

Robust model-based predictor of the finite population total

Yan Li^{1,2} and P. Lahiri¹

University of Maryland, College Park, MD 20742
National Cancer Institute, NIH, Rockville, MD 20851

Abstract

The prediction approach to finite population inference has received a considerable attention in recent years. Under this approach, the finite population is assumed to be a realization from a superpopulation described by a known probability model, usually a linear model. The prediction approach is often criticized for its lack of robustness against model misspecification. In this paper, we revisit this important issue and introduce a new robust prediction approach where the superpopulation model is chosen adaptively from the well-known Box-Cox class of probability distributions. The richness of the Box-Cox class ensures robustness in our model-based prediction approach. We demonstrate the robustness of our proposed predictors using a Monte Carlo simulation study and a real life example.

KEY WORDS: The Box-Cox transformation; Empirical best prediction; Superpopulation.

1. Introduction

The use of a superpopulation model to describe a finite population can be traced back at least to Cochran (1939). Brewer (1963) and Royall (1970) considered a prediction approach to estimate the finite population total, partly motivated by a superpopulation model. For a comprehensive review of the subject, see the books by Bolfarine and Zacks (1992), Korn and Graubard (1999) and Valliant et al. (2000), and the review paper by Graubard and Korn (2002). We refer to the book by Ghosh and Meeden (1997) for a related Bayes and empirical Bayes approach. Rao (2005) examined the interplay between sample survey theory and practice over the past 60 years or so.

Under the superpopulation prediction approach, the finite population is assumed to be a realization from a superpopulation described by a probability model. The superpopulation model is then used to predict the non-sampled units from the knowledge gained through the sample. The main criticism about this approach is that the prediction could be unreliable in case of a model misspecification (Hansen et al. 1983). Therefore, model robustness is important, a topic studied by researchers from different perspectives. Hartley and Rao (1968) proposed a “scale-load” nonparametric approach. Valliant (1985, 1986) extended Royall’s (1970) superpopulation approach to cover certain non-linear models. Ghosh and Lahiri (1987, 1988) derived linear empirical Bayes estimators of finite population means, replacing the normality assumption of Ghosh and Meeden (1986) by using certain moment

assumptions. Ghosh et al. (1989) put forward a nonparametric Bayesian approach using the Dirichlet process prior. Meeden (1999) proposed a noninformative Bayesian approach for two-stage cluster sampling. See Särndal et al. (1992) and Kott (2005) for a model-assisted approach.

In the mainstream statistics, transformations are often used to achieve normality, linearity, and homoscedasticity (Carroll and Ruppert, 1988), but the literature on transformations in finite population inference is not very rich. There is, however, a growing interest in developing methods that use an appropriate transformation with survey data. In some applications, especially in business and agricultural surveys, a linear model may not be appropriate for the study variable, but may be appropriate for a strictly monotonic transformation of the study variable. For the data set given in Royall and Cumberland (1981), Chen and Chen (1996) observed that the finite population distribution was severely skewed and that the log-transformation helped achieving symmetry. In addition, the log-transformation was useful in strengthening the linear relationship between the study and auxiliary variables. Korn and Graubard (1998) compared different confidence intervals, including intervals based on a logit-transformation, for proportions with small expected number of positive counts. Karlberg (2000) proposed an estimator based on a lognormal-logistic superpopulation model to predict the finite population total of a highly skewed survey variable. Their simulation results indicated that the lognormal-logistic model estimator offers a sensible alternative to other estimators, especially when the sample size is small. Chambers and Dorfman (2003) discussed the estimation of a finite population mean under certain general but known transformation on the continuous data.

Researchers find the transformation technique useful in analyzing survey data. However, the key step is the identification of an appropriate transformation that fits the survey data well. In many applications, the form of transformation is determined subjectively. However, a priori knowledge or theory may not suggest the transformation to be used. In such situations, it would be convenient to determine the transformation adaptively using the data.

The pioneering work of Box and Cox (1964) has led to the development of “data-decide-transformation” methods for constructing models with independently and identically distributed errors. Their paper and other papers on the subject, including Tukey (1957), John and Draper (1980), and Bickel and Doksum (1981), have inspired a large

volume of applied research. Spitzer (1976) examined the relationship between the demand for money and the liquidity trap with a generalized Box-Cox model. In the context of research related to malaria, Newman (1977) concluded that the Box-Cox functional specification was superior to earlier specifications. Miner (1982) and Davison et al. (1989) considered modeling of soybean yield functions and the U.S. soybean export respectively. They concluded that the Box-Cox transformation provides approximately normally distributed error terms. A bibliography of the published research related to the Box-Cox transformation can be found in a review paper by Sakia (1992).

In this paper, we use the Box-Cox transformation on the study variable to generate robust model-based predictors of a finite population total. Our approach deviates from the usual model-based approach that uses linear regression models with the normality assumption or a known transformation, such as the log-transformation or the square root transformation, on the study variable. The robustness is achieved because the appropriate transformation on the study variable is automatically determined by the data. The proposed research suggests a new way to achieve robustness in addressing various inferential issues in the prediction approach to the finite population theory.

We organize the paper in several sections. In Section 2, we propose our robust model-based approach to the finite population sampling. In Section 3, we evaluate our predictors using a real data analysis. Concluding remarks are provided in Section 4.

