REDUCED BIAS PREDICTION REGIONS AND ESTIMATORS OF THE ORIGINAL RESPONSE WHEN USING DATA TRANSFORMATIONS

by

MICHAEL L. WALKER

MARCUS B. PERRY, COMMITTEE CHAIR
B. MICHAEL ADAMS
SUBHA CHAKRABORTI
VOLODYMYR MELNYKOV
DAVID E. HARDY

A DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
in the Department of Information Systems,
Statistics, and Management Science
in the Graduate School of
The University of Alabama

TUSCALOOSA, ALABAMA

2015

Abstract

Initially motivated by electron microscopy experiments, we develop an approximate prediction interval on the univariate response variable Y, where it is assumed that a normal-theory linear model is fit using a transformed version of Y, and the transformation type is contained in the Box-Cox family. Further motivated by A-10 single-engine climb experiments, we then develop an approximate prediction interval on the univariate response Y, in which a linear model is fit using a transformed version of Y, contained in the Manly exponential family. For each case, we derive a closed-form approximation to the k^{th} moment of the original response variable Y, which is then used to estimate the mean and variance of Y, given parameter estimates obtained from fitting the model in the transformed domain. Chebychev's inequality is then used to construct a $100(1-\alpha)\%$ prediction interval estimator on Y based on these mean and variance estimators. Extended data obtained from the A-10 single-engine climb experiments motivates the development of prediction regions in the original domain of a q-variate response vector $\mathbf{Y}$ through the use of multivariate extensions of both the Box-Cox power transformation and the Manly exponential transformation. For each transformation, we derive closed-form approximations to the k^{th} moment of each original response Y_i $(i = 1, \ldots, q)$, as well as a closed-form approximation to $E(Y_i Y_{i'})$, $(i \neq i')$, which are used to estimate the mean and variance of Y_i and the covariance between Y_i and $Y_{i'}$, given parameter estimates obtained from fitting the model in the transformed domain. Exploiting two multivariate analogs of Chebyshev's inequality, we construct an approximate $100(1-\alpha)\%$ prediction sphere and ellipsoid on the original response vector $\mathbf{Y}$.

Acknowledgments

As the culmination of a long journey, this document represents the collective support of a number of people. I would first like to thank Marcus Perry for being both an excellent mentor and colleague. This would not have been possible without his ability to keep me focused, motivated, and on track. I would also like to thank my dissertation committee members of Mike Adams, Subha Chakraborti, Volodymyr Melnykov, and David Hardy for their perspective, experience, and insight. Each had a great impact on my professional growth and development long before joining and I am grateful for all of their continued support throughout this process. I am also very thankful to Uzma Raja and Burcu Keskin for providing me with pivotal computing resources at the very beginning of my research, Mike Porter for the recent insight into penalized regression models and cross-validation, and Chuck Sox for the encouragement and support for both research and conference travel. For the professional moral support of my current and future colleagues along the way, thank you to Semhar, Ibrahim, and my officemates in Bidgood 317. Finally, a very big thank you to my friends and family for keeping me grounded throughout this journey.

Contents

List of Tables

List of Figures

Part I

Introduction

Data transformations are common in statistical model fitting. When the assumption of normality in the model error terms is desired yet may not be reasonably assumed, a transformation may be performed on the response variable in an attempt to achieve approximate normality in the residual error terms of the fitted model. An appropriately applied response transformation will then result in a better model fit. However, when obtaining predictions in the original units of the response is the ultimate goal of the model, this method of data transformation can be problematic. As the model is fit in the transformed domain, predictions obtained from the model will also be in transformed space. As such, any predictions obtained from the model must be re-transformed to the original units, typically achieved by utilizing the inverse of the applied transformation. Additionally, if prediction intervals are also desired, intervals may be constructed in the transformed domain, with the upper and lower bounds then similarly re-transformed to the original units. The most popular and effective response transformations in practice are non-linear in nature, and thus will ultimately result in biased predictions and prediction intervals when using this approach, attributed generally to Jensen's inequality, e.g., see Land (1974).

The research presented in this dissertation presents proposed methodology to reduce this inherent bias in order to improve predictive ability from models constructed with a transformed response variable. The first article presents this methodology in the context of

the popular transformation technique of Box and Cox (1964). As a class of transformations that includes the natural log, square-root, and inverse transformations, its popularity is due in part to the ease of interpretation of the transformation. The second article extends this methodology to the shifted Box-Cox transformation before being presented in the context of the more flexible exponential transformation by Manly (1976). The Manly transformation is found to be more effective when using the proposed methodology as it is not as limited in its application as the Box-Cox transformations. Each of these first two articles involve models fit from a transformed univariate response. The final article of this dissertation therefore further extends the proposed methodology to models fit from a transformed multivariate response. Presented in the context of the multivariate extensions of both the Box-Cox and Manly transformations, the outlined methodology results not only in improved predictions, but also allows for the construction of prediction regions in the original domain.

Part II
A Prediction Interval Estimator for the Original Response when using Box-Cox Transformations

Abstract

Motivated by electron microscopy experiments, we develop an approximate prediction interval on the response variable Y, where it is assumed that a normal-theory linear model is fit using a transformed version of Y, and the transformation type is contained in the Box-Cox family. We derive a closed-form approximation to the k^{th} moment of the original response variable Y, which is then used to estimate the mean and variance of Y, given parameter estimates obtained from fitting the model in the transformed domain. Chebychev's inequality is then used to construct a $100(1-\alpha)\%$ prediction interval estimator on Y. Using Monte Carlo simulation, we assess the width performance of our proposed Chebychev prediction interval, relative to that obtained by employing a more common interval construction approach. General results suggest that, for a given level of expected coverage, the proposed interval estimator will achieve a smaller mean and variance of the interval width estimates, especially as the number of degrees of freedom beyond that required to estimate model terms is small. We apply our method to two experimental data sets, one involving a standard 2^k design, and the other involving a 2^k design with a split-plot error structure.

1 Introduction and Motivation

Data transformations are common in statistical model fitting. When symmetry or approximate normality of the response variable distribution is desired, a power transformation is often used. The technique of Box and Cox (1964) is one of the most popular methods for systematically determining the appropriate power transformation on the response variable Y. The Box-Cox family of power transformations is defined as

$$Y(\lambda) = \begin{cases} \frac{Y^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log_e(Y) & \lambda = 0 \end{cases} \tag{1}$$

where Y is the original response variable, λ is the transformation parameter, and $Y(\lambda)$ is the transformed response variable. Note this class of transformations includes the very popular natural log, square-root, and inverse transformations when $\lambda = 0$, $\lambda = \frac{1}{2}$, and $\lambda = -1$, respectively.

While there are a number of methods to estimate λ, the method of maximum likelihood is typically used. With λ adequately specified, the transformed response variable, $Y(\lambda)$, is assumed to follow a normal distribution with mean μ and variance σ^2, at least approximately[1]. Using the Box-Cox transformations in normal-theory linear model fitting exercises then ensures that the necessary model assumptions on the error term are at least approximately met.

When using the transformed response variable to fit a linear model, the subsequent predictions obtained are then specified in the transformed units rather than in the original units of measurement. To obtain a prediction in the original units of observation, one can simply invert the transformation in equation (1) so that the predicted value of the original

[1] $Y(\lambda)$ actually follows a truncated normal distribution with parameters μ, σ^2 and $-1/\lambda$, where $Y(\lambda) > -1/\lambda$ for $\lambda > 0$ and $Y(\lambda) < -1/\lambda$ for $\lambda < 0$.

response variable takes the form

$$\hat{Y} = \begin{cases} e^{\frac{\log_e(1+\lambda\hat{Y}(\lambda))}{\lambda}} & \lambda \neq 0 \\ e^{\hat{Y}(\lambda)} & \lambda = 0 \end{cases} \tag{2}$$

where $\hat{Y}(\lambda)$ is the predicted value in the transformed units. Unfortunately, this approach leads to inherent bias in the predicted values of the original response variable, even when both the transformation parameter, λ, and the linear model are properly specified. Theoretically, this bias stems from the fact that $E(Y^k)$ is a non-linear function of μ and σ^2, e.g., see Land (1974).

To illustrate this bias, consider the electron microscopy experiments presented in Perry et al. (2012). These authors discussed the process of machining copper bars and subsequently using electron microscopy to statistically examine the resulting microstructure, namely the distribution of grain sizes. They conducted a series of designed experiments to examine the grain structure in a chip[2] that has undergone severe plastic deformation by machining in efforts to relate the mean of the grain size distribution to the *rake angle* (factor A) and *cutting speed* (factor B) of the machine tool. The units of observation (i.e. surface grains) were obtained by randomly sampling a copper bar from a larger population of available bars and then machining the copper bar according to the specified levels of the experimental factors. Following chip formation, a chip was randomly selected and subjected to electron microscopy in order to measure the sizes of the resulting grains contained on the surface.

The practical questions answered by the experiments involved predicting material properties from the distribution of the nano-grains on the machined surface. In particular, the primary aim of their work was to create a predictive statistical model of ultra-fine-grained structures as a function of important machining parameters, which, in turn, could be used to determine the mechanical and other microstructure-controlled properties of the material.

[2]The word 'chip' refers to the material that has been removed from the surface during machining.

In light of this, the authors postulated the following linear mixed-effects model for grain size:

$$y_{ij}(\lambda) = \beta_0 + \beta_1 A + \beta_2 B + \beta_{12} AB + \delta_i + \epsilon_{ij} \tag{3}$$

for $i = 1, ..., N$ and $j = 1, ..., n_i$, where $y_{ij}(\lambda)$ denotes the j^{th} transformed grain size observation taken from the i^{th} bar-stock, N denotes the total number of experiments performed (i.e., total number of bar-stock used in the experiments), and n_i denotes the number of transformed grain size observations taken from the i^{th} bar-stock. Also, β_0, β_1, β_2, and β_{12} denote the fixed-effect model components and $\delta_i \sim N(0, \sigma_\delta^2)$ and $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$, where $Cov(\delta_i, \epsilon_{ij}) = 0$ for all i and j, denote the random model components.

The assumption that the same transformation works for both variance components in equation (3) is a strong one. However, as pointed out in Gurka et al. (2006), one can easily assess the validity of the assumption of normality of δ_i and ϵ_{ij} following a suitable transformation. Specifically, let $\gamma_{ij} = \delta_i + \epsilon_{ij}$ denote the total error term, which after a suitable transformation is assumed to be normally distributed. However, if δ_i and ϵ_{ij} are independent (as stipulated by the linear mixed model), then δ_i and ϵ_{ij} are also normally distributed (see Cramer (1970)) . This result extends to the multivariate case as well, suggesting the same transformation targets both variance components correctly.

Two replicates of a 2^2 factorial design were used to fit the model in equation (3), resulting in a total of eight experimental runs. For each experimental run, a vector of grain size observations of length n_i was recorded from the corresponding machined chip's surface using an electron microscope. We note that the number of grain size observations n_i can be quite large for any given i, often exceeding 1000 observations. Thus, using the actual data from these experiments, we randomly sampled (with replacement) 350 grain size observations from the larger pool of observations available at each of the design points. This was done to achieve balance in the design, i.e. $n_i = n$ for all i, so that known statistical tests on the fixed-effects components are available, but also to simplify the computational burden of

estimating the unknown model parameters. Figure 1 shows normal probability plots of the electron microscopy data for each replicate of the 2^2 design, clearly indicating the need for a transformation.

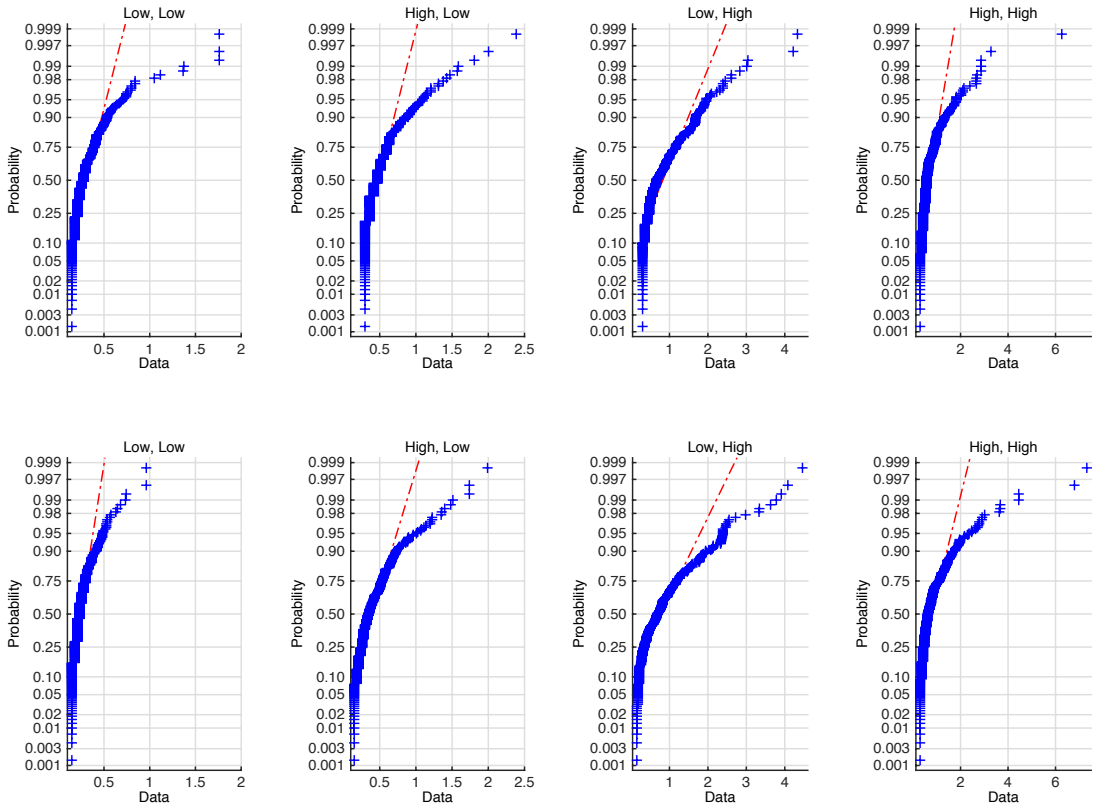

Figure 1: Normal probability plots of electron microscopy data from each replicate in the 2^2 design.

Estimation of the unknown parameters can be accomplished via maximum likelihood (e.g., see Solomon (1985)); however, in this paper a residual maximum likelihood (REML) framework was employed similar to that discussed in Gurka et al. (2006) and Perry et al. (2012), and also outlined in Section 4.2 of this manuscript, producing the following fitted model:

$$\hat{y}(\hat{\lambda}) = -0.9837 + 0.1819A + 0.4759B - 0.1718AB \tag{4}$$

with variance component estimates given by $\hat{\sigma}^2_\delta = 0.0364$ and $\hat{\sigma}^2_\epsilon = 0.5172$, and a transfor-

mation parameter estimate of $\hat{\lambda} = -0.3921$. Figure 2 shows the approximate 95% confidence interval on λ constructed as suggested by Box and Cox (1964), where the values of λ contained between the two vertical lines are potential candidates, or $\lambda \approx [-0.4375, -0.3475]$. Since $\lambda = 1$ is not contained in this interval, it is clear that a transformation is needed. Figure 3 shows normal probability plots of the transformed microscopy data at each design replicate using $\hat{\lambda} = -0.3921$, illustrating the effectiveness of the transformation. Table 1 shows the results of the analysis, suggesting the main effects and two-factor interaction are all at least marginally significant.

Out of interest, we also analyze the simpler 8-point design where the response values are averages of the 350 grain size observations taken at each of the design points. This design is shown in Table 2, while the ordinary least squares analysis results are shown in Table 3. We will note that the Box-Cox transformations were also applied to these response data, yielding a transformation parameter estimate of $\hat{\lambda} = 0.7190$. However, since the 95% confidence interval on λ contained 1, no transformation was performed. From the analysis results in Table 3, it is clear that the simpler 8-point average design failed to identify rake angle (i.e., factor A) as a significant main effect. However, more importantly is the fact that, for this design, inference is being drawn on *average* grain size, and thus all prediction intervals constructed from the analysis are appropriate for average grain size. Since the ultimate goal of the experiments was to construct prediction interval estimates on grain size at various levels of the experimental factors, this dominated our choice of procedure. Thus, the balanced subsample design was more appropriate.

Returning our focus to the results of the balanced subsample design in Table 1, note that the fitted model given in equation (4) is specified in the transformed units. Thus, in order to obtain predicted values in units of the original response variable, one can compute:

$$\hat{y} = \exp\left(\frac{\log_e(\hat{\lambda}\hat{y}(\hat{\lambda}) + 1)}{\hat{\lambda}}\right). \tag{5}$$

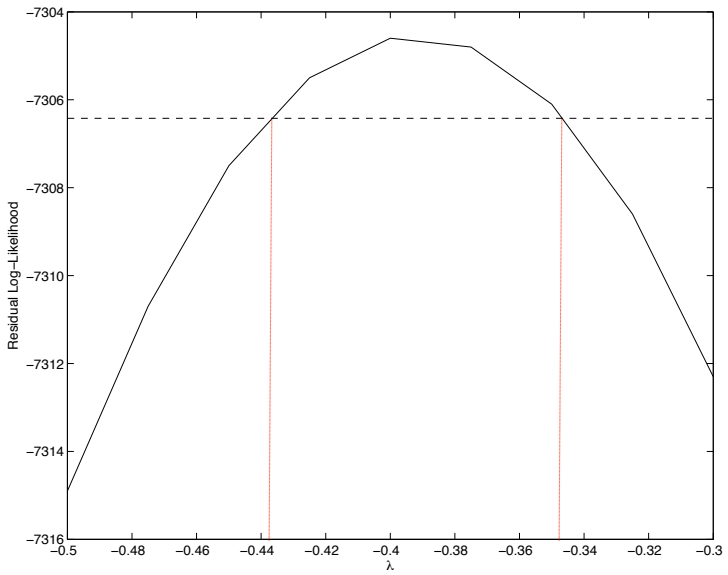

Figure 2: Approximate 95% confidence interval on λ for the electron microscopy data. All values of λ contained between the two vertical lines are potential candidates for the value of the transformation parameter.

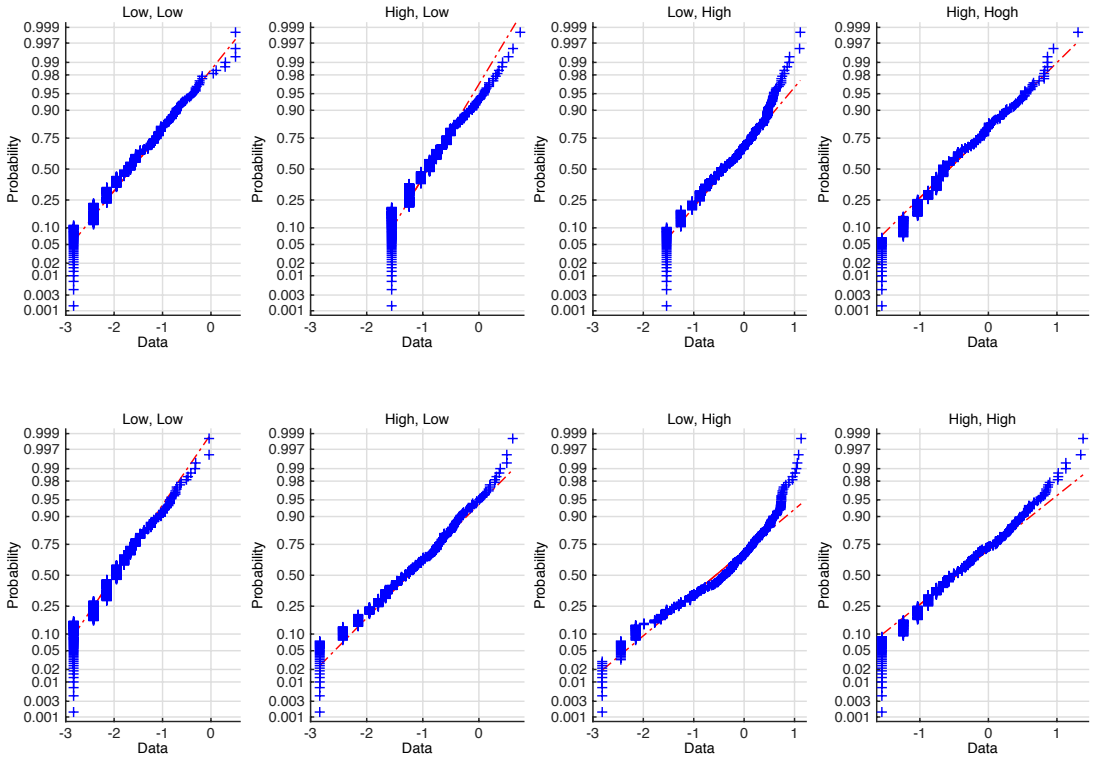

Figure 3: Normal probability plots of transformed electron microscopy data from each replicate in the 2^2 design with $\hat{\lambda} = -0.3921$.

Table 1: Analysis results for the balanced subsample design with $\hat{\sigma}^2_\delta = 0.0364$, $\hat{\sigma}^2_\epsilon = 0.5172$, and $\hat{\lambda} = -0.3921$.

	$\hat{\boldsymbol{\beta}}$	se($\hat{\boldsymbol{\beta}}$)	t	p
Int	-0.9837	0.0688	-14.3040	0.0001
A	0.1819	0.0688	2.6457	0.0572
B	0.4759	0.0688	6.9205	0.0023
AB	-0.1718	0.0688	-2.4984	0.0669

Table 2: 8-point design with average grain size response, where $\bar{y}_i$ is the average of the 350 grain size observations for a given treatment level combination.

A	B	AB	$\bar{y}_i$
-1	-1	1	0.3291
1	-1	-1	0.5297
-1	1	-1	0.8829
1	1	1	0.7589
-1	-1	1	0.2603
1	-1	-1	0.4301
-1	1	-1	0.9042
1	1	1	0.9010

Table 3: Analysis results for the 8-point averages design, where each design point is the average of the 350 grain size observations.

	$\hat{\boldsymbol{\beta}}$	se($\hat{\boldsymbol{\beta}}$)	t	p
Int	0.6245	0.0235	26.5922	0.0000
A	0.0304	0.0235	1.2944	0.2652
B	0.2372	0.0235	10.1010	0.0005
AB	-0.0622	0.0235	-2.6485	0.0571

Table 4 shows sample averages computed from *all* available experimental data at each design point. Further, Table 4 shows point predictions obtained from the fitted model at each design point. By the law of large numbers, one might view the sample averages obtained at any given design point as the true means. Then, clearly, the predicted values from the fitted model are biased downward. This should not be surprising as it is well known that the re-transformation back to the original units of observation given in equation (5) actually estimates the median of Y instead of the expected value of Y, e.g., see Carroll and Rupert (1981), Miller (1984) and Taylor (1986).

Table 4: Comparisons of sample averages and fitted model predictions of grain size at the four different experimental treatment level combinations.

Rake Angle	Cutting Speed	Sample Average	Predicted Value
low	low	0.2712	0.2542
high	low	0.4755	0.3991
low	high	0.8666	0.6241
high	high	0.8361	0.6347

In general, bias in the re-transformed mean has been addressed in the literature, e.g., see Miller (1984), Taylor (1986), Shumway et al. (1989), Sakia (1988) and Sakia (1990). Similar to our work is Taylor (1986) and Sakia (1990), where the authors consider the Box-Cox transformations and use a Taylor series approximation to $E(Y)$, yielding the estimator

$$\widehat{E(Y)} = \exp\left(\frac{\log_e(1+\hat{\lambda}\mathbf{x}_0'\mathbf{b})}{\hat{\lambda}}\right)\left[1 - \frac{\hat{\sigma}^2(\hat{\lambda}-1)}{2(1+\hat{\lambda}\mathbf{x}_0'\mathbf{b})^2}\right], \tag{6}$$

where $\hat{\lambda}$, $\hat{\sigma}^2$ and $\mathbf{b}$ are estimates of the transformation parameter, variance, and $p \times 1$ vector of fixed-effects $\boldsymbol{\beta}$, respectively. The $p \times 1$ vector $\mathbf{x}_0$ denotes a point in design space where a predicted value is of interest. Sakia (1988) and Sakia (1990) extended the results of Taylor (1986) to consider the balanced mixed ANOVA model, in which case $\hat{\sigma}^2 = \sum_\ell \hat{\sigma}^2_\ell$ in equation (6) and denotes the sum of the variance component estimates.

Applying the estimator in equation (6) to the microscopy experiments with $\hat{\lambda} = -0.3921$ and $\hat{\sigma}^2 = \hat{\sigma}^2_\delta + \hat{\sigma}^2_\epsilon = 0.5536$ yields the results shown in Table 5 for each treatment level combination. Notice the significant reduction in bias due to the use of the reduced-bias estimator, relative to the results given in Table 4.

Table 5: Comparisons of sample averages and reduced-bias predictions of grain size at the four different experimental treatment level combinations.

Rake Angle	Cutting Speed	Sample Average	Reduced-Bias Prediction
low	low	0.2712	0.2876
high	low	0.4755	0.4739
low	high	0.8666	0.7902
high	high	0.8361	0.8059

As noted above, the ultimate purpose of the experiments was to construct prediction interval estimates on grain size at various settings of the active experimental factors. When using the Box-Cox transformations, the typical way to construct prediction intervals on Y is to use standard methods of construction in the transformed domain. The upper and lower bounds of the prediction interval constructed in the transformed domain are then "re-transformed" using the inverse of the Box-Cox transformations, producing the interval estimator

$$\left[\exp\left(\frac{\log_e(\hat{\lambda}(\hat{Y}(\hat{\lambda})-\hat{H})+1)}{\hat{\lambda}}\right),\ \exp\left(\frac{\log_e(\hat{\lambda}(\hat{Y}(\hat{\lambda})+\hat{H})+1)}{\hat{\lambda}}\right)\right] \quad \text{if } \hat{\lambda} \neq 0$$

and

$$\left[\exp\left(\hat{Y}(\hat{\lambda}) - \hat{H}\right),\ \exp\left(\hat{Y}(\hat{\lambda}) + \hat{H}\right)\right] \quad \text{if } \hat{\lambda} = 0$$

where $\hat{H} = h\hat{\sigma}_{pred}$, $\hat{\sigma}^2_{pred} = \sum_\ell \hat{\sigma}^2_\ell + \mathbf{x}_0'(\mathbf{X}'\hat{\mathbf{V}}^{-1}\mathbf{X})^{-1}\mathbf{x}_0$ denotes the prediction variance estimate computed in the transformed domain at point $\mathbf{x}_0$, $\hat{\sigma}^2_\ell$ is the ℓ^{th} variance component estimate, $\hat{\mathbf{V}}$ is the estimated variance-covariance matrix of the transformed response observations, and the value of h is set to achieve some desired level of confidence. It is important to note that,

when $\hat{\lambda} \neq 0$, this interval estimator is only defined if

$$\begin{aligned} \hat{Y}(\hat{\lambda}) - \hat{H} &> -\tfrac{1}{\hat{\lambda}} \quad \text{for } \hat{\lambda} > 0 \\ \hat{Y}(\hat{\lambda}) + \hat{H} &< -\tfrac{1}{\hat{\lambda}} \quad \text{for } \hat{\lambda} < 0. \end{aligned} \tag{7}$$

Unfortunately, for small sample sizes, the parameters can be estimated quite poorly, which can often render this interval estimator undefined. Applying this re-transformation method to the electron microscopy data leads to the approximate 95% prediction intervals shown in Table 6, where $\hat{\mathbf{V}} = \hat{\sigma}_{\delta}^2 \mathbf{M}\mathbf{M}' + \hat{\sigma}_{\epsilon}^2 \mathbf{I}$, and $\mathbf{M}$ is a known indicator matrix taking the form $\mathbf{M} = blkdiag(\mathbf{1}_{m_1}, \mathbf{1}_{m_2}, ..., \mathbf{1}_{m_8})$, where $m_i = 350$ for all i, yielding $\hat{\sigma}_{pred}^2 = 0.5725$ at each of the design points. Also, $h = t_{4,0.975} = 2.7764$ was used as the reference value.

Notice that for these experiments, the prediction intervals constructed in this manner are quite poor, yielding large widths. In fact, when considering *all* of the grain size observations at each of the design points, one would expect the estimated prediction intervals shown in Table 6 to contain approximately 95% of these observations. Instead, the column showing the actual coverage indicates that 100% of the observations are contained in these intervals at each of the treatment level combinations (except for the *low*, *high* treatment combination where over 98% of the observations are contained). These results seem to suggest a need for improvement.

Table 6: Approximate 95% prediction intervals on grain size constructed using standard methods in the transformed domain, and then re-transformed to the original units of observation.

Rake Angle	Cutting Speed	Lower 95% Bound	Upper 95% Bound	Actual Coverage
low	low	0.0933	1.3566	100%
high	low	0.1254	3.5296	100%
low	high	0.1650	11.8481	98.2%
high	high	0.1667	12.5053	100%

In what follows, we discuss our approach to approximating higher-order moments of

the original response variable Y, and then exploit this to construct improved prediction intervals on Y. In our development we assume that a normal-theory linear model is fit using a transformed version of Y, where the transformation type is contained in the Box-Cox family. We derive a closed-form approximation to the k^{th} moment of the original response using a Taylor series approximation. This expression is then used to estimate its mean and variance, given the parameter estimates obtained from fitting the model in the transformed domain. We then exploit Chebychev's inequality to construct an approximate $100(1-\alpha)\%$ prediction interval estimate. Further, we discuss implications made from the results of Monte Carlo simulation studies used to assess the performance of the proposed prediction interval estimator, relative to that obtained by employing the more common re-transformation construction approach outlined above.

2 Proposed Methodology

Consider the class of transformations defined in equation (1). We assume that for some suitable λ, the transformed response $Y(\lambda)$ approximately follows the normal distribution with mean $\mu = \mathbf{x}'\boldsymbol{\beta}$ and variance $\sigma^2 = \sum_\ell \sigma_\ell^2$, where σ_ℓ^2 denotes the ℓ^{th} variance component. Draper and Cox (1969) suggested the condition $|\lambda\sigma_\ell/(1+\lambda\mathbf{x}'\boldsymbol{\beta})| << 1$ for all ℓ, which when satisfied, will ensure all observations are positive with high probability. In this case, the density function of the *untransformed* response Y is well approximated by

$$f_Y(y) \approx \begin{cases} \frac{y^{\lambda-1}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\frac{y^\lambda-1}{\lambda}-\mathbf{x}'\boldsymbol{\beta})^2}{2\sigma^2}\right) & \lambda \neq 0 \\ \frac{1}{y\sqrt{2\pi}\sigma} \exp\left(-\frac{(\log_e(y)-\mathbf{x}'\boldsymbol{\beta})^2}{2\sigma^2}\right) & \lambda = 0 \end{cases}$$

for $y > 0$. Obviously, when $\lambda = 0$, $f_Y(y)$ is the log-normal density function, and is thus exact. It then follows that, for $\lambda \neq 0$, an approximation to the k^{th} moment of Y can be

written as follows

$$E(Y^k) \approx \frac{1}{\sqrt{2\pi}} \int_R \exp\Big\{ \frac{\log_e\big(1+\lambda(\mathbf{x}'\boldsymbol{\beta}+\sigma u)\big)}{\lambda} \Big\}^k \exp\Big\{ -\frac{u^2}{2} \Big\} du \tag{8}$$

where $u = \frac{\frac{y^\lambda - 1}{\lambda} - \mathbf{x}'\boldsymbol{\beta}}{\sigma}$ and $du = \frac{y^{\lambda-1}}{\sigma} dy$. Note that R denotes the domain of the function $g(u) = \log_e\big(1+\lambda(\mathbf{x}'\boldsymbol{\beta}+\sigma u)\big)$, where if $\lambda > 0$, then $u > -\frac{(\mathbf{x}'\boldsymbol{\beta}+\frac{1}{\lambda})}{\sigma}$, and if $\lambda < 0$, then $u < -\frac{(\mathbf{x}'\boldsymbol{\beta}+\frac{1}{\lambda})}{\sigma}$.

