Studies in Avian Biology No. 34:45–54

A SMOOTHED RESIDUAL BASED GOODNESS-OF-FIT STATISTIC FOR NEST-SURVIVAL MODELS

RODNEY X. STURDIVANT, JAY J. ROTELLA, AND ROBIN E. RUSSELL

Abstract. Estimating nest success and identifying important factors related to nest-survival rates is an essential goal for many wildlife researchers interested in understanding avian population dynamics. Advances in statistical methods have led to a number of estimation methods and approaches to modeling this problem. Recently developed models allow researchers to include a covariate that varies by individual and time. These techniques improve the realism of the models, but they suffer from a lack of available diagnostic tools to assess their adequacy. The PROC NLMIXED procedure in SAS offers a particularly useful approach to modeling nest survival. This procedure uses Gaussian quadrature to estimate the parameters of a generalized linear mixed model. Using the SAS GLMMIX macro, we extend a goodness-of-fit measure that has demonstrated desirable properties for use in settings where quasi-likelihood estimation is used. The statistic is an unweighted sum of squares of the kernel-smoothed model residuals. We first verify the proposed distribution under the null hypothesis that the model is correctly specified using the new estimation procedure through simulation studies. We then illustrate the use of the statistic through an example analysis of daily nest-survival rates.

Key Words: binary response, generalized linear mixed model (GLMM), goodness-of-fit, kernel smoothing, logistic regression, nest survival.

UNA ESTADÍSITICA BASADA EN AJUSTE DE CALIDAD RESIDUAL SUAVIZADA PARA MODELOS DE SOBREVIVENCIA DE NIDO

Resumen. Estimar el éxito de nido e identificar factores importantes relacionados a las tasas de sobrevivencia de nido es una meta esencial para muchos investigadores de vida silvestre interesados en el entendimiento de las dinámicas poblacionales de aves Avances en métodos estadísticos han dirigido a un número de métodos de estimación y acercamiento para modelar este problema. Recientemente, modelos que han sido desarrollados permiten a los investigadores incluir una covariante que varia por individuo y tiempo. Estas técnicas mejoran la realidad de los modelos, pero padecen de la falta de disponibilidad de herramientas de diagnóstico para valorar qué tan adecuadas son. El procedimiento PROC NLMIXED en SAS ofrece un acercamiento particularmente útil para modelar la sobrevivencia de nido. Este procedimiento utiliza cuadratura Gaussiana para estimar los parámetros de un modelo generalizado linear mezclado. Usando el SAS GLMMIX macro aumentamos la medida de calidad de ajuste, la cual ha demostrado propiedades deseables para utilizar en ajustes donde la estimación de probabilidad aparente es utilizada. La estadística es una suma no cargada de cuadrados de residuos del modelo suavizado kernel. Primero verificamos la distribución propuesta bajo la hipótesis nula de que el modelo está correctamente especificado, utilizando el nuevo procedimiento de estimación a través de estudios de simulación. Después ilustramos el uso de la estadística por medio de un ejemplo de análisis de tasas de sobrevivencia de nido diarias.

Dinsmore et al. (2002), Stephens (2003), and Shaffer (2004a) concurrently developed methods for modeling daily nest-survival rates as a function of nest, group, and/or time-specific covariates using a generalized linear model (McCullagh and Nelder 1989) with binomial likelihood (see Rotella et al. [2004] for review). All of the methods use the likelihood presented by Dinsmore et al. (2002) and extend the model of Bart and Robson (1982). As with the commonly used Mayfield estimate (Mayfield 1975), overall nest success is estimated by raising daily survival rates to the power of n, where n is the number of days in the nesting cycle.

The model likelihood involves the probability a nest survives from day i to $i + 1$, denoted S_i (the daily survival rate). As an example, consider a nest found on day k, and active when revisited on day l, and last checked on day m ($k < l < m$). The nest survived the first interval and therefore contributes $S_k S_{k+1} \ldots S_{l-1}$ to the likelihood. The probability the nest failed would be one minus the product so that the likelihood is proportional to the product of probabilities of observed events for all nests in the sample (Dinsmore et. al. 2002).