2. Robust model-based predictor

Let $U = \{1, \dots, N\}$ be a finite population of N distinct units, each of which has a value of a study variable y associated with it. The population vector of y 's, i.e. $\mathbf{y} = (y_1, \dots, y_N)'$, is treated as a realization of a random vector $\mathbf{Y} = (Y_1, \dots, Y_N)'$. Let S be the set of all samples of size n , a sample s being a subset of U . Our goal is to predict the finite population total: $T = \sum_{i \in U} y_i$. Let $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ik})'$ be

a column vector of k known auxiliary variables associated with the i^{th} unit of the finite population and $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_N)'$.

For any sample s of size n , we can redefine \mathbf{y} and \mathbf{X} so that the first n rows of \mathbf{y} and \mathbf{X} correspond to those in the sample. We write

$$\mathbf{y} = \begin{pmatrix} \mathbf{y}_s \\ \mathbf{y}_r \end{pmatrix}, \quad \mathbf{X} = \begin{pmatrix} \mathbf{X}_s \\ \mathbf{X}_r \end{pmatrix},$$

where \mathbf{y}_s is a $n \times 1$ column vector of observed response variable; \mathbf{y}_r is a $(N-n) \times 1$ column vector of unobserved response variable; \mathbf{X}_s is a $n \times (k+1)$ matrix of known

auxiliary variables in the sample; \mathbf{X}_r is a $(N-n) \times (k+1)$ matrix of known auxiliary variables outside the sample.

In the prediction approach, the following standard linear model is often assumed for the superpopulation:

$$\mathbf{M}_1 : \mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2 \mathbf{I})$, a N -variate probability distribution with the mean vector $\mathbf{0}$ and variance covariance matrix $\sigma^2 \mathbf{I}$; \mathbf{I} is the usual $N \times N$ identity matrix; $\boldsymbol{\beta}$ is the $(k+1) \times 1$ column vector of regression coefficients. Both σ^2 and $\boldsymbol{\beta}$ are unknown superpopulation parameters. A concise summary about the prediction theory in finite population sampling using linear models can be found in Bolfarine and Zacks (1992), Lohr (1999), Valliant et al. (2000), Chambers and Skinner (2003), among others.

Tukey (1957) considered the following family of power transformations:

$$y^{(\lambda)} = \begin{cases} y^\lambda & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases},$$

where $y > 0$. In order to take care of the discontinuity at $\lambda = 0$, Box and Cox (1964) proposed the following family of transformations:

$$y^{(\lambda)} = \begin{cases} (y^\lambda - 1) / \lambda & \lambda \neq 0 \\ \log(y) & \lambda = 0 \end{cases}, \quad (2.1)$$

where $y > 0$. The parameter λ determines the nature of transformation. For example, $\lambda = 1, 0, 0.5, -1$ correspond to no transformation, log-transformation, square root transformation, and reciprocal transformation, respectively. The transformation parameter λ is estimated by the data. The Box-Cox analysis may lead to a log-transformation, but may equally lead to some other transformation in the above family – it depends on the actual data observed.

We consider the following superpopulation model for the transformed study variable:

$$\mathbf{M}_2 : \mathbf{Y}^{(\lambda)} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \text{ where } \boldsymbol{\varepsilon} \sim (\mathbf{0}, \sigma^2 \mathbf{I}).$$

2.1 Estimation of $\boldsymbol{\theta} = (\boldsymbol{\beta}, \lambda, \sigma)$

For computational advantages, Box and Cox (1964) suggested the following scaled transformation:

$$y^{*(\lambda)} = \begin{cases} (y^\lambda - 1) / \lambda \tilde{y}^{\lambda-1} & \lambda \neq 0 \\ \tilde{y} \cdot \log(y) & \lambda = 0 \end{cases}, \quad (2.2)$$

where $\tilde{y} = \left(\prod_{i=1}^n y_i \right)^{1/n}$, the geometric mean of the sample observations. Consider the following scaled model:

$$\mathbf{M}_3 : \mathbf{Y}^{*(\lambda)} = \mathbf{X}\boldsymbol{\beta}^* + \boldsymbol{\varepsilon}^*,$$

where $\boldsymbol{\varepsilon}^*$ is approximately $N(0, \sigma^{*2} \mathbf{I})$. The scaling avoids large numbers and simplifies the log-likelihood function, and thus the model \mathbf{M}_3 has computational advantages over the model \mathbf{M}_2 in estimating the respective parameters.

Note that this scaling is different from the one used by Zarembka (1968, note 8) who suggested dividing $y^{(\lambda)}$ by \tilde{y}^λ . Schlesselman (1971) showed that the maximum likelihood estimator of $\theta^* = (\beta^*, \lambda, \sigma^{*2})'$ is scale invariant so that rescaling the original observations y 's leads to the same log-likelihood function under model M_3 so long as the regression model contains an intercept term.