To simplify the integral in equation (6), one can expand the first term on the right-hand side of the integrand in a Taylor series about $u = 0$, producing the approximation

$$E(Y^k) \approx \frac{(1+\lambda\mathbf{x}'\boldsymbol{\beta})^{\frac{k}{\lambda}}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \left(1 + \sum_{i=1}^{\infty} \frac{k\sigma^i \prod_{j=1}^{i-1}(k-\lambda j)}{i!(1+\lambda\mathbf{x}'\boldsymbol{\beta})^i} u^i \right) \times e^{-\frac{u^2}{2}} du. \tag{9}$$

Retaining terms through order 4 in equation (9) and subsequently evaluating the resulting integral then produces the closed form approximation to the k^{th} moment of Y given by

$$E(Y^k) \approx \exp\left(\frac{k\log_e(1+\lambda\mathbf{x}'\boldsymbol{\beta})}{\lambda} \right) \left[1 - \frac{k}{2}\frac{(\lambda-k)\sigma^2}{(1+\lambda\mathbf{x}'\boldsymbol{\beta})^2} - \frac{k}{8}\frac{(\lambda-k)(2\lambda-k)(3\lambda-k)\sigma^4}{(1+\lambda\mathbf{x}'\boldsymbol{\beta})^4} \right], \tag{10}$$

where $\mathbf{x}'\boldsymbol{\beta} > -\frac{1}{\lambda}$ if $\lambda > 0$, $\mathbf{x}'\boldsymbol{\beta} < -\frac{1}{\lambda}$ if $\lambda < 0$ and $\sigma^2 > 0$. Note that equation (10) is defined for all values of λ except $\lambda = 0$, where, for this case, it is easily shown that the k^{th} moment of Y is given exactly by

$$E(Y^k) = \exp\Big\{ \frac{k}{2}(2\mathbf{x}'\boldsymbol{\beta} + k\sigma^2) \Big\} \tag{11}$$

for $\sigma > 0$ and $-\infty < \mathbf{x}'\boldsymbol{\beta} < \infty$.

Freeman and Modarres (2006) also provided an approximation to $E(Y^k)$, however, these authors only consider the case where $0 \leq \lambda \leq 1$. Clearly, the case where $\lambda < 0$ is also important, which is evidenced by the analysis of the microscopy experimental data discussed above. Point estimates for $E(Y)$ and $Var(Y)$ are subsequently found by substituting parameter estimates computed in the transformed domain into the expression given in equation

(10), producing the estimators

$$\widehat{E(Y)} = \hat{\mu}_Y = \begin{cases} \exp\left(\frac{\log_e(1+\hat{\lambda}\mathbf{x}'\hat{\boldsymbol{\beta}})}{\hat{\lambda}}\right)\left[1 - \frac{(\hat{\lambda}-1)\hat{\sigma}^2}{2(1+\hat{\lambda}\mathbf{x}'\hat{\boldsymbol{\beta}})^2} - \frac{(\hat{\lambda}-1)(2\hat{\lambda}-1)(3\hat{\lambda}-1)\hat{\sigma}^4}{8(1+\hat{\lambda}\mathbf{x}'\hat{\boldsymbol{\beta}})^4}\right] & \hat{\lambda} \neq 0 \\ \exp\{\mathbf{x}'\hat{\boldsymbol{\beta}} + \hat{\sigma}^2/2\} & \hat{\lambda} = 0 \end{cases} \tag{12}$$

and

$$\widehat{Var(Y)} = \hat{\sigma}_Y^2 = \widehat{E(Y^2)} - \widehat{E(Y)}^2 \tag{13}$$

where,

$$\widehat{E(Y^2)} = \begin{cases} \exp\left(\frac{2\log_e(1+\hat{\lambda}\mathbf{x}'\hat{\boldsymbol{\beta}})}{\hat{\lambda}}\right)\left[1 - \frac{(\hat{\lambda}-2)\hat{\sigma}^2}{(1+\hat{\lambda}\mathbf{x}'\hat{\boldsymbol{\beta}})^2} - \frac{(\hat{\lambda}-2)(2\hat{\lambda}-2)(3\hat{\lambda}-2)\hat{\sigma}^4}{4(1+\hat{\lambda}\mathbf{x}'\hat{\boldsymbol{\beta}})^4}\right] & \hat{\lambda} \neq 0 \\ \exp\{2\mathbf{x}'\hat{\boldsymbol{\beta}} + 2\hat{\sigma}^2\} & \hat{\lambda} = 0. \end{cases} \tag{14}$$

and

$$\begin{aligned} \mathbf{x}'\hat{\boldsymbol{\beta}} &> -\tfrac{1}{\hat{\lambda}} \quad \text{for } \hat{\lambda} > 0 \\ \mathbf{x}'\hat{\boldsymbol{\beta}} &< -\tfrac{1}{\hat{\lambda}} \quad \text{for } \hat{\lambda} < 0 \end{aligned} \tag{15}$$

are conditions on the parameter estimates that must be met in order for the estimators to be defined.

By expanding the estimators for μ_Y and σ_Y^2 in equations (9) and (10), respectively, in a Taylor series about $\hat{\lambda} = \lambda$, $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, and $\hat{\sigma} = \sigma$, one can show consistency. Specifically, it can be shown that if $\hat{\lambda}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\sigma}$ are consistent estimators of λ, $\boldsymbol{\beta}$, and σ, respectively, then $\hat{\mu}_Y$ and $\hat{\sigma}_Y$ are consistent estimators of μ_Y and σ_Y, up to a 4^{th}-order approximation.

It should be noted that the expressions for $E(Y)$ and $E(Y^2)$ given in equations (9) and (11), respectively, are equivalent to those suggested by Taylor (1986) and Sakia (1990), and Sakia (1990), respectively, with the exception of the last term contained in the brackets of the proposed estimators. Obviously, these extra terms are due to retaining terms up through 4^{th}-order in our Taylor series expansion of the first term on the right-hand side of the integrand in equation (6). The choice to retain terms through order 4 (as opposed

to order 2) in the Taylor series above permits one to estimate moments of higher order (i.e., $k > 1$) with increased precision. Since when constructing prediction intervals on Y an estimate for $E(Y^2)$ is needed, this choice is particularly justified.

As discussed in Section 1, prediction intervals are often preferred over a simple point estimate, and are typically constructed from a re-transformation of the standard prediction interval estimated in the transformed domain. However, when using the above expressions to estimate the mean and variance of Y, Chebychev's inequality can then be exploited to obtain an approximate $100(1-\alpha)\%$ prediction interval on the original response variable Y. Recall Chebychev's inequality

$$P(|Y - \mu_Y| \leq L\sigma_Y) \geq 1 - \frac{1}{L^2}, \tag{16}$$

which can be manipulated to produce the interval

$$\hat{\mu}_Y \pm L\hat{\sigma}_Y, \tag{17}$$

where $L = 4.47$ corresponds to an approximate 95% prediction interval on Y. Note that this interval is conservative, suggesting the probability that Y lies within L standard deviation units from its mean is *at least* $1 - \frac{1}{L^2}$ approximately. Also, since equations (9)-(11) are only defined if the conditions in equation (12) are satisfied, then the proposed interval estimator in equation (13) is also defined under these conditions. If we compare conditions in equation (12) for the proposed interval estimator to those in equation (7) for the traditional interval estimator, we see that the proposed interval estimator is defined regardless of the estimate for σ^2. This is significant, particularly for smaller sample sizes, as poorly estimated parameters can often render the traditional interval estimator undefined.

It should be noted that, if the parameters λ, $\boldsymbol{\beta}$ and σ^2 are accurately and precisely

estimated, then one can determine L from the relationship:

$$1-\alpha=\frac{1}{\sqrt{2\pi}}\int_{-L}^{L}\exp\{-u^2/2\}du=erf\left(\frac{\sqrt{2}}{2}L\right) \tag{18}$$

where $u=\frac{\frac{y^{\lambda}-1}{\lambda}-\mathbf{x}'\boldsymbol{\beta}}{\sigma}$, $du=\frac{y^{\lambda-1}}{\sigma}dy$ and $erf(\cdot)$ denotes the error function. It can be shown that the solution to equation (18) is obtained by setting $L=z_{\alpha/2}$, where $z_{\alpha/2}$ denotes the upper $\alpha/2$ quantile of the standard normal distribution. This implies that, for any given α, the value of L will lie in the interval $[z_{\alpha/2},\sqrt{\frac{1}{\alpha}}]$. In practice, if a large sample is available for estimating unknown model parameters, then one can justify settimg $L=z_{\alpha/2}$, otherwise we recommend using $L=\sqrt{\frac{1}{\alpha}}$, which will produce a prediction interval on Y with expected coverage of *at least* the nominal probability $1-\alpha$.

At this point we can revisit the electron microscopy data discussed in Section 1 and compute approximate 95% prediction intervals on grain size at each of the design points using the model parameter estimates and equation (13). These are shown in Table 7 with $L=z_{0.025}=1.96$. For reference purposes, Table 7 also shows the approximate 95% prediction intervals estimated directly from the untransformed response and using *all* available grain size observations at each of the design points. These are shown in parentheses.

The results in Table 7 suggest that, for the microscopy data, the proposed method yields reasonable estimates for the 95% prediction intervals on grain size at the various design points. This is because the estimates for the prediction intervals using the fitted model are close in value to the actual values computed directly using all available grain size observations at each design point. Further, it appears as if the proposed method yields a better prediction interval, relative to that obtained via the traditional re-transformation construction approach, i.e., the actual coverages of the proposed prediction intervals are much closer to the nominal 95% than those constructed using the traditional method shown in Table 6. Figure 4 shows the re-transformed intervals (corresponding to the error bars) and the proposed intervals (shown as the shaded regions) at each of the design points in the 2^2

Table 7: Estimated 95% prediction intervals on grain size computed at each of the design points using the fitted model and proposed methodology with $L = z_{0.025} = 1.96$. The 95% prediction intervals computed directly from the untransformed response using *all* available grain size observations at each of the design points are shown in parentheses. Negative lower interval bounds were set to zero.

Rake Angle	Cutting Speed	Lower 95% Bound	Upper 95% Bound	Actual Coverage
low	low	0.018	0.569	96.75%
		(0.018)	(0.524)	
high	low	0.000	1.046	95.37%
		(0.000)	(1.034)	
low	high	0.000	1.963	94.03%
		(0.000)	(2.166)	
high	high	0.000	2.011	94.76%
		(0.000)	(2.258)	

design. The black shaded region corresponds to the proposed interval with $L = z_{0.025} = 1.96$, while the red shaded region corresponds to the proposed interval with $L = 1/\sqrt{0.05} = 4.47$. Clearly, the proposed intervals are the better choice in this case.

Although the proposed interval estimator appears to perform quite well on the microscopy experimental data, it is not clear from this single example how the estimator will perform on the average. Therefore, in the next section we discuss results of a Monte Carlo simulation study designed to assess the average performance of $\hat{\mu}_Y$ and $\hat{\sigma}^2_Y$ in equations (9) and (10), respectively, as well as the proposed prediction interval in equation (13), relative to that obtained via the traditional re-transformation construction method

3 Performance Evaluation

In efforts to assess the expected performance of our proposed prediction interval estimator in equation (13), we used Monte Carlo simulation. In what follows, we describe the simulation model in detail.

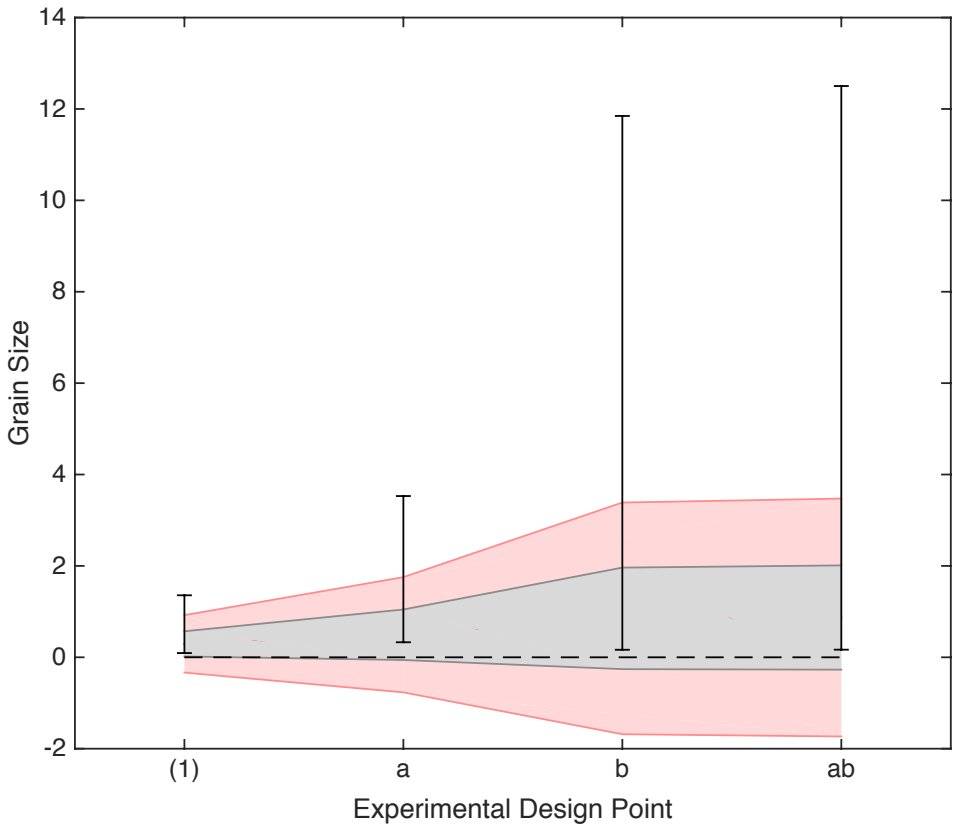

Figure 4: Approximate 95% prediction intervals on grain size at each of the design points in the 2^2 design. The vertical bars represent the traditional re-transformed intervals, while the shaded regions represent the proposed intervals. Proposed interval constructed using $L = z_{0.025} = 1.96$ is shown in black shaded region and $L = 1/\sqrt{0.05} = 4.47$ is shown in red shaded region. Re-transformed interval was constructed using the reference value $t_{4,0.025} = 2.7764$.

For any given simulation run, a random sample $\mathbf{y}$ of size n was generated from

$$y_i = \begin{cases} \exp\left[\frac{\log_e(\lambda y_i(\lambda)+1)}{\lambda}\right] & \text{for } \lambda \neq 0 \\ \exp\left[y_i(\lambda)\right] & \text{for } \lambda = 0 \end{cases} \tag{19}$$

where

$$y_i(\lambda) = \mathbf{x}_i^{'}\boldsymbol{\beta} + \epsilon_i \tag{20}$$

denotes the transformed response observation taken at $\mathbf{x}_i$, with $\mathbf{x}_i = [1, x_{i1}, x_{i2}, ..., x_{iq}]'$ denoting a $p \times 1$ vector of regressor variable settings ($p = q + 1$), $\boldsymbol{\beta} = [\beta_0, \beta_1, ..., \beta_q]'$ denotes a $p \times 1$ vector of fixed effects (or regression coefficients), and ϵ_i denotes a random error term where $\epsilon_i \sim i.i.d.N(0, \sigma^2)$ for all $i = 1, ..., n$. Once the response observations were generated, model parameters were then estimated using maximum likelihood estimation. Letting $\mathbf{X}$ denote an $n \times p$ design matrix and $\mathbf{y}(\lambda) = [y_1(\lambda), ..., y_n(\lambda)]'$ denote the $n \times 1$ vector of transformed responses, then the method of maximum likelihood leads to the following estimators for the unknown model parameters

$$\hat{\lambda} = \arg\max_{\lambda} \left[-n\log_e(\sqrt{2\pi}\hat{\sigma}_\epsilon) + (\lambda - 1)\sum_{i=1}^{n}\log_e(y_i) - \frac{1}{2\hat{\sigma}_\epsilon^2}\sum_{i=1}^{n}(y_i(\lambda) - \hat{\mu}_i)^2 \right], \tag{21}$$

where $y_i(\lambda) = \frac{y_i^{\lambda}-1}{\lambda}$ if $\lambda \neq 0$ and $y_i(\lambda) = \log_e(y_i)$ if $\lambda = 0$, with

$$\hat{\mu}_i = \mathbf{x}_i^{'}\hat{\boldsymbol{\beta}} = \mathbf{x}_i^{'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}(\lambda) \tag{22}$$

and

$$\hat{\sigma}_\epsilon^2 = \frac{\mathbf{y}(\lambda)'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}(\lambda)}{n - p}, \tag{23}$$

where Matlab's *fmincon* function was used for solving the optimization problem in equation (16). Once parameters were estimated, they were then substituted into equations (9) and (10) in order to produce estimates for, respectively, the mean and variance of the untransformed

response Y. At this point in the simulation model, prediction interval estimates on Y were also produced. In particular, the prediction interval estimate using equation (13) was constructed, as well as that using the traditional re-transformation construction method described in Section 1. Once these interval estimates were obtained, their widths were recorded by the simulation model. For a given simulation run, interval estimate coverage was computed by generating 50 million observations from the *true* distribution and then recording the proportion of these observations that fell within the estimated intervals. To make a fair comparison, we set the value of L in equation (13) so that the relative difference in the estimated expected coverages between the two methods over 500 Monte Carlo simulation runs was less than 1%. For all simulations, we chose $\alpha = 0.05$.

In our study, we considered fitting a main-effects plus two-factor interactions model using various replicates of a 2^3 factorial designed experiment. We considered values of $\lambda \in [-1, 1]$, $\mu = [-1, 0, 1]$ and $\sigma_\epsilon = 0.15$, so that $P[y_i(\lambda) > -\frac{1}{\lambda}|\lambda > 0] \approx 1$ and $P[y_i(\lambda) < -\frac{1}{\lambda}|\lambda < 0] \approx 1$. Specifically, for the simulations discussed below, the parameter vector $\boldsymbol{\beta}$ was simulated at $\boldsymbol{\beta} = [\mu, 0, \cdots, 0]'$ and the vector $\mathbf{x}$ was an arbitrarily chosen factorial point. It makes no difference which point in design space is used in the simulation model since the coefficients $\beta_1, ..., \beta_q$ were all set to zero. We should note that additional simulation runs were performed at other values for the parameters $\boldsymbol{\beta}$ and σ_ϵ, and at other points $\mathbf{x}$ within the design space (where the probabilities on $y_i(\lambda)$ mentioned above hold), and although not reported here, our general conclusions discussed below are not altered.

Table 8 shows estimated root mean square error (RMSE) performances of the estimators in equations (9) and (10), for the case where $\lambda \in [-1, 0.5]$ and $\mu = -1$, each obtained over 5000 independent Monte Carlo simulation runs. Similarly, Table 9 shows estimated RMSEs for the case where $\lambda \in [-1, 1]$ and $\mu = 0$, while Table 10 shows RMSEs for the case where $\lambda \in [-0.5, 1]$ and $\mu = 1$.

Notice in Tables 8-10 that, in general, as the number of design replicates R increases, a decrease in the RMSEs for both estimators is observed, regardless of the values of λ and

Table 8: Root mean square errors (RMSEs) of estimators in equations (9) and (10). Each RMSE entry in the table was computed over 5000 independent Monte Carlo simulation runs. For any given λ and number of design replicates R, the true mean was simulated at $\mu = \mathbf{x}'\boldsymbol{\beta} = -1$.

	$\widehat{E(Y)}$			$\widehat{Var(Y)}$		
λ	$R=2$	$R=3$	$R=4$	$R=2$	$R=3$	$R=4$
0.5	0.0514	0.0403	0.0359	0.0044	0.0030	0.0022
0.2	0.0417	0.0340	0.0293	0.0026	0.0019	0.0015
0.1	0.0387	0.0325	0.0278	0.0024	0.0017	0.0013
0	0.0373	0.0308	0.0266	0.0021	0.0015	0.0012
-0.1	0.0351	0.0293	0.0251	0.0019	0.0014	0.0011
-0.2	0.0339	0.0280	0.0243	0.0016	0.0012	0.0010
-0.5	0.0294	0.0245	0.0218	0.0013	0.0009	0.0008
-1	0.0255	0.0203	0.0181	0.0008	0.0006	0.0005

Table 9: Root mean square errors (RMSEs) of estimators in equations (9) and (10). Each RMSE entry in the table was computed over 5000 independent Monte Carlo simulation runs. For any given λ and number of design replicates R, the true mean was simulated at $\mu = \mathbf{x}'\boldsymbol{\beta} = 0$.

	$\widehat{E(Y)}$			$\widehat{Var(Y)}$		
λ	$R=2$	$R=3$	$R=4$	$R=2$	$R=3$	$R=4$
1	0.0986	0.0809	0.0706	0.0121	0.0086	0.0069
0.5	0.0993	0.0813	0.0719	0.0133	0.0094	0.0077
0.2	0.0999	0.0820	0.0714	0.0140	0.0105	0.0084
0.1	0.1021	0.0822	0.0723	0.0142	0.0107	0.0090
0	0.1033	0.0846	0.0728	0.0149	0.0115	0.0092
-0.1	0.1020	0.0846	0.0721	0.0167	0.0114	0.0095
-0.2	0.1016	0.0839	0.0736	0.0156	0.0124	0.0101
-0.5	0.1027	0.0842	0.0744	0.0168	0.0139	0.0110
-1	0.1064	0.0894	0.0780	0.0190	0.0170	0.0137

Table 10: Root mean square errors (RMSEs) of estimators in equations (9) and (10). Each RMSE entry in the table was computed over 5000 independent Monte Carlo simulation runs. For any given λ and number of design replicates R, the true mean was simulated at $\mu = \mathbf{x}'\boldsymbol{\beta} = 1$.

	$\widehat{E(Y)}$			$\widehat{Var(Y)}$		
λ	$R=2$	$R=3$	$R=4$	$R=2$	$R=3$	$R=4$
1	0.0997	0.0824	0.0701	0.0118	0.0082	0.0068
0.5	0.1487	0.1234	0.1064	0.0279	0.0199	0.0163
0.2	0.2136	0.1718	0.1477	0.0606	0.0429	0.0347
0.1	0.2367	0.2016	0.1701	0.0794	0.0612	0.0476
0	0.2767	0.2272	0.1913	0.1151	0.0806	0.0643
-0.1	0.3215	0.2678	0.2320	0.1613	0.1280	0.1007
-0.2	0.4051	0.3257	0.2822	0.2758	0.2059	0.1620
-0.5	0.9871	0.8006	0.6932	2.5825	1.8482	1.5349

μ. This is quite intuitive, and should be expected since with larger R more observations are available to estimate unknown model parameters. Also, note that, in Table 8 (where $\mu = -1$), as λ decreases, a decrease in the RMSEs of both estimators is observed for any given R. Further, in Tables 9 and 10 (where $\mu = 0$ and 1, respectively), as λ decreases, an increase in the RMSEs of the estimators for any given R is observed. These results are also intuitive since when $\mu \geq 0$, an increase in the variance of Y is observed as λ decreases, and when $\mu < 0$, a decrease in the variance of Y is observed as λ decreases. Thus, under these circumstances, one would expect to see the RMSEs behave in such a way.

We now focus our attention on the relative performances of the prediction interval estimators considered in our study; in particular, the interval estimator obtained via the traditional re-transformation method discussed in Section 1, as well as the proposed estimator given in equation (13). In order to effectively summarize the relative performances over the range of λ for any given number of design replicates R, we follow Han and Tsung (2006) and compute the relative mean index (RMI) defined by

$$RMI_{u_i} = \frac{1}{r}\sum_{j=1}^{r}\frac{u_{ij} - \min[u_{1j}, u_{2j}]}{\min[u_{1j}, u_{2j}]}, \tag{24}$$

where RMI_{u_i} denotes the relative mean index of method i with respect to performance measure u. Here, u_{ij} denotes a performance measure taken at the j^{th} level of λ for the i^{th} method being compared, where $i = 1, 2$, and $j = 1, 2, ..., r$, where r denotes the number of levels of λ considered. Thus, to compare expected width performance between the traditional and proposed interval estimators for a given number of design replicates R, let $\bar{\hat{w}}_{ij}$ denote the average of the width estimates obtained over the 5000 independent Monte Carlo simulation runs at the j^{th} level of λ and for the i^{th} method being compared, then $u_{ij} = \bar{\hat{w}}_{ij}$. Similarly, to compare the variability in the widths of the interval estimates, let $\hat{s}_{ij}$ denote the estimated standard error of $\bar{\hat{w}}_{ij}$ obtained over the 5000 Monte Carlo simulation runs at the j^{th} level of λ and for the i^{th} method being compared, then let $u_{ij} = \hat{s}_{ij}$. Among the two methods being compared, the one that achieves better relative performance will then have the smallest RMI, for both performance measures considered.

Table 11 shows the RMIs computed at values of $\mu = -1$, $\mu = 0$ and $\mu = 1$, at different numbers of design replicates R, and for both performance measures considered, i.e., $\bar{\hat{w}}_{ij}$ and $\hat{s}_{ij}$. Note that $RMI_{\bar{\hat{w}}_1}$ and $RMI_{\hat{s}_1}$ denote the RMIs corresponding to the traditional interval estimator, and $RMI_{\bar{\hat{w}}_2}$ and $RMI_{\hat{s}_2}$ denote those corresponding to the proposed interval estimator.

The results in Table 11 suggest that, if the two intervals yield approximately the same coverage probability, our proposed interval estimator will outperform that obtained via the traditional re-transformation construction method. In particular, our results show that, for all combinations of λ, μ and R considered in our study, our proposed interval estimator achieved an interval with a smaller expected width and a smaller variance of the width estimates. Large differences in performance are most evident when there are only small number of design replicates. Table 11 further suggests that, for any given $\mu = \mathbf{x}'\boldsymbol{\beta}$, as the number of design replicates increases, the performance of the traditional estimator appears to approach that of our proposed estimator.

We should note that, although in our simulations we set L so that the expected cov-

Table 11: Relative mean indices (RMIs) of performance results of the two prediction interval estimators for different values of $\mu = \mathbf{x}'\boldsymbol{\beta}$ and design replicates R. The method that produces the smaller RMI value suggests better relative performance across the range of λ.

	R	$RMI_{\bar{\hat{w}}_1}$	$RMI_{\bar{\hat{w}}_2}$	$RMI_{\hat{s}_1}$	$RMI_{\hat{s}_2}$
	2	0.14	0.00	2.70	0.00
	3	0.06	0.00	0.46	0.00
$\mu = -1$	4	0.01	0.00	0.14	0.00
	5	0.01	0.00	0.11	0.00
	50	0.01	0.00	0.03	0.00
	2	0.21	0.00	2.81	0.00
	3	0.08	0.00	0.56	0.00
$\mu = 0$	4	0.03	0.00	0.47	0.00
	5	0.03	0.00	0.20	0.00
	50	0.01	0.00	0.02	0.00
	2	0.44	0.00	21.51	0.00
	3	0.15	0.00	2.33	0.00
$\mu = 1$	4	0.05	0.00	0.50	0.00
	5	0.04	0.00	0.30	0.00
	50	0.02	0.00	0.05	0.00

erages between the two methods was approximately equal, it is not necessarily true that the nominal 95% coverage was obtained using either method (in fact, both methods produced less than 95% coverage). When using the proposed method with $L = 1/\sqrt{\alpha}$ one can achieve *at least* the nominal coverage of $100(1-\alpha)\%$, and thus it might be expected that the interval widths using the proposed construction method will be relatively larger on the average. However, it is still the case that the standard error of the interval width estimates using the proposed method is relatively smaller than that obtained using the re-transformation construction approach. Thus, the proposed method will permit better control of the type I error rate α, while also providing interval width estimates with greater relative precision. Therefore, based on these results, we recommend that the proposed prediction interval estimator be used as an alternative to that obtained via the re-transformation approach in practice, unless the number of design replicates (or, more generally, degrees of freedom beyond that required to estimate model parameters) is very large.

4 Application of Proposed Method to Other Experimental Data Sets

In this section we demonstrate the application of our proposed prediction interval estimator to other experimental data sets. In particular, we consider data from a 2^4 experimental design given in Daniel (1976) and reproduced in Montgomery (2009), as well as the wind tunnel split-plot experimental data considered in Simpson et al. (2004).

4.1 Daniel's Drill Advance Rate Experiments

Daniel (1976) describes a 2^4 factorial design used to study the advance rate of a drill as a function of drill load (A), flow rate (B), rotational speed (C), and drilling mud type (D). The design and response data are shown in Table 12.

Table 12: Drill advance rate data for 2^4 factorial design given in Daniel (1976).

Run	A	B	C	D	y
(1)	-1	-1	-1	-1	1.68
a	1	-1	-1	-1	1.98
b	-1	1	-1	-1	4.98
ab	1	1	-1	-1	5.70
c	-1	-1	1	-1	3.24
ac	1	-1	1	-1	3.44
bc	-1	1	1	-1	9.97
abc	1	1	1	-1	9.07
d	-1	-1	-1	1	2.07
ad	1	-1	-1	1	2.44
bd	-1	1	-1	1	7.77
abd	1	1	-1	1	9.43
cd	-1	-1	1	1	4.09
acd	1	-1	1	1	4.53
bcd	-1	1	1	1	11.75
abcd	1	1	1	1	16.30

The design in Table 12 was used to fit the full main-effects plus two-factor interaction

model. Unknown model parameters were estimated via maximum likelihood methods by way of equations (16), (17), and (18). Table 13 shows first iteration analysis results, namely, the t-tests on the model coefficients with $\hat{\lambda} = -0.7494$ and $\hat{\sigma}^2 = 0.000299$. Retaining those model terms in Table 13 that yielded p-values less than $\alpha = 0.05$ and subsequently performing a second iteration analysis of the experimental data yielded the results shown in Table 14 with $\hat{\lambda} = -0.4698$ and $\hat{\sigma}^2 = 0.000975$. Finally, retaining the terms in Table 14 that produced p-values less than 0.05 and subsequently performing a third iteration analysis on the resulting design produced the results in Table 15, yielding the final fitted model in the transformed units:

$$\hat{y}(\hat{\lambda}) = 1.1199 + 0.0349A + 0.3041B + 0.1579C + 0.0816D - 0.0488BC \tag{25}$$

with final estimates of the transformation parameter and experimental error variance given by $\hat{\lambda} = -0.4133$ and $\hat{\sigma}^2 = 0.001395$.