Using the logit link, the daily survival rate of a nest on day i is modeled:

$$\text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) = \beta_0 + \sum_k \beta_k x_{ik} \qquad (1)$$

where we let π_i denote the daily probability of nest survival and the x_{ik} are values of the K covariates. The outcome is modeled as a series

of Bernoulli trials, where the number of trials is *t* for a nest surviving an interval of *t* days, and one for a nest failing within the interval (Rotella et al. 2004). Stephens (2003) implements nest-survival models in PROC NLMIXED of SAS (SAS Institute 2004) using programming statements within the procedure to perform iterative logistic regression for each day in an interval. This implementation allows the modeler to include random as well as fixed effects, as do recent implementations (Dinsmore et al. 2002) of program MARK (White and Burnham 1999).

The random-effects logistic-regression model accounts for clustering structures inherent in the data. Variables whose observations can be thought of as random samples from the population of interest are candidates for inclusion into the model as random effects (Pinheiro and Bates 2000). Examples of covariates in nest-survival studies that might be treated as random effects are study site, year, or individual nest. In this case, with two levels, we might suppose that either or both coefficients (intercept and slope of the linear logit expression) vary randomly across groups. Suppose for simplicity that we have a single covariate. If we treat the intercept and slope as random, the logistic model of (1) becomes:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = \beta_{0j} + \beta_{1j}x_{ij} \qquad (2)$$

with $\beta_{0j} = \beta_0 + \mu_{0j}$, and $\beta_{1j} = \beta_1 + \mu_{1j}$. The random effects are typically assumed to have a normal distribution so that $\mu_{0j} \sim N(0, \sigma_0^2)$ and $\mu_{1j} \sim N(0, \sigma_1^2)$. Further, the random effects need not be uncorrelated so we have, in general, $\text{Cov}(\mu_{0j}, \mu_{1j}) = \sigma_{01}$.

Substituting the random effects into expression (2) and rearranging terms, the model is:

$$\text{logit}(\pi_{ij}) = \log\left(\frac{\pi_{ij}}{1-\pi_{ij}}\right) = (\beta_0 + \beta_1 x_{ij}) + (\mu_{0j} + \mu_{1j}x_{ij}) \qquad (3)$$

The model in (3) suggests a general matrix representation for the random effects logistic-regression model given by:

$$\mathbf{y} = \pi + \varepsilon$$

where $\mathbf{y}$ is an $N \times 1$ vector of the binary outcomes (survived or not), π the vector of probabilities, and ε the vector of errors.

The response is related to the data through the link function:

$$\text{logit}(\pi) = \mathbf{X}\beta + Z\mu \qquad (4)$$

Here, $\mathbf{X}$ is a design matrix for the fixed effects. For the model given in expression (4) this is an $N \times 2$ matrix with first column of ones and the second column the vector of values for the predictor variable x_{ij}. The vector β is the corresponding $p \times 1$ vector of parameters for the fixed portion of the model. In our example this is the 2×1 vector $(\beta_0, \beta_1)'$. Under the $BIN(\pi)$ assumption (*BIN* referring to the binomial distribution), the vector of level-one errors, ε, has mean zero and variance given by the diagonal matrix of binomial variances: $\text{Var}(\varepsilon) = \mathbf{W} = \text{diag}[\pi_{ij}(1 - \pi_{ij})]$.

The term $Z\mu$ in (4) introduces random effects and represents the difference between the random effects and standard logistic-regression models. The matrix $\mathbf{Z}$ is the design matrix for the random effects. In the example in (3), $\mathbf{Z}$ is an $N \times 2J$ matrix as there are two random effects. The matrix is block diagonal, with the blocks corresponding to the groups in the hierarchy (in this example, the level two groups indexed from $j = 1$ to J). The vector μ is a $2J \times 1$ vector of coefficients corresponding to the random effects. The elements are the random intercept and random slope for each group in the hierarchy. The vector has assumed distribution $\mu \sim N(0, \Omega)$ with block diagonal covariance matrix.

Several methods are available for estimating the parameters of the hierarchical logistic-regression model (Snijders and Bosker 1999). The methods include numerical integration (Rabe-Hesketh et al. 2002), use of the E-M algorithm (Dempster et al. 1977) or Bayesian techniques to optimize the likelihood (Longford 1993), and quasi-likelihood estimation (Breslow and Clayton 1993).