Box and Cox (1964) discussed the estimation of $\theta^* = (\beta^*, \lambda, \sigma^{*2})'$. The method requires the maximization of the approximate log-likelihood function, given by

$$l(\theta^*) = \log L = -\frac{1}{2} \sum_s \log(2\pi\sigma^{*2}) - \frac{1}{2\sigma^{*2}} \sum_s (y_i^{*(\lambda)} - \mathbf{x}'_i \beta^*)^2.$$

The above log-likelihood function is approximate because the distribution of the error term in model M_3 is not exactly normal. The maximum likelihood estimate of λ can be obtained by a grid search method. That is, for a large set of values of λ , model M_3 can be fit. This is a simple task since model M_3 is a linear model for a given λ . The computations and plotting of the log-likelihood values for the fitted model against the set of values for λ locate the maximum likelihood estimate, $\hat{\lambda}$, of the transformation parameter λ . The maximum likelihood estimators of β^* and σ^{*2} are then given by:

$$\hat{\beta}^* = (\mathbf{X}_s' \mathbf{X}_s)^{-1} \mathbf{X}_s' \mathbf{Y}_s^{*(\hat{\lambda})} \text{ and}$$

$$\hat{\sigma}^{*2} = \frac{1}{n} (\mathbf{Y}_s^{*(\hat{\lambda})} - \mathbf{X}_s' \hat{\beta}^*)' (\mathbf{Y}_s^{*(\hat{\lambda})} - \mathbf{X}_s' \hat{\beta}^*).$$

Using $\hat{\theta}^* = (\hat{\beta}^*, \hat{\lambda}, \hat{\sigma}^{*2})'$, the maximum likelihood estimator of θ under model M_2 is obtained as:

$\hat{\theta} = (\hat{\beta}, \hat{\lambda}, \hat{\sigma})' = (\tilde{y}^{\hat{\lambda}-1} \hat{\beta}^*, \hat{\lambda}, \tilde{y}^{\hat{\lambda}-1} \hat{\sigma}^{*2})'$. It is interesting to note that the maximum likelihood estimates of λ with respect to M_2 and M_3 are equivalent (see Li and Lahiri 2006). The above algorithm, originally proposed by Box and Cox (1964), supplements the four different algorithms of obtaining the maximum likelihood estimator of θ under model M_2 considered in Spitzer (1982a, b). Bickel and Doksum (1981) provided precise conditions under which the maximum likelihood estimator of θ is consistent. Gurka (2004, 2006) estimated the model and transformation parameters for the linear mixed model using the residual maximum likelihood (REML) approach in order to obtain more accurate estimate for theta.

2.2 Estimating the asymptotic variance-covariance matrix of $\hat{\theta}$

Note that for known λ , M_3 is simply the standard linear regression model and so one can suggest standard variance estimators for $\hat{\beta}^*$ and $\hat{\sigma}^{*2}$. However, these variance estimators underestimate the true uncertainties of

$\hat{\beta}^*$ and $\hat{\sigma}^{*2}$ since such variance estimators treat $\hat{\lambda}$ as the true value. See Bickel and Doksum (1981) and Hinkley and Runger (1984) for related discussions. In the context of estimation of $(\beta, \lambda)'$, Spitzer (1982a) incorporated the additional uncertainty due to estimation of λ by considering the inverse of the observed Fisher's information matrix. However, his model is different from the Box-Cox scaled model. Unlike Spitzer (1982a), we incorporate the additional uncertainties involving $\hat{\sigma}^{*2}$. Let $i(\hat{\theta}^*) = -(\partial^2 l(\theta^*) / \partial(\theta^*)^2)_{\hat{\theta}^*}$ be the observed Fisher information matrix, $(\partial^2 l(\theta^*) / \partial(\theta^*)^2)_{\hat{\theta}^*}$ being the $(k+3) \times (k+3)$ matrix of second partial derivatives of $l(\theta^*)$ with respect to θ^* . Then a standard consistent estimator of the asymptotic variance-covariance matrix of $\hat{\theta}^*$ is given by: $var(\hat{\theta}^*) = i(\hat{\theta}^*)^{-1}$. Explicit formulae for the derivatives are given in Li and Lahiri (2006).

Our goal is to obtain an estimate of the asymptotic variance-covariance matrix of $\hat{\theta} = (\hat{\beta}, \hat{\lambda}, \hat{\sigma})'$. Using the relationship $\hat{\beta} = \tilde{y}^{\hat{\lambda}-1} \hat{\beta}^*$, $\hat{\sigma} = \tilde{y}^{\hat{\lambda}-1} \hat{\sigma}^{*2}$ and the fact that $\hat{\theta}$ is the maximum likelihood estimate of $\theta = (\beta, \lambda, \sigma)'$ under model M_2 , we have: $var(\hat{\theta}) = \mathbf{J} var(\hat{\theta}^*) \mathbf{J}'$, where

$$\mathbf{J} = \begin{bmatrix} \frac{\partial}{\partial \hat{\beta}^*} \hat{\beta} & \frac{\partial}{\partial \hat{\lambda}} \hat{\beta} & \frac{\partial}{\partial \hat{\sigma}^{*2}} \hat{\beta} \\ \frac{\partial}{\partial \hat{\beta}^*} \hat{\lambda} & \frac{\partial}{\partial \hat{\lambda}} \hat{\lambda} & \frac{\partial}{\partial \hat{\sigma}^{*2}} \hat{\lambda} \\ \frac{\partial}{\partial \hat{\beta}^*} \hat{\sigma} & \frac{\partial}{\partial \hat{\lambda}} \hat{\sigma} & \frac{\partial}{\partial \hat{\sigma}^{*2}} \hat{\sigma} \end{bmatrix}.$$

Following Spitzer (1982a), one could have applied the Taylor series to obtain $var(\hat{\theta})$ from $var(\hat{\theta}^*)$. However, we note that such an argument is hard to justify since \mathbf{J} is a random matrix. Instead, we applied a direct method to obtain $var(\hat{\theta})$ (derivation is available upon request).

2.3 Prediction of the finite population total

We predict the finite population total by $\hat{T} = \sum_{i \in S} y_i + \sum_{i \in F} \hat{y}_i$, where \hat{y}_i denotes an arbitrary predicted value of the unobserved y_i . The difference $\hat{T} - T$ is known as the prediction error of the predictor \hat{T} .