Table 13: First iteration analysis results for drill advance rate experimental data given in Daniel (1976). Parameters estimated via maximum likelihood with $\hat{\lambda} = -0.7494$ and $\hat{\sigma}^2 = 0.000299$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	0.8778	0.0043	203.1279	0.0000
A	0.0224	0.0043	5.1866	0.0035
B	0.1857	0.0043	42.9616	0.0000
C	0.1009	0.0043	23.3517	0.0000
D	0.0487	0.0043	11.2787	0.0001
AB	-0.0092	0.0043	-2.1310	0.0863
AC	-0.0116	0.0043	-2.6758	0.0440
AD	0.0036	0.0043	0.8442	0.4371
BC	-0.0473	0.0043	-10.9430	0.0001
BD	-0.0055	0.0043	-1.2750	0.2583
CD	-0.0105	0.0043	-2.4396	0.0587

Using the fitted model in equation (25), and final estimates of λ and σ^2 given above, we compute the estimated means and variances of drill advance rate using equations (9) and (10), respectively, at each of the design points in the 2^4 design. These are shown in

Table 14: Second iteration analysis results for drill advance rate experimental data given in Daniel (1976). Parameters estimated via maximum likelihood with $\hat{\lambda} = -0.4698$ and $\hat{\sigma}^2 = 0.000975$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	1.0722	0.0078	137.3100	0.0000
A	0.0323	0.0078	4.1341	0.0025
B	0.2794	0.0078	35.7809	0.0000
C	0.1461	0.0078	18.7087	0.0000
D	0.0746	0.0078	9.5562	0.0000
AC	-0.0138	0.0078	-1.7654	0.1113
BC	-0.0493	0.0078	-6.3118	0.0001

Table 15: Third iteration analysis results for drill advance rate experimental data given in Daniel (1976). Parameters estimated via maximum likelihood with $\hat{\lambda} = -0.4133$ and $\hat{\sigma}^2 = 0.001395$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	1.1199	0.0093	119.9373	0.0000
A	0.0349	0.0093	3.7405	0.0038
B	0.3041	0.0093	32.5686	0.0000
C	0.1579	0.0093	16.9133	0.0000
D	0.0816	0.0093	8.7441	0.0000
BC	-0.0488	0.0093	-5.2226	0.0004

Table 16, while Figure 5 shows the estimated 95% prediction intervals at each of the design points computed via the re-transformation method, as well as from equation (13). It should be noted that we set the reference value for the re-transformation construction method to $t_{10,0.025} = 2.2281$, and for the proposed approach we used $L = 1.96$ (black-shaded region) and $L = 4.47$ (red-shaded region).

Table 16: Estimates for the means and variances of drill advance rate at each of the design points in the 2^4 design using the parameter estimates in Table 15 and proposed estimators in equations (9) and (10).

Run	$\hat{\mu}_Y$	$\hat{\sigma}^2_Y$
(1)	1.737	0.0067
a	1.899	0.0086
b	5.249	0.1527
ab	6.056	0.2290
c	3.118	0.0349
ac	3.497	0.0483
bc	8.470	0.5926
abc	10.096	0.9753
d	2.152	0.0122
ad	2.374	0.0161
bd	7.439	0.4101
abd	8.783	0.6568
cd	4.113	0.0765
acd	4.679	0.1103
bcd	13.054	2.0224
abcd	16.133	3.6908

From Figure 5 it can be observed that changes in the settings of the active experimental factors not only affects the mean of the response, but the variance as well. To assess the relationship between the mean and variance of drill advance rate, Figure 6 shows a plot of $\hat{\mu}_Y$ versus $\hat{\sigma}^2_Y$ using the estimates given in Table 16, exposing a nonlinear relationship. Figure 6 also plots $\log_e \hat{\mu}_Y$ versus $\log_e \hat{\sigma}^2_Y$, where the relationship is linear. If one is interested in a model for σ^2_Y as a function of μ_Y, then, for $\mu_Y \in [1.737, 16.133]$, one could use

$$\sigma^2_Y = e^{-6.5769+2.8336 \log_e \mu_Y}$$

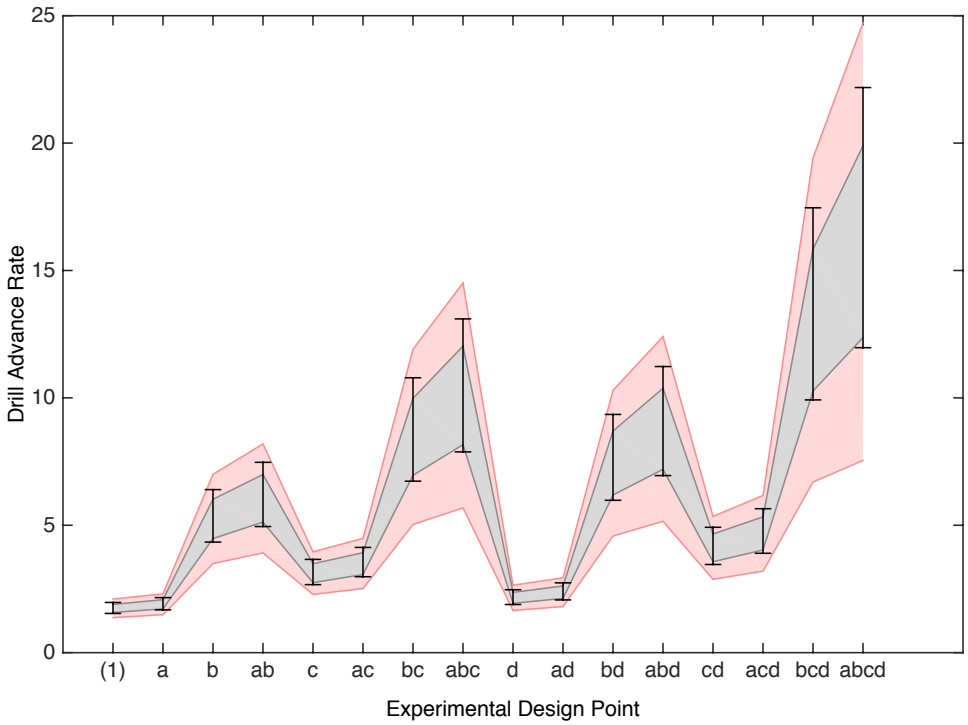

Figure 5: Approximate 95% prediction intervals on drill advance rate at each of the design points in the 2^4 design given in Daniel (1976). The vertical bars correspond to the traditional re-transformed intervals, while the shaded regions correspond the proposed intervals. Black shaded region used $L = z_{0.025} = 1.96$, while red shaded region used $L = 1/\sqrt{0.05} = 4.47$. Re-transformed interval was constructed using the reference value $t_{10,0.025} = 2.2281$.

where the coefficient vector $[-6.5769, 2.8336]'$ was estimated via ordinary least squares. Such a model can be useful when assessing the affect of changes in the levels of the design factors on the variance of Y.

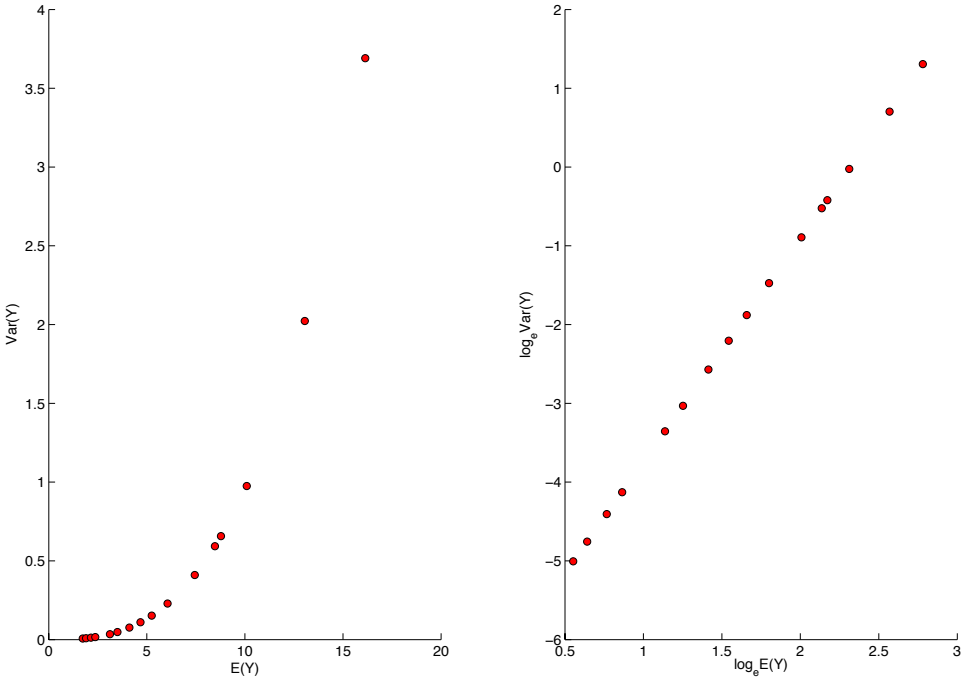

Figure 6: Scatter plots of $\hat{\mu}_Y$ versus $\hat{\sigma}^2_Y$ and $\log_e \hat{\mu}_Y$ versus $\log_e \hat{\sigma}^2_Y$ showing a non-linear relationship (left plot) and linear relationship (right plot) between the mean and variance of drill advance rate.

4.2 Simpson's Wind Tunnel Experiments

The process of automobile wind tunnel testing has the primary objective of characterizing aerodynamic performance in terms of important factors. In this section we analyze the experiments discussed in Simpson et al. (2004), where wind tunnel tests were conducted on a NASCAR Winston Cup Chevrolet Monte Carlo stock car to characterize changes in the *-Lift/Drag* (y) due to the factors: *front ride height* (x_1), *rear ride height* (x_2), *yaw angle* (x_3), and *grill tape* (x_4). Figure 7 shows a picture of the actual car used in the experiments,

where the factors *front ride height*, *rear ride height* and *yaw angle* are illustrated. Figure 8 illustrates the factor *grill tape*, where the high and low levels of the factor correspond to having, or not having, tape placed over the grill, respectively.

For these experiments, changing the ride heights was costly and time consuming, and so x_1 and x_2 were denoted hard-to-change factors, while x_3 and x_4 were denoted easy-to-change factors. Thus, the experiment was executed in a split-plot fashion, resulting in a nested error structure. Table 17 shows the factors studied in this experiment, along with their type and levels, while Table 18 shows the complete set of experiments and corresponding responses. For extensive detail about these experiments, the reader is referred to Simpson et al. (2004).

Table 17: Design factors studied in Simpson's wind tunnel experiments along with their type and levels.

Factor	Type	Low level	High level
Front Ride Height (x_1)	HTC	-0.5in	+0.5in
Rear Ride Height (x_2)	HTC	-1.0in	+1.0in
Yaw Angle (x_3)	ETC	$-3°$	$+1°$
Grill Tape (x_4)	ETC	0%	100%

To analyze the experiments in Table 18, we consider the linear model for the split-plot design, where it is assumed that, for some λ, we have

$$\mathbf{y}(\lambda) = \mathbf{X}\boldsymbol{\beta} + \mathbf{M}\boldsymbol{\delta} + \boldsymbol{\epsilon}$$

where

$\mathbf{y}(\lambda)$: $N \times 1$ vector of transformed responses,

$\mathbf{X}$: $N \times p$ known design matrix of full rank,

$\boldsymbol{\beta}$: $p \times 1$ unknown vector of fixed effects,

$\mathbf{M}$: $N \times w$ known indicator matrix,

Table 18: Wind tunnel split-plot design and -Lift/Drag response data given in Simpson et al. (2004). Replicated whole-plots have response entries under both the y_I and y_{II} columns, representing the 1^{st} and 2^{nd} replicates, respectively.

WP	x_1	x_2	x_3	x_4	y_I	y_{II}
	-1	-1	-1	-1	0.861	0.863
	-1	-1	1	-1	0.797	0.830
1	-1	-1	-1	1	1.033	1.051
	-1	-1	1	1	0.959	0.955
	-1	-1	0	0	0.887	0.932
	1	-1	-1	-1	0.721	0.715
	1	-1	1	-1	0.708	0.700
2	1	-1	-1	1	0.867	0.875
	1	-1	1	1	0.825	0.830
	1	-1	0	0	0.793	0.788
	-1	1	-1	-1	0.955	0.958
	-1	1	1	-1	0.852	0.872
3	-1	1	-1	1	1.118	1.124
	-1	1	1	1	1.061	1.074
	-1	1	0	0	1.006	1.012
	1	1	-1	-1	0.831	0.843
	1	1	1	-1	0.807	0.810
4	1	1	-1	1	0.996	0.994
	1	1	1	1	0.952	0.956
	1	1	0	0	0.926	0.918
	0	0	-1	-1	0.853	-
	0	0	1	-1	0.785	-
5	0	0	-1	1	1.004	-
	0	0	1	1	0.939	-
	0	0	0	0	0.912	-

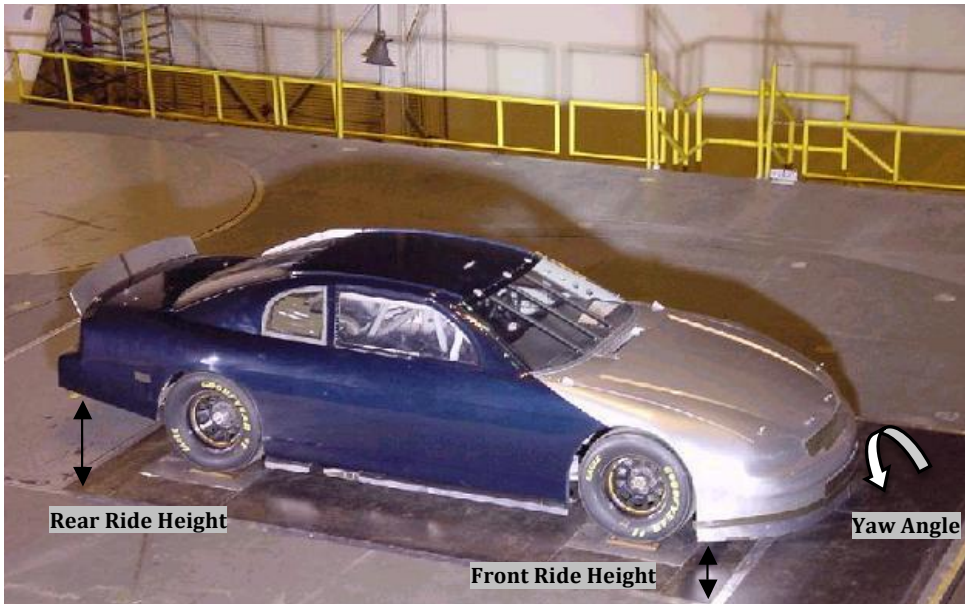

Figure 7: Stock car used in wind tunnel experiments illustrating the experimental factors studied.

$\boldsymbol{\delta} \sim MVN(\mathbf{0}_{w\times 1}, \sigma^2_\delta \mathbf{I}_{w\times w})$: $w \times 1$ random vector of errors,

$\boldsymbol{\epsilon} \sim MVN(\mathbf{0}_{N\times 1}, \sigma^2_\epsilon \mathbf{I}_{N\times N})$: $N \times 1$ random error vector

and w denotes the number of whole-plots, p denotes the number of fixed-effect terms, and $N = \sum_{i=1}^{w} m_i$ denotes the total number of subplots, where m_i is the number of subplots in whole-plot i. The matrix $\mathbf{M}$ takes the form

$$\mathbf{M} = blkdiag(\mathbf{1}_{m_1}, \mathbf{1}_{m_2}, ..., \mathbf{1}_{m_w}), \tag{26}$$

where *blkdiag* implies a block diagonal matrix. It is assumed that $Cov(\boldsymbol{\delta}, \boldsymbol{\epsilon}) = \mathbf{0}_{w\times N}$, and thus, it is easily shown that $E[\mathbf{y}(\lambda)] = \mathbf{X}\boldsymbol{\beta}$ and

$$Var[\mathbf{y}(\lambda)] = \sigma^2_\epsilon\{\mathbf{I}_{N\times N} + \eta \mathbf{M}\mathbf{M}^{'}\} = \sigma^2_\epsilon \mathbf{D}$$

where $\eta = \sigma^2_\delta/\sigma^2_\epsilon$ denotes the variance ratio.

To estimate the transformation parameter λ and variance ratio η using the exper-

Grill Tape (-1)

Grill Tape (+1)

Figure 8: **Top**: Setting for *grill tape* at low level (0% covered). **Bottom:** Setting for *grill tape* at high level (100% covered).

imental data, one can evaluate the residual log-likelihood of the untransformed response, or

$$\ell(\lambda, \eta|\mathbf{y}) = -\frac{(N-p)}{2}\log_e(\mathbf{e}'\mathbf{D}^{-1}\mathbf{e}) - \frac{1}{2}\log_e(|\mathbf{D}|) - \frac{1}{2}\log(|\mathbf{X}'\mathbf{D}^{-1}\mathbf{X}|) + (\lambda-1)\sum_{i=1}^{N}\log_e y_i \tag{27}$$

across the range of possible values of λ and η, and retain those values that achieve the maximum, or

$$[\hat{\lambda}, \hat{\eta}] = \arg\max_{\lambda,\eta}\{\ell(\lambda, \eta|\mathbf{y})\} \tag{28}$$

where $\mathbf{e} = \mathbf{y}(\lambda) - \mathbf{X}\hat{\boldsymbol{\beta}}$ and $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{D}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{D}^{-1}\mathbf{y}(\lambda)$, with the constraint $\eta > -\frac{1}{\nu^*}$, and ν^* is the maximum eigenvalue of $\mathbf{M}\mathbf{M}'$. The addition of this constraint on η ensures that the estimate for the matrix $\mathbf{D}$ is always positive definite. The estimates $\hat{\lambda}$ and $\hat{\eta}$ are then maximum (residual) likelihood estimates of the transformation parameter and variance ratio, respectively. Once λ and η have been estimated, the REML estimate for σ_ϵ^2 is given by

$$\hat{\sigma}_\epsilon^2 = \frac{(\mathbf{y}(\hat{\lambda}) - \mathbf{X}\mathbf{b})'(\mathbf{I} + \hat{\eta}\mathbf{M}\mathbf{M}')^{-1}(\mathbf{y}(\hat{\lambda}) - \mathbf{X}\mathbf{b})}{N-p},$$

where

$$\mathbf{b} = [\mathbf{X}'(\mathbf{I} + \hat{\eta}\mathbf{M}\mathbf{M}')^{-1}\mathbf{X}]^{-1}\mathbf{X}'(\mathbf{I} + \hat{\eta}\mathbf{M}\mathbf{M}')^{-1}\mathbf{y}(\hat{\lambda})$$

is the generalized least squares estimator for the parameter vector $\boldsymbol{\beta}$ evaluated at $\hat{\lambda}$ and $\hat{\eta}$. It then follows that $\hat{\sigma}^2$ in equations (9) and (10) is given by $\hat{\sigma}^2 = \hat{\sigma}_\epsilon^2(1 + \hat{\eta})$.

Similar to the analysis of Daniel's drill advance rate experiments, we performed iterative analyses on Simpson's wind tunnel experiments, each time re-estimating the unknown model parameters as terms with p-values greater than $\alpha = 0.05$ are dropped from the model. This yielded the final results given in Table 19. Notice that inference on $\boldsymbol{\beta}$ was accomplished

via t-tests (i.e., $t_j = \hat{\beta}_j / \hat{\sigma}_{\hat{\beta}_j}$), with degrees of freedom equal to 5 for whole-plot terms (i.e., x_1, x_2, and x_1x_2), and 33 for subplot terms (i.e., x_3, x_4, and x_1x_3). Based upon the estimates given in Table 19, the final fitted model is given by

$$\hat{y}(\hat{\lambda}) = -0.1108 - 0.0652x_1 + 0.0582x_2 + 0.0135x_1x_2 - 0.0287x_3 + 0.0879x_4 + 0.0093x_1x_3 \quad (29)$$

with $\hat{\lambda} = 0.0363$, $\hat{\eta} = 0.0518$, and $\hat{\sigma}^2_\epsilon = 0.000219$.

Table 19: Final analysis results for wind tunnel experimental data given in Simpson et al. (2004). Transformation parameter, and variance parameters estimated as $\hat{\lambda} = 0.0363$, $\hat{\eta} = 0.0518$ and $\hat{\sigma}^2 = 0.000219$, respectively.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	-0.1108	0.0025	-44.76	0.0000
x_1	-0.0652	0.0026	-24.84	0.0000
x_2	0.0582	0.0026	22.15	0.0000
x_1x_2	0.0135	0.0026	5.13	0.0037
x_3	-0.0287	0.0025	-11.64	0.0000
x_4	0.0879	0.0025	35.64	0.0000
x_1x_3	0.0093	0.0026	3.54	0.0007

Using the fitted model in equation (29), and corresponding estimates for λ, η and σ^2_ϵ given above, Table 20 shows estimates for the mean and variance of Y = *-Lift/Drag* at each design point in the 2^4 design computed from equations (9) and (10), respectively. Further, Figure 9 shows approximate 95% prediction intervals at each of the design points computed via the re-transformation approach with $h = t_{5,0.025} = 2.5706$, as well as from the proposed approach with $L = 1.96$ (black-shaded region) and $L = 4.47$ (red-shaded region). If one's goal is to maximize *-Lift/Drag*, then Figure 9 suggests setting experimental factors to the levels shown in Table 21.

It should be clear that one can maximize *-Lift/Drag* by applying the settings given in Table 21; however, the approximate 95% prediction intervals in Figure 9 also suggest that

Table 20: Estimates for the means and variances of -Lift/Drag ($-L/D$) at each of the design points in the 2^4 design using the parameter estimates in Table 19 and proposed estimators in equations (9) and (10).

Design Point	x_1	x_2	x_3	x_4	$\hat{\mu}_{-L/D}$	$\hat{\sigma}^2_{-L/D}$
(1)	-1	-1	-1	-1	0.8689	0.00018
a	1	-1	-1	-1	0.7276	0.00012
b	-1	1	-1	-1	0.9505	0.00021
ab	1	1	-1	-1	0.8409	0.00016
c	-1	-1	1	-1	0.8049	0.00015
ac	1	-1	1	-1	0.6996	0.00012
bc	-1	1	1	-1	0.8807	0.00018
abc	1	1	1	-1	0.8087	0.00015
d	-1	-1	-1	1	1.0363	0.00025
ad	1	-1	-1	1	0.8688	0.00018
bd	-1	1	-1	1	1.1329	0.00029
abd	1	1	-1	1	1.0031	0.00023
cd	-1	-1	1	1	0.9604	0.00021
acd	1	-1	1	1	0.8355	0.00016
bcd	-1	1	1	1	1.0502	0.00025
abcd	1	1	1	1	0.9649	0.00022

Table 21: Design factor settings that yields the greatest expected -Lift/Drag.

Factor	Level
Front Ride Height (x_1)	-0.5in
Rear Ride Height (x_2)	1.0in
Yaw Angle (x_3)	$-3°$
Grill Tape (x_4)	100%

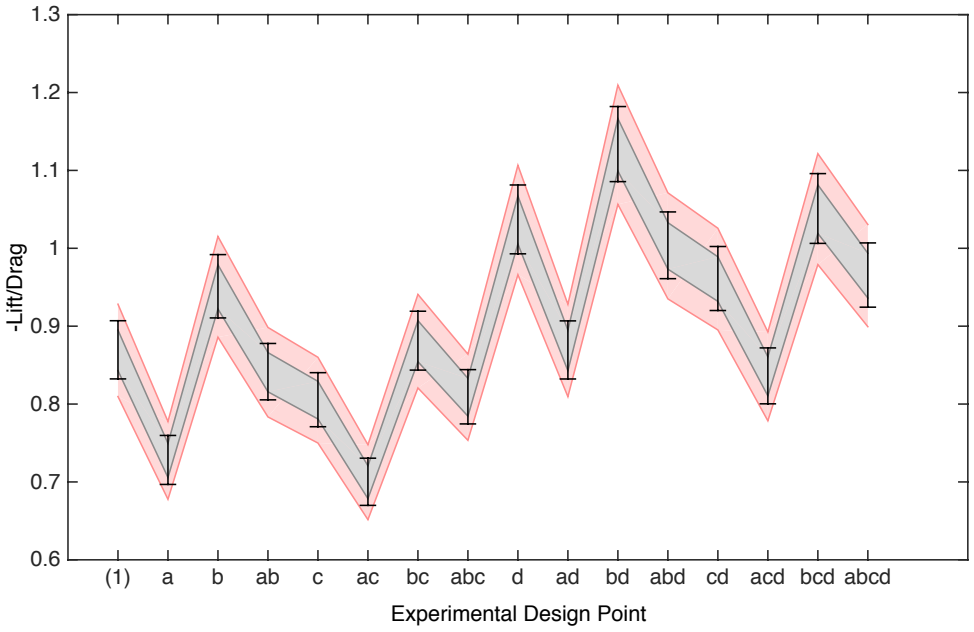

Figure 9: Approximate 95% prediction intervals on $-Lift/Drag$ at each of the design points in the 2^4 design given in Simpson et al. (2004). The vertical bars correspond to the traditional re-transformed intervals, while the shaded regions correspond the proposed intervals. Black shaded region used $L = z_{0.025} = 1.96$, while red shaded region used $L = 1/\sqrt{0.05} = 4.47$. Re-transformed interval was constructed using the reference value $t_{5,0.025} = 2.5706$.

the variance of -*Lift/Drag* is also maximized at these factor level settings. To assess the relationship between the mean and variance of -*Lift/Drag*, Figure 10 shows a scatter plot of $\hat{\mu}_Y$ versus $\hat{\sigma}^2_Y$ given in Table 20, suggesting a linear relationship between the mean and variance of -*Lift/Drag*. Thus, changes in the settings of the active experimental factors not only influences the mean of -*Lift/Drag*, but its variance as well. A simple model for the relationship between $\mu_Y \in [0.6996, 1.1329]$ and σ^2_Y is then given by

$$\sigma^2_Y = -0.00018 + 0.00041\mu_Y$$

where, as before, the coefficient vector $[-0.00018, 0.00041]'$ was estimated using ordinary least squares.

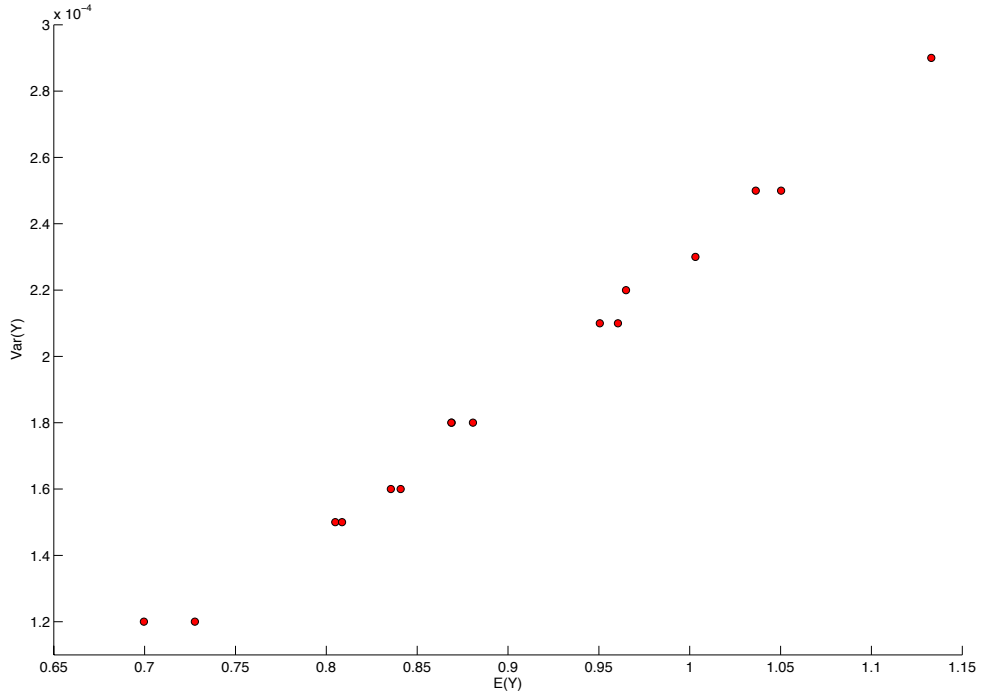

Figure 10: Scatter plot of $\hat{\mu}_Y$ versus $\hat{\sigma}^2_Y$ showing a linear relationship between the mean and variance of -Lift/Drag.

Our analyses of the experimental data given in Daniel (1976) and Simpson et al.

(2004) in the last two subsections illustrates some significant practical value in our proposed approach. Having estimates for both the mean *and* variance of the original response variable permits the construction of a reliable prediction interval estimate on this variable, which relative to a simple point estimate, is much more useful and informative in practice. Further, although the original experiments are likely conducted to draw inference on $E(Y)$, the proposed method allows one to assess the relationship between $E(Y)$ and $Var(Y)$, permitting inference to be drawn on $Var(Y)$ from the same set of experiments.

5 Summary and Discussion

Motivated by electron microscopy experiments, in this paper we developed an approximate $100(1-\alpha)\%$ prediction interval on the response variable Y, where it was assumed that a linear model was fit using a transformed version of Y, with the transformation type contained in the Box-Cox family. We derived a closed-form approximation to the k^{th} moment of Y, which was then used to estimate the mean and variance of Y, given parameter estimates obtained from fitting the model in the transformed units. Chebychev's inequality was then exploited to construct a prediction interval estimate on Y.

Using Monte Carlo simulation, we assessed the performance of the proposed estimators for μ_Y and σ^2_Y, as well as the proposed prediction interval estimator, relative to that obtained by employing the more traditional re-transformation interval construction approach. Results suggest that our proposed estimators for μ_Y and σ^2_Y yield acceptable performance with respect to the RMSE, with values that decrease with the variance of Y. Further, our simulation results suggest that, on average, when both prediction interval estimators yield approximately the same expected coverage, then the proposed method will yield smaller interval widths. Additionally, regardless of whether or not the expected coverages are the same, on average, the proposed interval estimator will generally yield width estimates with smaller variance than that provided by the traditional interval estimator. Finally, in our simulations

we found that the re-transformed intervals generally provide less than the nominal coverage probability of $1-\alpha$. Thus, constructing prediction intervals using the proposed method with $L = 1/\sqrt{\alpha}$ will yield a conservative interval, permitting better control of the type I error rate α, while also providing interval width estimates with greater relative precision. As a result, we recommend using the proposed interval estimator in practice.

One obvious advantage to using our proposed methodology is that data analyses can be performed for any λ without sacrificing interpretation. This is because the fitted model can always be converted back to the original units of measurement. We demonstrated this on the microscopy experiments discussed in Section 1, as well as on the experimental data sets given in Daniel (1976) and Simpson et al. (2004). A second major advantage to employing the proposed methodology is that reduced-bias estimates for μ_Y and σ_Y^2 are available and can be used to empirically determine the type of relationship that exists between the mean and variance of Y. This was also demonstrated using the experimental data sets in Daniel (1976) and Simpson et al. (2004), where a nonlinear and linear relationship was found to exist between the mean and variance of Y, respectively. A model capturing the relationship between the mean and variance of Y could allow for inference to be drawn on both of these parameters, at given levels of the experimental factors, using the same set of experiments.

Acknowledgements

We would like to thank the Editor and an anonymous reviewer for their helpful comments that led to great improvements to this paper.

References

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B 26*, 211–252.

Carroll, R. J. and D. Rupert (1981). On prediction and the power transformation family. *Biometrika 68*, 609–615.

Cramer, H. (1970). *Random Variables and Probability Distributions*. Cambridge University Press.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*. Wiley, New York.

Draper, N. and D. Cox (1969). On distributions and their transformations to normality. *Journal of the Royal Statistical Society, Series B 31*, 472–476.

Freeman, J. and R. Modarres (2006). Inverse box-cox: The power-normal distribution. *Statistics and Probability Letters 76*, 764–772.

Gurka, M. J., L. J. Edwards, K. E. Muller, and L. L. Kupper (2006). Extending the box-cox transformation to the linear mixed model. *Journal of the Royal Statistical Society, Series A 169*, 273–288.

Han, D. and F. Tsung (2006). A reference free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *Journal of the American Statistical Association 101*, 368–386.

Land, C. E. (1974). Confidence interval estimation for means after data transformation to normality. *Journal of the American Statistical Association 69*, 795–802.

Miller, D. (1984). Reducing transformation bias in curve fitting. *The American Statistician 38*, 124–126.

Montgomery, D. C. (2009). *Design and Analysis of Experiments*. John Wiley & Sons.

Perry, M., J. Kharoufeh, S. Shekhar, J. Cai, and M. R. Shankar (2012). Statistical characterization of nanostructured materials from severe plastic deformation in machining. *IIE Transactions 44*, 534–550.

