POWER TRANSFORMATIONS IN COMPONENTS OF VARIANCE MODELS FOR SMALL AREA ESTIMATION

Getachew Asfaw Dagne
Dept. of Epidemiology and Biostatistics, University of South Florida, 13201 Bruce B. Downs Blvd., MDC 56, Tampa, FL 33612

Key Words: Box-Cox transformations; components of variance; small area estimation; mixed-effects models.

1 INTRODUCTION

Small area estimation has recently received much attention in the literature due to growing demand for reliable small area estimation or prediction. Brackstone (1987) gives an interesting survey of supply and demand for small area estimation. Very recently, Ghosh and Rao (1994) gave an appraisal of several small area estimation methodologies, such as ratio-synthetic, sample size dependent, empirical Bayes and hierarchical Bayes estimation; they also discussed several examples of small area applications.

Several results in small area estimation follow from the assumption that the response variable, conditional on covariates and small area effects, is normally distributed with a common variance and additive error structure (see Kacker and Harville, 1984; Battese, Harter and Fuller, 1988; Dagne and Press, 1997). In situations where these assumptions are seriously violated, Box and Cox (1964) proposed a parametric power transformation technique in order to improve the agreement between the observations and the assumptions in a model.

Little or no investigation has been done on power transformations for components of variance models in the context of small area estimation. It is important to study whether the power transformations can contribute towards a more reliable prediction of small area means or totals of a variable of interest. The purpose of this study is, therefore, to investigate small area estimation problems for cases in which a linear mixed effects model holds after an unknown power transformation of the response variable has been identified.

2 Mixed-effects Model

Consider a large area which is divided into m smaller areas. Within the jth small area, there are N_j units, $j = 1, \ldots, m$. Assume there is a response variable, denoted by y_{kj}, that could be measured on the kth unit in the jth small area, along with other covariates pertinent to it, denoted by X_{kj}.

Thus, for m small areas, each with N_j units,

$$y_{kj}^{(\lambda)} = X_{kj}\beta + v_j + e_{kj}, \quad k = 1, \ldots, N_j, \qquad (1)$$

where,

$$y_{kj}^{(\lambda)} = \begin{cases} \frac{y_{kj}^{\lambda}-1}{\lambda}, & \text{if } \lambda \neq 0, \\ log(y_{kj}), & \text{if } \lambda = 0, \end{cases}$$

or using the more compact notation of matrix algebra,

$$Y^{(\lambda)} = X\beta + Zv + e,$$

where, $Y^{(\lambda)} = (Y_1^{(\lambda)'}, Y_2^{(\lambda)'}, \cdots, Y_m^{(\lambda)'})'$,
$X = (X_1', \cdots, X_m')'$, and
$Z = (Z_1', \cdots, Z_m')'$; $Y_j^{(\lambda)} = (Y_{1j}^{(\lambda)}, \cdots, Y_{N_jj}^{(\lambda)})'$, $X_j = (X_{1j}', \cdots, X_{N_jj}')'$, and $Z_j = (Z_{1j}', \cdots, Z_{N_jj}')'$; $v =$

$(v_1, \cdots, v_m)'$;e $= (e_1', e_2', e_N')'$; $Z_j = 1_{N_j} \bigotimes I_p$, $\beta = (\beta_1, \ldots, \beta_p)'$, a $(p \times 1)$ column vector of fixed effects parameters associated with the covariates; v_j = a random effect associated with the jth small area; 1_{N_j} is an N_j-dimensional column vector of unity.

The small area effect, v_j, and the error terms, e_{kj}, are assumed to be independently and normally distributed with zero means and variances σ_v^2 and σ_e^2, respectively. Letting $u_{kj} = v_j + e_{kj}$ and given the above assumptions, the covariance structure for the random variable u_{kj} is given by

$$cov(u_{kj}, u_{k'j'}) = \begin{cases} \sigma_v^2 + \sigma_e^2 & \text{if } k = k',\ j = j' \\ \sigma_v^2 & \text{if } k = k',\ j \neq j' \\ 0 & \text{otherwise.} \end{cases}$$

Based on the model (1) the jth small area mean, given the realized small area effect v_j, is given by $\mu_j = \bar{X}_j\beta + v_j$, assuming that N_j, the number of population units in the jth area is large, where $\bar{X}_j = \sum_k^{N_j} X_{kj}/N_j$ which are known for each area. Our objective is to predict μ_j, for $j = 1, \cdots, m$. In order to estimate μ_j, one has to obtain estimates of β and v_j; and also the components of variance, σ_v^2 and σ_e^2.

3 ESTIMATION

In this section the likelihood function of the sample, y, and estimators of the parameters of the model (1) will be discussed. Under the assumption that there exits some λ for which u_{kj} is approximately normally distributed with mean zero and variance $\sigma_v^2 + \sigma_e^2$, the likelihood function is given by

$$L(\beta, \lambda, \Sigma) \propto |\Sigma|^{-1/2} exp\{-1/2A\} \prod_j^m \prod_k^{n_j} y_{kj}^{\lambda-1}$$

whereg $A = (Y^{(\lambda)} - X\beta)'\Sigma^{-1}(Y^{(\lambda)} - X\beta)$ and $y_{kj}^{\lambda-1}$ is the Jacobian of the transformation from u to y.