Let $B(\hat{T} - T) = E(\hat{T} - T)$, $Var(\hat{T} - T) = E(\hat{T} - E(T))^2$ and $MSPE(\hat{T} - T) = E(\hat{T} - T)^2$ denote the bias, variance and mean squared prediction error (MSPE) of the prediction error respectively, where all the expectations are taken with respect to the model M_2 . Note that $MSPE(\hat{T} - T) = Var(\hat{T} - T) + B^2(\hat{T} - T)$. For an unbiased

predictor, i.e., for a predictor with $B(\hat{T}-T)=0$, we have $MSPE(\hat{T}-T) = Var(\hat{T}-T)$.

The best predictor (BP) of T , i.e., the predictor which minimizes the MSPE is obtained when $\hat{y}_i = \hat{y}_i^{BP}(\theta)$,

where $\hat{y}_i^{BP}(\theta) = E(y_i) = \int_{-\infty}^{\infty} [\lambda(\mathbf{x}_i'\beta + \sigma z) + 1]^{\frac{1}{\lambda}} \phi(z) dz$, and $\phi(z)$ is the density of the standard normal deviate. We can evaluate the above integral by numerical integration or by the following Monte Carlo approximation:

$$\hat{y}_i^{BP}(\theta) \approx \frac{1}{M} \sum_{j=1}^M (\lambda(\mathbf{x}_i'\beta + \varepsilon_{ij}) + 1)^{\frac{1}{\lambda}}, \tag{2.3}$$

where M denotes the number of independent simulation runs and $\varepsilon_{ij} \sim N(0, \sigma^2)$.

In practice, θ is unknown. Replacing θ by $\hat{\theta}$ in $\hat{y}_i^{BP}(\theta)$, we obtain an empirical best predictor (EBP) of T . In EBP, we use $\hat{y}_i = \hat{y}_i^{EBP}(\hat{\theta})$, where

$$\hat{y}_i^{EBP}(\hat{\theta}) = \int_{-\infty}^{\infty} [\hat{\lambda}(\mathbf{x}_i'\hat{\beta} + \hat{\sigma} z) + 1]^{\frac{1}{\hat{\lambda}}} \phi(z) dz \approx \frac{1}{M} \sum_{j=1}^M (\hat{\lambda}(\mathbf{x}_i'\hat{\beta} + \hat{\varepsilon}_{ij}) + 1)^{\frac{1}{\hat{\lambda}}} \tag{2.4}$$

and $\hat{\varepsilon}_{ij} \sim iid N(0, \hat{\sigma}^2)$.

When σ^2 is small, we can approximate $\hat{y}_i^{BP}(\theta)$ by the Taylor Series expansion and obtain the approximate best predictor (ABP) of T when $\hat{y}_i = \hat{y}_i^{ABP}(\theta)$ with

$$\hat{y}_i^{ABP}(\theta) = (\lambda(\mathbf{x}_i'\beta) + 1)^{1/\lambda} \approx \hat{y}_i^{BP}(\theta). \tag{2.5}$$

Replacing θ by $\hat{\theta}$ in $\hat{y}_i^{ABP}(\theta)$, we obtain the following approximated empirical best predictor (AEBP) of T when $\hat{y}_i = \hat{y}_i^{AEBP}(\hat{\theta})$, where

$$\hat{y}_i^{AEBP}(\hat{\theta}) = (\hat{\lambda}(\mathbf{x}_i'\hat{\beta}) + 1)^{1/\hat{\lambda}} \tag{2.6}$$

The ABP or AEBP is easier to implement than BP or EBP in terms of the CPU time.

2.4 Estimation of the prediction variance of the population total predictor

First note that for each $i \in r$, any arbitrary predictor \hat{y}_i is a function of y_i , $i \in s$ and hence independent of all y_i , $i \in r$ under model M_2 . Thus, using the fact that $\hat{T} - T = \sum_{i \in r} \hat{y}_i - \sum_{i \in r} y_i$, we have

$$Var(\hat{T} - T) = Var(\sum_{i \in r} y_i) + Var(\sum_{i \in r} \hat{y}_i). \tag{2.7}$$

If $\hat{y}_i = \hat{y}_i^{BP}(\theta)$ or $\hat{y}_i^{ABP}(\theta)$, the second term of right side of (2.7) is zero since \hat{y}_i is non-stochastic. Thus, for both BP and ABP we have the identical prediction variance given by

$$Var(\hat{T} - T) = \sum_{i \in r} Var(y_i) = \sum_{i \in r} \left[\int_{-\infty}^{\infty} [\lambda(\mathbf{x}_i'\beta + \sigma z) + 1]^{\frac{2}{\lambda}} \phi(z) dz - E^2(y_i) \right].$$

The BP and ABP differ in terms of their prediction biases and the MSPE. Evidently, for the BP the prediction bias is zero and thus the MSPE is the same as the prediction variance. On the other hand, ABP suffers from prediction bias, but as noted in Section 2.3, for small σ^2 , the bias is negligible.