By conditioning on the random effects and then integrating them out, an expression for the maximum likelihood is available. Although this integral is difficult to evaluate, estimation techniques involving numerical integration, such as adaptive Gaussian quadrature, recently have been implemented in many software packages including SAS PROC NLMIXED (SAS Institute 2004). These methods are computationally intensive and do not always result in a solution. As a result, in most cases, the technique cannot handle larger models (such as data with more than two hierarchical levels or a large number of groups or random effects).

In this paper, we wish to extend the goodness-of-fit measure introduced in the next section beyond the quasi-likelihood estimation approach (Breslow and Clayton 1993) used in its development and testing. Specifically, the SAS GLMMIX (SAS Institute 2004) macro was used

in simulation studies to verify the theoretical distribution of the statistic (note that since that study, SAS has implemented the SAS GLMMIX procedure which can be used to obtain the same results). SAS GLMMIX implements a version of quasi-likelihood estimation which SAS refers to as PL or pseudo-likelihood (Wolfinger and O'Connell 1993). For the logistic-hierarchical model, a Taylor approximation is used to linearize the model. The estimation is then iterative between fixed and random parameters. These procedures suffer from known bias in parameter estimates (Rodriguez and Goldman 1995). In this paper, we extend the statistic to the less biased estimation approach of Gaussian quadrature available in PROC NLMIXED (SAS Institute 2004).

The Goodness-of-fit Measure

Various goodness-of-fit statistics are available for use in the standard logistic-regression setting, but none have been developed for use in the random effects version of the model. Recently, two approaches have been proposed that might extend to the nest-survival models discussed above. Pan and Lin (2005) suggest statistics to test each fixed effect and the link function in generalized linear mixed models (GLMM) which, taken together, would address overall model fit. Studying their approach in this setting is worthy of future research. The approach we examine here is a single statistic designed to measure overall model fit outlined by Sturdivant (2005) and Sturdivant and Hosmer (in press). They extend a residual based goodness-of-fit statistic used in standard logistic models to the case of the hierarchical logistic model. This statistic is based on the unweighted sum of squares (USS) statistic proposed by Copas (1989) for the standard logistic-regression model.

In the random effects logistic model, the statistic uses kernel-smoothed residuals. These smoothed residuals are a weighted average of the residuals given by:

$$\hat{\mathbf{e}}_{\mathbf{s}} = \boldsymbol{\Lambda}\hat{\mathbf{e}},$$

where Λ is the matrix of smoothing weights:

$$\Lambda = \begin{bmatrix} \lambda_{11} & & \lambda_{1n} \\ & \ddots & \\ \lambda_{n1} & & \lambda_{nn} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\lambda}_1 \\ \\ \boldsymbol{\lambda}_n \end{bmatrix}.$$

The weights, λ_{ij}, produced using the kernel density are:

$$\lambda_{ij} = \frac{\mathbf{K}\left(\frac{\left|\hat{\pi}_1 - \hat{\pi}_j\right|}{h}\right)}{\sum_j \mathbf{K}\left(\frac{\left|\hat{\pi}_1 - \hat{\pi}_j\right|}{h}\right)} \quad (5)$$

where K(ξ) is the Kernel density function and h is the bandwidth.

Previous research has explored three kernel-density functions commonly used in studies of standard logistic-regression models, and all three densities produced acceptable results (Sturdivant 2005, Sturdivant and Hosmer, in press). The uniform density used in a study of a goodness-of-fit measure in standard logistic regression (le Cessie and van Houwelingen 1991) is defined as:

$$K(\xi) = \begin{cases} 1 \text{ if } |\xi| < 0.5 \\ 0 \text{ otherwise} \end{cases}$$

A second choice used in standard logistic studies involving smoothing in the y-space (Hosmer et. al. 1997, Fowlkes 1987) is the cubic kernel given by:

$$K(\xi) = \begin{cases} 1 \text{ if } |\xi|^3 \text{ if } |\xi| < 1 \\ 0 \text{ otherwise} \end{cases}$$

The final choice was the Gaussian kernel density (Wand and Jones 1995) defined:

$$K(\xi) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2}\xi^2\right)$$

The choice of kernel function is considered less critical than that of the bandwidth (Hardle 1990). The bandwidth, h, controls the number of observations weighted in the case of the uniform and cubic densities. The choice of bandwidth for the kernel-smoothed USS statistic is related, as well, to the number of subjects per cluster. Previous studies suggest a bandwidth weighting $0.5\sqrt{n}$ of the n residuals for relatively large clusters (>20 subjects) and weighting only $0.25\sqrt{n}$ for situations with smaller cluster sizes (Sturdivant 2005). For the Gaussian density, all observations are weighted. However, observations that are 2–3 SE outside of the mean effectively receive zero weight. The bandwidth then determines how many residuals are effectively given zero weight in the Gaussian case. Thus, the bandwidth choices for the Gaussian kernel place the selected number of observations within two standard deviations of the mean of the $N(0,1)$ density.

