A RAND NOTE

The Rescaling of a Transformed Outcome Variable and Its Interpretations on a Predictive Scale

Naihua Duan, Ker-Chau Li

June 1987

RAND

The research described in this report was supported by a cooperative agreement between the SIAM Institute for Mathematics and Society (SIMS) and the U.S. Environmental Protection Agency, and by The RAND Corporation's own research funds.

Published by The RAND Corporation
1700 Main Street, P.O. Box 2138, Santa Monica, CA 90406-2138

A RAND NOTE

N-2397-SIMS/RC

The Rescaling of a Transformed Outcome Variable and Its Interpretations on a Predictive Scale

Naihua Duan, Ker-Chau Li

June 1987

Prepared for
SIAM Institute for Mathematics and Society

RAND

PREFACE

This Note provides new statistical methods for rescaling a transformed outcome variable so as to obtain useful interpretations on a predictive scale. It is part of an ongoing research effort at The RAND Corporation and the University of California at Los Angeles to develop new statistical methods applicable to environmental and health sciences.

This research is supported principally by a cooperative agreement between the SIAM (Society for Industrial and Applied Mathematics) Institute for Mathematics and Society (SIMS) and the U.S. Environmental Protection Agency, and by The RAND Corporation's own research funds. Other funds were provided by National Science Foundation Grant No. MCS 83-01587 and by Contract No. DAAG 29-80-C-0041 between the Mathematics Research Center, University of Wisconsin-Madison, and the United States Army.

This Note is being issued simultaneously as a RAND Note and as a SIMS Technical Report.

Naihua Duan is a Statistician at The RAND Corporation. Ker-Chau Li is an Associate Professor at the University of California at Los Angeles. Part of this research was done while the authors were visiting the Mathematics Research Center, University of Wisconsin-Madison, supported by the United States Army.

SUMMARY

In studying the relationship between an outcome variable y and regressor variables x, we assume that the regression analysis is made after a transformation of y, and that it is desired to reinterpret the results in terms of a predictive scale. For example, we might consider imposing a global treatment to all individuals in a target population and attempt to estimate the average change in the expected outcome as measured on the predictive scale, averaged over individuals in the target population.

We propose two methods to rescale the transformed outcome variable so that the estimated slope vector based on the *rescaled* transformation can be interpreted on the predictive scale. Under mild regularity conditions, the estimated slope vector based on the *rescaled* transformation is consistent for the *average slope vector* on the predictive scale. For a given target population, the average slope vector measures the average rate of change in the mean response when the regressor variables are perturbated and can be used to approximate the average change in the expected outcome. Under further assumptions, the prediction based on the *rescaled* transformation is consistent for the best linear approximation to the response surface.

For the power transformation $\eta = (y^{\lambda}-1)/\lambda$, if the predictive scale coincides with the observed scale, then one of our rescaling methods is analogous to the Jacobian normalization that Box and Cox (1964, 1982) proposed; our method uses the harmonic mean of $\{y^{\lambda-1}\}$, while Box and Cox's z transformation uses the geometric mean. Our method has the useful interpretation in terms of the average slope vector on the predictive scale.

ACKNOWLEDGMENTS

We appreciate David Wallace and Sanford Weisberg for useful discussions on the motivation of linearizing power transformations. We also appreciate Alan Abrahamse, David Hinkley, James Hodges, Stephen Stigler, and two anonymous referees for useful comments on earlier versions of this Note.

CONTENTS

FIGURE

TABLES

I. INTRODUCTION

We consider an outcome variable y, which is related to a row vector of regressor variables x according to a linear model after a transformation:

$$y = g(\alpha + x\beta + \varepsilon), \quad \varepsilon \sim F(\varepsilon), \tag{1.1}$$

where g is the transformation function, assumed to be known and invertible, α is the unknown intercept, β is the unknown slope vector, and F is the unknown error distribution. We have assumed that the stochastic errors ε are identically distributed, but are not necessarily normal or symmetric.

Given the transformation model (1.1), it is natural to transform the observed outcome by the inverse transformation

$$\eta = g^{-1}(y) = \alpha + x\beta + \varepsilon, \tag{1.2}$$

and regress the transformed outcome η on the regressor x. For example, we might use the ordinary least squares (OLS) estimate, or we might use robust regression estimates such as the M-estimates.