If $\hat{y}_i = \hat{y}_i^{EBP}(\hat{\theta})$ or $\hat{y}_i^{AEBP}(\hat{\theta})$, the prediction variance of \hat{T} is given by

$$Var(\hat{T} - T) = Var(\sum_{i \in r} y_i) + Var\left[\sum_{i \in r} \hat{y}_i(\hat{\theta})\right], \tag{2.8}$$

where $Var(\sum_{i \in r} \hat{y}_i(\hat{\theta})) = \sum_{i \in r} \sum_{j \in r} Cov(\hat{y}_i(\hat{\theta}), \hat{y}_j(\hat{\theta}))$. Note the second term of (2.8), $Var(\sum_{i \in r} \hat{y}_i(\hat{\theta}))$, captures the variability

due to the estimation of $\hat{\theta}$, which can be ignored for large sample. Using the Taylor Series expansion argument, we can propose the following variance estimator:

$$var(\sum_{i \in r} \hat{y}_i(\hat{\theta})) \approx \left(\sum_{i \in r} \frac{\partial}{\partial \hat{\theta}} \hat{y}_i(\hat{\theta}) \right) var(\hat{\theta}) \left(\sum_{j \in r} \frac{\partial}{\partial \hat{\theta}} \hat{y}_j(\hat{\theta}) \right).$$

For $\hat{y}_i(\hat{\theta}) = \hat{y}_i^{AEBP}(\hat{\theta})$, we have

$$\begin{aligned} \frac{\partial}{\partial \hat{\lambda}} \hat{y}_i^{AEBP}(\hat{\theta}) &= \hat{\lambda}^{-1} (\mathbf{x}_i'\hat{\beta}) (\hat{\lambda} \mathbf{x}_i'\hat{\beta} + 1)^{1/\hat{\lambda}-1} - \\ &\quad \hat{\lambda}^{-2} (\hat{\lambda} \mathbf{x}_i'\hat{\beta} + 1)^{1/\hat{\lambda}} \log(\hat{\lambda} \mathbf{x}_i'\hat{\beta} + 1) \\ \frac{\partial}{\partial \hat{\beta}} \hat{y}_i^{AEBP}(\hat{\theta}) &= [\hat{\lambda}(\mathbf{x}_i'\hat{\beta}) + 1]^{1/\hat{\lambda}-1} \mathbf{x}_i. \end{aligned}$$

For $\hat{y}_i(\hat{\theta}) = \hat{y}_i^{EBP}(\hat{\theta})$, we have

$$\begin{aligned} \frac{\partial}{\partial \hat{\lambda}} \hat{y}_i^{EBP}(\hat{\theta}) &= \int_{-\infty}^{\infty} \hat{\lambda}^{-2} \hat{w}_i^{1/\hat{\lambda}} (1 - \hat{w}_i^{-1} - \log \hat{w}_i^{-1}) \phi(z) dz \\ \frac{\partial}{\partial \hat{\beta}} \hat{y}_i^{EBP}(\hat{\theta}) &= \int_{-\infty}^{\infty} \hat{w}_i^{1/\hat{\lambda}-1} \mathbf{x}_i \phi(z) dz \\ \frac{\partial}{\partial \hat{\sigma}} \hat{y}_i^{EBP}(\hat{\theta}) &= \int_{-\infty}^{\infty} \hat{w}_i^{1/\hat{\lambda}-1} z \phi(z) dz \end{aligned}$$

where

$$\hat{w}_i = \hat{\lambda}(\mathbf{x}_i'\hat{\beta} + \hat{\sigma}z) + 1.$$

2.5 The Box-Cox Transformation on Both the Study and Auxiliary Variables

Box and Cox (1964) considered a general model that allows transformation on both the study and auxiliary variables. Khan and Ross (1977), Spitzer (1976), Zarembka (1968), Boylan et al. (1980), Mills (1978), among others, considered a particular case of the general Box-Cox model when a common transformation parameter is assumed for the study and all auxiliary variables. However, this approach restricts the class of models. Gemmill (1980) argued that more robust analysis is

possible by using the general Box-Cox model. Boylan et al. (1982) applied a fully generalized Box-Cox transformation to import functions, and found a substantial effect on the parameter estimates. Ozuna et al. (1993) developed a truncated generalized Box-Cox model and compared their model with truncated double-log and semi-log model using the data from three Texas coastal recreation sites. Their empirical results showed that neither the truncated double-log nor the truncated semi-log model is appropriate for their application.

In Sections 2.1-2.4, for the sake of notational simplicity we develop methods under the model with transformation only on the study variable y . But since the transformation on the auxiliary variables does not affect the distribution of the study variable, the extension of Sections 2.1-2.4 to the general Box-Cox model with transformations on both the study and auxiliary variables is straightforward. The theoretical framework under the general Box-Cox model differs from the one based on the transformation only on the study variable in that the derivatives of the log-likelihood function with respect to the transformation parameters on different auxiliary variables are also needed. See Li and Lahiri (2006) for details on the derivative computations. In Section 3, we implement the general Box-Cox model via a real data analysis.

3. Real data analysis

In this section, we treat the actual survey data from the AAGIS as an artificial finite population of $N=431$ farms. For each farm, we have information on the number of beef cattles (study variable, y) and the farm area (auxiliary variable, x). In Figure 1, we plot the histograms of y and $y^{(\hat{\lambda}_1)}$, where $\hat{\lambda}_1$ is the maximum likelihood estimate of λ_1 , the transformation parameter on the study variable in the Box-Cox model. It is clear that the distribution of y is highly skewed and that the Box-Cox transformation is useful in achieving nearly normal distribution. In Figure 2, we display the scatter plots of y vs. x and $y^{(\hat{\lambda}_1)}$ vs. $x^{(\hat{\lambda}_2)}$, where $\hat{\lambda}_2$ is the maximum likelihood estimate of λ_2 . The Box-Cox transformation is exhibiting a better linear fit. The adjusted R^2 for the Box-Cox transformed data is .74 compared to .45 for the original data.