Regardless of the bandwidth criteria, a different bandwidth h_i is used for each $\hat{\pi}_i$ (Fowlkes 1987). The weights are then standardized so that they sum to one for each $\hat{\pi}_i$ by dividing by the total weights for the observation as shown in expression (5).

The goodness-of-fit statistic is then the USS statistic but using the smoothed rather than raw residuals:

$$\hat{S}_s = \sum_{i=1}^{n} \hat{e}_{si}^2 = \hat{\mathbf{e}}_\mathbf{s}{}' \hat{\mathbf{e}}_\mathbf{s}$$

The distribution of this statistic is extremely complicated due to the smoothing and the complexity of the hierarchical logistic model. Using an approach similar to that of Hosmer et. al. (1997), Sturdivant (2005) produced expressions to approximate the moments of the statistic. These moments are used to form a standardized statistic which, under the null hypothesis that the model is correctly specified, should have an asymptotic standard normal distribution:

$$Z_{\hat{S}} = \frac{\hat{S}_s - \mathrm{E}(\hat{S}_s)}{\sqrt{\mathrm{Var}(\hat{\mathrm{S}}_s)}} \qquad (6)$$

where:

$$\begin{aligned} E(\hat{S}_s) &= \mathbf{E[e'(I-\hat{M})'\Lambda'\Lambda(I-\hat{M})e} + \\ &\quad \mathbf{2\hat{g}'\Lambda'\Lambda(I-\hat{M})e + \hat{g}'\Lambda'\Lambda\hat{g}]} \\ &= \mathbf{trace\,[(I-\hat{M})'\Lambda'\Lambda(I-\hat{M})W]} + \\ &\quad \mathbf{\hat{g}'\Lambda'\Lambda\hat{g}} \end{aligned}$$

and:

$$\mathrm{Var}(\hat{S}_s) = \sum_{i=1}^{n}\left[a_{ii}^2 w_i(1-6w_i)\right] + 2\,\mathrm{trace}\,\mathbf{(A\hat{W}A\hat{W})} + \mathbf{b'\hat{W}b} + 2\sum_i a_{ii}b_i\pi_i(1-\pi_i)(1-2\pi_i)$$

In these expressions $\mathbf{A} = \mathbf{(I-\hat{M})'\Lambda'\Lambda(I-\hat{M})}$, $\mathbf{b'} = \mathbf{2\hat{g}'\Lambda'\Lambda(I-\hat{M})}$, $\mathbf{M} = \mathbf{WQ[Q'WQ + R]^{-1}Q'}$ and $\mathbf{g} = \mathbf{WQ[Q'WQ + R]^{-1}R\delta}$. Further, $\mathbf{Q} = [\mathbf{X}\ \ \mathbf{Z}]$ is the design matrix for both fixed and random effects, and

$$\hat{\boldsymbol{\delta}} = \begin{pmatrix} \hat{\boldsymbol{\beta}} \\ \hat{\boldsymbol{\mu}} \end{pmatrix}$$

the vector of estimated fixed and random effects. The other matrix in the expression involves the estimated random-parameter covariances and is defined:

$$\mathbf{R} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \hat{\boldsymbol{\Omega}}^{-1} \end{bmatrix}$$

While complicated, the matrix expressions are easily implemented in standard statistical software packages using output of the random effects estimation (Sturdivant et al. 2006; Appendix 1).

To test model fit, the moments are evaluated using the estimated quantities from the model where necessary in expression (6). The standardized statistic is compared to the standard normal distribution. A large (absolute) value leads to rejecting the null hypothesis and calls into question the correctness of the specified model.

The asymptotic distribution is complicated but expected to be standard normal under a central-limit-theorem argument. Previous simulations studies have shown that the distribution holds under the null distribution not just for large samples, but for smaller samples likely to occur in practice (to include small cluster sizes) (Sturdivant 2005, Sturdivant and Hosmer, in press).