In many situations, our ultimate goal is not just estimating the linear model (1.2) on the transformed scale; instead, we want to reinterpret the results in terms of a predictive scale

$$y^* = h(y) = k(\eta), \tag{1.3}$$

where h is a known transformation function (it might be the identity function), and k denotes the composition of h and g. (See, e.g., Rubin, 1984; Carroll and Ruppert, 1984; Morris, 1984; and Rubin, 1984.)

In this Note we assume that a fixed treatment, denoted as a row vector t, is to be applied to all members in a given target population, and we want to estimate the expected change in the outcome y^*, averaged over individuals in the target population. For an individual with characteristics x prior to the treatment, we assume that his characteristics after the treatment would be x + t. We denote by $\psi(t)$ the expected outcome with treatment t, averaged over the target population $x \sim Q(x)$:

$$\psi(t) = E[k(\alpha + (x + t)\beta + \varepsilon)], \tag{1.4}$$

where the expectation is taken with respect to the joint distribution $(x,\varepsilon) \sim Q(x)F(\varepsilon)$. (We assume that the regressor x is independent of error ε in the target population.) The expected outcome without the treatment is given by $\psi(0)$. The quantity of interest is $D(t) = \psi(t) - \psi(0)$, which can be approximated by

$$D(t) = \psi(t) - \psi(0) \simeq t\nabla\psi(0). \tag{1.5}$$

The gradient vector $\nabla\psi(0)$ on the righthand side of (1.5) has several important properties. First, it can be seen from (1.5) that it is the local rate of change for the expected outcome $\psi(t)$, as the treatment vector t varies near the null vector 0. Second, it can be interpreted as the average of the slope vectors on the response surface, as to be discussed in Section 2. Since this is a crucial interpretation, we will refer to the gradient vector $\nabla\psi(0)$ as the *average slope*. Third, the average slope is usually proportional to the slope vector β on the transformed scale. Assuming that the integration and differentiation can be interchanged, we have

$$\begin{aligned}\nabla\psi(0) &= \beta \bullet \iint k'(\alpha + x\beta + \varepsilon)\, dF(\varepsilon)dQ(x) \\ &= \beta \bullet E[k'(\alpha + x\beta + \varepsilon)].\end{aligned} \tag{1.6}$$

1.1 Examples

There are many examples in which it is important to reinterpret the regression results in terms of a predictive scale. As an example, in acid precipitation studies, the acidity is usually measured and analyzed on the pH scale, the logarithm of the concentration of the hydrogen ion. It is very often necessary to reinterpret the results in terms of the concentration scale.

As another example, we consider the RAND Health Insurance Experiment (HIE), a longitudinal social experiment designed to study how the generosity of health insurance affects medical expenditure. (See, e.g., Duan, 1983; Duan et al., 1983; Newhouse et al., 1981.) In this experiment, medical expenditures were observed on the dollar scale, while the regression analysis was carried out on the logarithmic scale. In order to assess the policy implications of the results, it is necessary to reinterpret the regression analysis in terms of the dollar scale.

The HIE enrolls a random sample of 2,756 families from six sites across the United States, and assigns them to 14 different insurance plans with varying generosity in coverage. The experiment reimburses participants' insurance claims, thereby obtaining a measure of their demand for health care. The outcome variable is the individual annual medical expenditure. Since the outcome variable is fairly skewed, a logarithmic transformation is taken. (A constant of \$5 is added to the annual expenditure to avoid taking the logarithm of zero.) The model is therefore given as follows:

$$\eta = \log(y + \$5) = \alpha + x\beta + \varepsilon. \tag{1.7}$$

(The final analysis in the HIE was based on a model consisting of four regression equations. We have used a substantially simpler model here.)

The first column of Table 1 gives the OLS estimate for the slope vector β in the model (1.7) based on a subset of the HIE sample. In order to avoid dealing with intracluster correlation in this example, we have restricted to the first year of the experiment and have restricted to one randomly chosen member from each family. The main explanatory variable of interest is the log coinsurance rate (LC), which measures the generosity of the insurance coverage. (More precisely, the opposite of the generosity.) The variable is defined to be the logarithm of 100% plus the coinsurance rate, and ranges from zero (zero coinsurance rate) to 0.69368 (100% coinsurance).