The benefit of taking Box-Cox transformation on both y and x is obvious. Note that no transformation and the log-transformation belong to the class of Box-Cox transformations when $\lambda = 1$ and 0 respectively. Figure 3 plots the histograms of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ using 1,000 bootstrap samples, each bootstrap sample (of size 431) being selected using a simple random sampling with replacement from the finite population. The variability of $\hat{\lambda}_1$ and $\hat{\lambda}_2$ over different samples supports the adaptive transformation approach over the fixed subjectively chosen transformation approach.

Next we study the predictive power of different models by the well-known cross-validation method in which we

drop one unit at a time and using the remaining units we predict the unit deleted. For details on the cross-validation approach, we refer the interested readers to Efron and Tibshirani (1993). We compare four predictors (NTP, LTP, AEBP and EBP) of the total number of beef cattles in $N = 431$ farms based on the three different models: no transformation model M_1 , log-transformation model, and the Box-Cox transformation model M_2 . We estimate the prediction variance $Var(\hat{T} - T)$, construct 95% asymptotic CI for T , and calculate the length of confidence interval (I) for each of the 431 possible cross-validation samples from the finite population.

In order to evaluate the four predictors, the following evaluation statistics for each predictor are calculated:

- Average Absolute Relative Deviation

$$(AARD) = \frac{1}{431} \sum_{i=1}^{431} \left| \frac{\hat{T}_i - T}{T} \right|,$$

- Average Relative Deviation

$$(ARD) = \frac{1}{431} \sum_{i=1}^{431} \frac{\hat{T}_i - T}{T},$$

- Average length of 95% confidence interval

$$(ALCI) = \frac{1}{431} \sum_{i=1}^{431} I_i, \text{ and}$$

- Proportion of times the true value included in the 95% confidence interval

$$(P) = \frac{1}{431} \sum_{i=1}^{431} I\{T \in CI_i\},$$

where the subscript i denotes the i^{th} sample selected from the beef population, $I\{\}$ is an indicator function which is equal to one if the true value T is included in the CI, and zero otherwise.

Table 1 reports the $AARD$, ARD , $ALCI$, and P for four predictors based on different models. Two predictors based on the Box-Cox model achieve the smallest $AARD$ (.0011), and ARD of EBP is zero. EBP and AEBP give the shorter length of confidence interval, compared to the other two predictors. The coverage rates for all four predictors are about 95%, very close to the nominal 95% confidence interval.

4. Concluding Remarks

Unlike the model-based approaches considered previously in the literature, in this paper we consider a new adaptive approach where the model is automatically determined by the survey data. Thus our method advances the model-based prediction approach to finite population sampling. We propose an AUTOGREG estimator that is doubly robust in the sense that it has the desirable design-consistency property, irrespective of the model that is used to generate the estimator, in addition to the robustness property already achieved by the use of the Box-Cox model. Our AUTOGREG can be viewed as a robust version of the popular GREG estimator that uses a linear model. Our numerical results suggest that our approach utilizes

available auxiliary variables in an efficient manner and offers a potential attractive alternative to the relatively more expensive design-based methods that require more samples to achieve the same level of precision.

Acknowledgements

The first author is grateful to her Ph.D. supervisory committee members Professors Katherine Abraham, Wolfgang Jank, Fritz Scheuren, Paul Smith and Mr. Paul D. Williams for their generous guidance throughout her dissertation research.

References:

- Bickel, P. J. and Doksum, K. A. (1981), An Analysis of Transformations Revisited, *Journal of the American Statistical Association*, **76**, 296-311.
- Bolfarine, H. and Zacks, S. (1992), *Prediction Theory for Finite Populations*, Springer: Verlag.
- Box, G. E. and Cox, D. R. (1964), An Analysis of Transformations, *Journal of the Royal Statistical Society, Series B* **26**, 211-52.
- Boylan, T. A., Cuddy, M. P., and O’Muircheartaigh, I. G. (1980), The Functional Form of the Aggregate Import Demand Equation: a Comparison of Three European Economies, *Journal of international economics*, **10**, 561-566.
- Boylan, T. A., Cuddy, M. P., and O’Muircheartaigh, I. G. (1982), Import Demand Equations: An Application of A Generalized Box-Cox Methodology, *International statistical review*, **50**, 103-112.
- Brewer, K. R. W. (1963), Ratio Estimation and Finite Populations: Some Results Deducible from the Assumption of an Underlying Stochastic Process, *Australian Journal of Statistics*, **5**, 93-105.
- Carroll, R. J. and Ruppert, D. (1988), *Transformations and Weighting in Regression*, London: Chapman and Hall.
- Chambers, R. L. and Dorfman, A. H. (2003), Transformed Variables in Survey Sampling, Technical Report.
- Chambers, R. L. and Skinner, C. J. (2003), *Analysis of Survey Data*, Wiley: Chichester.
- Chen, G. and Chen, J. (1996), A Transformation Method for Finite Population Sampling Calibrated with Empirical Likelihood, *Survey Methodology*, **22**, 139-46.
- Cochran, W. G. (1939), The Use of Analysis of Variance in Enumeration by Sampling, *Journal of the American Statistical Association*, **34**, 492-510.
- Davison, C. W., Arnade, C. A. and Hallahan, C. B. (1989), Box-Cox Estimation of U.S. Soyabean Exports, *Journal of Agricultural Economics Research*, **41**, 8-15.
- Efron, B. and Tibshirani, R. J. (1993), *An Introduction to the Bootstrap*, Chapman & Hall.
- Gemmil, G. (1980), Using the Box-Cox Form for Estimating Demand: a Comment, *The review of economics and statistics*, **62**, 147-148.
- Ghosh, M. and Lahiri, P. (1987), Robust Empirical Bayes Estimation of Means from Stratified Samples, *Journal of the American Statistical Association*, **82**, 1153-62.
- Ghosh, M. and Lahiri, P. (1988), Bayes and Empirical Bayes Analysis in Multistage Sampling, *Statistical Decision Theory and Related Topics IV*, **1**, 195-212.
- Ghosh, M., Lahiri, P. and Tiwari, R.C. (1989), Nonparametric Bayes and Empirical Bayes Estimation of the Distribution Function and the Mean, *Communications in Statistics: Theory and Methods*, **18**, 121-146.
- Ghosh, M., and Meeden, G. (1986), Empirical Bayes Estimation in Finite Population Sampling, *Journal of the American Statistical Association*, **81**, 1058-1062.
- Ghosh, M. and Meeden, G. (1997), *Bayesian Methods for Finite Population Sampling*, Chapman and Hall, London.
- Graubard, B. and Korn, E. (2002), Inference for Superpopulation Parameters using Sample Surveys, *Statistical Science*, **17**, 73-96.
- Gurka, M. (2004), The Box-Cox Transformation in the General Linear Mixed Model for Longitudinal Data, Ph.D. Dissertation.
- Gurka, M. (2006), Expanding the Box-Cox transformation to the linear mixed model, *Journal of the Royal Statistical Society, Series A*, **169**, 273-88.
- Hansen, M. H., Madow, W. G. and Tepping, B. J. (1983), An Evaluation of Model-Dependent and Probability Sampling Inferences in Sample Surveys, *Journal of the American Statistical Association*, **78**, 776-793.
- Hartley, H.O. and Rao, J.N.K. (1968), A New Estimation Theory for Sample Surveys, *Biometrika*, **55**, 547-557.
- Hinkley, D. V. and Runger, G. (1984), The Analysis of Transformed Data (with discussion). *Journal of the American Statistical Association*, **79**, 302-20.
- John, J. A. and Draper, N. R. (1980), An Alternative Family of Transformations, *Applied Statistics*, **29**, 190-97.
- Karlberg, F. (2000), Survey Estimation for Highly Skewed Population in the Presence of Zeroes, *Journal of Official Statistics*, **16**, 229-41.
- Khan, M. S. and Ross, K. Z. (1977), The Functional Form of the Aggregate Import Demand Equation, *Journal of international economics*, **7**, 149-160.
- Korn, E. and Graubard, B. (1998), Confidence Intervals for Proportions with Small Expected Number of Positive Counts Estimated from Survey Data, *Survey Methodology*, **24**, 193-201.
- Korn, E. and Graubard, B. (1999), *Analysis of Health Surveys*, New York: John Wiley & Sons, Inc.
- Kott, P.S. (2005), Randomized-Assisted Model-Based Survey Sampling, *Journal of Statistical Planning and Inference*, **129**, 263-277.
- Li, Y. and Lahiri, P. (2006), Robust Model-Based and Model-Assisted Predictors of the Finite Population Mean, Unpublished Manuscript.

- Lohr, S. L. (1999), *Sampling: Design and Analysis*, Pacific Grove, California: Duxbury.
- Meeden, G. (1999), A Noninformative Bayesian Approach for Two-Stage Cluster Sampling, *Sankhya, Series B*, **61**, 133-144.
- Miner, A. G. (1982), The Contribution of Weather and Technology to U.S. Soybean Yield, *Unpublished Dissertation*, University of Minnesota.
- Mills, T.C. (1978), Estimation of Parameters in the Case of Unconstrained Box-Cox Transformations, Submitted to *J.R. Statist. Soc. C*.
- Newman, P. (1977), Malaria and Mortality, *Journal of the American Statistical Association*, **72**, 257-63.
- Ozuna T., Jones, L. L., and Capps, O. (1993), Functional Form and Welfare Measures in Truncated Recreation Demand Models, *American Journal of agricultural economics*, **75**, 1030-1035.
- Rao, J. N. K. (2005), Interplay Between Sample Survey Theory and Practice: An Appraisal, *Survey Methodology*, **31**, 117-138.
- Royall, R. M. (1970), On Finite Population Sampling Under Certain Linear Regression Models, *Biometrika*, **57**, 377-87.
- Royall, R. M. and Cumberland, W. G. (1981), The Finite Population Linear Regression Estimator and Estimators of its Variance – An Empirical Study, *Journal of the American Statistical Association*, **76**, 924-30.
- Sakia, R. M. (1992), The Box-Cox Transformation Technique: a review, *The Statistician*, **41**, 169-78.
- Särndal, C. E., Swensson, B., and Wretman, J. (1992), *Model Assisted Survey Sampling*, New York: Springer-Verlag.
- Schlesselman, J. (1971), Power Families: A Note on the Box and Cox Transformation, *Journal of the Royal Statistical Society, Series B* **33**, 307-11.
- Spitzer, J. J. (1976), The Demand for Money, the Liquidity Trap and Functional Forms, *The International Economics Review*, **17**, 220-27.
- Spitzer, J. J. (1982a), A Primer on Box-Cox Estimation, *Review of the Economics and Statistics*, **64**, 307-13.
- Spitzer, J. J. (1982b), A Fast and Efficient Algorithm for the Estimation of Parameters in Models with the Box and Cox Transformation, *Journal of the American Statistical Association*, **77**, 760-66.
- Tukey, J. W. (1957), The Comparative Anatomy of Transformation, *Annals of Mathematical Statistics*, **28**, 601-32.
- Valliant, R. (1985), Nonlinear Prediction Theory and the Estimation of Proportions in a Finite Population, *Journal of the American Statistical Association*, **80**, 631-41.
- Valliant, R. (1986), Mean Squared Error Estimation Finite Populations under Nonlinear Models, *Communications in Statistics A*, **15**, 1975-93.
- Valliant, R., Dorfman, A. H. and Royall, R. M. (2000), *Finite Population Sampling and Inference: a Prediction Approach*, New York: Wiley.
- Zarembka, P. (1968), Functional Form in the Demand for Money, *Journal of the American Statistical Association*, **63**, 502-11.