Simulation Study Results

The proposed goodness-of-fit statistic was developed and tested in hierarchical logistic-regression models fit using penalized quasi-likelihood (PQL) estimation. Stephens (2003) implements nest-survival models in PROC NLMIXED (SAS Institute 2004) using programming statements within the procedure to perform iterative logistic regression for each day in an interval. Rotella et al. (2004) demonstrate the value of this approach as it accounts for the time-varying covariates, in essence performing a discrete-time survival analysis. In addition, the estimation uses the less biased Gaussian quadrature estimation approach (SAS NLMIXED procedure) rather than PQL estimation.

Before accepting the kernel-smoothed USS statistic for use in such models, we performed simulations to validate its use with the different estimation schemes and in models with time-varying covariates. Theoretically, a residual-based goodness-of-fit measure would not be affected by the form of the model or the estimation method. However, the complexity of the models and the statistic, particularly in the presence of random effects and clustering, leads to the need to validate the theory when using a different procedure.

We were interested in examining the rejection rates of the statistic in settings similar to those of nest survival data for which Rotella et al. (2004) propose using PROC NLMIXED (SAS Institute 2004). Previous extensive simulations using the GLMMIX macro have shown that the statistic rejects at the desired significance (Sturdivant 2005). Here, we wish to confirm that this continues in the new setting and estimation scheme.

The simulations involve models typical of those found in nest-survival studies. In particular, the simulated data included a standard continuous fixed effect as well as a time- varying fixed effect. For nest-survival models, a random intercept or, in some instances, a random slope for the time varying covariate may be deemed appropriate. Thus, we simulated both situations. In each case, we created 1,000 replicates using a data structure with 20 clusters (sites) each including 20 subjects (nests). The simulated time intervals between nest visits were 5–8 d (chosen at random from a uniform distribution). The kernel-smoothed statistic was calculated using the cubic kernel and a bandwidth weighting $0.5\sqrt{n}$ of the residuals (the choice of bandwidth is discussed in the section on the goodness-of-fit measure). In this case, with N = 400 subjects (20 clusters of 20 subjects), this bandwidth choice means that 10 observations were weighted to produce each smoothed residual value.

The estimated moments for the statistic from the two simulation runs approximate the observed moments of the simulated statistical values (Table 1). Further, the empirical rejection rates at the 0.01, 0.05, and 0.10 significance levels are similar. The 95% confidence regions for the rejection rates at these three significance levels are 0.6%, 1.4%, and 1.9%, respectively. Only in the case of the 0.01 significance level for the random-slope model is the observed rejection rate outside of this interval. In that instance, the statistic rejects slightly more often than expected.

The case where the statistic appears to reject slightly too often deserves several comments. First, when the results of the goodness-of-fit test indicate a lack of model fit, the analyst should be prompted to further investigate the data and model, and not necessarily reject the model outright. Therefore, the slightly higher than expected rejection rate merely results in periodically investigating model fit under circumstances when researchers might not ordinarily do so. Further, the actual number of models used in the simulation of the random-slope model was 470 (of the 1,000 replications)—the NLMIXED procedure failed to converge (a well documented issue with the Gaussian quadrature estimation scheme in practice and not related to the goodness of fit). Thus, it is possible that with more simulations the actual rejection rate would converge to a value within the confidence region. In fact, the 95% confidence interval with only 470 replications is wider (0.9%) so that the observed rejection rate is even less; for a very sensitive rejection rate (0.01).

We conclude that the simulation results reported here confirm earlier papers (Sturdivant 2005, Sturdivant and Hosmer, in press) and suggest that the change in estimation method and the inclusion of time-varying covariate does not hurt the performance of the kernel smoothed USS statistic.

Example

To illustrate the use of the statistic in a fitted model, we use data for Mallard (*Anas platyrhynchos*) nests monitored in 2000 in the Coteau region of North Dakota (Rotella et al. 2004). The data set we used contains 1,585 observations of 565 nests collected as part of a larger study. Rotella et al. (2004) analyzed the data using various techniques to account for the time varying covariates, in essence performing a discrete-time survival analysis. They estimated parameters for random effects models using Gaussian quadrature in PROC NLMIXED (SAS Institute 2004). We fit the same models and produced the kernel-smoothed USS statistic to measure overall model fit (Sturdivant et al. 2006; Appendix 1). The fixed effects of interest here include: nest age (1–35 d) and the proportion of grassland cover on the site containing the nest. The clusters or groups in this case are the 18 sites monitored during a 90-d nesting season.