Sakia, R. M. (1988). Application of the box-cox transformation technique to linear balanced mixed analysis of variance models with a multi-error structure. *Unpublished PhD Thesis, Universitaet Hohenheim, FRG*.

Sakia, R. M. (1990). Retransformation bias: A look at the box-cox transformation to linear balanced mixed anova models. *Metrika 37*, 345–351.

Shumway, R., A. Azari, and P. Johnson (1989). Transformation for environmental data with detection limits. *Technometrics 31*, 347–356.

Simpson, J., S. Kowalski, and D. Landman (2004). Experimentation with randomization restrictions: Targeting practical implementation. *Quality and Reliability Engineering International 20*, 481–495.

Solomon, P. J. (1985). Transformations for components of variance and covariance. *Biometrika 31*, 233–239.

Taylor, J. (1986). The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association 81*, 114–118.

Part III
A Prediction Interval Estimator for the Original Response when using Manly's Exponential Transformations

Abstract

Motivated by A-10 single-engine aircraft climb experiments, we develop an approximate prediction interval on the response variable Y, where it is assumed that a normal-theory linear model is fit using a transformed version of Y, and the transformation type is contained in the Manly exponential family. We derive a closed-form approximation to the k^{th} moment of the original response variable Y, which is then used to estimate the mean and variance of Y, given parameter estimates obtained from fitting the model in the transformed domain. Chebychev's inequality is then used to construct a $100(1-\alpha)\%$ prediction interval estimator on Y. Using Monte Carlo simulation, we assess the width performance of our proposed Chebychev prediction interval, relative to that obtained by employing a more common interval construction approach. General results suggest that, for a given level of expected coverage, the proposed interval estimator will achieve a smaller mean and variance of the interval width estimates, especially as the number of degrees of freedom beyond that required to estimate model terms is small.

1 Introduction and Motivation

Data transformations are common in statistical model fitting. When symmetry or approximate normality of the model error terms is desired, a power transformation on the response variable is often used. The technique of Box and Cox (1964) is one of the most popular methods for systematically determining the appropriate power transformation on the response variable Y. The Box-Cox family of power transformations is defined as

$$Y(\lambda) = \begin{cases} \frac{Y^{\lambda}-1}{\lambda} & \lambda \neq 0 \\ \log_e(Y) & \lambda = 0 \end{cases} \tag{1}$$

where Y is the original response variable, λ is the transformation parameter, and $Y(\lambda)$ is the transformed response variable.

While there are a number of methods to estimate λ, the method of maximum likelihood is typically used. With λ adequately specified, the transformed response variable, $Y(\lambda)$, is assumed to follow a normal distribution with mean μ and variance σ^2, at least approximately. Using the Box-Cox transformations in normal theory linear model fitting exercises then ensures that the necessary model assumptions on the error terms are at least approximately met.

Although this method can lead to an improved model fit in the transformed domain, it can be problematic when the ultimate goal of the model is prediction in the original units of observation. Since the model is fit in transformed space, any predictions obtained from the model will also be in transformed space. Typically, the predicted values are "re-transformed" back to the original units using the inverse of the applied transformation. Additionally, if prediction intervals are desired, the re-transformation approach is also typically used, i.e., the lower and upper bounds of the prediction interval are constructed in the transformed space and then re-transformed to the original units.

Unfortunately, for a nonlinear transformation, a re-transformation can lead to signif-

icant bias in both the predicted values and the prediction intervals in the original units of observation. Theoretically, this bias stems from the fact that $E(Y^k)$ is a nonlinear function of μ and σ^2, e.g., see Land (1974). This bias was investigated by several authors, including Taylor (1986), Shumway et al. (1989), Sakia (1990), and more recently by Perry and Walker (2015), where the latter authors proposed reduced-bias prediction intervals on the original response Y when fitting linear models using the Box-Cox transformations. In particular, these authors derived a closed-form approximation to $E(Y^k)$ $(k = 1, 2, ...)$, and then used this estimator, along with estimates of λ, μ and σ^2 obtained from fitting the model in the transformed domain, to find estimates for the mean and variance of Y, or μ_Y and σ_Y^2. Using moment approximations, Chebychev's inequality was then exploited to construct approximate $100(1-\alpha)\%$ prediction intervals on the original response variable.

Historically, the Box-Cox transformations have proven to be quite popular and effective. However, this class of transformations is limited in that it requires $Y > 0$. To accommodate negative values of the response variable, Box and Cox also proposed the shifted Box-Cox transformations. This approach introduces a second parameter that shifts the original response values to positive space, and then applies the one-parameter Box-Cox transformation to the shifted response data. This class of transformations is defined as

$$Y(\lambda) = \begin{cases} \frac{(Y+\lambda_s)^\lambda - 1}{\lambda} & \lambda \neq 0, Y > -\lambda_s \\ \log_e(Y+\lambda_s) & \lambda = 0, Y > -\lambda_s \end{cases} \tag{2}$$

where, Y is the original response variable, λ is the transformation parameter, λ_s is the shift parameter, and $Y(\lambda)$ is the transformed response variable. For a shift parameter equal to 0, the method reduces to the one-parameter Box-Cox transformations.

While the joint estimation of λ and λ_s can be problematic (e.g., see Atkinson (1983)), the shift parameter itself can essentially be viewed as a constant for the purposes of transformation (e.g., see Sakia (1992)) and thus chosen by setting $\lambda_s = -\min(Y) + \delta$, if $\min(Y) \leq 0$, where $\delta > 0$ is some small value to ensure that $Y + \lambda_s$ is strictly positive. The value of λ_s

can also be chosen based on pre-existing knowledge of the process of interest, just as long as $Y + \lambda_s$ is strictly positive for all Y. When viewed this way, the approximations developed in Perry and Walker (2015) can still be used to obtain estimates for μ_Y and σ_Y^2. The only difference being the need to subtract from $E(Y)$ the value of λ_s, since a shift in the distribution will leave the variance unchanged.

To demonstrate, consider the A-10 single-engine climb experiments performed by Hutto and Simpson (2013). In these experiments, the *ClimbRate* of A-10 jets is studied as a function of changes in the experimental factors *Temperature* (factor A), *Altitude* (factor B), *GearWeight* (factor C), *Flaps* (factor D), and *LandingGear* (factor E). The experiment consisted of 82 runs across various levels of the experimental factors, though run 44 contained missing data, and thus, was omitted from the analysis, leaving a sample size of $n = 81$. We note that *ClimbRate* can take on both positive and negative values.

For this set of experiments, a linear model is desired of the form $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. While not conducted as an orthogonal design, the solution to $(\mathbf{X}'\mathbf{X})^{-1}$ exists and thus classic regression analysis can be carried out to produce the fitted model $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ where the vector of regression coefficients is estimated by ordinary least squares (OLS), or $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$. Iterative OLS analysis of the untransformed response yielded the following final fitted model

$$\hat{y} = -0.9831 + 1.6210B - 2.6991B^2 \tag{3}$$

where a linear and quadratic term in *Altitude* were the only significant factors at the 5% level. Table 1 shows the results of the analysis for the significant model, while Figure 1 shows a normal probability plot of the residuals, and a scatter plot of the residuals versus the factor *Altitude*. Notice that the residuals appear to be left skewed and the constant variance assumption appears to be violated, i.e., the variance of *ClimbRate* appears to decrease as *Altitude* increases. The Shapiro-Wilk normality test on the residuals yielded a p-value of 0.0135, suggesting the residuals are non-normal.

Table 1: OLS analysis results of the model with significant terms. Experimental error variance estimate is $\hat{\sigma}^2 = 2.3796$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	-0.9831	0.2480	-3.9636	0.0001
B	1.6210	0.2746	5.9029	0.0000
B^2	-2.6991	0.4576	-5.8978	0.0000

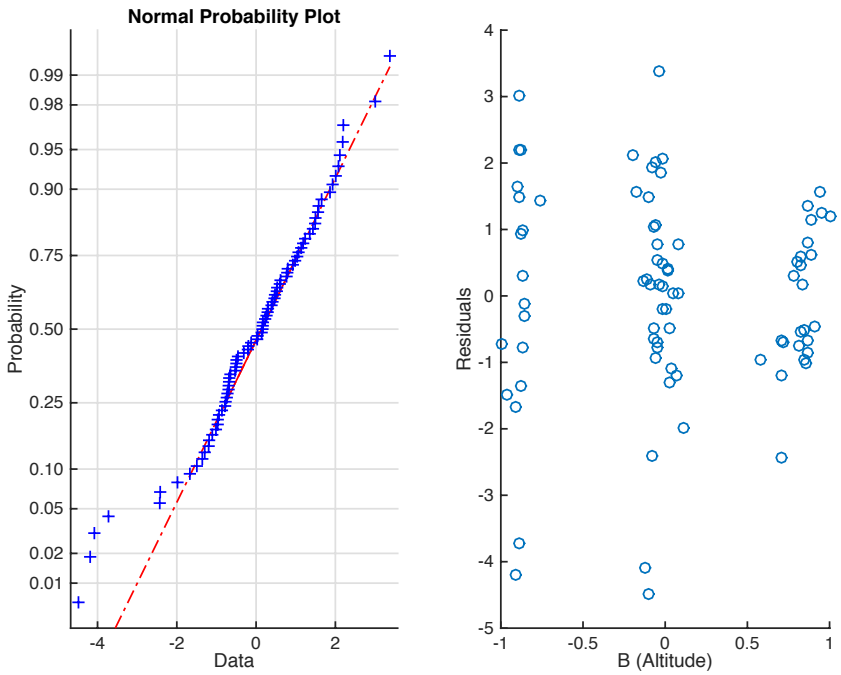

Figure 1: Normal probability plot of residuals and residuals versus factor *Altitude* from analysis of *ClimbRate* without transformation. NPP of residuals suggests a negative skew, and residuals versus *Altitude* suggests non-constant variance, i.e., a decrease in the variance of *ClimbRate* as *Altitude* increases.

To correct for violations in model assumptions, the Box-Cox transformations are often applied. For these experiments, since the response variable is not strictly positive, the shifted Box-Cox transformations defined in equation (2) must be employed. As discussed earlier, the shift parameter, λ_s, will be estimated by $-\min(Y) + \delta$ where δ is some very small positive value necessary to ensure $Y + \lambda_s > 0$ for all Y. For this analysis we use $\delta = 2 \times 10^{-16}$, and thus the shift parameter is estimated as $\hat{\lambda}_s = 8.8970 + 2 \times 10^{-16}$. The analysis is carried out by first estimating the full main-effects model plus a quadratic term in *Altitude* (results shown in Table 2), then retaining only the significant terms at the 5% level and estimating the final fitted model (results shown in Table 3).

Table 2: First iteration analysis results for Hutto-Simpson climb rate experimental data using proposed methodology with the Shifted Box-Cox transformation. Parameters estimated via maximum likelihood with $\lambda = 0.7416$, $\lambda_s = 8.8970 + 2 \times 10^{-16}$, and $\hat{\sigma}^2 = 1.1101$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	4.8631	0.1806	29.9340	0.0000
A	0.0963	0.3105	0.3100	0.7574
B	1.1900	0.3087	3.8546	0.0002
B^2	-1.6665	0.3593	-4.6387	0.0000
C	0.1713	0.2243	0.7638	0.4475
D	0.1270	0.1244	1.0208	0.3108
E	-0.0602	0.1615	-0.3727	0.7105

Table 3: Second iteration analysis results for Hutto-Simpson climb rate experimental data using proposed methodology with the Shifted Box-Cox transformation. Parameters estimated via maximum likelihood with $\hat{\lambda} = 0.7554$, $\hat{\lambda}_s = 8.8970 + 2 \times 10^{-16}$, and $\hat{\sigma}^2 = 1.1190$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	4.9765	0.1701	29.2601	0.0000
B	1.1268	0.1883	5.9836	0.0000
B^2	-1.7971	0.3138	-5.7265	0.0000

The final fitted model using the shifted Box-Cox transformations is then given by

$$\hat{y}(\hat{\lambda}, \hat{\lambda}_s) = 4.9765 + 1.1268B - 1.7971B^2$$

with a variance parameter estimate of $\hat{\sigma}^2 = 1.1190$ and transformation parameter estimates of $\hat{\lambda} = 0.7554$ and $\hat{\lambda}_s = 8.8970 + 2 \times 10^{-16}$. To justify use of the transformation, Figure 2 shows the 95% confidence interval on the transformation parameter λ for the final fitted model. Notice that this interval does not contain 1, which suggests the need for a data transformation.

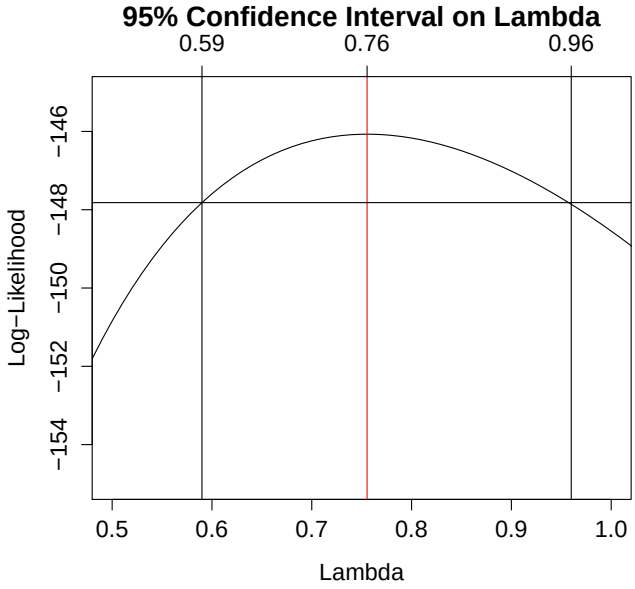

Figure 2: 95% Confidence Interval on the transformation parameter λ.

Although the confidence interval on λ in Figure 2 suggests that a transformation is appropriate, the residual plots in Figure 3 suggest that the shifted Box-Cox transformation was not effective in achieving approximate normality of the residuals, i.e., the residuals remain left-skewed. This is further evidenced by the Shapiro-Wilk normality test on the residuals, which suggests that the resulting residuals cannot be assumed to be normally distributed (p-value = 0.0003). Additionally, this transformation was not effective at cor-

recting the non-constant variance across the factor *Altitude*. The transformation's inability to correct for violations in the underlying assumptions may be attributed to the underlying distributional assumptions of Box-Cox transformations. The transformation inherently assumes the original distribution as power-normal (e.g., see Freeman and Modarres (2006)), and, given the parameter space for λ suggested by Box and Cox (1964), i.e., $\lambda \in [-2, 2]$, is then appropriate to use primarily in cases of positively skewed distributions. Given the negatively skewed nature still observed in the final residual plots, it is apparent the shifted Box-Cox transformation may not be appropriate for this particular data set.

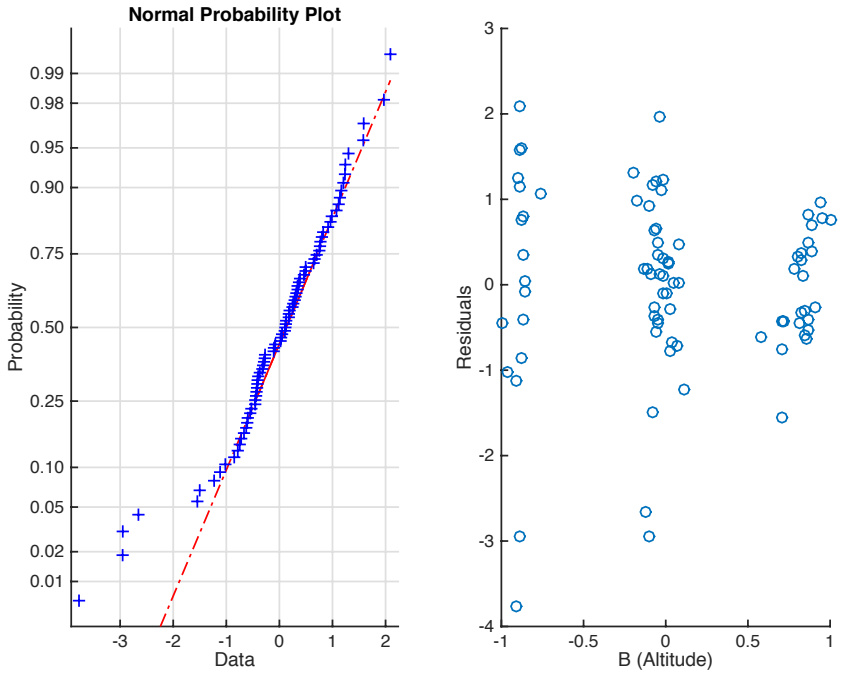

Figure 3: Residual plots from final model fit using the shifted Box-Cox transformations. Normal probability plot of residuals (left plot) and residuals versus altitude (right plot).

As an alternative to the two-parameter shifted Box-Cox transformations, Manly (1976) proposed a one-parameter exponential transformation that permits both positive and negative response values. The idea behind Manly's transformation is to first take the natural exponential of Y, and then apply the one-parameter Box-Cox transformations to

this variable. In particular, this class of transformations is defined by

$$Y(\gamma) = \begin{cases} \frac{e^{\gamma Y}-1}{\gamma} & \gamma \neq 0 \\ Y & \gamma = 0 \end{cases} \tag{4}$$

where Y is the original response variable, γ is the transformation parameter, and $Y(\gamma)$ is the transformed response variable. The special case of $\gamma = 0$ results in no transformation.

Manly's class of exponential transformations is not of the same family of transformations as the Box-Cox power transformations, though it can be used similarly to achieve approximate normality of the error terms. This transformation addresses the issue of non-positive Y values and can be applied to all real-valued response variables without limitation, using only one transformation parameter γ. Additionally, this transformation does not rely on the underlying power-normal distributional assumption as with the Box-Cox transformations, and thus can be used as an appropriate transformation for both positively and negatively skewed data.

These characteristics point towards the Manly exponential transformation being a more robust alternative to the Box-Cox transformations. Unfortunately, this method is rarely used in practice today. This might simply be due to the fact that this class of transformations is not as well known as the Box-Cox family. Or, perhaps it's due to the assumed rarity of encountering non-positive and/or negatively skewed variables of interest that would justify its use over the Box-Cox transformations, as well as the lack of interpretation of the transformation parameter γ that is inherent in the Box-Cox transformation parameter λ. While many values of λ in the Box-Cox transformations have easy interpretation, e.g., natural log transformation ($\lambda = 0$), square root transformation ($\lambda = 1/2$), inverse transformation ($\lambda = -1$), the parameter γ in the Manly case lacks this interpretation as there is no direct relationship between specific values of λ and values of γ. However, values of γ do have some meaningful interpretation, i.e., values of $\gamma < 0$ are sufficient to transform positively skewed data, values of $\gamma > 0$ are sufficient to transform negatively skewed data, and larger

magnitudes of γ in either direction imply a more exaggerated underlying skewness in the data.

We now apply Manly's exponential transformation to the single-engine climb data, where the analysis is performed similarly to those earlier. The full main-effects model plus quadratic term in *Altitude* is first estimated (results given in Table 4), then retaining only the significant terms at the 5% level and re-estimating produces the results in Table 5 and the final fitted model given by

$$\hat{y}(\hat{\gamma}) = -0.8374 + 1.0849B - 2.0475B^2 \tag{5}$$

with a variance parameter estimate of $\hat{\sigma}^2 = 1.3783$ and a transformation parameter estimate of $\hat{\gamma} = 0.1111$. A 95% confidence interval constructed on the transformation parameter γ is shown in Figure 4 and reveals that, since $\gamma = 0$ is not contained in the interval, the transformation may be appropriate. Additionally, the confidence interval contains values of $\gamma > 0$, suggesting the transformation was necessary to account for negative skew.

Table 4: First iteration analysis results for Hutto-Simpson climb rate experimental data using proposed methodology with the Manly exponential transformation. Parameters estimated via maximum likelihood with $\hat{\gamma} = 0.1129$ and $\hat{\sigma}^2 = 1.4343$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	-0.8054	0.2052	-3.9244	0.0002
A	0.0897	0.3529	0.2542	0.8000
B	1.1609	0.3509	3.3081	0.0015
B^2	-2.1171	0.4084	-5.1840	0.0000
C	0.1077	0.2549	0.4224	0.6739
D	-0.0312	0.1414	-0.2204	0.8262
E	-0.0914	0.1836	-0.4977	0.6202

The resulting normal probability plot of the residuals obtained using the final fitted model is illustrated in Figure 5, and suggests the approximate normality assumption on the error term appears to be reasonably met. This is further evidenced by the Shapiro-Wilk

Table 5: Second iteration analysis results for Hutto-Simpson climb rate experimental data using proposed methodology with the Manly exponential transformation. Parameters estimated via maximum likelihood with $\gamma = 0.1111$ and $\hat{\sigma}^2 = 1.3783$.

	$\hat{\beta}$	$se(\hat{\beta})$	t	p
Constant	-0.8374	0.1888	-4.4363	0.0000
B	1.0849	0.2090	5.1908	0.0000
B^2	-2.0475	0.3483	-5.8787	0.0000

normality test performed on these residuals, which does not conclude any significant deviation from normality (p-value = 0.1709). Additionally, from the plot of residuals versus the factor *Altitude* in Figure 5, it appears as if the non-constant variance issue has been nearly resolved as a result of performing the transformation. For reference, Figure 6 shows a comparison of the residuals versus altitude plots for the untransformed case and the exponentially transformed case.

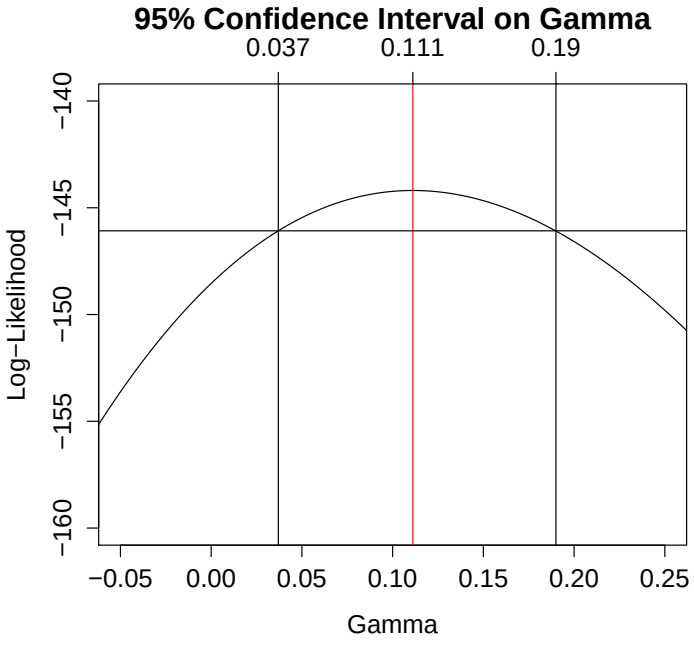

Figure 4: 95% Confidence Interval on the transformation parameter γ.

Based on the above discussions, we conclude that Manly's exponential transforma-

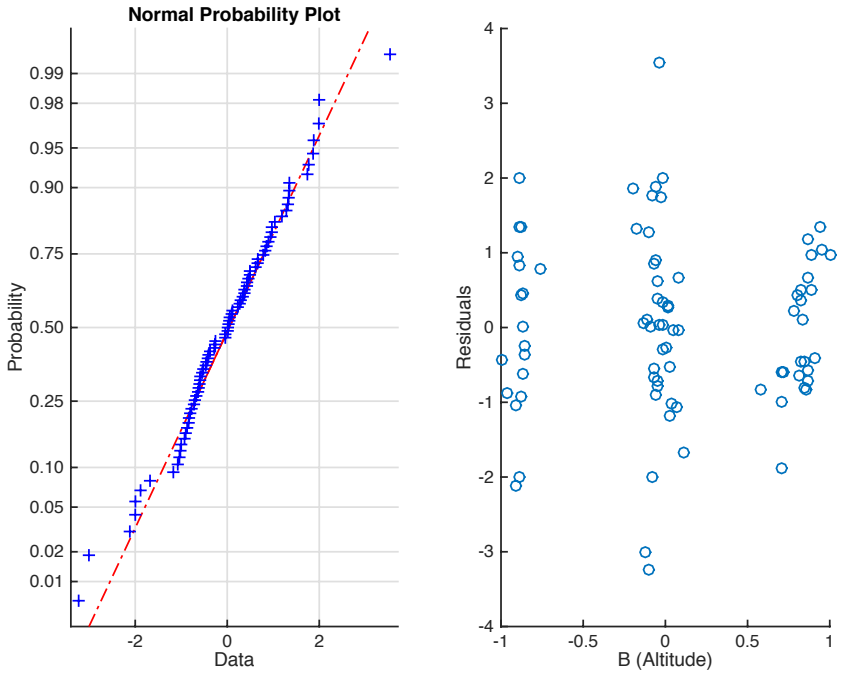

Figure 5: Residual plots from final model fit using the Manly transformations. Normal probability plot of residuals (left plot) and residuals versus altitude (right plot).

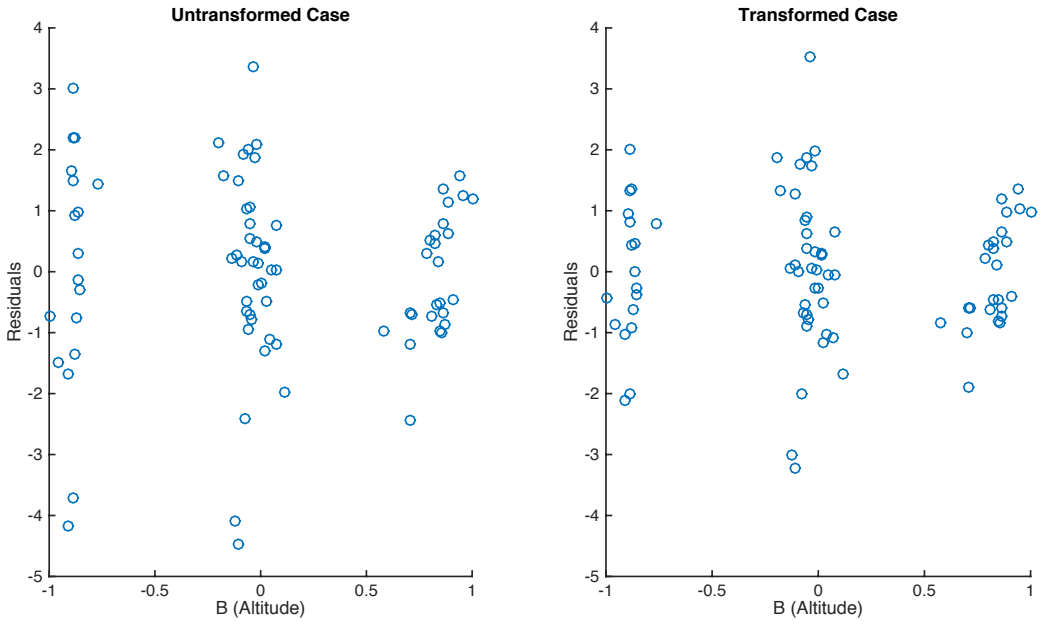

Figure 6: Residuals versus altitude for untransformed case (left panel) and transformed case using Manly transformation (right panel).

tion is a more appropriate transformation for these data. At this point we are interested in obtaining response predictions in the original units of observation and at various levels of interest of the factor *Altitude*. Unfortunately, the exponential transformation suffers the same bias issues as the Box-Cox transformations when a re-transformation of the fitted model back to the original units is performed. In order to more accurately estimate the mean and variance of *ClimbRate* in the original units of measurement, as well as construct more accurate and precise prediction intervals on *ClimbRate* at various levels of *Altitude*, a methodology similar to that proposed by Perry and Walker (2015) for the Box-Cox transformations will be developed.

In what follows, we discuss our approach to approximating higher-order moments of the original response variable Y, where it is assumed that a normal-theory linear model is fit using a transformed version of Y, and the transformation type is contained in the Manly exponential class. We derive a closed-form approximation to the k^{th} moment of Y that is applicable for all values of transformation parameter γ. This expression is then used to estimate the mean and variance of Y, given the parameter estimates obtained from fitting the model in the transformed domain. We then exploit Chebychev's inequality to construct an approximate $100(1-\alpha)\%$ prediction interval estimator on Y. Further, we discuss implications made from the results of Monte Carlo simulation studies used to assess the performance of the proposed prediction interval estimator, relative to that obtained by employing the more common re-transformation construction approach outlined above.

2 Proposed Methodology

Consider the Manly class of exponential transformations defined in equation (3). We may assume that for some value of the transformation parameter γ, the transformed response $Y(\gamma)$ approximately follows a normal distribution with mean $\mu = \mathbf{x}'\boldsymbol{\beta}$ and variance σ^2. The density function of the untransformed response Y can then be approximated by

$$f_Y(y) \approx \frac{e^{\gamma y}}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(\frac{e^{\gamma y}-1}{\gamma} - \mathbf{x}'\boldsymbol{\beta})^2}{2\sigma^2}\right), \gamma \neq 0$$

for all Y. Note that when $\gamma = 0$, $f_Y(y)$ is the normal density function. It then follows that, for $\gamma \neq 0$, an approximation to the k^{th} moment of Y can be written by the following,

$$E(Y^k) \approx \frac{1}{\sqrt{2\pi}} \int_R \left\{\frac{\log_e\big(1+\gamma(\mathbf{x}'\boldsymbol{\beta}+\sigma u)\big)}{\gamma}\right\}^k \exp\left\{-\frac{u^2}{2}\right\} du \tag{6}$$

where $u = \frac{\frac{e^{\gamma y}-1}{\gamma} - \mathbf{x}'\boldsymbol{\beta}}{\sigma}$ and $du = \frac{e^{\gamma y}}{\sigma} dy$. Note that R denotes the domain of the function $g(u) = \log_e\big(1+\gamma(\mathbf{x}'\boldsymbol{\beta}+\sigma u)\big)$, where if $\gamma > 0$, then $u > -\frac{(\mathbf{x}'\boldsymbol{\beta}+\frac{1}{\gamma})}{\sigma}$, and if $\gamma < 0$, then $u < -\frac{(\mathbf{x}'\boldsymbol{\beta}+\frac{1}{\gamma})}{\sigma}$.

To simplify the integral in equation (6), one can expand the first term on the right-hand side of the integrand in a Taylor series about $u = 0$, producing the approximation

$$E(Y^k) \approx \frac{\log_e(a)^k}{\sqrt{2\pi}\gamma^k} \int_{-\infty}^{\infty} \left(1 + \sum_{i=1}^{\infty} \frac{k\gamma^i\sigma^i \prod_{j=1}^{i-1}[k-(1+\log_e a)j]}{i! a^i \log_e(a)^i} u^i\right) \times e^{-\frac{u^2}{2}} du \tag{7}$$

where $a = 1+\gamma\mathbf{x}'\boldsymbol{\beta}$. Retaining terms through order 4 in equation (7) and subsequently evaluating the resulting integral then produces a closed form approximation to the k^{th} moment of Y, or

$$E(Y^k) \approx \left(\frac{\log_e(a)}{\gamma}\right)^k \left(1 - \frac{kb\gamma^2\sigma^2}{4\log_e(a)a^4} + \frac{k(k-1)c\gamma^2\sigma^2}{8\log_e(a)^2a^4} - \frac{3k(k-1)(k-2)\gamma^4\sigma^4}{4\log_e(a)^3a^4} + \frac{k(k-1)(k-2)(k-3)\gamma^4\sigma^4}{8\log_e(a)^4a^4}\right) \tag{8}$$

where $b = 2a^2 + 3\gamma^2\sigma^2$ and $c = 4a^2 + 11\gamma^2\sigma^2$, with $\sigma > 0$, $\mathbf{x}'\boldsymbol{\beta} > -\frac{1}{\gamma}$ if $\gamma > 0$ and $\mathbf{x}'\boldsymbol{\beta} < -\frac{1}{\gamma}$ if $\gamma < 0$.

