial and temporal analysis of BBS data is that the assumption of consistency in detectability can be plausibly advanced in considering counts for individual observers taken at the same site, through time; this assumption is less plausible for comparisons of counts among routes at a large geographic scale. Regional differences in observer ability and basic regional differences in detectability of birds may exist, with the potential to bias estimates of relative abundance (Sauer et al. 1995). Unfortunately, little information is available on counts for individual observers at distant sites; hence, this component of detectability has never been modeled.

CONCLUSIONS

Analysis of count data should begin with the frank acknowledgment that counts are not necessarily very good surrogates for population sizes. Pattern in counts is reflective not only of corresponding pattern in population sizes, but also of corresponding patterns in the proportion of animals counted. Surveys producing count data should be designed, inasmuch as possible, to minimize variation in the proportion of animals counted. Complete removal of this variation is not likely to be possible, hence, analysts must be aware of potential sources of pattern in this proportion, and must design analytic methods accordingly. Thus, analysis of count data requires a delicate interaction among sta-

tistical modeling, biological intuition, and familiarity with the methods of the surveys producing the counts.

For the BBS, we have shown that several factors produce pattern in the proportion of birds counted; these include differences among observers and changes through time in the ability of individual observers. These factors can be incorporated into analyses through use of covariates for within-observer effects (e.g., Kendall et al. 1996) and by conditioning to accommodate among-observer effects (Sauer et al. 1994, Link and Sauer 1994). Missing data from inconsistently surveyed routes further complicate analysis. Geographic structuring in the data raises the issue of appropriate scale for composite analyses.

The overdispersed multinomial analysis (Link and Sauer 1997a) provides a means of accommodating these constraints. This method conditions on observer totals, allows for the incorporation of start-up and other effects as covariates, provides valid estimates of variance of trend, and can be applied at the scale of individual routes or regions. It also allows more complicated covariable analysis, as shown in the intervention analysis of the example. We feel that modeling on individual routes is unlikely to provide sufficient information for adequate modeling of start-up effects or higher level polynomials; hence, we apply the analysis at a relatively low geographic scale (physiographic strata within states) and then use weighted averages of these estimates to summarize trends at higher scales.

LITERATURE CITED

Barker, R. J., and J. R. Sauer. 1992. Modelling population change from time series data. Pages 182–194 in D. R. McCullough and R. Barrett, editors. Wildlife 2001: populations. Elsevier, New York, New York, USA.

Flather, C. H., and J. R. Sauer. 1996. Using landscape ecology to test hypotheses about large-scale abundance patterns in migratory songbirds. Ecology 77:28–35.

Geissler, P. H., and B. R. Noon. 1981. Estimates of avian population trends from the North American Breeding Bird Survey. Pages 42–51 in C. J. Ralph and J. M. Scott, editors. Estimating numbers of terrestrial birds. Studies in Avian Biology, 6.

Geissler, P. H., and J. R. Sauer. 1990. Topics in route-regression analysis. Pages 54–57 in J. R. Sauer and S. Droege, editors. Survey designs and statistical methods for the estimation of avian population trends. U.S. Fish and Wildlife Service, Biological Report 90(1).

Hastie, T. J., and R. J. Tibshirani. 1990. Generalized additive models. Chapman and Hall, New York, New York, USA.

James, F. C., C. E. McCulloch, and D. A. Wiedenfeld. 1996. New approaches to the analysis of population trends in land birds. Ecology 77:13–21.

Kendall, W. L., B. G. Peterjohn, and J. R. Sauer. 1996. First-time observer effects in the North American Breeding Bird Survey. Auk 113:823–829.

Lancia, R. A., J. D. Nichols, and K. H. Pollock. 1994. Estimating the number of animals in wildlife populations.

Pages 215–253 *in* T. A. Bookhout, editor. Research and management techniques for wildlife and habitats. Fifth edition. Wildlife Society, Bethesda, Maryland, USA.

Link, W. A., and J. R. Sauer. 1994. Estimating equations estimates of trends. Bird Populations **2**:23–32.

Link, W. A., and J. R. Sauer. 1996. Extreme values in ecological studies: on avoiding the misleading effects of sampling variation in summary analyses. Ecology **77**:1633–1640.

Link, W. A., and J. R. Sauer. 1997*a*. Estimation of population trajectories from count data. Biometrics **53**:63–72.

Link, W. A., and J. R. Sauer. 1997*b*. "New approaches to the analysis of population trends in land birds": a comment on statistical methods. Ecology **78**:2632–2634.

Maurer, B. A. 1994. Geographical population analysis. Blackwell Scientific, Oxford, UK.

McCullagh, P., and J. A. Nelder. 1989. Generalized linear models. Second edition. Chapman and Hall, New York, New York, USA.

Peterjohn, B. G. 1994. The North American Breeding Bird Survey. Birding **26**:386–398.

Peterjohn, B. G., J. R. Sauer, and W. A. Link. 1994. The 1992 and 1993 summary of the North American Breeding Bird Survey. Bird Populations **2**:46–51.

Peterjohn, B. G., J. R. Sauer, and C. S. Robbins. 1995. The North American Breeding Bird Survey and population trends of neotropical migrant birds. Pages 3–39 *in* T. E. Martin and D. Finch, editors. Neotropical migrant birds. Cambridge University Press, New York, New York, USA.

Robbins, C. S., B. Bruun, and H. S. Zim. 1983. Birds of North America. Golden Press, New York, New York, USA.

Robbins, C. S., D. Bystrak, and P. H. Geissler. 1986. The Breeding Bird Survey: its first fifteen years, 1965–1979. U. S. Fish and Wildlife Service, Resource Publication 157.

Robbins, C. S., J. R. Sauer, R. S. Greenberg, and S. Droege. 1989. Recent declines in populations of birds that migrate to the neotropics. Proceedings of the National Academy of Sciences (USA) **86**:7658–7662.

Sauer, J. R., G. W. Pendleton, and S. Orsillo. 1995. Mapping of bird distributions from point count surveys. Pages 151–160 *in* C. J. Ralph, J. R. Sauer, and S. Droege, editors. Monitoring bird populations by point counts. USDA Forest Service, Pacific Southwest Research Station, General Technical Report **PSW-GTR-149**.

Sauer, J. R., G. W. Pendleton, and B. G. Peterjohn. 1996. Evaluating causes of population change in North American insectivorous songbirds. Conservation Biology **10**:465–478.

Sauer, J. R., B. G. Peterjohn, and W. A. Link. 1994. Observer differences in the North American Breeding Bird Survey. Auk **111**:50–62.