The best random-effects model (Rotella et. al. 2004) included both nest age (treated as time varying) and proportion of grassland cover with a random intercept. With 18 nest sites (clusters) and 1,585 total observations, the bandwidth weighting more observations $(\frac{1}{2}\sqrt{1585})$ is

Table 1. Simulation study results using PROC NLMIXED with time varying covariates (N = 1,000 replications for random intercept and 470 for random slope) and cubic kernel USS statistic.

	Kernel-smoothed statistic		Moment estimates				
Model	Mean	SD [a]	EV [b]	SD [c]	Rejection rates [c]		
					0.01	0.05	0.10
Random Intercept	12.5	2.0	12.1.	1.8	0.013	0.047	0.105
Random Slope	2.3	0.8	2.3	0.7	0.021	0.045	0.085

[a] SD = standard deviation.
[b] EV = expected value.
[c] Significance levels.

preferred. Using this bandwidth and the cubic kernel, the calculated statistic and moments are as follows: $\hat{S}_s = 15.3$, $E(\hat{S}_s) = 13.8$, $Var(\hat{S}_s) = 1.8$, $Z_{\hat{s}} = 0.80$, and P-value = 0.423.

Comparing the statistic to the standard normal distribution, we fail to reject the hypothesis of model fit (P = 0.42). Thus, we can conclude that the overall model specification has no problems and that this model is reasonable in terms of goodness-of-fit. Clearly, other possible considerations are possible in fitting models (such as the best model). The goodness-of-fit statistic is useful as shown in this example when the model building exercise is complete and the analyst wishes to verify the appropriateness of the final model selected. Note that, if desired, the goodness-of-fit statistic can be used as one would any other such statistic. For example, one might use it to evaluate the fit of the global model—such a procedure is often recommended when using an information-theoretic approach to model selection, and especially when model-averaging is done, in which case there may not be a clear choice of the model for which fit should be evaluated (Burnham and Anderson 2002). As discussed by Burnham and Anderson (2004), goodness-of-fit theory about the selected best model is a subject that has been almost totally ignored in the model-selection literature. In particular, if the global model fits the data, does the selected model also fit? Burnham and Anderson (2004) explored this question and provide evidence that in the case of AIC-based model selection that the selected best model typically does fit if the global model fits. However, they also point out that results can vary with the information criterion used to select among models as well as other particulars of the study in question. The goodness-of-fit statistic provided here should prove useful to future development of goodness-of-fit theory with regards to nest-survival data.

DISCUSSION

Our results suggest that the kernel-smoothed USS statistic is a reasonable measure of overall model fit in random effects logistic-regression models involving time-varying covariates and using Gaussian quadrature for estimation. This work is an important extension demonstrating that the USS statistic is valid in settings beyond the PQL procedures used in its development. Further, no other available tools exist to assess overall model fit in models which offer great value to wildlife researchers modeling nest survival. This statistic is easily implemented in software packages and is currently available for use with PROC NLMIXED (SAS Institute 2004) as well as the GLMMIX macro (SAS Institute 2004).

The power of the USS statistic deserves further exploration (Sturdivant and Hosmer, in press); this statistic has reasonable power to detect issues of fixed-effect specification in the presence of random effects (Sturdivant and Hosmer, in press). However, exactly how much power and what sort of model misspecification is detected is an area of current research.

Goodness-of-fit measures are designed to warn of potential problems with the selected model. However in using our methods, if the model fit is rejected it is currently not clear what an analyst should do to address issues with the model. In practice, the analyst should re-examine the model and the data to identify reasons (such as outliers or inaccurate data) for why the null hypothesis of model fit was rejected. This exploration will often offer insights leading to a more appropriate model. The use of the statistic in studies which fit a variety of models will provide information regarding the causes of null hypothesis rejection, and allow researchers to develop methods for improving model fit.

ACKNOWLEDGMENTS

We thank D. W. Hosmer for his role in development in the methods presented here, the Northern Great Plains Office of Ducks Unlimited, Inc. for financial support, and W. L. Thompson for facilitating our application of the method to nest-survival data. We thank K. Gerow and G. White for helpful review comments on an earlier draft of this manuscript.