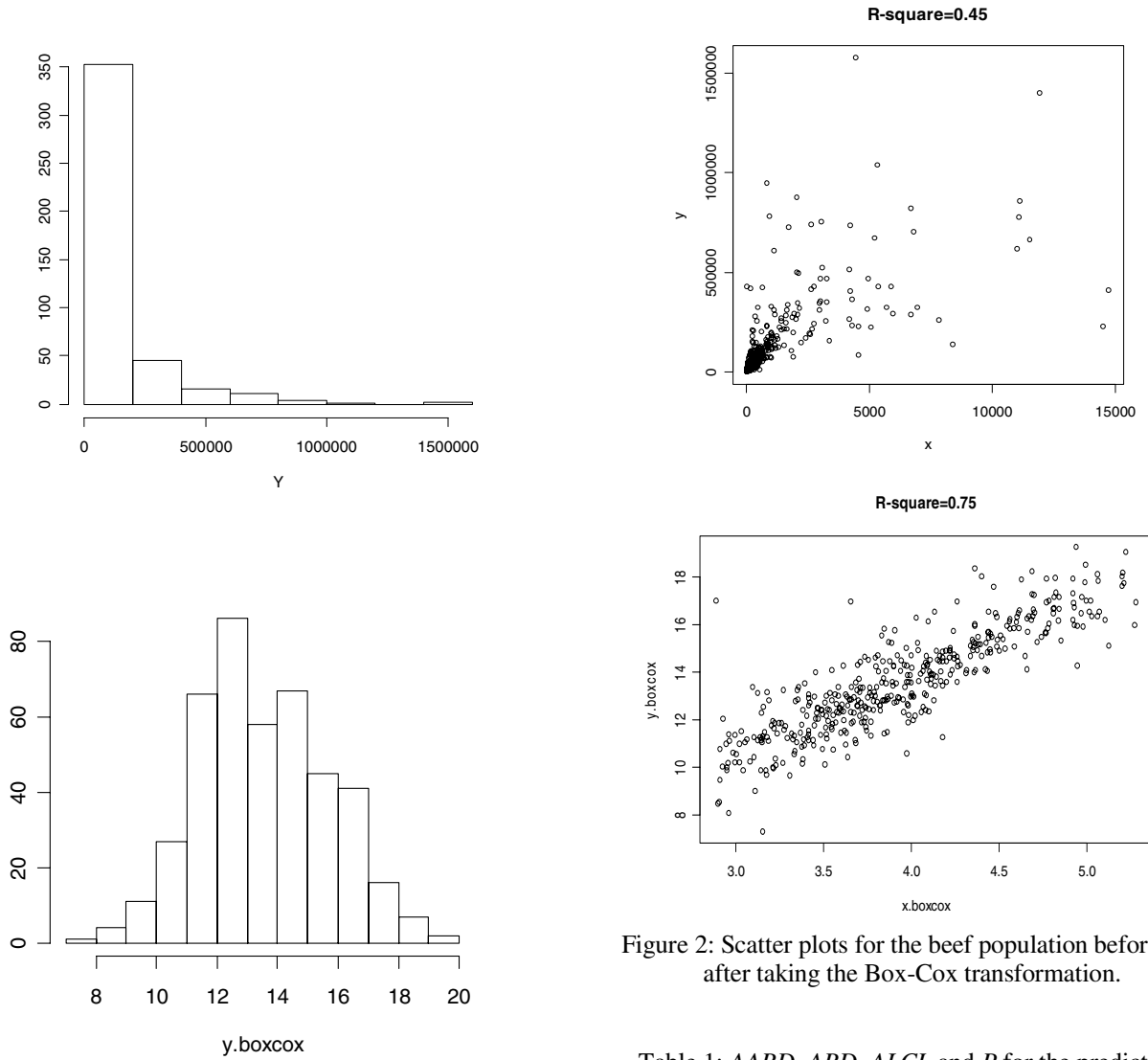


Figure 1: Histograms for the beef population before and after taking the Box-Cox transformation.

Figure 2: Scatter plots for the beef population before and after taking the Box-Cox transformation.

Table 1: *AARD*, *ARD*, *ALCI*, and *P* for the predictors.

	<i>AARD</i>	<i>ARD</i>	<i>ALCI</i>	<i>P</i>
NTP	0.0015	0.0000	563,656	95%
LTP	0.0014	0.0005	468,563	95%
EBP	0.0011	0.0000	352,421	95%
AEBP	0.0011	0.0004	351,687	94%