Point estimates for $E(Y)$ and $Var(Y)$ are subsequently found by substituting parameter estimates computed in the transformed domain (i.e., $\hat{\boldsymbol{\beta}}$, $\hat{\sigma}^2$ and $\hat{\gamma}$) into the expression

given in equation (8), producing the estimators

$$\widehat{E(Y)} = \hat{\mu}_Y = \frac{\log_e(\hat{a})}{\hat{\gamma}} \left[1 - \frac{\hat{b}\hat{\gamma}^2\hat{\sigma}^2}{4\log_e(\hat{a})\hat{a}^4} \right], \quad \hat{\gamma} \neq 0 \tag{9}$$

and

$$\widehat{Var(Y)} = \hat{\sigma}_Y^2 = \widehat{E(Y^2)} - \widehat{E(Y)}^2 \tag{10}$$

where,

$$\widehat{E(Y^2)} = \frac{\log_e(\hat{a})^2}{\hat{\gamma}^2} \left[1 - \frac{\hat{b}\hat{\gamma}^2\hat{\sigma}^2}{2\log_e(\hat{a})\hat{a}^4} + \frac{\hat{c}\hat{\gamma}^2\hat{\sigma}^2}{4\log_e(\hat{a})^2\hat{a}^4} \right], \quad \hat{\gamma} \neq 0 \tag{11}$$

and

$$\begin{aligned} \mathbf{x}'\hat{\boldsymbol{\beta}} &> -\tfrac{1}{\hat{\gamma}} \quad \text{for } \hat{\gamma} > 0 \\ \mathbf{x}'\hat{\boldsymbol{\beta}} &< -\tfrac{1}{\hat{\gamma}} \quad \text{for } \hat{\gamma} < 0 \end{aligned} \tag{12}$$

are conditions on the parameter estimates that must be met in order for the estimators to be defined.

By expanding the estimators for μ_Y and σ_Y^2 in equations (9) and (10), respectively, in a Taylor series about $\hat{\gamma} = \gamma$, $\hat{\boldsymbol{\beta}} = \boldsymbol{\beta}$, and $\hat{\sigma} = \sigma$, one can show consistency. Specifically, it can be shown that if $\hat{\gamma}$, $\hat{\boldsymbol{\beta}}$, and $\hat{\sigma}$ are consistent estimators of γ, $\boldsymbol{\beta}$, and σ, respectively, then $\hat{\mu}_Y$ and $\hat{\sigma}_Y$ are consistent estimators of μ_Y and σ_Y, up to a 4^{th}-order approximation. Since reduced-bias estimates of μ_Y and σ_Y^2 are available via the above expressions, one can exploit Chebychev's inequality to produce an approximate $100(1-\alpha)\%$ prediction interval estimator on Y, or

$$\hat{\mu}_Y \pm L\hat{\sigma}_Y, \tag{13}$$

where for any given α the value of L will lie in the interval $[z_{\alpha/2}, \sqrt{\frac{1}{\alpha}}]$. In practice, if a large sample is available for estimating unknown model parameters, then one can justify setting $L = z_{\alpha/2}$, otherwise the authors recommend using $L = \sqrt{\frac{1}{\alpha}}$, which will produce a prediction interval on Y with expected coverage of *at least* the nominal probability $1-\alpha$, approximately.

We can now revisit the single-engine climb experiments discussed in Section 1 using the model parameter estimates given in Table 5. Estimates for the mean and variance of the original response can be obtained from equations (9) and (10), with prediction intervals constructed from equation (13), at each of three specific levels of interest of *Altitude*, i.e., 1000, 3000, and 5000 feet. These are shown in Table 6, while Figure 7 illustrates them graphically. Note in Figure 7 that the red shaded region corresponds to $L = z_{\alpha/2} = 1.96$, where the black-shaded region corresponds to $L = 1/\sqrt{\alpha}$.

Table 6: Estimated means, variances, and approximate 95% prediction intervals on *ClimbRate* computed with *Altitude* at 1000, 3000, and 5000 feet using the fitted model with the exponential transformation and proposed methodology with $L = z_{0.025} = 1.96$.

Altitude	$\hat{\mu}_Y$	$\hat{\sigma}^2_Y$	Lower Bound	Upper Bound	Interval Width
1000	-5.4645	4.9607	-9.8299	-1.0991	8.7308
3000	-1.0454	1.7903	-3.6679	1.5771	5.2450
5000	-1.7530	2.1045	-4.5964	1.0903	5.6867

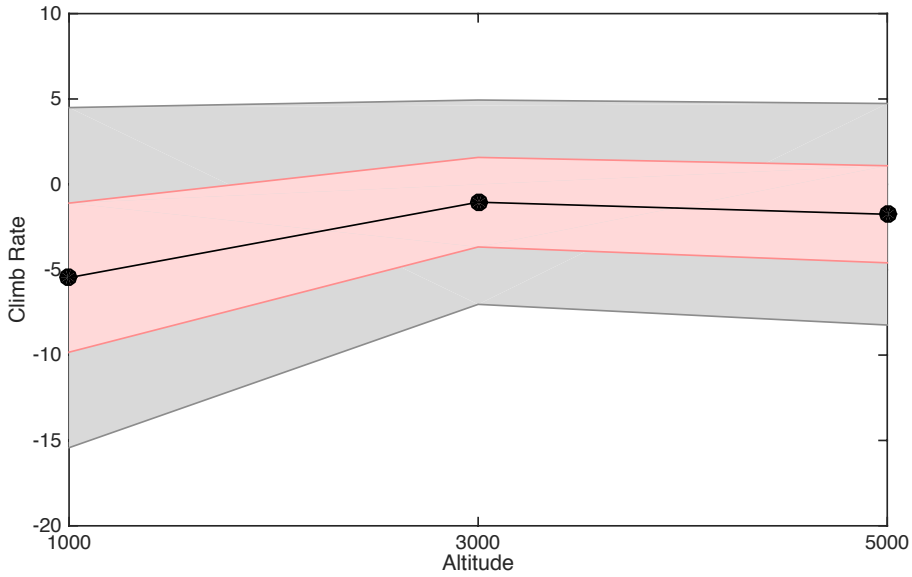

Figure 7: Prediction interval estimates across the factor *Altitude* using the proposed method. The black-shaded region corresponds to $L = 1/\sqrt{\alpha}$, while the red-shaded region corresponds to $L = z_{\alpha/2} = 1.96$.

For comparison purposes, we also constructed prediction intervals on *ClimbRate* using standard construction methods with the untransformed response and the fitted model in equation (3). These are shown in Table 7. Notice that the widths for the intervals constructed from the untransformed response are all approximately equal, whereas those constructed using the proposed methodology vary depending on the level of the experimental factor *Altitude*. The proposed intervals can account for the non-constant variance across *Altitude*, and thus, would seem to be most appropriate for these data.

Table 7: Approximate 95% prediction intervals on *ClimbRate* computed with *Altitude* at 1000, 3000, and 5000 feet using the fitted model without transformation and $\hat{\sigma}^2 = 2.3796$.

Altitude	Predicted Value	Lower Bound	Upper Bound	Interval Width
1000	-5.2770	-8.4699	-2.0840	6.3859
3000	-1.0751	-4.1856	2.0353	6.2209
5000	-1.6839	-4.8279	1.4601	6.2880

Although application of the Manly exponential transformation and the proposed methodology to the Hutto-Simpson experimental data appears to yield reasonable performance, it is not clear from this single example how the proposed estimators for μ_Y and σ_Y^2, as well as the prediction interval estimator will perform on average. Thus, in the next section we discuss results of a Monte Carlo simulation study designed to assess the average performance of $\hat{\mu}_Y$ and $\hat{\sigma}_Y^2$ in equations (9) and (10), respectively, as well as the performance of the prediction interval estimator in equation (13), relative to that obtained via the more commonly-used re-transformation construction method mentioned earlier.

3 Performance Evaluation

In efforts to assess the expected performance of our proposed estimators for μ_Y and σ_Y^2 in equations (9) and (10), respectively, as well as the proposed prediction interval estimator in equation (13), we used Monte Carlo simulation. In what follows, we describe the simulation

model in detail.

For any given simulation run, a random sample $\mathbf{y}$ of size n was generated from

$$y_i = \begin{cases} \frac{\log_e(1+\gamma y_i(\gamma))}{\gamma} & \text{for } \gamma \neq 0 \\ y_i & \text{for } \gamma = 0 \end{cases} \tag{14}$$

where

$$y_i(\lambda) = \mathbf{x}_i^{'}\boldsymbol{\beta} + \epsilon_i \tag{15}$$

denotes the transformed response observation taken at $\mathbf{x}_i$, with $\mathbf{x}_i = [1, x_{i1}, x_{i2}, ..., x_{iq}]'$ denoting a $p \times 1$ vector of regressor variable settings ($p = q + 1$), $\boldsymbol{\beta} = [\beta_0, \beta_1, ..., \beta_q]'$ denotes a $p \times 1$ vector of fixed effects (or regression coefficients), and ϵ_i denotes a random error term where $\epsilon_i \sim i.i.d.N(0, \sigma^2)$ for all $i = 1, ..., n$. Once the response observations were generated, model parameters were then estimated using maximum likelihood estimation. Letting $\mathbf{X}$ denote an $n \times p$ design matrix and $\mathbf{y}(\gamma) = [y_1(\gamma), ..., y_n(\gamma)]'$ denote the $n \times 1$ vector of transformed responses, then the method of maximum likelihood leads to the following estimators for the unknown model parameters

$$\hat{\gamma} = \arg\max_{\gamma} \left[-n \log_e(\sqrt{2\pi}\hat{\sigma}_\epsilon) + \gamma \sum_{i=1}^{n} y_i - \frac{1}{2\hat{\sigma}_\epsilon^2} \sum_{i=1}^{n} (y_i(\gamma) - \hat{\mu}_i)^2 \right], \tag{16}$$

where $y_i(\gamma) = \frac{e^{\gamma y_i} - 1}{\gamma}$ if $\gamma \neq 0$ and $y_i(\gamma) = y_i$ if $\gamma = 0$, with

$$\hat{\mu}_i = \mathbf{x}_i^{'}\hat{\boldsymbol{\beta}} = \mathbf{x}_i^{'}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}(\gamma) \tag{17}$$

and

$$\hat{\sigma}_\epsilon^2 = \frac{\mathbf{y}(\gamma)'(\mathbf{I}_n - \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}')\mathbf{y}(\gamma)}{n - p}, \tag{18}$$

where the optimization problem in equation (16) was solved using the *optimize* function in

R. Once parameters were estimated, they were then substituted into equations (9) and (10) in order to produce estimates for, respectively, the mean and variance of the untransformed response Y. At this point in the simulation model, prediction interval estimates on Y were also produced. In particular, the prediction interval estimate using equation (13) was constructed, as well as that using a traditional re-transformation construction method given by

$$\left[\frac{\log_e(\hat{\gamma}(\hat{Y}(\hat{\gamma}) - \hat{H}) + 1)}{\hat{\gamma}}, \frac{\log_e(\hat{\gamma}(\hat{Y}(\hat{\gamma}) + \hat{H}) + 1)}{\hat{\gamma}}\right], \quad \text{if } \hat{\gamma} \neq 0 \tag{19}$$

where $\hat{H} = h\hat{\sigma}_{pred}$, and $\hat{\sigma}_{pred} = \hat{\sigma}\sqrt{1 + \mathbf{x}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{x}}$ denotes the estimate of the standard deviation of prediction computed in the transformed domain. The value of h is set to achieve some desired level of confidence, and is most often chosen from a t_v distribution, where v denotes the degrees of freedom. It is important to note that, when $\hat{\gamma} \neq 0$, this interval estimator is only defined if

$$\begin{aligned} \hat{Y}(\hat{\gamma}) - \hat{H} &> -\tfrac{1}{\hat{\gamma}} \quad \text{for } \hat{\gamma} > 0 \\ \hat{Y}(\hat{\gamma}) + \hat{H} &< -\tfrac{1}{\hat{\gamma}} \quad \text{for } \hat{\gamma} < 0. \end{aligned} \tag{20}$$

Once these interval estimates were obtained, their widths were recorded by the simulation model. For a given simulation run, interval estimate coverage was computed by generating 50 million observations from the *true* distribution and then recording the proportion of these observations that fell within the estimated intervals. To make a fair comparison, we set the value of L in equation (13) so that the relative difference in the estimated expected coverages between the two methods over 500 Monte Carlo simulation runs was less than 1%. For all simulations, we chose $\alpha = 0.05$.

In our study, we considered fitting a main-effects plus two-factor interactions model using various replicates of a 2^3 factorial designed experiment. We considered values of $\gamma \in [-0.22, 0.22]$, $\mu = [-1, 0, 1]$ and $\sigma_\epsilon = 1$, so that $P[y_i(\gamma) > -\frac{1}{\gamma}|\gamma > 0] \approx 1$ and $P[y_i(\gamma) < -\frac{1}{\gamma}|\gamma < 0] \approx 1$. Specifically, for the simulations discussed below, the parameter vector $\boldsymbol{\beta}$ was simulated at $\boldsymbol{\beta} = [\mu, 0, \cdots, 0]'$ and the vector $\mathbf{x}$ was an arbitrarily chosen factorial point. It

makes no difference which point in design space is used in the simulation model since the coefficients $\beta_1, ..., \beta_q$ were all set to zero. We should note that additional simulation runs were performed at other values for the parameters $\boldsymbol{\beta}$ and σ_ϵ, and at other points $\mathbf{x}$ within the design space (where the probabilities on $y_i(\gamma)$ mentioned above hold), and although not reported here, our general conclusions discussed below are not altered.

Table 8 shows estimated root mean square error (RMSE) performances of the estimators in equations (9) and (10), for the case where $\gamma \in [-0.22, 0.15]$ and $\mu = -1$, each obtained over 5000 independent Monte Carlo simulation runs. Similarly, Table 9 shows estimated RMSEs for the case where $\gamma \in [-0.18, 0.18]$ and $\mu = 0$, while Table 10 shows RMSEs for the case where $\gamma \in [-0.15, 0.22]$ and $\mu = 1$.

Table 8: Root mean square errors (RMSEs) of estimators in equations (9) and (10). Each RMSE entry in the table was computed over 5000 independent Monte Carlo simulation runs. For any given γ and number of design replicates R, the true mean was simulated at $\mu = \mathbf{x}'\boldsymbol{\beta} = -1$.

	$\widehat{E(Y)}$			$\widehat{Var(Y)}$		
γ	$R=2$	$R=3$	$R=4$	$R=2$	$R=3$	$R=4$
0.15	0.3945	0.3510	0.3293	0.7947	0.6033	0.4953
0.10	0.3209	0.2710	0.2401	0.6285	0.4756	0.3843
0.05	0.2726	0.2257	0.1916	0.5380	0.3937	0.3224
0.00	0.2502	0.2077	0.1768	0.4797	0.3447	0.2776
-0.05	0.2466	0.1971	0.1767	0.4379	0.3204	0.2584
-0.10	0.2475	0.2033	0.1844	0.4120	0.3041	0.2476
-0.15	0.2575	0.2130	0.1954	0.3917	0.2994	0.2481
-0.18	0.2602	0.2279	0.2073	0.3883	0.2926	0.2421
-0.22	0.2647	0.2410	0.2166	0.3860	0.2897	0.2389

Notice in Tables 8-10 that, in general, as the number of design replicates R increases, a decrease in the RMSEs for both estimators is observed, regardless of the values of γ and μ. This is quite intuitive, and should be expected since with larger R more observations are available to estimate unknown model parameters. Also, note that as $\gamma \to 0$, a decrease in the RMSE of the mean estimator is observed for any given R. These results are intuitive

Table 9: Root mean square errors (RMSEs) of estimators in equations (9) and (10). Each RMSE entry in the table was computed over 5000 independent Monte Carlo simulation runs. For any given γ and number of design replicates R, the true mean was simulated at $\mu = \mathbf{x}'\boldsymbol{\beta} = 0$.

	$\widehat{E(Y)}$			$\widehat{Var(Y)}$		
γ	$R = 2$	$R = 3$	$R = 4$	$R = 2$	$R = 3$	$R = 4$
0.18	0.2813	0.2411	0.2127	0.5542	0.4188	0.3464
0.15	0.2781	0.2287	0.2014	0.5278	0.3957	0.3228
0.10	0.2623	0.2110	0.1844	0.4999	0.3682	0.3020
0.05	0.2599	0.2056	0.1760	0.4760	0.3573	0.2897
0.00	0.2526	0.2066	0.1767	0.4760	0.3473	0.2776
-0.05	0.2525	0.2110	0.1812	0.4839	0.3499	0.2846
-0.10	0.2640	0.2108	0.1881	0.4928	0.3687	0.3002
-0.15	0.2748	0.2244	0.2008	0.5282	0.4009	0.3223
-0.18	0.2898	0.2452	0.2124	0.5439	0.4152	0.3417

Table 10: Root mean square errors (RMSEs) of estimators in equations (9) and (10). Each RMSE entry in the table was computed over 5000 independent Monte Carlo simulation runs. For any given γ and number of design replicates R, the true mean was simulated at $\mu = \mathbf{x}'\boldsymbol{\beta} = 1$.

	$\widehat{E(Y)}$			$\widehat{Var(Y)}$		
γ	$R = 2$	$R = 3$	$R = 4$	$R = 2$	$R = 3$	$R = 4$
0.22	0.2650	0.2384	0.2178	0.3945	0.2902	0.2399
0.18	0.2631	0.2225	0.2040	0.3928	0.2932	0.2423
0.15	0.2524	0.2191	0.1918	0.3999	0.2999	0.2435
0.10	0.2503	0.2034	0.1850	0.4108	0.3038	0.2441
0.05	0.2435	0.2003	0.1767	0.4401	0.3244	0.2591
0.00	0.2538	0.2004	0.1762	0.4751	0.3476	0.2816
-0.05	0.2706	0.2202	0.1950	0.5384	0.3979	0.3172
-0.10	0.3188	0.2733	0.2376	0.6226	0.4677	0.3821
-0.15	0.3989	0.3617	0.3274	0.7841	0.5989	0.4936

since an increase in $|\gamma|$ implies greater skewness in the underlying distribution of Y and thus an increase in the variance of Y is observed. Additionally, in Table 8 (where $\mu = -1$), as γ decreases, a decrease in the RMSE of the variance estimator is observed for any given R. Further, in Table 10 (where $\mu = 1$), as γ decreases, an increase in the RMSE of the variance estimator for any given R is observed. These results are also intuitive since when $\mu > 0$, an increase in the variance of Y is observed as γ decreases, and when $\mu < 0$, a decrease in the variance of Y is observed as γ decreases. Thus, under these circumstances, one would expect to see the RMSEs behave in such a way.

We now focus our attention on the relative performances of the prediction interval estimators considered in our study; in particular, the interval estimator obtained via the traditional re-transformation method in equation (19), as well as the proposed estimator given in equation (13). In order to effectively summarize the relative performances over the range of γ for any given number of design replicates R, we follow Perry and Walker (2015) and compute the relative mean index (RMI) defined by

$$RMI_{u_i} = \frac{1}{r}\sum_{j=1}^{r}\frac{u_{ij} - \min[u_{1j}, u_{2j}]}{\min[u_{1j}, u_{2j}]}, \tag{21}$$

where RMI_{u_i} denotes the relative mean index of method i with respect to performance measure u. Here, u_{ij} denotes a performance measure taken at the j^{th} level of λ for the i^{th} method being compared, where $i = 1, 2$, and $j = 1, 2, ..., r$, where r denotes the number of levels of γ considered. Thus, to compare expected width performance between the traditional and proposed interval estimators for a given number of design replicates R, let $\bar{\hat{w}}_{ij}$ denote the average of the width estimates obtained over the 5000 independent Monte Carlo simulation runs at the j^{th} level of γ and for the i^{th} method being compared, then $u_{ij} = \bar{\hat{w}}_{ij}$. Similarly, to compare the variability in the widths of the interval estimates, let $\hat{s}_{ij}$ denote the estimated standard error of $\bar{\hat{w}}_{ij}$ obtained over the 5000 Monte Carlo simulation runs at the j^{th} level of γ and for the i^{th} method being compared, then let $u_{ij} = \hat{s}_{ij}$. Among the two methods

being compared, the one that achieves better relative performance will then have the smallest RMI, for both performance measures considered.

Table 11 shows the RMIs computed at values of $\mu = -1$, $\mu = 0$ and $\mu = 1$, at different numbers of design replicates R, and for both performance measures considered, i.e., $\bar{\hat{w}}_{ij}$ and $\hat{s}_{ij}$. Note that $RMI_{\bar{\hat{w}}_1}$ and $RMI_{\hat{s}_1}$ denote the RMIs corresponding to the traditional interval estimator, and $RMI_{\bar{\hat{w}}_2}$ and $RMI_{\hat{s}_2}$ denote those corresponding to the proposed interval estimator. The results in Table 11 suggest that, if the two intervals yield

Table 11: Relative mean indices (RMIs) of performance results of the two prediction interval estimators for different values of $\mu = \mathbf{x}'\boldsymbol{\beta}$ and design replicates R. The method that produces the smaller RMI value suggests better relative performance regardless of underlying skewness within the data across the range of γ.

	R	$RMI_{\bar{\hat{w}}_1}$	$RMI_{\bar{\hat{w}}_2}$	$RMI_{\hat{s}_1}$	$RMI_{\hat{s}_2}$
	2	0.13	0.00	0.46	0.00
	3	0.07	0.00	0.26	0.00
$\mu = -1$	4	0.04	0.00	0.14	0.00
	5	0.03	0.00	0.10	0.00
	50	0.00	0.00	0.02	0.00
	2	0.12	0.00	0.42	0.00
	3	0.06	0.00	0.22	0.00
$\mu = 0$	4	0.04	0.00	0.13	0.00
	5	0.03	0.00	0.09	0.00
	50	0.00	0.00	0.02	0.00
	2	0.15	0.00	0.52	0.00
	3	0.08	0.00	0.31	0.00
$\mu = 1$	4	0.05	0.00	0.16	0.00
	5	0.04	0.00	0.11	0.00
	50	0.01	0.00	0.03	0.00

approximately the same coverage probability, our proposed interval estimator will outperform that obtained via the traditional re-transformation construction method. In particular, our results show that, for all combinations of γ, μ and R considered in our study, our proposed interval estimator achieved an interval with a smaller expected width and a smaller variance of the width estimates. Large differences in performance are most evident when there are only a small number of design replicates. Table 11 further suggests that, for any given $\mu = \mathbf{x}'\boldsymbol{\beta}$,

as the number of design replicates increases, the performance of the traditional estimator appears to approach that of our proposed estimator. Therefore, based on these results, we recommend that the proposed prediction interval estimator be used as an alternative to that obtained via the re-transformation approach in practice, unless the number of design replicates (or, more generally, degrees of freedom beyond that required to estimate model parameters) is very large.

4 Summary and Discussion

Motivated by A-10 single engine aircraft climb experiments, in this paper we demonstrated that when a transformation of a negatively skewed response variable is desired, the class of Manly exponential transformations is more appropriate, relative to the more popular Box-Cox family of power transformations. We developed an approximate prediction interval estimator on the response variable Y, where it was assumed that a normal-theory linear model was fit using a transformed version of Y, and the transformation type was contained in the Manly exponential class. We derived a closed-form approximation to the k^{th} moment of the original response variable Y, which was then used to estimate the mean and variance of Y, given parameter estimates obtained from fitting the model in the transformed domain. Chebychev's inequality was then exploited to construct an approximate $100(1-\alpha)\%$ prediction interval estimator on the original response variable.

Using Monte Carlo simulation, we assessed the performance of the proposed estimators for μ_Y and σ_Y^2, as well as the prediction interval estimator, relative to that obtained by employing a more traditional re-transformation interval construction approach. Results suggest that our proposed estimators for μ_Y and σ_Y^2 yield acceptable performance with respect to the RMSE, with values that decrease with the variance of Y. Further, our simulation results suggest that, in general, when both prediction interval estimators yield approximately the same expected coverage, then the proposed prediction interval estimator will yield a

smaller estimate for the expected interval width, as well as greater precision in this estimate, relative to the traditional interval estimator.

One obvious advantage to using our proposed methodology is that it alleviates the lack of interpretation of the transformation parameter γ that has inherently made it a less preferable alternative to the Box-Cox transformations. Data analysis may be performed for any γ, with the fitted model able to be converted back to the original units with reduced transformation bias that could otherwise significantly affect the prediction. Further, since reduced-bias estimates for μ_Y and σ_Y^2 are available, they can be used to approximate the type of relationship that exists between the mean and variance of Y. A model capturing the relationship between the mean and variance of Y could allow for inference to be drawn on both of these parameters, at given levels of the experimental factors, using the same set of experiments.

References

Atkinson, A. C. (1983). Diagnostic regression analysis and shifted power transformations. *Technometrics 25*, 23–332.

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B 26*, 211–252.

Freeman, J. and R. Modarres (2006). Inverse box-cox: The power-normal distribution. *Statistics and Probability Letters 76*, 764–772.

Hutto, G. T. and J. R. Simpson (2013). A-10 single engine climb using design of experiments. *96 Test Wing Technical Document, Eglin Air Force Base, Florida*.

Land, C. E. (1974). Confidence interval estimation for means after data transformation to normality. *Journal of the American Statistical Association 69*, 795–802.

Manly, B. F. J. (1976). Exponential data transformations. *Journal of the Royal Statistical Society, Series D 25*, 37–42.

Perry, M. B. and M. L. Walker (2015). A prediction interval estimator for the original response when using box-cox transformations. *Journal of Quality Technology*. To Appear.

Sakia, R. M. (1990). Retransformation bias: A look at the box-cox transformation to linear balanced mixed anova models. *Metrika 37*, 345–351.

Sakia, R. M. (1992). The box-cox transformation technique: A review. *Journal of the Royal Statistical Society, Series D 41*, 169–178.

Shumway, R., A. Azari, and P. Johnson (1989). Transformation for environmental data with detection limits. *Technometrics 31*, 347–356.

Taylor, J. (1986). The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association 81*, 114–118.

Part IV
Multivariate Predictions and Prediction Regions in the Original Units of Observation when Modeling Transformations of the Response

Abstract

Motivated by A-10 single engine climb experiments, we demonstrate the use of both the multivariate Box-Cox power transformation and the multivariate Manly exponential transformation in fitting normal-theory linear models to a transformed version of a q-variate response vector $\mathbf{Y}$. We derive closed-form approximations to the k^{th} moment of each original response Y_i $(i = 1, \ldots, q)$, as well as a closed-form approximation to $E(Y_i Y_{i'})$, $(i \neq i')$, which are used to estimate the mean and variance of Y_i and the covariance between Y_i and $Y_{i'}$, given parameter estimates obtained from fitting the model in the transformed domain. Exploiting two multivariate analogs of Chebyshev's inequality, we construct an approximate $100(1-\alpha)\%$ prediction sphere and ellipsoid, respectively, on the original response vector $\mathbf{Y}$. Using Monte Carlo simulation, we assess the performance of the proposed estimators for the means, variances, and covariances, as well as the coverage performance of the two Chebyshev prediction regions constructed using the proposed estimators. The methodology is further applied to two data examples, one consisting of a four variable response of voting results, and one consisting of experimental machining process data with a bivariate response.

1 Introduction and Motivation

Statistical linear models are based upon normal distribution theory. In a linear model constructed on a univariate response, a transformation may be performed on the response variable in an attempt to approximate normality in the residual error terms of the model. In the case of a multivariate response, the condition of multivariate normality is also desired and may similarly be approximated through the use of a transformation of the response variables. For instance, the popular power transformation proposed by Box and Cox (1964) can easily be extended for use with a q-variate response, given by

$$Z_i = \begin{cases} \frac{Y_i^{\lambda_i}-1}{\lambda_i} & \lambda_i \neq 0 \\ \log_e(Y_i) & \lambda_i = 0 \end{cases} \tag{1}$$

for $i = 1, 2, \ldots, q$, where Y_i is the i^{th} original variable, λ_i is the transformation parameter associated with the i^{th} variable, and Z_i is the i^{th} transformed variable. The underlying condition of the Box-Cox transformation, a strictly positively valued original variable, holds in the multivariate case as well, with each $Y_i > 0$ in order to apply the transformation. In cases of non-positively valued variables, Box and Cox proposed adding a shift parameter to the original variable in order to ensure this condition is met for the purposes of transformation. The Shifted Box-Cox transformation, as applied to the multivariate case, is then given by

$$Z_i = \begin{cases} \frac{(Y_i+\lambda_i^{(s)})^{\lambda_i}-1}{\lambda_i} & \lambda_i \neq 0, Y_i > -\lambda_i^{(s)} \\ \log_e(Y_i + \lambda_i^{(s)}) & \lambda_i = 0, Y_i > -\lambda_i^{(s)} \end{cases} \tag{2}$$

for $i = 1, 2, \ldots, q$, where Y_i is the original variable, λ_i is the transformation parameter associated with the i^{th} variable, $\lambda_i^{(s)}$ is the shift parameter associated with the i^{th} variable, and Z_i is the transformed variable. For a shift parameter equal to 0, these equations revert to the single parameter transformations in equation (1).

While quite effective and widely used in practice, the Box-Cox transformation does have limitations in its ability to appropriately transform variables that may be negatively

skewed. A more flexible alternative to the Box-Cox transformation is the exponential transformation proposed by Manly (1976). The Manly transformation has an advantage in its ability to transform both negatively and positively skewed variables to approximate normality. Additionally, this transformation does not require a secondary shift parameter in order to apply the transformation. In the multivariate case, the Manly transformation is given by

$$Z_i = \begin{cases} \frac{e^{\gamma_i Y_i} - 1}{\gamma_i} & \gamma_i \neq 0 \\ Y_i & \gamma_i = 0 \end{cases} \tag{3}$$

for $i = 1, 2, \ldots, q$, where Y_i is the original variable, γ_i is the transformation parameter associated with the i^{th} original variable, and Z_i is the transformed variable. The special case of $\gamma_i = 0$ results in no transformation.

While both the Box-Cox and Manly transformations can be useful in model building, problems arise when prediction is desired in the original units of observation. With a model fit in transformed space, any predicted values will also be in transformed space. Thus, in order to predict in the original units, a re-transformation of any predicted values would need to be performed, typically achieved by using the inverse of the applied transformation. As neither of these transformations are linear in nature, obtaining predictions in this manner inherently results in biased predicted values, investigated by several authors, including Taylor (1986), Shumway et al. (1989), and Sakia (1990). The issue becomes more complicated if a prediction region is also desired in the original units. For a univariate response, the limits of a prediction interval may be obtained in transformed space, then similarly re-transformed through the use of the inverse in the transformation, similarly resulting in biased interval limits. However, for a multivariate response, there is no similar method to apply such a re-transformation to the equation of an ellipse constructed in transformed space. Recent methodology has been presented to reduce the bias in predicted values and prediction intervals of a univariate response using the Box-Cox transformation in Perry and Walker (2015) and the Manly transformation in Walker and Perry (2015). The purpose of the research presented in this

paper is to extend this methodology for use with the multivariate response transformations defined in equations (1), (2), and (3). This proposed methodology will ultimately result in the ability to obtain reduced bias predictions, as well as the ability to construct prediction regions in the original units of observation.

To demonstrate the use of both the multivariate Box-Cox and the multivariate Manly transformations in model building and prediction, consider the A-10 single-engine climb experiments performed by Hutto and Simpson (2013). In these experiments, the climb rate of A-10 jets is studied as a function of changes in the experimental factors $Temperature$ (factor A), $Altitude$ (factor B), $GearWeight$ (factor C), $Flaps$ (factor D), and $LandingGear$ (factor E). Climb rate as initially recorded in the experiments is a multivariate response consisting of three separate measurements of climb rate, $Climb(FLT)$ (flight measured), $Climb(Chart)$ (chart measured), and $Climb(-1)$ (-1 chase around). For reference, a summary of the specific levels of interest for each of the experimental factors is given in Table 1. The experiment consisted of 82 runs across the various levels of the experimental factors, though run 44 contained missing data, and thus, was omitted from the analysis, leaving a sample size of $n = 81$.