APPENDIX 1. SAS CODE FOR THE EXAMPLE DATA ANALYSIS USED IN THIS PAPER IS AVAILABLE (STURDIVANT ET AL. 2006).

```
* MACRO used to produce the USS kernel-smoothed statistic ;
%MACRO u1kern1 ;
PROC IML ;
  USE piest ;
    read all var {ifate} into yvec ;   * RESPONSE VARIABLE NAME HERE ;
    read all var {pred} into pihat ;
  CLOSE piest ;
  USE west ;
    read all var {pred} into wvec ;
  CLOSE west ;

  ehat = yvec-pihat ;
  what = diag(wvec) ;
  n = nrow(pihat) ;
  getwt =  ceil(0.5*sqrt(n))+1 ;   * SELECT THE BANDWIDTH HERE ;

*  KERNEL SMOOTH ROUTINE ;

  wtmat = J(n,n) ;
  rx=J(n,1) ;
  do i=1 to n ;
       x = abs(pihat[i] - pihat);
       rx[rank(x)]=x;
       bw = rx[getwt] ;
       if bw = 0 then do ;
              bw = 0.0000000000001 ;
       end ;
       wtmat[,i] = x / bw ;
  end ;
  * Get Kernel density values and weights;
    * UNIFORM (-a,a) ;
  ukern = t(wtmat<1) ;
  icolsum = 1/ukern[,+] ;
  uwt = ukern # icolsum ;

    * CUBIC  ;
  ctemp = 1 - (t(wtmat))##3 ;
  ckern = ukern # ctemp ;
  icolsum = 1/ckern[,+] ;
  cwt = ckern # icolsum ;

    * NORMAL ;
  nkern = pdf('norm',t(2*wtmat)) ;
  icolsum = 1/nkern[,+] ;
  nwt = nkern # icolsum ;

* MOMENTS and TEST STATISTICS;

  USE mall ;          * NAMES OF FIXED DESIGN MATRIX HERE and data set ;
    read all var{lv3 sage PctGr4} into x ;   * Note: here lv3 is all ones so
              used as int ;
    read all var {site} into groups ; * NAME OF LEVEL2 VARIABLE HERE ;
  CLOSE mall ;
  zmat = design(groups) ;
  Q = x||zmat ;

  USE betahat ;
    read all var {Estimate} into betahat ;
  CLOSE betahat ;
  USE Randeff ;
    read all var {estimate} into muhat ;
  CLOSE Randeff ;
```

```
  USE Sigmahat ;
    read all var {estimate} into cov2 ;
  CLOSE Sigmahat ;
  icov2 = 1/cov2 ;
  icov2d=diag(icov2) ;
  icov2a = BLOCK(icov2d,icov2d,icov2d) ;
  icovbl2 = BLOCK(icov2a,icov2a,icov2a,icov2a,icov2a,icov2a);
* BLOCKS SAME NUMBER AS GROUPS ;

  faketop = j(ncol(x),ncol(x)+ncol(zmat),0) ;
  fakeleft = j(ncol(zmat),ncol(x),0) ;
  comb1 = fakeleft||icovbl2 ;
  R = faketop//comb1 ;

  dhat = betahat//muhat ;

  * CREATE g vector and M matrix ;

  mymat = inv( t(Q)*what*Q + R ) ;

  g = what * Q * mymat * R * dhat ;
  M = what * Q * mymat * t(Q) ;

  * CALCULATE TEST STATISTICS;

  im = I(nrow(M))-M ;

  midunif = t(uwt)*uwt ;
  midcube = t(cwt)*cwt ;
  midnorm = t(nwt)*nwt ;

  aunif = t(im)*midunif*im ;
  acube = t(im)*midcube*im ;
  anorm = t(im)*midnorm*im ;

  bunif = 2*t(im)*midunif*g ;
  bcube = 2*t(im)*midcube*g ;
  bnorm = 2*t(im)*midnorm*g ;

  Tuni = t(ehat)*midunif*ehat ;
  Tc = t(ehat)*midcube*ehat ;
  Tn = t(ehat)*midnorm*ehat ;

  * CALCULATE EXPECTED VALUES ;
  eunif = trace( aunif*what) + t(g)*midunif*g ;
  ecube = trace( acube*what) + t(g)*midcube*g ;
  enorm = trace( anorm*what) + t(g)*midnorm*g ;

  * CALCULATE VARIANCE ;