Based on the results reported in Table 1, we can predict that when the log coinsurance rate increases by Δ, the expected log annual expenditure will decrease by 0.96Δ. However, the main interest in the HIE is not in the expected log annual expenditure. The main question of interest for the policy decisions is how much the expected annual expenditure (on the dollar scale) would decrease if the log coinsurance rate increases by Δ.

In terms of the notations given earlier, the treatment vector is given by $t = \langle\Delta,0,\ldots,0\rangle$, the goal is to estimate the expected change $D(t) = \psi(t) - \psi(0)$. It is necessary to specify a target population, $x \sim Q(x)$, in order to answer the policy question posed above: for the same

Table 1

SLOPE VECTORS FOR THE HIE DATA

Regressor	β	Aβ[a]	Bβ[b]
LC[c]	-0.959	-195.75	-169.83
Age	0.015	3.01	2.61
Female	0.500	102.19	88.66
Log income[d]	0.507	103.53	89.82
Health[e]	-2.945	-601.37	-521.74

[a]The average slope vector based on the d transformation.
[b]The average slope vector based on the δ transformation.
[c]The logarithm of (100% + coinsurance rate).
[d]The logarithm of annual family income.
[e]A general measure of health.

change in the log coinsurance rate, different target populations might have different changes in the expected expenditure. For this example, we take the target population to be the subset of HIE participants enrolled in an insurance plan that provides free medical care.

One way to estimate D(t) is to go through a formal retransformation procedure for each individual in the target population. For each individual, we predict his expected annual expenditure twice, first based on his original log coinsurance rate, then based on the perturbated log coinsurance rate. This can be done, e.g., using the smearing estimate proposed in Duan (1983) or the approximation method proposed in Taylor (1986). The difference is the change in his expected annual expenditure. The average of the differences over all individuals in the target population is the change in the expected annual expenditure for the target population. The first column of Table 2 gives the predicted changes based on the smearing estimate when the log coinsurance rate increases from zero to Δ.

Table 2

PREDICTED CHANGES IN THE EXPECTED MEDICAL EXPENDITURE

Δ	Smearing ($)	d transformation	δ transformation
0.01	-2.01	-1.96	-1.70
0.05	-9.88	-9.79	-8.49
0.10	-19.29	-19.58	-16.98
0.25	-44.98	-48.94	-42.46
0.50	-80.37	-97.88	-84.92
1.00	-130.13	-195.75	-169.83

The retransformation approach described above can be laborious if the target population is large. It might even be infeasible if we do not have access to individual level data in the target population. For the logarithmic transformation, the effect of varying the regressor variables is multiplicative on the original scale, therefore we can calculate the average difference without using individual level data. For other transformations, we usually need individual level data in order to calculate the average difference.

1.2 Rescaling

As an alternative to the retransformation approach, we can approximate the expected change in the outcome, $D(t)$, using the righthand side of (1.5), $t\nabla\psi(0)$. In the sampling case, we need to estimate the average slope $\nabla\psi(0)$. Since the average slope $\nabla\psi(0)$ is usually proportional to β according to (1.6), we need only estimate β from the linear model (1.2), then multiply the estimated slope vector by the appropriate proportionality constant. This can also be achieved by multiplying the transformed outcome variable η by the same constant. In order to consider predictions along with interpreting slope vectors, we will also consider location shifts. More specifically, we define a rescaling of the transformed outcome variable η to be a linear transformation of the following form:

$$d(\eta) = L + \eta/S,$$

where L and S are scalars to be determined. For example, Box and Cox (1964, 1982) proposed to rescale the power transformation

$$\eta = (y^{\lambda}-1)/\lambda$$

by the Jacobian $\dot{y}^{\lambda-1}$, where $\dot{y}$ denotes the geometric mean of the observed outcome data $\{y_1,\ldots,y_n\}$. This is equivalent to the rescaling defined above with $L = 0$ and $S = \dot{y}^{\lambda-1}$.

According to (1.6), the appropriate scalar S should be an estimate for $1/E(k')$. The scalar $S = \dot{y}^{\lambda-1}$ in Box and Cox's Jacobian rescaling usually does not satisfy this requirement, therefore the slope vector for their rescaled transformation does not estimate the average slope and cannot be used to approximate the expected outcome change D(t).