A feasible predictor for the mean of the jth small area μ_j is given by

$$\hat{\mu}_j = \bar{X}_j\hat{\beta} + \hat{v}_j,$$

where, $\hat{\beta} = (X'\hat{\Sigma}^{-1}X)^{-1}X'\hat{\Sigma}^{-1}Y^{(\hat{\lambda})}$, $\bar{X}_j = \sum_k^{N_j} X_{kj}/N_j$, $\hat{v}_j = \hat{u}_j\hat{g}_j$, $\hat{g}_j = \frac{\hat{\sigma_v^2}}{\hat{\sigma_v^2} + \hat{\sigma_e^2}/n_j}$, $\hat{u}_j = \bar{y}_j^{(\hat{\lambda})} - \bar{X}_j\hat{\beta}$. where, $\hat{\lambda}, \hat{\Sigma}, \hat{\sigma_v^2}$ and $\hat{\sigma_e^2}$ are maximum likelihood estimators of $\lambda, \Sigma, \sigma_v^2$ and σ_e^2, respectively.

Inferences on small area means could also be carried out on the original scale of the dependent variable after the Box-Cox procedure has been performed. Thus, an approximate method of estimating means conditional on the realized small area effects, v, and covariates is given below.

$$\tilde{\mu} = (1 + \hat{\lambda}(X\hat{\beta} + \hat{v})^{1/\hat{\lambda}}[1 + \frac{(1 - \hat{\lambda})\hat{\sigma_e^2}}{2(1 + \hat{\lambda}(X\hat{\beta} + \hat{v})^2}].$$

4 Numerical Example

To illustrate the application of the proposed method we give an example of crop acreage estimation for counties in Iowa, which was originally analyzed by Battese, Harter and Fuller (1988) based on a nested-error regression model. Two sources of data for 12 counties in Iowa were used in the estimation process. One source was the 1978 June Enumerative Survey of the U.S. Department of Agriculture. The number of hectares of corn and soybeans in 37 segments for the 12 counties was obtained by interviewing farmers. The Landsat satellites provided another useful data source for county estimates of crop acreage. The number of pixels classified as corn and soybeans for each sampled segment and as well as the county mean number of pixels per segment classified as corn and soybeans were obtained. General information about the data is summarized in Table 1.

Table 1: Summary statistics on reported hectares

Crop	Min	Median	Mean	Max
Corn	64.75	116.4	120.3	206.4
Soybean	4.47	102.6	95.35	174.3

Transformation is likely to be helpful when the ratio of largest data value to the smallest data

value in a given data set is large, but may not be helpful when the ratio is less than 2 (Hoaglin, Mosteller and Tukey, 1983, p. 125). As can be seen in Table 1, the ratio of the maximum reported hectares under corn to the smallest reported hectares is 3.19 and the ratio for that of soybeans is 26.94. In both cases the ratios may be considered large which suggest the utility of a transformation.

Transforming the response variable in an attempt to obtain a better fit can be carried out by using the Box-Cox power transformations given in (1). Thus, in line with model (1), the reported hectares of a crop, y_{kj}, is modeled as

$$y_{kj}^{(\lambda)} = X_{kj}'\beta + v_j + e_{kj} \qquad (2)$$

where y_{kj} is the number of reported hectares of corn (or soybeans) in the kth unit within the jth county; $X_{kj} = (1, x_{1kj}, x_{2kj})$; x_{1kj} (x_{1kj}) is the number of pixels classified as corn (soybeans) in the kth unit within the jth county.

We refer to the model (2) as the Box-Cox model. Battese, Harter and Fuller (1988) employed a model similar to (2) except that the response variable, y_{kj}, was untransformed. We will compare the performance of the Box-Cox model with that of Batteseg Harter and Fuller's (BHF) model.

From Table 2 we can see that the ratio of maximum to minimum values of reported hectares of soybeans is 26.94, which is relatively large thereby warranting transformation. As a result there is much gain in terms of efficiency since the average relative efficiency of the Box-Cox estimates to the BHF estimates is 1.53. That is, the Box-Cox method is more efficient than the untransformed BHF's method. In the case of corn datag for instance, the ratio is 3.19, thus little gain was obtained by transforming. Actually, in this case estimating the extra parameter of transformation λ gives slightly less efficiency. Note that the numbers given in the third column of Table 2 are averages over the 12 counties.

Table 2: Relative Efficiency of the Box-Cox predictors to the BHF predictors

Crop	Ratio	Efficiency
Corn	3.19	0.78
Soybeans	26.94	1.53

In summary, since the proposed model is based on the Box-Cox family of power transformations, the advantage of this approach is its applicability to a larger class of problems where it is required to achieve normality of distributions, constancy of error variance and/or simplicity of the model structure. The results of this study indicate that the use of a power transformation of the response variable in components of variance model improves the quality of prediction of small area means. The proposed model provides a reasonable (if not perfect) fit to the data considered in this study. Therefore one should be comfortable to use the Box-Cox power transformations as the main analysis.

References

Battese, G.E., Harter, R.M. and Fuller, W.A. (1988). An error components model for prediction of county crop areas using survey and satellite data, *J. Amer. Stat. Assoc.*, 83, 28-36.

Box, G.E.P and Cox, D.R. (1964). An analysis of transformations, *J. Royal Statist. Society, Series B*, 26, 211-252.

Brackstone, G. J. (1987). Small area data: policy issues and technical Challenges, *Small Area Statistics*, (R. Platek, J.N.K. Rao, C.E. Sarndal and M.P. Singh, Eds.), Wiley, New York, 3-20.

Dagne, G. A. and Press, S.J. (1997). Bayesian prediction for small areas using SUR models *Commun. in Statistics.-Theory and Methods*, 26, 1355-1379.

Hoaglin, D.C., Mosteller, F. and Tukey, J.W. (1983). *Understanding robust and exploratory data analysis*, John Wiley, New York.

Ghosh, M. and Rao, J.N.K. (1994). Small area estimation: An appraisal, *Statistical Science*, 9, 55-93.

Kackar, R.N. and Harville, D.A. (1984). Approximations for standard errors of estimators of fixed and random effects in Mixed linear models, *J. Amer. Stat. Assoc.*, 79, 853-861.