Table 1: Summary of experimental factors with levels of interest

Factor	Description	Low	Medium	High	Units
A	Temperature	66	72	80	degrees fahrenheit
B	Altitude	1000	3000	5000	feet
C	Gear Weight	37500	41250	45000	lbs.
D	Flaps	0		7	degrees
E	Landing Gear	down		up	NA

As all three response variables are measurements of climb rate, there is expected to be at least some amount of correlation between them. This correlation among the climb rate variables can be observed in Table 2. Of note is the very high correlation between $Climb(FLT)$ and $Climb(Chart)$. This is due to the climb rate being similarly measured by

two separate instruments, and can be viewed, apart from some small measurement error, as being two instances of the same variable. Thus, only one of these two variables will need to be included in the model. Removing $Climb(FLT)$, leaves $Y_1 = Climb(Chart)$ and $Y_2 = Climb(-1)$ to be used as a bivariate response in the model. The two remaining climb rate measurements are still relatively highly correlated, however they are distinct measurements of climb rate, with $Climb(-1)$ typically viewed as a measure of maneuverability as opposed to a straight climb.

Table 2: Correlation among response variables, A-10 climb data

	Climb(FLT)	Climb(Chart)	Climb(-1)
Climb(FLT)	1.000	0.998	0.456
Climb(Chart)	0.998	1.000	0.424
Climb(-1)	0.456	0.424	1.000

A multivariate regression analysis of the experimental data was performed on the reduced bivariate response

$$\mathbf{Y} = \begin{bmatrix} Climb(Chart) \\ Climb(-1) \end{bmatrix}$$

using main effects and two-way interactions of the experimental factors detailed in Table 1 to estimate a linear model, i.e. $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$, with a fitted model of the form $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$. After removing insignificant interaction terms, the final fitted model is found to contain all five main effects as well as four of the two-way interactions, observed in Table 3. Given the fitted model, prediction regions can be constructed at specific points in design space. To demonstrate, prediction ellipses are constructed at two separate design points, arbitrarily chosen as the 9^{th} and 75^{th} design points in the design matrix $\mathbf{X}$, illustrated in Figure 1 with the predicted observations and ellipses shown in orange. The data observations originally obtained at these design points are also shown in red for reference. Of note is that these prediction regions, as well as prediction regions that may be obtained at any design point,

are each constructed by utilizing the single estimate $\hat{\boldsymbol{\Sigma}}_\epsilon$ obtained from the model residuals, and therefore will all have identical shape, size, and orientation.

Table 3: MANOVA results for the fitted model

Factor	Wilks' λ	Approx. F	Num. DF	Den. DF	$P(>F)$
A	0.63312	20.282	2	70	1.127×10^{-7}
B	0.34248	67.196	2	70	$< 2.2 \times 10^{-16}$
C	0.37019	59.546	2	70	7.850×10^{-16}
D	0.91825	3.116	2	70	0.050530
E	0.91961	3.059	2	70	0.053232
AB	0.85115	6.121	2	70	0.003550
AC	0.91404	3.291	2	70	0.043033
AE	0.83024	7.156	2	70	0.001487
BD	0.89403	4.149	2	70	0.019827

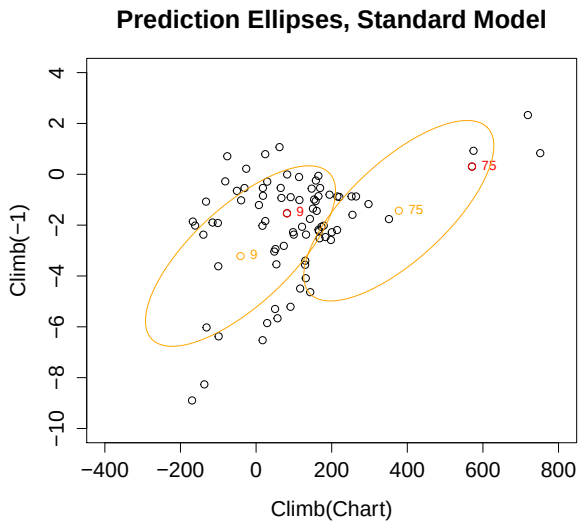

Figure 1: Prediction ellipses at design points 9 and 75

However, a multivariate Shapiro-Wilk normality test performed on the residuals (W=0.9486, p-value=0.002674) suggests that bivariate normality may not be reasonably assumed. This is further illustrated by the plot of the Mahalanobis distances in Figure 2.

These results indicate that a transformation of the bivariate response may be appropriate to achieve approximate normality desired of the model's error terms and ultimately

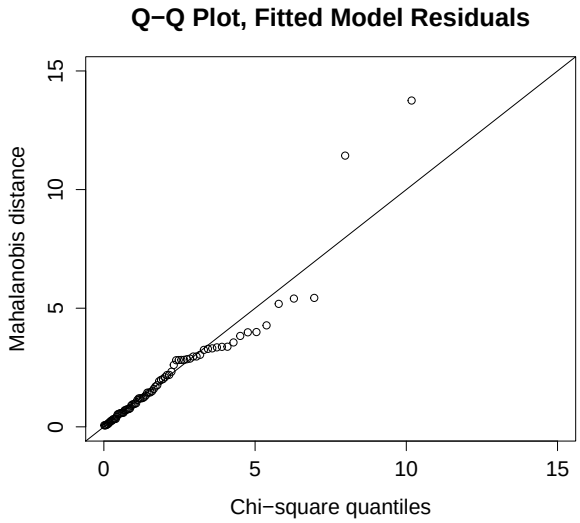

Figure 2: Q-Q Plot of Mahalanobis distances from the fitted model residuals

result in a better model fit. In what follows, each of the aforementioned multivariate data transformations on the response, the Box-Cox transformation and the Manly transformation, are investigated.

1.1 Box-Cox Transformation Analysis

Given that both of the response variables can take negative values, the Shifted Box-Cox transformation defined in equation (2) can be applied to the data. As the joint estimation of the shift parameter and transformation parameter can be problematic even for univariate response transformations (e.g., see Atkinson (1983)), it is common practice to set the shift parameter beforehand (e.g., see Sakia (1992)) to $\lambda_i^{(s)} = -\min(Y_i) + \delta$, where δ is some constant value used to ensure $Y_i + \lambda_i^{(s)} + \delta > 0$ for all Y_i. For this experimental data, these shift parameters are estimated at $\hat{\lambda}_1^{(s)} = 169.303$ and $\hat{\lambda}_2^{(s)} = 8.897$. With the shift parameters specified, all other model parameters can be obtained through maximum likelihood, as discussed in Andrews et al. (1971), using the *optim* function in R for the maximization. Iteratively removing insignificant terms results in a final fitted model retaining all five main effects and five of the interaction terms. With the transformation parameters estimated as

$\hat{\lambda}_1 = 0.563$ and $\hat{\lambda}_2 = 0.827$, the significance of these model factors can be observed in Table 4.

Table 4: MANOVA results for the fitted model using the Shifted Box-Cox transformation

Factor	Wilks' λ	Approx. F	Num. DF	Den. DF	$P(> F)$
A	0.58271	24.706	2	69	8.095×10^{-9}
B	0.31495	75.040	2	69	$< 2.2 \times 10^{-16}$
C	0.28906	84.851	2	69	$< 2.2 \times 10^{-16}$
D	0.87387	4.979	2	69	0.0095495
E	0.86590	5.343	2	69	0.0069592
AB	0.77059	10.271	2	69	0.0001246
AC	0.89632	3.991	2	69	0.0229108
AE	0.75503	11.193	2	69	6.164×10^{-5}
BD	0.88764	4.367	2	69	0.0163748
CE	0.76666	10.500	2	69	0.0001044

Figures 3 and 4 illustrate the marginal 95% confidence intervals for each of the transformation parameters. Of note is the 95% confidence interval for $\hat{\lambda}_1$, which does not contain the nominal value of $\lambda = 1$ (no transformation). This suggests a marginal transformation on the response $Climb(Chart)$ may be justified. The 95% confidence interval for $\hat{\lambda}_2$ does contain $\lambda = 1$, suggesting that a marginal transformation on $Climb(-1)$ may not be necessary. However, given that a transformation was found to be necessary for $Climb(Chart)$ and the transformation parameters are jointly estimated in the fitted model along with all model parameters, an isolated transformation of a single response would result in no gain or loss in the model building process.

Further diagnostics on the fitted model residuals reveal that the application of the Box-Cox transformation was able to achieve approximate multivariate normality. The results of the multivariate Shapiro-Wilk test (W=0.9782, p-value=0.1816) suggest that normality can at least be reasonably assumed. The plot of the Mahalanobis distances in Figure 5 illustrate an improvement over those obtained from the standard model fit in Figure 2.

The diagnostics suggest that the multivariate Box-Cox transformation was able to

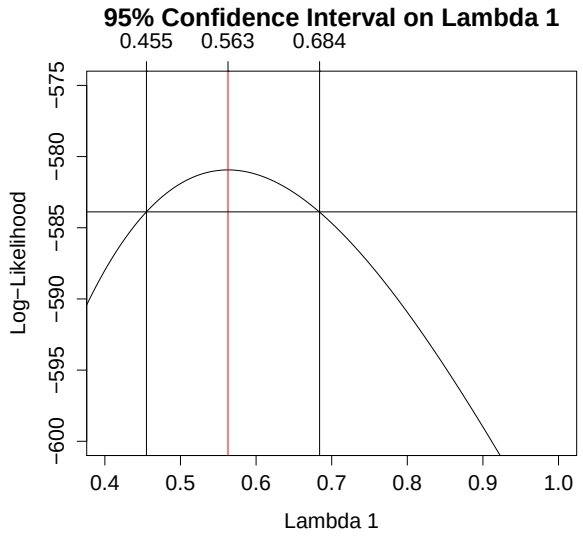

Figure 3: 95% Confidence Interval on λ_1, with $\hat{\lambda}_2 = 0.827$

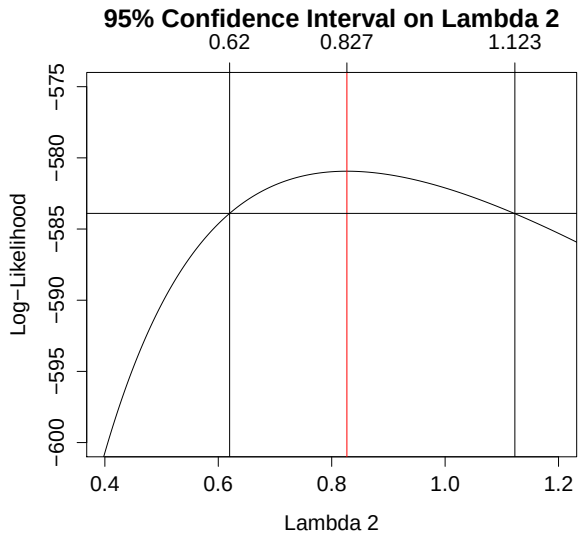

Figure 4: 95% Confidence Interval on λ_2, with $\hat{\lambda}_1 = 0.563$

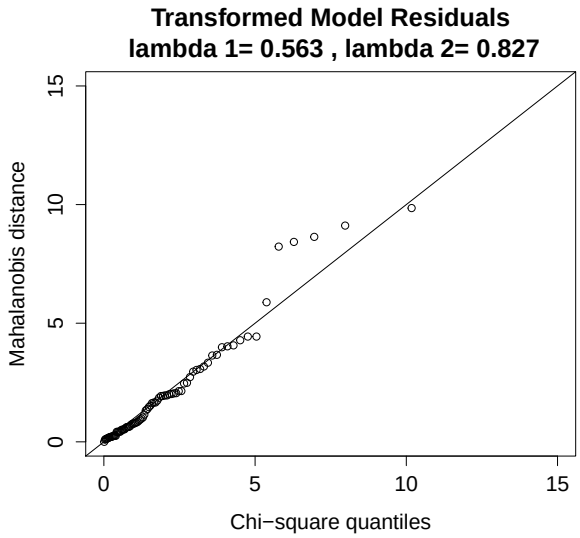

Figure 5: Q-Q Plot of Mahalanobis distances from the fitted model residuals using the Shifted Box-Cox transformation

reasonably achieve multivariate normality in the residuals. However, the Box-Cox transformation is generally limited in its inability to transform negatively skewed variables, given that it inherently assumes the original distribution as power-normal (e.g., see Freeman and Modarres (2006)) and is therefore only appropriate for correcting positively skewed residuals. This underlying assumption may be problematic in general as the direction of skewness in the model residuals is rarely known beforehand. We next turn our attention to the Manly transformation due to its flexibility and ease of use in comparison, as it makes no comparable distinct underlying distributional assumption and can be employed regardless of the direction of underlying skewness.

1.2 Manly Transformation Analysis

While additional shift parameters were necessary in the previous application of the multivariate Box-Cox transformation due its constraint $Y_i > 0$, the Manly transformation in equation (3) has no similar constraint and therefore can be easily applied to the experimental data without estimating any additional shift parameters a priori. Thus, all model parameters can

be obtained through maximum likelihood estimation, again using the *optim* function in R for the optimization. After removing insignificant terms, the final fitted model is found to contain all five main effects, as well as six of the interaction terms, observed in Table 5. The transformation parameters are estimated as $\hat{\gamma}_1 = -0.0019$ and $\hat{\gamma}_2 = 0.0283$.

Table 5: MANOVA results for the fitted model using the Manly transformation

Factor	Wilks' λ	Approx. F	Num. DF	Den. DF	$P(> F)$
A	0.52176	31.164	2	68	2.478×10^{-10}
B	0.34052	65.848	2	68	$< 2.2 \times 10^{-16}$
C	0.27078	91.565	2	68	$< 2.2 \times 10^{-16}$
D	0.81058	7.945	2	68	0.0007927
E	0.83458	6.739	2	68	0.0021373
AB	0.74728	11.498	2	68	4.994×10^{-05}
AC	0.88903	4.244	2	68	0.0183283
AE	0.72116	13.146	2	68	1.490×10^{-05}
BC	0.91164	3.295	2	68	0.0430593
BD	0.89341	4.056	2	68	0.0216655
CE	0.69181	15.147	2	68	3.627×10^{-06}

While the estimates of the transformation parameters, $\hat{\gamma}_1 = -0.0019$ and $\hat{\gamma}_2 = 0.0283$, both seem very close to the nominal value of $\gamma = 0$ (no transformation) at first glance, the values of the Manly transformation parameters are very data dependent and can potentially be extremely large or extremely small. Figures 6 and 7 illustrate the marginal 95% confidence intervals for each of these Manly transformation parameter estimates. As the 95% confidence interval for $\hat{\gamma}_1$ does not contain the nominal value of $\gamma = 0$, this suggests a marginal transformation on the response $Climb(Chart)$ may be justified. Also, since $\hat{\gamma}_1 < 0$, this implies the transformation was necessary to correct for significant positive skew. The 95% confidence interval for $\hat{\lambda}_2$ does contain $\lambda = 0$, suggesting that a marginal transformation on $Climb(-1)$ may not be necessary. Given that a transformation was found to be necessary for $Climb(Chart)$ and the transformation parameters are jointly estimated along with all model parameters, the multivariate transformation is justified.

Diagnostics on the fitted model residuals reveal that the application of the Manly

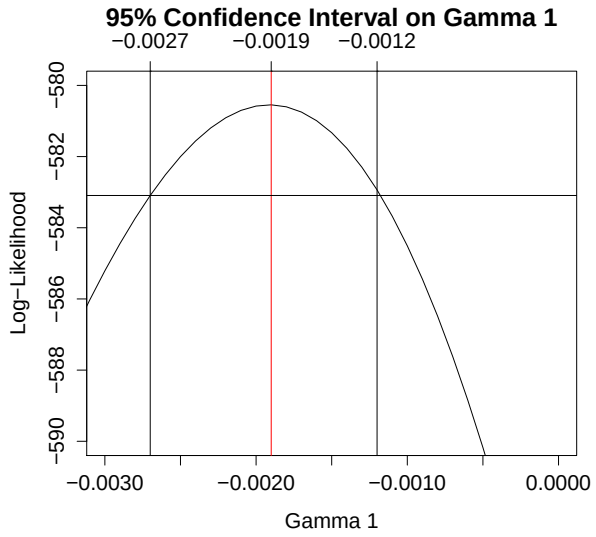

Figure 6: 95% Confidence Interval on γ_1, with $\hat{\gamma}_2 = 0.002$

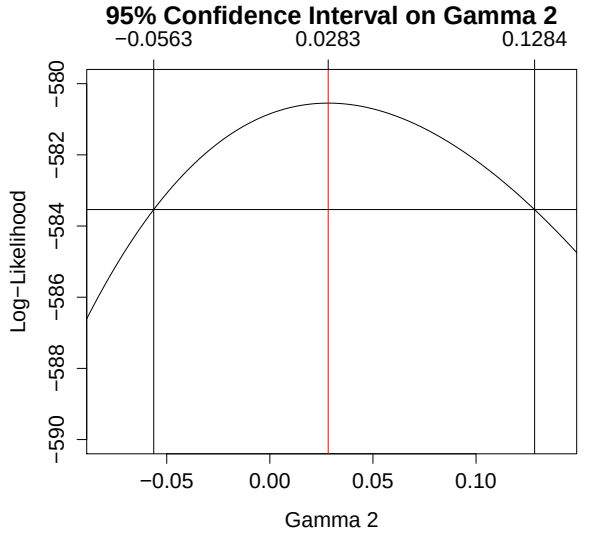

Figure 7: 95% Confidence Interval on γ_2, with $\hat{\gamma}_1 = -0.002$

transformation was able to achieve approximate multivariate normality. The results of the multivariate Shapiro-Wilk test (W=0.9862, p-value=0.5364) suggest that normality can at least be reasonably assumed. Furthermore, the p-value is much higher than that obtained from the model fit using the Box-Cox transformation, suggesting that the Manly transformation may have resulted in residual terms that are more approximately normally distributed by comparison. The plot of the Mahalanobis distances in Figure 8 illustrate an improvement over those obtained from both the standard model fit and the model fit using the Box-Cox transformation observed in Figures 2 and 5, respectively.

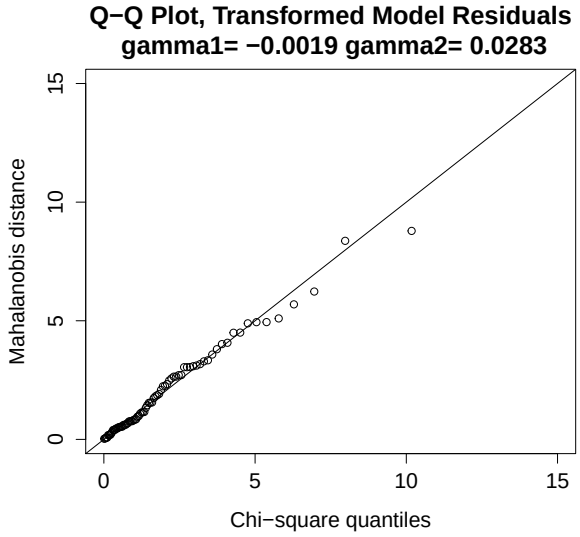

Figure 8: Q-Q Plot of Mahalanobis distances from the fitted model residuals using the Manly transformation

As previously discussed, while multivariate models fit using a transformed response in order to meet model assumptions on the error terms will result in a better model fit over standard analysis, problems arise when predictions and prediction regions are desired in the original units of observation. In what follows, we discuss an approach to approximating higher-order moments of the original q-variate response vector $\mathbf{Y}$, where it is assumed that a normal-theory linear model is fit using a transformed version of $\mathbf{Y}$, with both the Box-Cox transformation and the Manly transformation investigated. We derive a closed-form

approximation to the k^{th} moment of each Y_i $(i = 1, 2, \ldots, q)$, as well as a closed-form approximation for $E(Y_i Y_{i'})$ $(i \neq i')$, in order to estimate the mean vector and covariance matrix of $\mathbf{Y}$, given parameter estimates obtained from fitting the model in the transformed domain. We then exploit multivariate analogs to Chebyshev's inequality to construct approximate $100(1-\alpha)\%$ prediction regions on $\mathbf{Y}$, a q-dimensional spheroid based on the estimators for the mean and variance of each Y_i, and a q-dimensional ellipsoid utilizing the full covariance matrix estimator. Further, we discuss the results of Monte Carlo simulation studies used to assess the performance of the proposed estimators as well as the resulting prediction regions constructed from these estimators.

2 Proposed Methodology

In the preceding example, both the multivariate Box-Cox transformation and the multivariate Manly transformation result in residual terms that may be assumed to follow the multivariate normal distribution. As such, it is of interest to derive methodologies to improve prediction accuracy in the original units for each of these families of transformations.

2.1 Multivariate Box-Cox Transformation

Recall the multivariate extension of the Box-Cox transformation given generally by equation (1),

$$Z_i = \begin{cases} \frac{Y_i^{\lambda_i} - 1}{\lambda_i} & \lambda_i \neq 0 \\ \log_e(Y_i) & \lambda_i = 0 \end{cases}$$

for $i = 1, 2, \ldots, q$, where $Y_i > 0$ is the i^{th} original variable, λ_i is the transformation parameter associated with the i^{th} variable, and Z_i is the i^{th} transformed variable such that the vector $\mathbf{Z}$ can be assumed to follow a q-variate normal distribution with mean vector $\boldsymbol{\mu}_{\mathbf{Z}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$, i.e.

$$\mathbf{Z} \sim N_q(\boldsymbol{\mu}_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}}).$$

The joint density function of the Z_i's is then given by

$$f(z_1, z_2, \ldots, z_q) = \frac{1}{(2\pi)^{\frac{q}{2}}|\boldsymbol{\Sigma}_{\mathbf{z}}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})'\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})).$$

It follows that the joint density function of the original untransformed Y_i's is then given by

$$f(y_1, y_2, \ldots, y_q) = \frac{\prod_{i=1}^{q} y_i^{\lambda_i - 1}}{(2\pi)^{\frac{q}{2}}|\boldsymbol{\Sigma}_{\mathbf{z}}|^{\frac{1}{2}}} \exp(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})'\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})) \tag{4}$$

where the additional term in the numerator, $\prod_{i=1}^{q} y_i^{\lambda_i - 1}$, is the Jacobian of the multivariate Box-Cox transformation.

Using this density of the vector $\mathbf{Y}$, given that a multivariate Box-Cox transformation has been appropriately applied, we are first interested in deriving the moments of the Y_i's. To accomplish this, we first perform a $\mathbf{U}$-substitution, itself a linear transformation. Recall that for a continuous 1-to-1 transformation from, say, $(Y_1, Y_2, \ldots, Y_q)$ to $(U_1, U_2, \ldots, U_q)$ we have

$$\int \cdots \int_R f_{Y_1,\ldots,Y_q}(y_1, \ldots, y_q) dy_1 \cdots dy_q = \int \cdots \int_{R'} f_{Y_1,\ldots,Y_q}(u_1, \ldots, u_q) \left| \frac{\partial(y_1, \ldots, y_q)}{\partial(u_1, \ldots, u_q)} \right| du_1 \cdots du_q$$

where the term $\left| \frac{\partial(y_1,\ldots,y_q)}{\partial(u_1,\ldots,u_q)} \right|$ is the Jacobian of the transformation. Defining $\mathbf{U}$ as $\mathbf{U} = \boldsymbol{\Sigma}_{\mathbf{Z}}^{-\frac{1}{2}}(\mathbf{Z} - \boldsymbol{\mu}_{\mathbf{Z}})$, note that

$$\left| \frac{\partial(u_1, u_2, ..., u_q)}{\partial(y_1, y_2, ..., y_q)} \right| = \left| \boldsymbol{\Sigma}_{\mathbf{z}}^{-\frac{1}{2}} \right| \times \left| \frac{\partial(z_1, z_2, ..., z_q)}{\partial(y_1, y_2, ..., y_q)} \right| = |\boldsymbol{\Sigma}_{\mathbf{z}}|^{-\frac{1}{2}} \prod_{i=1}^{q} y_i^{\lambda_i - 1}$$

so that the density of $\mathbf{U}$ is then given by

$$f_{U_1,\ldots,U_q}(u_1, \ldots, u_q) = \frac{\exp\left(-\frac{1}{2}\mathbf{u}'\mathbf{u}\right)}{(2\pi)^{\frac{q}{2}}}$$

and also note that $\mathbf{Y}$ can now be defined in terms of $\mathbf{U}$ by

$$y_i = \exp\left(\frac{\log_e(1+\lambda_i(\mu_{z_i}+\sum_{j=1}^{q} r_{ij}u_j))}{\lambda_i}\right)$$

where the r_{ij}'s are the elements of $\mathbf{R} = \mathbf{\Sigma}_{\mathbf{z}}^{1/2}$.

We may now define the moments of the i^{th} original untransformed $\mathbf{Y}$ variable in terms of $\mathbf{U}$ as

$$E(Y_i^k) = \int\int\cdots\int_R \exp\left(\frac{\log_e(1+\lambda_i(\mu_{z_i}+\sum_{j=1}^{q} r_{ij}u_j))}{\lambda_i}\right)^k \times \frac{\exp\left(-\frac{1}{2}\mathbf{u}'\mathbf{u}\right)}{(2\pi)^{\frac{q}{2}}} du_1 du_2 \cdots du_q$$

which does not generally have a closed-form solution. However, we can approximate $E(Y_i^k)$ by expanding y_i in a 4^{th}-order multivariate Taylor series about $u_j = 0$ for each j and evaluate the above integral. This ultimately leads to a final closed-form approximation for the moments of Y_i given by

$$E(Y_i^k) \approx (1+\lambda_i\mu_{z_i})^{\frac{k}{\lambda_i}}\left[1 + \frac{k(k-\lambda_i)}{2(1+\lambda_i\mu_{z_i})^2}\sum_{j=1}^{q} r_{ij}^2 + \frac{k(k-\lambda_i)(k-2\lambda_i)(k-3\lambda_i)}{8(1+\lambda_i\mu_{z_i})^4}(\sum_{j=1}^{q} r_{ij}^2)^2\right] \tag{5}$$

for $i = 1, 2, \ldots, q$, where $Y_i > 0$ is the i^{th} original variable, λ_i is the transformation parameter associated with the i^{th} variable, μ_{z_i} is the mean of the i^{th} transformed variable, and the r_{ij}'s are the elements of $\mathbf{R} = \mathbf{\Sigma}_{\mathbf{z}}^{1/2}$.

Additionally, a general expression for $E(Y_iY_{i'})$, $i \neq i'$, may be similarly derived using a 2^{nd}-order multivariate Taylor series expansion and is given by

$$E(Y_iY_{i'}) = a_i^{\frac{1}{\lambda_i}-2} a_{i'}^{\frac{1}{\lambda_{i'}}-2}\left[a_i^2a_{i'}^2 + a_ia_{i'}\sum_{j=1}^{q} r_{ij}r_{i'j} - \frac{a_{i'}^2c_i}{2}\sum_{j=1}^{q} r_{ij}^2 - \frac{a_i^2c_{i'}}{2}\sum_{j=1}^{q} r_{i'j}^2\right]. \tag{6}$$

with $a_i = (1+\lambda_i\mu_{z_i})$ and $c_i = \lambda_i - 1$, for $i \neq i' = 1, 2, \ldots, q$, where $Y_i > 0$ is the i^{th} original variable, λ_i is the transformation parameter associated with the i^{th} variable, the r_{ij}'s are the elements of $\mathbf{R} = \mathbf{\Sigma}_{\mathbf{z}}^{1/2}$, and μ_{z_i} is the mean of the i^{th} transformed variable.

The approximations derived above in equation (5) ultimately allows for the estimation of means and variances for each of the original untransformed Y_i's, given by

$$\hat{\mu}_{Y_i} = E(Y_i)$$

and

$$\hat{\sigma}^2_{Y_i} = E(Y_i^2) - [E(Y_i)]^2,$$

respectively. The approximation derived in equation (6) additionally allows for the estimation of the covariance between Y_i and $Y_{i'}$, given by

$$\hat{\sigma}_{Y_i Y_{i'}} = E(Y_i Y_{i'}) - E(Y_i)E(Y_{i'}).$$

These estimators are then used as the components in the final result, reliable estimators for the mean vector $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$ in the original units of observation. This process of deriving the estimators of interest for the multivariate Box-Cox transformation may be similarly followed for the multivariate Manly transformation.

2.2 Multivariate Manly Transformation

Recall the multivariate extension of the Manly exponential transformation given generally by equation (3),

$$Z_i = \begin{cases} \frac{e^{\gamma_i Y_i} - 1}{\gamma_i} & \gamma_i \neq 0 \\ Y_i & \gamma_i = 0 \end{cases} \tag{7}$$

for $i = 1, 2, \ldots, q$, where Y_i is the original variable, γ_i is the transformation parameter associated with the i^{th} original variable, and Z_i is the transformed variable such that the vector $\mathbf{Z}$ can be assumed to follow a q-variate normal distribution with mean vector $\boldsymbol{\mu}_{\mathbf{Z}}$ and covariance matrix $\boldsymbol{\Sigma}_{\mathbf{Z}}$, i.e.

$$\mathbf{Z} \sim N_q(\boldsymbol{\mu}_{\mathbf{Z}}, \boldsymbol{\Sigma}_{\mathbf{Z}})$$

while the special case of $\gamma_i = 0$ results in no transformation.

The joint density function of the Z_i's is then given by

$$f(z_1, z_2, \ldots, z_q) = \frac{1}{(2\pi)^{\frac{q}{2}}|\boldsymbol{\Sigma}_{\mathbf{z}}|^{\frac{1}{2}}} \exp\big(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})'\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})\big).$$

It follows that the joint density function of the original untransformed Y_i's is then given by

$$f(y_1, y_2, \ldots, y_q) = \frac{\prod_{i=1}^{q} e^{\gamma_i y_i}}{(2\pi)^{\frac{q}{2}}|\boldsymbol{\Sigma}_{\mathbf{z}}|^{\frac{1}{2}}} \exp\big(-\frac{1}{2}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})'\boldsymbol{\Sigma}_{\mathbf{z}}^{-1}(\mathbf{z}-\boldsymbol{\mu}_{\mathbf{z}})\big) \tag{8}$$

where the additional term in the numerator, $\prod_{i=1}^{q} e^{\gamma_i y_i}$, is the Jacobian of the multivariate Manly transformation.

Again utilizing a $\mathbf{U}$-substitution, $\mathbf{Y}$ can now be defined in terms of $\mathbf{U}$ by

$$y_i = \frac{\log_e(1 + \gamma_i(\mu_{z_i} + \sum_{j=1}^{q} r_{ij}u_j))}{\gamma_i}$$

where the r_{ij}'s are the elements of $\mathbf{R} = \boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}$.