Alternatively, we propose two methods in Section 3 to rescale the transformation, based on two different estimators for $E(k')$. For the HIE, the estimated average slope vectors based on the two rescaling methods are given as the second and third columns of Table 1. The approximate predicted outcome changes based on (1.5) are given as the second and third columns of Table 2. For small and moderate perturbations in the log coinsurance rate ($\Delta < 0.25$), the approximation is fairly accurate, especially for the first rescaling method. A key condition for the second rescaling method, (A5), does not hold for the HIE data (see, e.g., Duan et al., 1983). For large perturbations in the log coinsurance rate ($\Delta \geq 0.25$), the approximations can be poor, since the linear approximation in (1.5) is poor.

The rescaling methods proposed in Section 3 also impose location shifts on the transformed outcome variable. In the latter part of Section 3, we discuss using the rescaled transformations to estimate the best linear approximation to the response surface.

II. AVERAGE SLOPE

In this section, we consider the population case: we observe (y,x), which follows the transformation model (1.1), and attempt to reinterpret the linear model (1.2) in terms of the predictive scale. First, we discuss several concepts related to reinterpreting (1.2).

The response surface $\nu(x) = E(y^*|x)$ is the conditional expectation of the outcome on the predictive scale for a given design point x. (We are interested here in predicting the mean outcome instead of some other summary of the conditional distribution of y^* given x. For the HIE, this is the relevant estimand of interest. For budgetary considerations, we are interested in the per capita expenditure, which suggests a focus on the prediction of the mean expenditure.)

If a treatment t is applied to an individual with characteristics x, the expected change in his outcome is $\nu(x + t) - \nu(x)$, which can be approximated by

$$\nu(x + t) - \nu(x) \simeq t\nabla\nu(x).$$

The gradient vector $\nabla\nu(x)$ can be interpreted as the local effect of the regressor variables on the outcome y^* at the design point x. This vector is referred to as the *pointwise slope* at the design point x. Figure 1 illustrates the response surface, the pointwise slopes, and the average slope, which will be defined later.

A comprehensive way to understand the effects of the regressor x on the outcome y^* is to give a complete description of all pointwise slopes. However, this might be laborious. Furthermore, even if a complete description is available, it would still be useful to summarize the pointwise slopes for easy comprehension. We therefore consider the *average slope,* which is defined to be the expectation of the pointwise slopes with respect to a given target population. More specifically, the average slope is $E[\nabla\nu(x)]$, where the expectation is taken over the target population $x \sim Q(x)$. The average slope is therefore the *average* of the local effects of the regressor on the outcome y^*.