  temp1 = wvec#(1-6*wvec) ;
  temp3 = pihat#(1-pihat)#(1-2*pihat) ;

  tempu = (vecdiag(aunif))##2 ;
  tempc = (vecdiag(acube))##2 ;
  tempn = (vecdiag(anorm))##2 ;

  v1unif = sum(tempu#temp1) ;
  v2unif = 2* trace(aunif*what*aunif*what) ;
  v3unif = t(bunif)*what*bunif ;
  v4unif = 2*sum( (vecdiag(aunif))#bunif#temp3 ) ;

  v1cube = sum(tempc#temp1) ;
  v2cube = 2* trace(acube*what*acube*what) ;
  v3cube = t(bcube)*what*bcube ;
  v4cube = 2*sum( (vecdiag(acube))#bcube#temp3 ) ;
```

```
   v1norm = sum(tempn#temp1) ;
   v2norm = 2* trace(anorm*what*anorm*what) ;
   v3norm = t(bnorm)*what*bnorm ;
   v4norm = 2*sum( (vecdiag(anorm))#bnorm#temp3 ) ;

   vunif = v1unif + v2unif + v3unif + v4unif ;
   vcube = v1cube + v2cube + v3cube + v4cube ;
   vnorm = v1norm + v2norm + v3norm + v4norm ;

   cubestat = (Tc-ecube)/sqrt(vcube) ;
   normstat = (Tn-enorm)/sqrt(vnorm) ;
   unifstat = (Tuni-eunif)/sqrt(vunif) ;

   punif = 2*(1-probnorm(abs(unifstat))) ;
   pcube = 2*(1-probnorm(abs(cubestat))) ;
   pnorm = 2*(1-probnorm(abs(normstat))) ;

   print Tc ecube vcube cubestat pcube ;
   print Tn enorm vnorm normstat pnorm ;
   print Tuni eunif vunif unifstat punif ;

quit ; run ;
%MEND ;

* NEST DATA ;

data Mall; set Nests.mall2000nd;
 LV3 =1 ;                              * ADD A COLUMN OF ONES FOR INTERCEPT ;
 if ifate=0 then ness+1;
 else if ifate=1 then ness+t;
/* create indicator variables for different nesting habitats */
    if hab=1 then NatGr=1; else NatGr=0;                    /* Native Grassland */
    if hab=2 or hab=3 or hab=9 then CRP=1; else CRP=0;     /* CRP & similar    */
    if hab=7 or hab=22 then Wetl=1; else Wetl=0;           /* Wetland sites    */
    if hab=20 then Road=1; else Road=0;                    /* Roadside sites   */
run;

Proc Sort data=Mall;
by site; run;

* FIT MODEL USING PROC NLMIXED;

PROC NLMIXED DATA=Mall tech=quanew method=gauss maxiter=1000;
parms B0=2.42, B2=0.019, B4=0.38, s2u=0.1;
  p=1;
     do i=0 TO t-1;
     if i=0 then Ob=1;
     else Ob=0;
          logit=(B0+u)+B2*(sage+i)+B4*PctGr4 ;
        p=p*(exp(logit)/(1+exp(logit)));
     end;
model ifate~binary(p);
random u~normal(0,s2u) subject=site out=randeff;
predict p*(1-p) out=west;
predict p out=piest ;
ods output ParameterEstimates=betahat
     (where=(Parameter=:"B")) ;
ods output ParameterEstimates=sigmahat
     (where=(Parameter=:"s2")) ;
ods output ParameterEstimates=B0Hat
     (where=(Parameter='B0') rename=(Estimate=Est_B0));
ods output ParameterEstimates=B1Hat
     (where=(Parameter='B1') rename=(Estimate=Est_B1));
ods output ParameterEstimates=B2Hat
```

```
      (where=(Parameter='B2') rename=(Estimate=Est_B2));
ods output ParameterEstimates=s2uhat
      (where=(Parameter='s2u') rename=(Estimate=Est_s2u));
run ;

%ulkern1 ;  * CALL KERNEL SMOOTHED STATISTIC MACRO ;
```