We may now define the moments of the i^{th} original untransformed $\mathbf{Y}$ variable in terms of $\mathbf{U}$ as

$$E(Y_i^k) = \int\int\cdots\int_R \left[\frac{\log_e(1 + \gamma_i(\mu_{z_i} + \sum_{j=1}^{q} r_{ij}u_j))}{\gamma_i}\right]^k \frac{1}{(2\pi)^{\frac{q}{2}}} e^{-\frac{1}{2}\mathbf{u}'\mathbf{u}} du_1 \cdots du_q$$

which does not generally have a closed-form solution. However, we can approximate $E(Y_i^k)$ by expanding y_i in a 4^{th}-order multivariate Taylor series about $u_j = 0$ for each j and evaluate the above integral. This ultimately leads to a final closed-form approximation for the moments of Y_i given by

$$E(Y_i^k) \approx \left[\frac{b_i}{\gamma_i}\right]^k \left[1 - \frac{k}{2}\frac{\gamma_i^2(b_i - k + 1)}{a_i^2 b_i^2}\sum_{j=1}^{q} r_{ij}^2 - \frac{k}{8}\frac{\gamma_i^4 c_i}{a_i^4 b_i^4}\left[\sum_{j=1}^{q} r_{ij}^2\right]^2\right] \tag{9}$$

for $i = 1, 2, \ldots, q$, where $Y_i > 0$ is the i^{th} original variable, λ_i is the transformation parameter associated with the i^{th} variable, μ_{z_i} is the mean of the i^{th} transformed variable, and the r_{ij}'s are the elements of $\mathbf{R} = \boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}$. The terms $a_i = (1 + \gamma_i \mu_{z_i})$, $b_i = \log_e(1 + \gamma_i \mu_{z_i})$, and $c_i = 6b_i^3 - 11(k-1)b_i^2 + 6(k-1)(k-2)b_i - (k-1)(k-2)(k-3)$, where μ_{z_i} is the mean of the i^{th} transformed variable, are incorporated in order to condense an otherwise lengthy expression. As we are typically interested in estimating only the first and second moments of Y_i, it is of note that the term c_i reduces to $c_i = 6b_i^3$ in estimation of the first moment and $c_i = 6b_i^3 - 11(k-1)b_i^2$ in estimation of the second moment.

A general expression for $E(Y_i Y_{i'})$, $i \neq i'$, involving q variables may be similarly derived using a 2^{nd}-order multivariate Taylor series expansion and is given by

$$E(Y_i Y_{i'}) \approx \frac{1}{\gamma_i \gamma_{i'} a_i^2 a_{i'}^2} \left[a_i^2 b_i a_{i'}^2 b_{i'} + a_i a_{i'} \gamma_i \gamma_{i'} \sum_{j=1}^{q} r_{ij} r_{i'j} - \frac{a_i^2 b_i \gamma_{i'}^2}{2} \sum_{j=1}^{q} r_{i'j}^2 - \frac{a_{i'}^2 b_{i'} \gamma_i^2}{2} \sum_{j=1}^{q} r_{ij}^2 \right]. \tag{10}$$

with $a_i = (1 + \gamma_i \mu_{z_i})$ and $b_i = \log_e(1 + \gamma_i \mu_{z_i})$, for $i = 1, 2, \ldots, q$, $i \neq i'$, where $Y_i > 0$ is the i^{th} original variable, γ_i is the transformation parameter associated with the i^{th} variable, the r_{ij}'s are the elements of $\mathbf{R} = \boldsymbol{\Sigma}_{\mathbf{z}}^{1/2}$, and μ_{z_i} is the mean of the i^{th} transformed variable.

The approximations derived above in equations (9) and (10) ultimately allow for the estimation of means, variances, and covariances for each of the original Y_i's, given by

$$\hat{\mu}_{Y_i} = E(Y_i),$$

$$\hat{\sigma}_{Y_i}^2 = E(Y_i^2) - [E(Y_i)]^2,$$

and

$$\hat{\sigma}_{Y_i Y_{i'}} = E(Y_i Y_{i'}) - E(Y_i) E(Y_{i'}).$$

These estimators are then used as the components for the mean vector $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$ and the covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$ in the original units of observation.

2.3 Chebyshev Prediction Regions

Each transformation's respective proposed estimators derived above will ultimately yield reliable estimates for the mean vector, $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$, and the covariance matrix, $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$, of the original untransformed variable $\mathbf{Y}$. When prediction is the ultimate goal, these estimators may be used to construct reliable $100(1-\alpha)\%$ joint prediction regions about specific points of interest in the original domain by utilizing the multivariate extension of Chebyshev's inequality discussed in Navarro (2014). Specifically, it may be of interest to construct one of two separate prediction regions, a spheroid obtained by using strictly the individual variance estimators $\hat{\sigma}_{Y_i}$, $i = 1, 2, \ldots, q$, given by

$$||\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}||^2 \leq L \sum_{i=1}^{q} \sigma^2_{Y_i} \tag{11}$$

or an ellipsoid obtained by incorporating the entire estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$, given by

$$(\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}})' \boldsymbol{\Sigma}^{-1}_{\mathbf{Y}} (\mathbf{Y} - \boldsymbol{\mu}_{\mathbf{Y}}) \leq Lq. \tag{12}$$

The value L in equations (11) and (12) theoretically corresponds to $L = \frac{1}{\alpha}$ in order to meet Chebyshev's inequality. As such, this value is associated with achieving *at least* $100(1-\alpha)\%$ coverage from each prediction region. In practice, it may be more effective to attain approximate expected coverage of $100(1-\alpha)\%$, in which case L can be shown via simulation to asymptotically approach $L = \chi^2_{q,1-\alpha}$ for a q-variate $\mathbf{Y}$ vector. Thus, for any desired α, the value L will lie in the interval

$$\left[\chi^2_{(q,1-\alpha)}, \frac{1}{\alpha}\right]. \tag{13}$$

2.4 Motivating Example

With the proposed estimators for the mean vector and covariance matrix of $\mathbf{Y}$ derived for both the Box-Cox transformation and the Manly transformation, we may now return to the motivating example discussed in Section 1. The fitted models previously described based on standard OLS analysis, the Box-Cox transformation, and the Manly transformation can now be compared in terms of predictive ability in the original units of the response. To compare the predictive ability of the models, we can investigate the root mean square error (RMSE) of the fitted values,

$$RMSE = \sqrt{\frac{\text{tr}(\mathbf{Y} - \hat{\mathbf{Y}})(\mathbf{Y} - \hat{\mathbf{Y}})'}{nq}}. \tag{14}$$

The RMSE obtained from the fitted values of each of the three models is given in Table 6. Each of the transformation models utilizing the proposed methodology results in a lower RMSE than that from the standard analysis model. Additionally, the Manly transformation seems to perform better than the Box-Cox model. As each of these models incorporates similar methodology to attain fitted values, this is most likely due to the ability of the Manly transformation to more effectively transform the residual error terms to approximate normality.

Table 6: RMSEs of model fitted values in original units

Method	RMSE
OLS	72.648
Box-Cox	69.163
Manly	64.244

Since the proposed methodology allows us to estimate the means and variances of each original response variable as well as the full covariance matrix, $\hat{\Sigma}_{\mathbf{Y}}$, we are able to also construct prediction regions in the original domain, specifically those given by equations (11) and (12). As initially demonstrated in Figure 1, the standard analysis allows us to construct

prediction regions by utilizing the estimate $\hat{\mathbf{\Sigma}}_\epsilon$ obtained from the model residuals, and as such will result in prediction regions of identical shape, size, and orientation for any predicted point in space. By utilizing equations (5) and (6) for the Box-Cox transformation model and equations (9) and (10) for the Manly transformation model, we are able to estimate the mean vector and the covariance matrix of the original response $\mathbf{Y}$ at *each* point in design space. Thus, the proposed methodology is able to capture changing variability in the response associated with certain point in the design matrix. This ultimately results in prediction regions of varying shape and size, as well as orientation when utilizing the covariance estimators, dependent on which point in design space a prediction is desired.

The scatter plots of the bivariate response from the experimental data in Figures 9 and 10 illustrate the prediction regions constructed from the Box-Cox transformation model at the 9^{th} and 75^{th} design points, respectively. Each of these figures contains the resulting oval prediction region and resulting prediction ellipse constructed around the predicted value in blue, along with the previously illustrated standard analysis prediction region in orange for reference. The proposed prediction regions illustrated for the 75^{th} design point suggest that there may be much more expected variation in observations obtained from this point in the design space.

Similarly, the scatter plots in Figures 11 and 12 illustrate the prediction regions constructed from the Manly transformation model at the 9^{th} and 75^{th} design points, respectively. Each of these figures contains the resulting oval prediction region and resulting prediction ellipse constructed around the predicted value in blue, along with the previously illustrated standard analysis prediction region in orange for reference. As was witnessed by the prediction regions obtained from the Box-Cox model, the proposed prediction regions illustrated for the 75^{th} design point obtained from the Manly model also suggest that there may be much more expected variation in observations obtained from this point in the design space.

While the application of the Box-Cox and Manly transformations using the proposed estimators to the Hutto-Simpson experimental data appears to yield acceptable increased

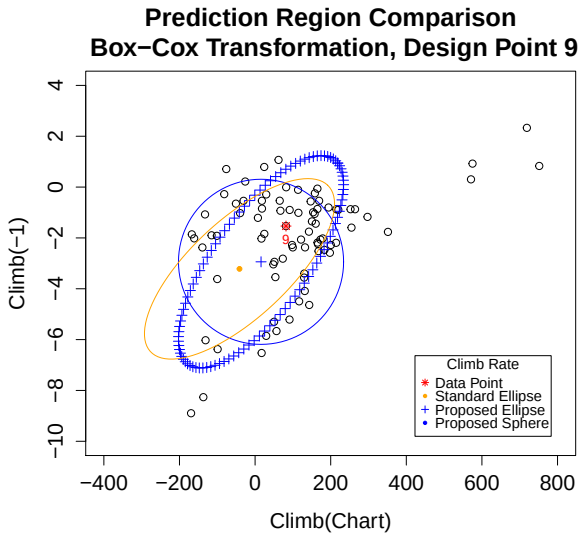

Figure 9: Proposed prediction regions at design point $\mathbf{x}_9$, Box-Cox Transformation

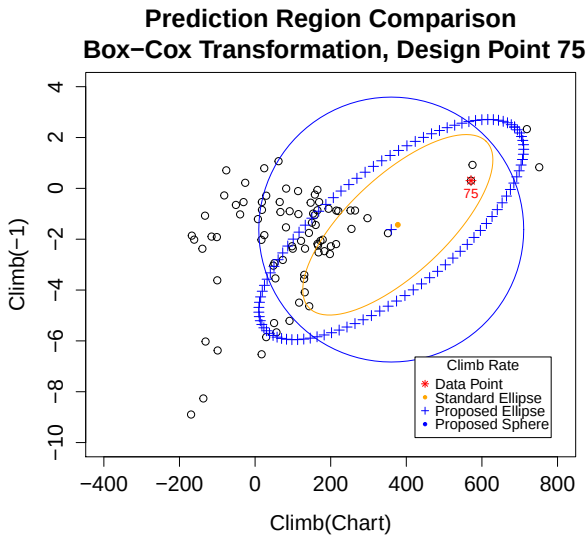

Figure 10: Proposed prediction regions at design point $\mathbf{x}_{75}$, Box-Cox Transformation

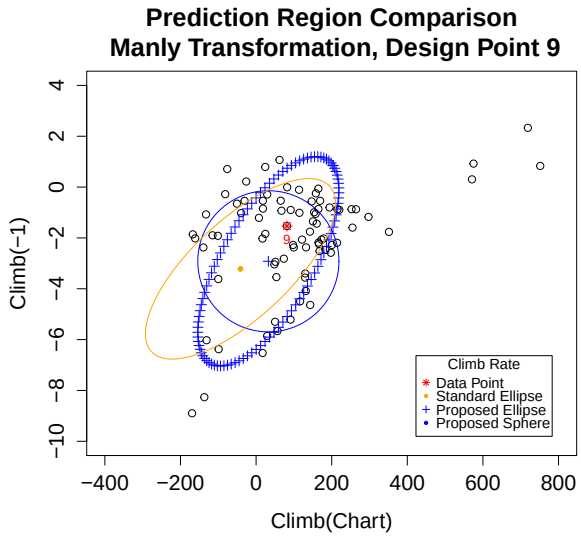

Figure 11: Proposed prediction regions at design point $\mathbf{x}_9$, Manly Transformation

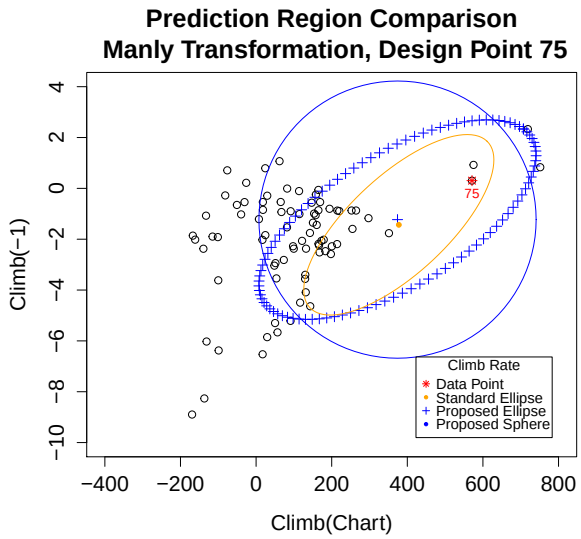

Figure 12: Proposed prediction regions at design point $\mathbf{x}_{75}$, Manly Transformation

performance in predictive ability, it is not clear from this single example how these proposed estimators, $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$ and $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$, and the Chebyshev prediction regions constructed from these estimators will perform on average. In the following section we discuss results of a Monte Carlo simulation study designed to assess the average performance of $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$ and $\hat{\boldsymbol{\Sigma}}_{Y}$ as well as the performance of each of the two resulting prediction regions in equations (11) and (12).

3 Performance Evaluation

To assess the expected performance of our proposed estimators for $\boldsymbol{\mu}_{\mathbf{Y}}$ and $\boldsymbol{\Sigma}_{\mathbf{Y}}$, as well as the relative performance of the multivariate Chebyshev prediction regions, we used Monte Carlo simulation. In what follows, we describe the simulation models in detail, beginning with the multivariate Box-Cox transformation case, followed by the case of the multivariate Manly transformation.

3.1 Box-Cox Evaluation

For any given simulation run, a bivariate random sample $\mathbf{Y}$ of dimension $n \times 2$ was generated from

$$y_{ij} = \begin{cases} \exp\left(\frac{\log_e(1+\lambda_j z_{ij})}{\lambda_j}\right) & \text{for } \lambda_j \neq 0 \\ \exp(z_{ij}) & \text{for } \lambda_j = 0 \end{cases} \tag{15}$$

with $i = 1, 2, \ldots, n$ and $j = 1, 2$ where

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{16}$$

denotes the transformed $n \times 2$ matrix of bivariate responses obtained from the $n \times p$ design matrix $\mathbf{X}$, with $\mathbf{x}_i = [1, x_{i1}, x_{i2}, ..., x_{ik}]$ denoting the i^{th} vector of regressor variable settings ($p = k+1$), the $p \times 2$ matrix of of regression coefficients $\boldsymbol{\beta}$, with $\boldsymbol{\beta}_j = [\beta_0, \beta_1, ..., \beta_k]'$ denoting the $p \times 1$ vector of fixed effects associated with the j^{th} transformed variable Z_j, and the $n \times 2$

matrix $\boldsymbol{\epsilon}$ of random error term where $\boldsymbol{\epsilon} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$. Once the response observations were generated, model parameters were then estimated using maximum likelihood estimation, leading to the following estimators for the unknown model parameters

$$\hat{\boldsymbol{\lambda}} = \arg\max_{\boldsymbol{\lambda}} \Big[-n\log_e(2\pi) - \frac{n}{2}\log_e(|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}|) + \sum_{j=1}^{2}(\lambda_j - 1)\sum_{i=1}^{n}\log_e(y_{ij}) - \frac{1}{2}\mathbf{tr}\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1}(\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}})'(\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}})\Big], \tag{17}$$

where the elements of the matrix $\mathbf{Z}$ are given by $z_{ij} = \frac{y_{ij}^{\lambda_j} - 1}{\lambda_j}$ if $\lambda_j \neq 0$ and $z_{ij} = \log_e(y_{ij})$ if $\lambda_j = 0$, with the row elements of $\hat{\boldsymbol{\mu}}_{\mathbf{Z}}$ given by

$$\hat{\boldsymbol{\mu}}_{\mathbf{Z}_i} = \mathbf{x}_i\hat{\boldsymbol{\beta}} = \mathbf{x}_i(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z} \tag{18}$$

and

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \frac{1}{n - 2p}(\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}})'(\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}}) \tag{19}$$

where the optimization problem in equation (17) was solved using the *optim* function in R. Once parameters were estimated, they were then substituted into the moment approximations in equations (5) and (6) in order to produce estimates for the mean vector, $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$, and covariance matrix, $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$, of the untransformed response vector.

At this point in the simulation model, two prediction regions on $\mathbf{Y}$ were also produced for comparison. The first, based on the spheroid defined by equation (11), ignores the estimated covariance between Y_1 and Y_2 and results in a symmetric oval shaped region. The second, based on equation (12), utilizes the full estimated covariance matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$, resulting in an ellipse rotated based on the level of estimated covariance between Y_1 and Y_2. Additionally, each of these regions were constructed using both the asymptotic estimate and the conservative theoretical estimate of L given in equation (13) for further comparison of

the practical implications of the choice of L. Once these interval estimates were obtained, their performance was observed by the simulation model, over 5000 Monte Carlo simulation runs. For a given simulation run, interval estimate coverage was computed by generating one million observations from the *true* distribution and then recording the proportion of these observations that fell within the estimated intervals. For all simulations, we chose $\alpha = 0.05$.

In our study, we considered fitting a main-effects plus two-factor interactions model using various replicates of a 2^3 factorial designed experiment. We considered values of the transformation vector $\boldsymbol{\lambda}$ at $\lambda_j = [-0.5, 0, 0.5]$ and the response $\mathbf{Z}$ was centered at $\boldsymbol{\mu}_{\mathbf{Z}} = \mathbf{0}$ with the elements of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ set to $\sigma_1^2 = \sigma_2^2 = 0.15^2$ and $\sigma_{12} = 0.2 \times 0.15^2$ to ensure some level of estimable covariance, such that $P[z_{ij} > -\frac{1}{\lambda_j} | \lambda_j > 0] \approx 1$ and $P[z_{ij} < -\frac{1}{\lambda_j} | \lambda_j < 0] \approx 1$, ensuring constraints on the transformation are sufficiently met. Specifically, for the simulations discussed below, the parameter matrix $\boldsymbol{\beta}$ was simulated at $\boldsymbol{\beta}_j = [\mu_{z_j}, 0, \ldots, 0]'$ and the vector $\mathbf{x}_i$ was an arbitrarily chosen factorial point. It makes no difference which point in design space is used in the simulation model since the coefficients $\beta_1, ..., \beta_k$ were all set to zero. We should note that additional simulation runs were performed at other values for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, and at other points $\mathbf{x}_i$ within the design space, and although not reported here, our general conclusions discussed below are not altered.

Table 7 shows estimated root mean square error (RMSE) performances of the estimators for the means and elements of the covariance matrix obtained from equations (5) and (6) for $3^2 = 9$ combinations of $\lambda_j = [-0.5, 0, 0.5]$ with $R = 3$ replicates, with each row of results obtained over 5000 independent Monte Carlo simulation runs. Similarly, Tables 8, 9, and 10 further show the RMSE performance of these estimators for $R = 4$, $R = 5$, and $R = 50$ replicates, respectively.

Notice in Tables 7-10 that, in general, as the number of design replicates R increases, a decrease in the RMSEs for all estimators is observed, regardless of the values of λ_j. This is quite intuitive, and should be expected since with larger R more observations are available to estimate unknown model parameters. Also, note that generally as λ_j increases, a decrease

Table 7: Root mean square errors (RMSEs) of Box-Cox estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 3$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 0.15^2$, and $\sigma_{12} = \sigma_{21} = 0.2 \times 0.15^2$.

λ_1	λ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.5	0.5	0.034	0.033	0.023	0.010	0.022
0.5	0	0.033	0.035	0.022	0.011	0.025
0.5	-0.5	0.033	0.038	0.023	0.012	0.027
0	0.5	0.035	0.033	0.025	0.011	0.023
0	0	0.034	0.035	0.024	0.011	0.025
0	-0.5	0.035	0.039	0.024	0.012	0.027
-0.5	0.5	0.039	0.032	0.028	0.012	0.023
-0.5	0	0.039	0.036	0.027	0.012	0.025
-0.5	-0.5	0.039	0.039	0.028	0.014	0.027

Table 8: Root mean square errors (RMSEs) of Box-Cox estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 4$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 0.15^2$, and $\sigma_{12} = \sigma_{21} = 0.2 \times 0.15^2$.

λ_1	λ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.5	0.5	0.027	0.027	0.013	0.007	0.013
0.5	0	0.027	0.028	0.013	0.007	0.014
0.5	-0.5	0.027	0.029	0.013	0.007	0.016
0	0.5	0.028	0.027	0.014	0.007	0.013
0	0	0.028	0.028	0.014	0.007	0.014
0	-0.5	0.028	0.030	0.014	0.008	0.016
-0.5	0.5	0.030	0.027	0.016	0.007	0.013
-0.5	0	0.031	0.029	0.016	0.008	0.014
-0.5	-0.5	0.031	0.030	0.016	0.009	0.016

Table 9: Root mean square errors (RMSEs) of Box-Cox estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 5$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 0.15^2$, and $\sigma_{12} = \sigma_{21} = 0.2 \times 0.15^2$.

λ_1	λ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.5	0.5	0.024	0.024	0.010	0.005	0.010
0.5	0	0.024	0.025	0.010	0.005	0.011
0.5	-0.5	0.024	0.026	0.010	0.006	0.012
0	0.5	0.025	0.024	0.010	0.005	0.010
0	0	0.025	0.025	0.010	0.006	0.011
0	-0.5	0.025	0.026	0.010	0.006	0.012
-0.5	0.5	0.026	0.024	0.012	0.006	0.010
-0.5	0	0.026	0.024	0.012	0.006	0.011
-0.5	-0.5	0.026	0.026	0.012	0.007	0.012

Table 10: Root mean square errors (RMSEs) of Box-Cox estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 50$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 0.15^2$, and $\sigma_{12} = \sigma_{21} = 0.2 \times 0.15^2$.

λ_1	λ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.5	0.5	0.008	0.008	0.002	0.001	0.002
0.5	0	0.008	0.008	0.002	0.001	0.002
0.5	-0.5	0.007	0.008	0.002	0.001	0.002
0	0.5	0.008	0.008	0.002	0.001	0.002
0	0	0.008	0.008	0.002	0.001	0.002
0	-0.5	0.008	0.008	0.002	0.002	0.002
-0.5	0.5	0.008	0.008	0.002	0.001	0.002
-0.5	0	0.008	0.008	0.002	0.002	0.002
-0.5	-0.5	0.008	0.008	0.002	0.002	0.002

in the RMSE of the estimators is observed for any given R. These results follow from the underlying skewness of the untransformed response, with more skewed distributions transformed by smaller values of λ_j. While the above results illustrate the effectiveness of the estimators for each element of the estimated covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{Y}}$, we also investigated the performance of the full covariance matrix estimator by comparison of matrix norms, specifically the ratio of Frobenius norms as these are 2×2 matrices in the simulation model. This ratio is then defined by

$$\frac{||\hat{\mathbf{\Sigma}}_{\mathbf{Y}}||_F}{||\mathbf{\Sigma}_{\mathbf{Y}}||_F} = \frac{\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2}|\hat{\sigma}_{ij}|^2}}{\sqrt{\sum_{i=1}^{2}\sum_{j=1}^{2}|\sigma_{ij}|^2}} \tag{20}$$

where $\mathbf{\Sigma}_{\mathbf{Y}}$ is calculated from the *true* distribution at each of the 5000 Monte Carlo runs at a given combination of λ_1, λ_2. Table 11 generally illustrates that, as the number of replicates increases, the ratio of the norms defined in equation (20) approaches 1 for all combinations of λ_1 and λ_2. Since the RMSE's of the elements of the covariance matrix estimator decrease as R increases, the estimator for the full covariance matrix will tend to better estimate the true covariance matrix of $\mathbf{Y}$ as R increases, as observed through the ratio of the respective matrix norms.

Table 11: Covariance matrix evaluation by ratio of matrix norms defined in equation (20). Each entry in the table was computed over 5000 independent Monte Carlo runs simulated with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 0.15^2$, and $\sigma_{12} = \sigma_{21} = 0.2 \times 0.15^2$.

λ_1	λ_2	$R = 3$	$R = 4$	$R = 5$	$R = 50$
0.5	0.5	1.842	1.451	1.311	1.022
0.5	0	1.848	1.464	1.315	1.017
0.5	-0.5	1.853	1.455	1.310	1.011
0	0.5	1.867	1.454	1.312	1.021
0	0	1.864	1.457	1.308	1.017
0	-0.5	1.865	1.464	1.309	1.009
-0.5	0.5	1.869	1.454	1.306	1.012
-0.5	0	1.881	1.461	1.313	1.009
-0.5	-0.5	1.886	1.459	1.306	1.002

With the performance of the proposed estimators evaluated, we may now turn our attention to the performance of the prediction regions constructed using our proposed estimators. Table 12 illustrates the estimated coverage obtained in the simulation model from the symmetric oval prediction region constructed using equation (11) with $L = \chi^2(2, 1-\alpha)$ ($\alpha = 0.05$) from equation (13) for each combination of λ_1 and λ_2 over $R = 3, 4, 5, 50$ replicates. Similarly, Table 13 illustrates the estimated coverage obtained from the elliptical prediction region incorporating the full estimated covariance matrix using equation (12) with $L = \chi^2(2, 1-\alpha)$ from equation (13). The oval prediction region tends to be quite conservative, with estimated coverage of ≈ 1.00 for smaller replicates, slightly decreasing as the number of replicates increases. Also of note is the higher estimated coverage obtained for larger values of λ_j, due to the higher levels of skewness in the underlying distribution in cases of smaller values of λ_j previously discussed. As the nominal expected coverage in the simulation model is $1 - \alpha = 0.95$, these results suggest that the spheroidal prediction region definition this region is constructed from may be too conservative in practice. This is due to its lack of incorporating estimates of the covariances between the original response variables. When constructing the elliptical prediction regions, the proposed estimator for these covariances is utilized and the results from this prediction region in Table 13 illustrate that this added information is beneficial in more accurately attaining the expected coverage. The elliptical prediction region still attains estimated coverage of at least 95% for smaller numbers of replicates, but is less conservative than the oval. Additionally, as the number of replicates increases, the estimated coverage obtained from the ellipse approximates the nominal expected coverage, as seen for $R = 50$ replicates. It should also be noted that, while the choice of $L = \frac{1}{\alpha}$ was also investigated in the construction of both prediction regions, it proved to be an extremely conservative choice and led to estimated coverages of 100% for most of the simulation model settings.

Table 12: Estimated coverage of oval prediction regions constructed from the Box-Cox estimators using the spheroid defined in equation (11) with $L = \chi^2(2, 1-\alpha)$. Each entry in the table was computed over 5000 independent Monte Carlo runs simulated with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 0.15^2$, and $\sigma_{12} = \sigma_{21} = 0.2 \times 0.15^2$.

λ_1	λ_2	$R = 3$	$R = 4$	$R = 5$	$R = 50$
0.5	0.5	0.999	0.998	0.998	0.996
0.5	0	0.999	0.998	0.997	0.995
0.5	-0.5	0.998	0.996	0.995	0.993
0	0.5	0.999	0.998	0.997	0.995
0	0	0.998	0.997	0.996	0.994
0	-0.5	0.997	0.996	0.995	0.992
-0.5	0.5	0.998	0.996	0.995	0.993
-0.5	0	0.997	0.996	0.995	0.991
-0.5	-0.5	0.997	0.994	0.993	0.989

Table 13: Estimated coverage of elliptical prediction regions constructed from the Manly estimators using the ellipsoid defined in equation (12) with $L = \chi^2(2, 1-\alpha)$. Each entry in the table was computed over 5000 independent Monte Carlo runs simulated with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 0.15^2$, and $\lambda_{12} = \lambda_{21} = 0.2 \times 0.15^2$.

γ_1	γ_2	$R = 3$	$R = 4$	$R = 5$	$R = 50$
0.5	0.5	0.974	0.966	0.962	0.951
0.5	0	0.973	0.965	0.961	0.951
0.5	-0.5	0.969	0.962	0.959	0.950
0	0.5	0.974	0.964	0.961	0.951
0	0	0.971	0.964	0.959	0.950
0	-0.5	0.968	0.961	0.957	0.949
-0.5	0.5	0.970	0.962	0.959	0.950
-0.5	0	0.968	0.961	0.957	0.949
-0.5	-0.5	0.965	0.957	0.955	0.947

3.2 Manly Evaluation

In evaluating the Manly transformation, for any given simulation run, a bivariate random sample $\mathbf{Y}$ of dimension $n \times 2$ was generated from

$$y_{ij} = \begin{cases} \frac{\log_e(1+\gamma_j z_{ij})}{\gamma_j} & \text{for } \gamma_j \neq 0 \\ y_{ij} & \text{for } \gamma_j = 0 \end{cases} \tag{21}$$

with $i = 1, 2, \ldots, n$ and $j = 1, 2$ where

$$\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon} \tag{22}$$

denotes the transformed $n \times 2$ matrix of bivariate responses obtained from the $n \times p$ design matrix $\mathbf{X}$, with $\mathbf{x}_i = [1, x_{i1}, x_{i2}, ..., x_{ik}]$ denoting the i^{th} vector of regressor variable settings ($p = k+1$), the $p \times 2$ matrix of of regression coefficients $\boldsymbol{\beta}$, with $\boldsymbol{\beta}_j = [\beta_0, \beta_1, ..., \beta_k]'$ denoting the $p \times 1$ vector of fixed effects associated with the j^{th} transformed variable Z_j, and the $n \times 2$ matrix $\boldsymbol{\epsilon}$ of random error term where $\boldsymbol{\epsilon} \sim N_2(\mathbf{0}, \boldsymbol{\Sigma})$. Once the response observations were generated, model parameters were then estimated using maximum likelihood estimation, leading to the following estimators for the unknown model parameters

$$\hat{\boldsymbol{\gamma}} = \arg\max_{\boldsymbol{\gamma}} \left[-n \log_e(2\pi) - \frac{n}{2} \log_e(|\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}|) + \sum_{j=1}^{2} \gamma_j \sum_{i=1}^{n} y_{ij} - \frac{1}{2} \mathbf{tr} \hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}}^{-1} (\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}})'(\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}}) \right], \tag{23}$$

where the elements of the matrix $\mathbf{Z}$ are given by $z_{ij} = \frac{e^{\gamma_j y_{ij}} - 1}{\gamma_j}$ if $\gamma_j \neq 0$ and $z_{ij} = y_{ij}$ if $\gamma_j = 0$, with the row elements of $\hat{\boldsymbol{\mu}}_{\mathbf{Z}}$ given by

$$\hat{\boldsymbol{\mu}}_{\mathbf{Z}_i} = \mathbf{x}_i \hat{\boldsymbol{\beta}} = \mathbf{x}_i (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{Z} \tag{24}$$

and

$$\hat{\boldsymbol{\Sigma}}_{\boldsymbol{\epsilon}} = \frac{1}{n-2p}(\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}})'(\mathbf{Z} - \hat{\boldsymbol{\mu}}_{\mathbf{Z}}) \tag{25}$$

where the optimization problem in equation (23) was solved using the *optim* function in R. Once parameters were estimated, they were then substituted into the moment approximations in equations (9) and (10) in order to produce estimates for the mean vector, $\hat{\boldsymbol{\mu}}_{\mathbf{Y}}$, and covariance matrix, $\hat{\boldsymbol{\Sigma}}_{\mathbf{Y}}$, of the untransformed response vector.

At this point, the two prediction regions on $\mathbf{Y}$ were produced for comparison. Each of these regions were also constructed using both the asymptotic estimate and the conservative theoretical estimate of L given in equation (13). Once these interval estimates were obtained, their performance was observed over 5000 Monte Carlo simulation runs. For a given simulation run, interval estimate coverage was computed by generating one million observations from the *true* distribution and then recording the proportion of these observations that fell within the estimated intervals. For all simulations, we chose $\alpha = 0.05$.