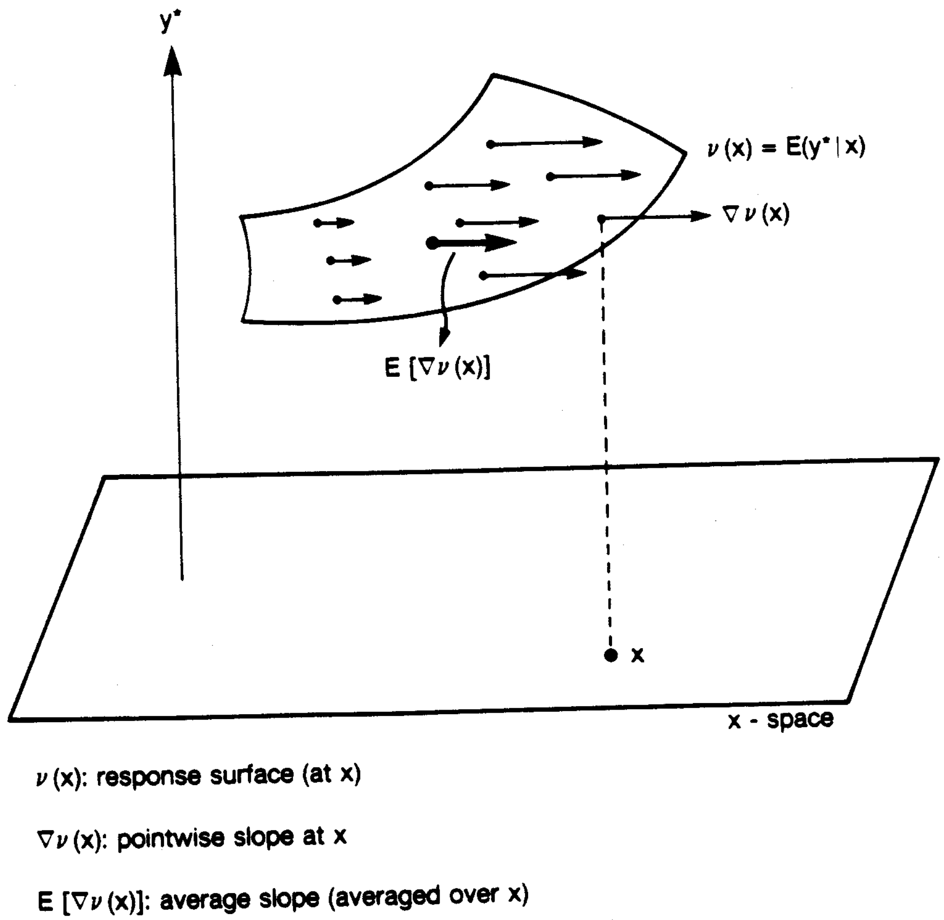

Fig. 1 -- Response Surface, Pointwise Slopes, and Average Slope

A result will be established later in this section (Proposition 2.1) that relates the average slope as defined above with the gradient vector $\nabla\psi(0)$ given in (1.5). We will also give an alternative expression for the proportionality constant $E(k')$ in (1.6). The results require several regularity conditions, which are listed below.

Assumptions

(A1) The transformations g and h in (1.1) and (1.3) are known; g is invertible; the composition

$$k(\eta) = h(g(\eta)) \tag{2.1}$$

is differentiable, where $\eta = g^{-1}(y)$ is the transformed scale on which the linear model (1.2) holds; the expectation $E(y^*|x)$ exists, where $y^* = h(y)$ is the predictive scale (1.3).

(A2) The expectation $E[k'(\eta)|x]$ exists; integration and expectation can be interchanged below:

$$D_u \int k(u + \varepsilon)\, dF(\varepsilon) = \int k'(u + \varepsilon)\, dF(\varepsilon), \tag{2.2}$$

where D_u denotes the partial derivative with respect to u.

(A3) In the target population, the regressor variables x and the error ε are sampled randomly from a joint distribution $Q(x)\bullet F(\varepsilon)$; the expectation $E[k'(\eta)]$ exists, where the expectation is taken over the joint distribution $(x,\varepsilon) \sim Q(x)\bullet F(\varepsilon)$.‖

Proposition 2.1: For a random vector (y,x), which follows the transformation model (1.1), we have the following results.

(1) Under assumption (A1), the response surface depends on x only through $x\beta$:

$$\nu(x) \equiv E(y^*|x) = \tau(x\beta), \tag{2.3}$$

where $\tau(u) \equiv \int k(\alpha + u + \varepsilon)\, dF(\varepsilon)$.

(2) Under assumptions (A1)-(A2), the pointwise slope on the response surface has the following expression:

$$\nabla\nu(x) = \beta \bullet E[k'(\eta)|x]. \tag{2.4}$$

(3) Under assumptions (A1)-(A3), the average slope on the response surface has the following expression:

$$E[\nabla\nu(x)] = \beta \bullet E[k'(\eta)]. \| \quad (2.5)$$

(The proof is straightforward.)

It follows from the proposition that both the pointwise slopes and the average slope on the predictive scale are proportional to the slope vector β on the transformed scale, therefore the ratio β_j/β_k is identical to the corresponding ratio in either the pointwise slope or the average slope, provided that the proportionality constant $E[k'(\eta)|x]$ or $E[k'(\eta)]$ is nonzero. (The latter condition is satisfied, e.g., if the transformation k is strictly monotonic.) In other words, those ratios can be interpreted directly and simultaneously on the transformed scale, the original scale, and the predictive scale.

According to (3) and (1.6), the average slope $E[\nabla\nu(x)]$ as defined above is identical to the gradient vector $\nabla\psi(0)$ in (1.5), therefore it can be interpreted as the local rate of change for the expected outcome $\psi(t)$ and can be used to approximate the expected outcome change using (1.5).

The following corollary gives an alternative expression for the proportionality constant in (2.5).

Corollary 2.1: The proportionality constant in (2.5) has the following expression

$$E[k'(\eta)] = Cov(y^*,\eta)/Var(\eta) \quad (2.6)$$

under conditions (A1)-(A3) and the following conditions:

(A4) The transformation k in (2.1) is absolutely continuous with respect to the Lebesgue measure.

(A5) The transformed dependent variable η is normal.‖

(Proof) The result is a direct application of the following integration-by-parts lemma (see, e.g., Stein, 1981):

Lemma 2.1: Let u be a $N(\mu,\sigma^2)$ real random variable and let the real-valued function g(u) be the indefinite integral of the Lebesgue measurable function g'(u). Assume that $E|g'(u)| < \infty$. Then

$E[g'(u)] = Cov(u,g(u))/Var(u)$.‖

The alternative expression for the proportionality constant given by the corollary is the regression of y^* on η. In other words, the average slope can be interpreted as a combination of two regressions: first we regress y^* on η, then we regress η on x. Duan and Li (1985, Theorem 1) generalized a result in Brillinger (1982) and showed that the two regressions can actually be combined into one:

Proposition 2.2: The OLS regression of $y^* = h(y)$ on x is consistent for $\beta \bullet Cov(y^*,\eta)/Var(\eta)$, under conditions (A1)-(A3) and the following condition:

(A6) The regressor variables x have linear conditional expectations, i.e., for any linear combination xb, the conditional expectation $E(xb|x\beta)$ is linear in $x\beta$.

(The proof is given in the appendix.)

Condition (A6) is satisfied if the distribution of x is normal or elliptically symmetric.

This result is not necessarily restricted to the regression of y^* on x, and is valid for any transformation of y, say, $q = q(y)$. For example, we might have used the wrong transformation q instead of g^{-1} to linearize the model; as a result, we no longer obtain a consistent estimate of β. However, the OLS regression of q on x is consistent for $\beta \bullet Cov(q,\eta)/Var(\eta)$, under conditions (A1)-(A3) and (A6). Hence, we still obtain a consistent estimate for the direction of β.

Remark 2.1: The identity $\nabla\psi(0) = E[\nabla\nu(x)]$ is not restricted to the transformation model (1.1). Actually, it requires only that integration and differentiation be interchangeable. Therefore, the average slope $E[\nabla\nu(x)]$ can be used to approximate the expected outcome change D(t) using (1.5) in very general situations.

Remark 2.2: By Lebesgue's dominated convergence theorem, the differentiation and integration in (2.2) can be interchanged if, for any constant a, we can find a positive constant b and a function q such that

$$|k'(u + \varepsilon)| \leq q(\varepsilon), \quad E[q(\varepsilon)] < \infty,$$

for all u that satisfies $a - b < u < a + b$. (See, e.g., Burrill, 1972, p. 119.) Consider, for example, the power transformations $k(\eta) = \eta^c$. For $c - 1 > 0$, we have the upper bound

$$|k'(u + \varepsilon)| \leq c2^{c-1} \bullet \{\max[|a-b|^{c-1}, |a + b|^{c-1}] + |\varepsilon|^{c-1}\};$$

the upper bound has a finite expectation if the $(c - 1)$-st moment for ε exists.

III. RESCALING

In this section, we consider the sampling case: a sample $\{(y_i,x_i), i = 1,\ldots,n\}$ from the transformation model (1.1) is observed. We take the transformation $\eta = g^{-1}(y)$ on the outcome variable, and regress η on x. It is assumed that

(A7) the estimated slope b based on regressing the transformed outcome η on the regressors x is consistent for the slope β on the transformed scale.

This condition would be satisfied under appropriate regularity conditions on the error distribution F and the design matrix $X = [x_1',\ldots,x_n']'$. For example, for the OLS estimate, the condition is satisfied if $E(\varepsilon)$ exists and $X'X/n$ converges to a nonsingular matrix.

In order to estimate the average slope on the predictive scale, we also need to estimate the proportionality constant in (2.5) or (2.6). We assume that we have a random sample from the target population, $\{(y_{0j},x_{0j}), j = 1,\ldots,m\}$, which follows the transformation model (1.1) with $x \sim Q(x)$. We can then estimate the proportionality constants from this sample.