We again considered fitting a main-effects plus two-factor interactions model using various replicates of a 2^3 factorial designed experiment. The response $\mathbf{Z}$ was centered at $\boldsymbol{\mu}_{\mathbf{Z}} = \mathbf{0}$ with the elements of $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$ set to $\sigma_1^2 = \sigma_2^2 = 1$ and $\sigma_{12} = 0.2$ to ensure some level of estimable covariance. We considered values of the transformation vector $\boldsymbol{\gamma}$ at $\gamma_j = [-0.1, 0, 0.1]$, allowing for a positively skewed, symmetric, and negatively skewed j^{th} untransformed variable, respectively, such that $P[z_{ij} > -\frac{1}{\gamma_j}|\gamma_j > 0] \approx 1$ and $P[z_{ij} < -\frac{1}{\gamma_j}|\gamma_j < 0] \approx 1$, ensuring constraints on the transformation are sufficiently met. Specifically, for the simulations discussed below, the parameter matrix $\boldsymbol{\beta}$ was simulated at $\boldsymbol{\beta}_j = [\mu_{z_j}, 0, \ldots, 0]'$ and the vector $\mathbf{x}_i$ was an arbitrarily chosen factorial point. It makes no difference which point in design space is used in the simulation model since these coefficients $\beta_1, ..., \beta_k$ were all set to zero. Additional simulation runs were performed at other values for the parameters $\boldsymbol{\beta}$ and $\boldsymbol{\Sigma}_{\boldsymbol{\epsilon}}$, and at other points $\mathbf{x}_i$ within the design space.

Table 14 shows estimated root mean square error (RMSE) performances of the estimators for the means and elements of the covariance matrix obtained from equations (9)

and (10) for $3^2 = 9$ combinations of $\gamma_j = [-0.1, 0, 0.1]$ with $R = 3$ replicates, with each row of results obtained over 5000 independent Monte Carlo simulation runs. Similarly, Tables 15, 16, and 17 further show the RMSE performance of these estimators for $R = 4$, $R = 5$, and $R = 50$ replicates, respectively.

Table 14: Root mean square errors (RMSEs) of Manly estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 3$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = \sigma_{21} = 0.2$.

γ_1	γ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.1	0.1	0.227	0.232	1.035	0.416	1.037
0.1	0	0.226	0.218	1.026	0.421	0.977
0.1	-0.1	0.230	0.228	1.026	0.429	1.025
0	0.1	0.219	0.228	0.994	0.425	1.044
0	0	0.216	0.216	1.014	0.430	1.000
0	-0.1	0.221	0.229	1.009	0.428	1.036
-0.1	0.1	0.229	0.226	1.032	0.427	1.050
-0.1	0	0.231	0.218	1.043	0.428	0.997
-0.1	-0.1	0.227	0.230	1.017	0.423	1.039

Notice in Tables 14-17 that, in general, as the number of design replicates R increases, a decrease in the RMSEs for all estimators is observed, regardless of the values of γ_j. This is quite intuitive, and should be expected since with larger R more observations are available to estimate unknown model parameters. Also, note that generally for $\gamma_j = 0$, a decrease in the RMSE of the estimators is observed for any given R. These results are also intuitive since an increase in $|\gamma_j|$ implies greater skewness in the underlying distribution of Y_j and thus an increase in the variance of Y_j is observed. Under these circumstances, one would expect to see the RMSEs behave in such a way.

While the above results illustrate the effectiveness of the estimators for each element of the estimated covariance matrix $\hat{\mathbf{\Sigma}}_{\mathbf{Y}}$, we again investigated the performance of the full covariance matrix estimator by investigating the ratio of Frobenius norms as previously defined in equation (20). Table 18 generally illustrates that, as the number of replicates

Table 15: Root mean square errors (RMSEs) of Manly estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 4$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = \sigma_{21} = 0.2$.

γ_1	γ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.1	0.1	0.186	0.188	0.612	0.279	0.607
0.1	0	0.189	0.181	0.603	0.280	0.569
0.1	-0.1	0.187	0.186	0.604	0.289	0.603
0	0.1	0.181	0.188	0.569	0.283	0.607
0	0	0.182	0.182	0.580	0.284	0.574
0	-0.1	0.179	0.184	0.578	0.283	0.600
-0.1	0.1	0.185	0.185	0.597	0.285	0.613
-0.1	0	0.188	0.178	0.602	0.285	0.574
-0.1	-0.1	0.184	0.185	0.597	0.281	0.609

Table 16: Root mean square errors (RMSEs) of Manly estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 5$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = \sigma_{21} = 0.2$.

γ_1	γ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.1	0.1	0.164	0.164	0.444	0.227	0.441
0.1	0	0.162	0.163	0.457	0.227	0.423
0.1	-0.1	0.162	0.162	0.442	0.224	0.439
0	0.1	0.158	0.164	0.430	0.225	0.449
0	0	0.157	0.158	0.423	0.224	0.421
0	-0.1	0.161	0.164	0.416	0.224	0.448
-0.1	0.1	0.164	0.164	0.446	0.226	0.436
-0.1	0	0.162	0.161	0.436	0.223	0.426
-0.1	-0.1	0.164	0.165	0.446	0.224	0.434

Table 17: Root mean square errors (RMSEs) of Manly estimators. Each RMSE entry in the table was computed over 5000 independent Monte Carlo runs simulated using $R = 50$ replicates with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = \sigma_{21} = 0.2$.

γ_1	γ_2	$\widehat{E(Y_1)}$	$\widehat{E(Y_2)}$	$\widehat{Var(Y_1)}$	$\widehat{Cov(Y_1, Y_2)}$	$\widehat{Var(Y_2)}$
0.1	0.1	0.050	0.051	0.080	0.052	0.081
0.1	0	0.051	0.050	0.080	0.053	0.076
0.1	-0.1	0.050	0.050	0.078	0.052	0.079
0	0.1	0.051	0.050	0.074	0.052	0.081
0	0	0.050	0.050	0.074	0.052	0.075
0	-0.1	0.050	0.051	0.074	0.052	0.080
-0.1	0.1	0.051	0.051	0.078	0.052	0.078
-0.1	0	0.051	0.051	0.079	0.051	0.076
-0.1	-0.1	0.051	0.051	0.079	0.052	0.080

increases, the ratio of the norms defined in equation (20) approaches 1 for all combinations of γ_1 and γ_2. Since the RMSE's of the elements of the covariance matrix estimator decrease as R increases, the estimator for the full covariance matrix will tend to better estimate the true covariance matrix of $\mathbf{Y}$ as R increases, as observed through the ratio of the respective matrix norms.

Table 18: Covariance matrix evaluation by ratio of matrix norms defined in equation (20). Each entry in the table was computed over 5000 independent Monte Carlo runs simulated with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = \sigma_{21} = 0.2$.

γ_1	γ_2	$R = 3$	$R = 4$	$R = 5$	$R = 50$
0.1	0.1	1.839	1.460	1.309	1.017
0.1	0	1.840	1.451	1.315	1.019
0.1	-0.1	1.837	1.456	1.305	1.019
0	0.1	1.852	1.451	1.313	1.019
0	0	1.857	1.456	1.310	1.020
0	-0.1	1.849	1.452	1.308	1.021
-0.1	0.1	1.851	1.455	1.309	1.017
-0.1	0	1.849	1.454	1.307	1.022
-0.1	-0.1	1.839	1.450	1.307	1.018

We now turn our attention to the performance of the prediction regions constructed

using our proposed estimators. Table 19 illustrates the estimated coverage obtained in the simulation model from the symmetric oval prediction region constructed using equation (11) with $L = \chi^2(2, 1-\alpha)$ ($\alpha = 0.05$) from equation (13) for each combination of γ_1 and γ_2 over $R = 3, 4, 5, 50$ replicates. Similarly, Table 20 illustrates the estimated coverage obtained from the elliptical prediction region incorporating the full estimated covariance matrix using equation (12) with $L = \chi^2(2, 1-\alpha)$ from equation (13). The oval prediction region tends to be quite conservative, with estimated coverage of ≈ 1.00 for smaller replicates, slightly decreasing to ≈ 0.995 as the number of replicates increases to $R = 50$. As the nominal expected coverage in the simulation model is $1 - \alpha = 0.95$, these results reaffirm that the general spheroidal prediction region may be too conservative in practice. The results from the elliptical prediction region in Table 20 again illustrate that the added covariance information is beneficial in more accurately attaining the expected coverage.

Table 19: Estimated coverage of oval prediction regions constructed from the Manly estimators using the spheroid defined in equation (11) with $L = \chi^2(2, 1-\alpha)$. Each entry in the table was computed over 5000 independent Monte Carlo runs simulated with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = \sigma_{21} = 0.2$.

γ_1	γ_2	$R = 3$	$R = 4$	$R = 5$	$R = 50$
0.1	0.1	0.999	0.998	0.997	0.995
0.1	0	0.999	0.998	0.998	0.996
0.1	-0.1	0.999	0.998	0.997	0.995
0	0.1	0.999	0.998	0.998	0.996
0	0	1.000	0.999	0.999	0.997
0	-0.1	0.999	0.998	0.998	0.996
-0.1	0.1	0.999	0.998	0.997	0.995
-0.1	0	0.999	0.998	0.998	0.996
-0.1	-0.1	0.999	0.998	0.997	0.995

Table 20: Estimated coverage of elliptical prediction regions constructed from the Manly estimators using the ellipsoid defined in equation (12) with $L = \chi^2(2, 1-\alpha)$. Each entry in the table was computed over 5000 independent Monte Carlo runs simulated with $\mu_1 = \mu_2 = 0$, $\sigma_1^2 = \sigma_2^2 = 1$, and $\sigma_{12} = \sigma_{21} = 0.2$.

γ_1	γ_2	$R = 3$	$R = 4$	$R = 5$	$R = 50$
0.1	0.1	0.974	0.966	0.960	0.950
0.1	0	0.976	0.966	0.961	0.950
0.1	-0.1	0.974	0.965	0.960	0.950
0	0.1	0.976	0.966	0.961	0.950
0	0	0.977	0.967	0.962	0.951
0	-0.1	0.975	0.966	0.961	0.950
-0.1	0.1	0.975	0.965	0.961	0.949
-0.1	0	0.976	0.966	0.962	0.950
-0.1	-0.1	0.975	0.965	0.961	0.950

4 Application

To further illustrate the application of our proposed methodology in practice, in what follows we investigate two real data examples. The first involves data obtained from the British general elections of 1974, used to predict observations of a four variable response consisting of the votes allotted to certain political parties by constituency. The second example uses experimental data of a machining process to predict a bivariate response.

4.1 Scottish Election Data

The Scottish election data originally presented by Brown (1980) and further investigated by Breiman and Friedman (1997) considers electoral results for all 71 constituencies in the British general elections of February and October of 1974. The raw data consist of total votes for each of the four political parties - Conservative (C), Labour (S), Liberal (L), and Nationalist (N) - for both the February and the October elections, along with the total size of the electorate (E) of each constituency and a categorical variable listing the region of the constituency (R). The ultimate goal is to build a predictive model in order to predict

October voting results based on the data obtained from the February election.

From this data, four response variables and seven independent variables are generated as in Brown (1980) with $x_1 = C_{Feb}/E$, $x_2 = S_{Feb}/E$, $x_3 = L_{Feb}/E$, and $x_4 = N_{Feb}/E$ representing the proportion of total votes for each party in the February election, three additional coded independent variables given by

$$x_5 = \begin{cases} 0.5 & \text{if } L_{Oct} > 0 \text{and } L_{Feb} = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$x_6 = \begin{cases} 0.5 & \text{if } R = 5, 6 \\ 0 & \text{otherwise} \end{cases}$$

$$x_7 = \begin{cases} 0.5 & \text{if Labour or Nationalist top party in February and } |x_2 - x_4| \leq 0.2 \\ 0 & \text{otherwise,} \end{cases}$$

and $y_j = w_j/E - x_j$, $j = 1, 2, 3, 4$, representing the difference in the proportion of total votes for each party from the February to the October elections, where $w_1 = C_{Oct}$, $w_2 = S_{Oct}$, $w_3 = L_{Oct}$, and $w_4 = N_{Oct}$. After centering the x variables, these variables are then used to build a predictive model of the form $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$, with final predictions of the votes cast in the October election given by $\hat{w}_j = (\hat{y}_j + x_j)E$, $j = 1, 2, 3, 4$.

Table 21, listing the correlation matrix among the four response variables, reveals that each of the response variables is relatively correlated with the others. Additionally, a standard OLS model constructed from the data result in residuals that cannot be assumed to follow multivariate normality (W = 0.8485, p-value = 5.203×10^{-07}). As such, it may be advantageous to incorporate a transformation of the response vector $\mathbf{Y}$ in model building, and since we ultimately are interested in predicting the total votes cast for the political parties in the October election, we need predicted values obtained from the model in the original units of $\mathbf{Y}$.

Utilizing the methodology outlined in this paper, we fit models to the data using

Table 21: Correlation among response variables, Scottish election data

	y_1	y_2	y_3	y_4
y_1	1.000	0.537	-0.424	-0.495
y_2	0.537	1.000	-0.380	-0.393
y_3	-0.424	-0.380	1.000	-0.381
y_4	-0.495	-0.393	-0.381	1.000

the Shifted Box-Cox transformation and the Manly transformation, along with the standard OLS model for comparison. While the prediction regions defined in equations (11) and (12) may be constructed for each of the Box-Cox and Manly models, these regions are impossible to truly visualize in the context of this data given the 4-variable response. Therefore, we are only concerned with observable increases in predictive ability obtained from the Box-Cox and Manly models using the proposed estimators for each transformation. To accomplish this, we incorporate cross-validation methodology, specifically leave-one-out cross-validation (LOOCV) in which n models are fit, each using $n-1$ observations with the i^{th} observation omitted from the model in order to then predict the omitted i^{th} observation ($i = 1, 2, \ldots, n$). Two metrics are used to gauge the predictive ability of the three modeling methods, the root mean square error (RMSE) of prediction and the total number of correctly predicted winning parties in the October election (CPred). The results of the cross-validation analysis, given in Table 22, reveal that each of the transformed models result in a lower RMSE than the OLS model, with the Manly model performing slightly better than the Box-Cox model. All three models correctly predicted 66 of the 71 winning parties in the October election, each incorrectly predicting the results of the five constituencies at observations 8, 15, 37, 43, and 49. Each of the election results from these five constituencies were ultimately decided by a relatively small margin in close races between two of the parties.

The LOOCV analysis also allows a closer look at the transformation parameters, estimated at each step of the LOOCV analysis for both the Box-Cox and Manly models. The average $\hat{\boldsymbol{\gamma}}$ vector values obtained from the 71 fitted Manly models are $\bar{\hat{\gamma}}_1 = -9.59$,

Table 22: LOOCV results, Scottish election data

	RMSE	CPred
OLS	1077.354	66
Box-Cox	1067.298	66
Manly	1066.300	66

$\bar{\hat{\gamma}}_2 = -1.79$, $\bar{\hat{\gamma}}_3 = 4.19$, and $\bar{\hat{\gamma}}_4 = -0.14$. The positive value for $\bar{\hat{\gamma}}_3$ indicates the Manly transformation was attempting to control for a negative skew associated with the residuals of the third variable. As previously discussed, the Box-Cox transformation is generally not appropriate for transforming underlying negative skew, which is most likely the reason we see a slight improvement in the RMSE for the Manly model over the Box-Cox model.

4.2 Machining Experiments

The machining data attributed to Yeo et al. (1989) consists of experimental results of a turning operation, relating Tool Life (in minutes) and Surface Roughness (μm) to 3 factors, Speed (m/min), Feed Rate (mm/rev), and Depth of Cut (mm). For reference, all 24 observations of the experiment are detailed in Table 23.

For this experiment, three linear models are fit using the standard OLS method, the Box-Cox transformation utilizing our proposed methodology, and the Manly transformation also utilizing our proposed methodology. In a machining process such as this, it is typically of interest to determine the factor settings used to minimize an observable variable such as surface roughness, as it is a common measure of the quality of the machined surface. As each of the three models determine that the factor settings at design point 2 result in the lowest predicted surface roughness, the results of the proposed methodology will be illustrated for this specific point in design space, $x_1 = 173.74$, $x_2 = 0.16$, and $x_3 = 0.533$. While the fitted models need not be presented here, it is of note that all three of the independent variables are found to be significant in each of the models, similarly with no interactions

Table 23: Machining data

Obs	x_1 Speed	x_2 Feed Rate	x_3 Depth of Cut	y_1 Tool Life	y_2 = Surface Roughness
1	103.63	0.16	0.533	70	2.24
2	173.74	0.16	0.533	29	1.93
3	103.63	0.36	0.533	60	6.58
4	173.74	0.36	0.533	28	4.93
5	103.63	0.16	1.016	64	2.67
6	173.74	0.16	1.016	32	2.08
7	103.63	0.36	1.016	44	6.86
8	173.74	0.36	1.016	24	6.35
9	134.11	0.23	0.737	35	3.12
10	134.11	0.23	0.737	31	3.45
11	134.11	0.23	0.737	38	3.30
12	134.11	0.23	0.737	35	3.07
13	92.96	0.23	0.737	52	4.04
14	193.55	0.23	0.737	23	2.92
15	134.11	0.12	0.737	40	1.96
16	134.11	0.44	0.737	28	8.23
17	134.11	0.23	0.343	46	2.90
18	134.11	0.23	1.156	33	5.46
19	92.96	0.23	0.737	46	3.53
20	193.55	0.23	0.737	27	2.82
21	134.11	0.12	0.737	37	1.55
22	134.11	0.44	0.737	34	8.64
23	134.11	0.23	0.343	41	3.25
24	134.11	0.23	1.156	28	5.89

found to be significant in any of the fitted models. Table 24 lists the predicted values for Tool Life and Surface Roughness obtained from each modeling method with the original data values observed from the experiment at design point 2 for reference. Also listed are the estimators for the variances of each of the response variables and the covariance between them at this point in design space using the proposed methodology with the Box-Cox and Manly transformations.

Table 24: Machining data predictions and variance/covariance estimators, design point 2

Method	y_1 Tool Life	y_2 = Surface Roughness	$\hat{y}_1$	$\hat{y}_2$	$\hat{\sigma}_1^2$	$\hat{\sigma}_2^2$	$\hat{\sigma}_{12}$
OLS	29	1.93	31.96	1.35	-	-	-
Box-Cox	29	1.93	30.71	1.77	11.93	0.12	-0.34
Manly	29	1.93	30.86	1.69	11.67	0.17	-0.39

To further illustrate the results of the modeling methods at this point in design space, the prediction regions defined by equations (11) and (12) can be constructed for the Box-Cox and the Manly models. The scatter plot in Figure 13 illustrates both of these prediction regions constructed using the Box-Cox transformation model estimators compared to the standard OLS prediction ellipse. Similarly, Figure 14 illustrates these prediction regions constructed using the Manly transformation model estimators, also compared to the standard OLS prediction ellipse

Each of these figures reveal that there is less expected variation in observations obtained at this design point than the standard OLS model suggests, particularly in that of Tool Life. While the Chebyshev prediction regions are expected to yield at least 95% expected coverage for future observations at this design point, they cover much less area than the traditionally constructed prediction ellipse obtained from the OLS model, and thus should be much more useful to practitioners.

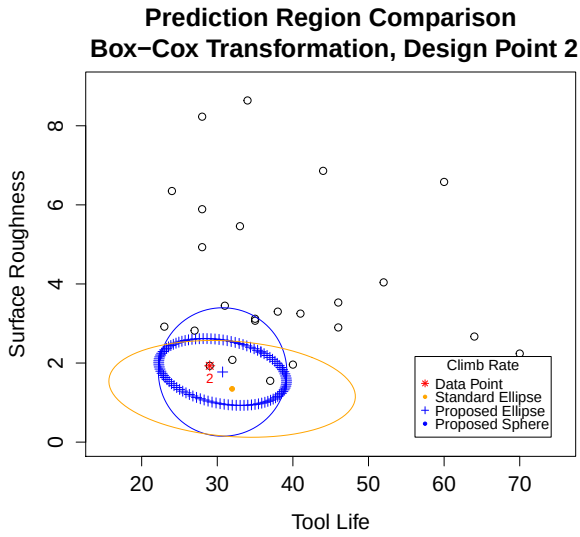

Figure 13: Proposed prediction regions at design point $\mathbf{x}_2$ of the machining data, Box-Cox Transformation

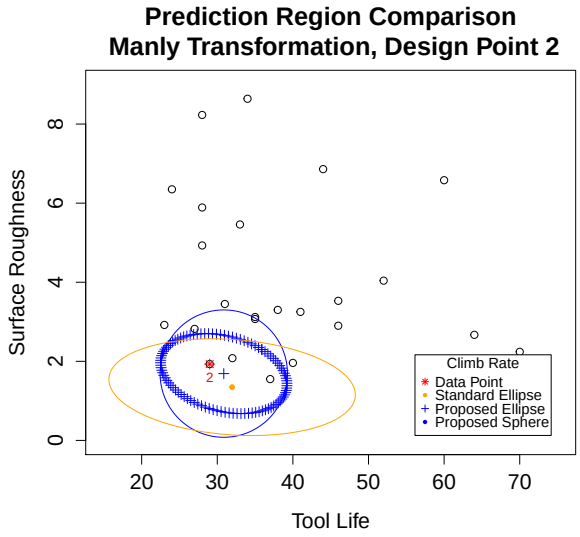

Figure 14: Proposed prediction regions at design point $\mathbf{x}_2$ of the machining data, Manly Transformation

5 Summary and Discussion

Motivated by A-10 single engine aircraft climb experiments, in this paper we demonstrated the use of both the multivariate Box-Cox power transformation and the multivariate Manly exponential transformation in fitting normal-theory linear models to a transformed version of a q-variate response vector $\mathbf{Y}$. We derived closed-form approximations to the k^{th} moment of each original response Y_i $(i = 1, \ldots, q)$, as well as a closed-form approximation to $E(Y_i Y_{i'})$, $(i \neq i')$, which are then used to estimate the mean and variance of Y_i and the covariance between Y_i and $Y_{i'}$, respectively, given parameter estimates obtained from fitting the model in the transformed domain. Exploiting two different multivariate analogs of Chebyshev's inequality, we constructed an approximate $100(1-\alpha)\%$ prediction sphere and ellipsoid, respectively, on the original response vector $\mathbf{Y}$. Using Monte Carlo simulation, we assessed the performance of the proposed estimators for the means, variances, and covariances, as well as the coverage performance of the two Chebyshev prediction regions constructed using the proposed estimators. General results suggest that the proposed estimators for the means, variances, and covariances yield acceptable mean square error performances, and the prediction ellipsoid incorporating the estimated covariance matrix will yield a closer estimate to the nominal coverage, especially as the number of degrees of freedom beyond that required to estimate model terms is small.

There are several advantages to using our proposed methodology with multivariate linear models. Typically, transformations of a response are avoided if possible whenever predictions are desired due to a loss of interpretability in the model and resulting bias in predicted values obtained in the original units of observation. Our methodology is specifically formulated to result in reduced biased predicted values and can be applied for all values of the transformation parameters without loss of interpretation. Further, a check for normality of residuals before application of the Manly transformation is generally unnecessary as the transformation parameter $\gamma_i = 0$ applies no transformation to the i^{th} variable whenever appropriate, and all transformed variables can easily be transformed back to the original

units. Similarly, the Box-Cox transformation also includes a transformation parameter value associated with no transformation, $\lambda_i = 1$, however we recommend that this particular transformation be used with more attention due to its necessary constraints and general inability to adjust for negative skewness. An even more significant advantage lies in the ability to estimate the covariance matrix of the original response at any point in the data or design space. This obviously can have many benefits over standard analysis, as it allows us the ability to make more appropriate inference at points in design space which inherently result in more variation, as well as the ability to construct prediction regions in the original units of varying shape, size, and orientation at these points in space.

References

Andrews, D. F., R. Gnanadesikan, and J. L. Warner (1971). Transformations of multivariate data. *Biometrics 27*, 825–840.

Atkinson, A. C. (1983). Diagnostic regression analysis and shifted power transformations. *Technometrics 25*, 23–332.

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B 26*, 211–252.

Breiman, L. and J. H. Friedman (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society, Series B 59*, 3–54.

Brown, P. J. (1980). Aspects of multivariate regression. *Trabajos de estadística y de investigación operativa 31*, 249–265.

Freeman, J. and R. Modarres (2006). Inverse box-cox: The power-normal distribution. *Statistics and Probability Letters 76*, 764–772.

Hutto, G. T. and J. R. Simpson (2013). A-10 single engine climb using design of experiments. *96 Test Wing Technical Document, Eglin Air Force Base, Florida.*

Manly, B. F. J. (1976). Exponential data transformations. *Journal of the Royal Statistical Society, Series D 25*, 37–42.

Navarro, J. (2014). A very simple proof of the multivariate chebyshev's inequality. *Communications in Statistics - Theory and Methods*. doi:10.1080/03610926.2013.873135.

Perry, M. B. and M. L. Walker (2015). A prediction interval estimator for the original

response when using box-cox transformations. *Journal of Quality Technology*. To Appear.

Sakia, R. M. (1990). Retransformation bias: A look at the box-cox transformation to linear balanced mixed anova models. *Metrika 37*, 345–351.

Sakia, R. M. (1992). The box-cox transformation technique: A review. *Journal of the Royal Statistical Society, Series D 41*, 169–178.

Shumway, R., A. Azari, and P. Johnson (1989). Transformation for environmental data with detection limits. *Technometrics 31*, 347–356.

Taylor, J. (1986). The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association 81*, 114–118.

Walker, M. L. and M. B. Perry (2015). Reduced-bias prediction intervals for the original response when using the manly exponential transformation. *The American Statistician*. Under Review.

Yeo, S. H., M. Rahman, and Y. S. Wong (1989). Towards enhancement of machinability data by multiple regression. *Journal of Mechanical Working Technology 19*, 85–99.

Part V
Overall Conclusion

The research presented in the three previous articles generally details methods proposed to improve predictive ability from models fit using a transformed response. The two distinct data transformations incorporated as the basis of these methods are the very popular Box-Cox power transformation and the more flexible Manly exponential transformation. With models fit using one of these transformations on a single response, we derived estimators for the mean and variance of the original response variable and utilized Chebyshev's inequality to construct prediction intervals in the original domain based on these estimators. The primary advantage to using these proposed estimators over the traditional inverse transformation approach is in the reduction of inherent bias in both the predicted values and the prediction intervals. In addition, the proposed method allows for estimation of not only the mean, but also the variance at each point in the design space. This added benefit then gives us the ability to observe a change in expected variation at specific levels of the experimental factors and thus may result in prediction intervals of varying width.

For models fit using one of these transformations on a multivariate response, we derived estimators for the mean and variance of each individual original response, as well as the covariances between them. We then utilized multivariate analogs of Chebyshev's inequality to we construct prediction regions in the original domain. The extended methodology allows for estimation of the mean vector and full covariance matrix of the original multivariate response at each level of the experimental factors. The ability to estimate the covariance

matrix in this way allows us the ability to make more precise inference at points in the design space. This additionally gives us the ability to construct prediction regions in the original units of varying shape, size, and orientation at each point in space.

In general, the methods presented also address the problem of interpretability when using response transformations. As these methods result in reduced bias predictions in the original domain for all values of the respective transformation parameters, each of these transformations may be used without loss of interpretation. With the lack of interpretation addressed when used in conjunction with our methodology, our general conclusion is that the Manly transformation has increased advantages over the Box-Cox transformation. While each of the two transformations of interest have increased predictive ability over standard methods, the flexibility of the Manly transformation is more easily applicable for general use. When applicable, the Box-Cox transformation is as effective, but this technique should be used with greater care and attention due to its limitations.

References

Andrews, D. F., R. Gnanadesikan, and J. L. Warner (1971). Transformations of multivariate data. *Biometrics 27*, 825–840.

Atkinson, A. C. (1983). Diagnostic regression analysis and shifted power transformations. *Technometrics 25*, 23–332.

Box, G. and D. Cox (1964). An analysis of transformations. *Journal of the Royal Statistical Society, Series B 26*, 211–252.

Breiman, L. and J. H. Friedman (1997). Predicting multivariate responses in multiple linear regression. *Journal of the Royal Statistical Society, Series B 59*, 3–54.

Brown, P. J. (1980). Aspects of multivariate regression. *Trabajos de estadística y de investigación operativa 31*, 249–265.

Carroll, R. J. and D. Rupert (1981). On prediction and the power transformation family. *Biometrika 68*, 609–615.

Cramer, H. (1970). *Random Variables and Probability Distributions*. Cambridge University Press.

Daniel, C. (1976). *Applications of Statistics to Industrial Experimentation*. Wiley, New York.

Draper, N. and D. Cox (1969). On distributions and their transformations to normality. *Journal of the Royal Statistical Society, Series B 31*, 472–476.

Freeman, J. and R. Modarres (2006). Inverse box-cox: The power-normal distribution. *Statistics and Probability Letters 76*, 764–772.

Gurka, M. J., L. J. Edwards, K. E. Muller, and L. L. Kupper (2006). Extending the box-cox transformation to the linear mixed model. *Journal of the Royal Statistical Society, Series A 169*, 273–288.

Han, D. and F. Tsung (2006). A reference free cuscore chart for dynamic mean change detection and a unified framework for charting performance comparison. *Journal of the American Statistical Association 101*, 368–386.

Hutto, G. T. and J. R. Simpson (2013). A-10 single engine climb using design of experiments. *96 Test Wing Technical Document, Eglin Air Force Base, Florida*.

Land, C. E. (1974). Confidence interval estimation for means after data transformation to normality. *Journal of the American Statistical Association 69*, 795–802.

Manly, B. F. J. (1976). Exponential data transformations. *Journal of the Royal Statistical Society, Series D 25*, 37–42.

Miller, D. (1984). Reducing transformation bias in curve fitting. *The American Statistician 38*, 124–126.

Montgomery, D. C. (2009). *Design and Analysis of Experiments*. John Wiley & Sons.

Navarro, J. (2014). A very simple proof of the multivariate chebyshev's inequality. *Communications in Statistics - Theory and Methods*. doi:10.1080/03610926.2013.873135.

Perry, M., J. Kharoufeh, S. Shekhar, J. Cai, and M. R. Shankar (2012). Statistical characterization of nanostructured materials from severe plastic deformation in machining. *IIE Transactions 44*, 534–550.

Perry, M. B. and M. L. Walker (2015). A prediction interval estimator for the original response when using box-cox transformations. *Journal of Quality Technology*. To Appear.

Sakia, R. M. (1988). Application of the box-cox transformation technique to linear balanced mixed analysis of variance models with a multi-error structure. *Unpublished PhD Thesis, Universitaet Hohenheim, FRG*.

Sakia, R. M. (1990). Retransformation bias: A look at the box-cox transformation to linear balanced mixed anova models. *Metrika 37*, 345–351.

Sakia, R. M. (1992). The box-cox transformation technique: A review. *Journal of the Royal Statistical Society, Series D 41*, 169–178.

Shumway, R., A. Azari, and P. Johnson (1989). Transformation for environmental data with detection limits. *Technometrics 31*, 347–356.

Simpson, J., S. Kowalski, and D. Landman (2004). Experimentation with randomization restrictions: Targeting practical implementation. *Quality and Reliability Engineering International 20*, 481–495.

Solomon, P. J. (1985). Transformations for components of variance and covariance. *Biometrika 31*, 233–239.

Taylor, J. (1986). The retransformed mean after a fitted power transformation. *Journal of the American Statistical Association 81*, 114–118.

Walker, M. L. and M. B. Perry (2015). Reduced-bias prediction intervals for the original response when using the manly exponential transformation. *The American Statistician*. Under Review.

Yeo, S. H., M. Rahman, and Y. S. Wong (1989). Towards enhancement of machinability data by multiple regression. *Journal of Mechanical Working Technology 19*, 85–99.