The two samples described above might be the same, but they need not be so. In order to distinguish between the two, we will refer to the first sample on which we estimate the slope as the *estimation sample*, and refer to the second sample on which we estimate the proportionality constant as the *prediction sample*. In the HIE example, participants from all insurance plans were used to estimate the slope, but we restricted to the participants enrolled in the free plan to estimate the proportionality constants, i.e., the free plan participants are taken as the target population.

The proportionality constant $E[k'(\eta)]$ in (2.5) can be estimated from the prediction sample consistently by the sample average

$$A = m^{-1}\Sigma_j k'(\eta_{0j}), \tag{3.2}$$

where $\eta_{0j} = g^{-1}(y_{0j})$. It follows that the rescaled estimated slope Ab is consistent for the average slope on the predictive scale.

Similarly, the righthand side of (2.6) can be estimated consistently by the OLS regression of y^* on η:

$$B = C(y^*,\eta)/S(\eta), \tag{3.3}$$

where $C(y^*,\eta)$ denotes the sample covariance between y^* and η, and $S(\eta)$ denotes the sample variance of η, both based on the prediction sample. It follows that the rescaled estimated slope Bb is consistent for the average slope on the predictive scale.

Instead of rescaling the estimated slope, we can also rescale the transformed outcome $\eta = g^{-1}(y)$ by the multiplicative scalar A in (3.2) or B in (3.3) before applying the regression analysis. The rescaled transformation is referred to as the d and δ transformations:

$$d_i = M(y^*) + A\bullet(\eta_i - \bar{\eta}), \tag{3.4}$$

$$\delta_i = M(y^*) + B\bullet(\eta_i - \bar{\eta}), \tag{3.5}$$

where $M(y^*)$ denotes the sample average of y^*. (We discuss the location shifts in (3.4) and (3.5) in the latter part of this section.) The rescaled transformations can be interpreted as linear approximations to the transformation k. Note that the measurement units for the d and δ transformations are the same as the unit for the predictive scale y^*. Note also that the δ transformation is the OLS prediction for y^* based on the regression of y^* on η.

In order to relate the estimated slopes based on the rescaled transformations d and δ to the estimated slope based on the transformation η, we assume that

(A8) the regression estimate preserves location and scale, i.e., when we regress the rescaled outcome $L + \eta/S$ on the regressor x, the estimated intercept and slope are, respectively, $L + a/S$ and b/S, where a and b are the estimated intercept and slope based on regressing η on x.

This condition is obviously satisfied for the OLS estimate. It is also satisfied for robust regression estimates that allow an adaptive scale parameter.

Under condition (A8), the estimated slopes based on d and δ are, respectively, Ab and Bb. We have assumed in (A7) that the estimated slope on the transformed scale is consistent for β. The sample estimates A and B are consistent for the corresponding proportionality constants. Therefore, the estimated slopes based on d and δ are consistent for the average slope. The following proposition summarizes the above discussions.

Proposition 3.1: Given an estimation sample $\{(y_i,x_i),i = 1,\ldots,n\}$ from the transformation model (1.1), and a prediction sample $\{(y_{0j},x_{0j}),j = 1,\ldots,m\}$ from model (1.1) with $x \sim Q(x)$, we have the following results.

(1) When we regress the rescaled transformation d in (3.4) on the regressor x, the estimated slope b(d) is consistent for the average slope on the predictive scale under conditions (A1)-(A3) and (A7)-(A8).

(2) When we regress the rescaled transformation δ in (3.5) on the regressor x, the estimated slope b(δ) is consistent for the average slope on the predictive scale under conditions (A1)-(A5) and (A7)-(A8).‖

For the power transformation $\eta = (y^\lambda - 1)/\lambda$, Box and Cox (1964) proposed to rescale η by the Jacobian $\dot{y}^{\lambda-1}$ into the z transformation:

$$z_i = (y_i^\lambda - 1)/\lambda\dot{y}^{\lambda-1},$$

where $\dot{y}$ denotes the geometric mean of $\{y_i\}$. Box and Cox (1982), Box and Fung (1983), and Hinkley and Runger (1984) gave further recommendations on the use of the z transformation in their discussions of Bickel and Doksum (1981). Duan (1986) proposed a location shift to the z transformation:

$$\zeta_i = \dot{y} + (y_i^\lambda - \dot{y}^\lambda)/\lambda\dot{y}^{\lambda-1}.$$

Note that ζ has the same unit as y, while z does not have a meaningful unit. (The difference between z and ζ is irrelevant for estimating the slope.)

The rescaling in the z and ζ transformations implicitly assumes that the observed scale y is the predictive scale, i.e., h is the identity function in (1.3), and that the prediction sample is identical to the estimation sample. Under these assumptions, the d transformation (3.4) for the power transformation is given as follows:

$$d_i = \bar{y} + [y_i^{\lambda} - M(y^{\lambda})] \bullet M[(\lambda y^{\lambda-1})^{-1}],$$

where $M(f(y)) = n^{-1}\Sigma_i f(y_i)$.

Instead of rescaling the power transformation $\eta = (y^{\lambda}-1)/\lambda$ by the geometric mean of $\{y^{\lambda-1}\}$ as in the z and ζ transformations, in the d transformation we rescale η by the harmonic mean of $\{y^{\lambda-1}\}$. According to Proposition 3.1, the d and the δ transformations have the useful interpretation that the estimated slopes based on those rescaled transformations consistently estimate the average slope on the predictive scale; the z and ζ transformations do not have such interpretations.

We have adjusted the location as well as the scale in the d and δ transformations. The location shift is irrelevant for the estimation of slopes; under condition (A8), the estimated slopes are invariant to location shifts. We now consider the implications of the location shift in terms of making predictions.

For any given design point x_0, the linear prediction based on the d or δ transformation is consistent for the following linear approximation to the response surface (2.3):

$$\rho(x_0) = E(y^*) + [x_0-E(x)] \bullet \beta \bullet E[k'(\eta)], \qquad (3.6)$$

under the conditions in Proposition 3.1 and the following condition:

(A9) The regression estimate preserves the mean asymptotically, i.e., $\bar{\eta}$ - (a + $\bar{x}$b) converges to zero, where a and b are the estimated intercept and slope based on regressing η on x; $\bar{\eta}$ and $\bar{x}$ are the sample averages taken over the prediction sample.

Condition (A9) is obviously satisfied for the OLS estimate. It is not usually satisfied for robust regression estimates. However, the robust regression estimates can be modified to satisfy this condition: We can use the estimated slope b based on the robust method and estimate the intercept by $\bar{\eta} - \bar{x}b$.

The hyperplane $\rho(x_0)$ passes through the center of mass $(E(y^*),E(x))$; its slope is the average slope on the response surface. The following proposition indicates that it is a good linear approximation to the response surface.

Proposition 3.2: Under conditions (A1)-(A6), the hyperplane (3.6) is the best linear approximation to the response surface $\nu(x) = E(y^{*}|x)$ in the sense of minimizing the mean squared error

$$E[(\nu(x) - (a + xb))^{2}]$$

over a and b, where the expectation is taken over the target population $x \sim Q(x)$.‖

(The proof is given in the appendix.)

Remark 3.1: It is not necessary for the error distribution $F(\varepsilon)$ to be the same in the estimation and the prediction samples. It is not necessary for the transformations to be the same in the two samples, either. It is necessary only that the two samples have the same slope on their respective transformed scales. All we need, in essence, is to estimate β on the estimation sample's transformed scale and estimate the proportionality constants using (3.2) or (3.3) for the prediction sample.

Remark 3.2: If we observe the regressor x_{0j}'s for the prediction sample, but do not observe the corresponding outcome y_{0j}'s, we can still estimate the proportionality constants if we assume that the error distributions in the two samples are the same, so that we can combine the regressor distribution in the prediction sample with the error distribution based on the estimation sample. For example, we can estimate the proportionality $E(k')$ in (2.5) as follows:

$$A^{*} = (mn)^{-1} \Sigma_{ij} k'(a + x_{0j}b + e_{i}),$$

where a and b are the estimates for α and β, and e_i is the i-th residual, based on fitting the linear model (1.2) to the estimation sample.

Remark 3.3: We would like to thank David Wallace for his suggestion to consider secant instead of tangent linearizations of the power transformations. Note that ζ is a tangent linearization to the curve (y,y^{λ}), while d and δ are both secant linearizations.

APPENDIX

We provide a new proof for Propositions 2.2 and 3.2 below that is shorter than the earlier proof in Duan and Li (1985) and the related proof in Brillinger (1982).

The OLS regression of $q = q(y)$ on x is consistent for the solution to the following minimization problem:

Minimize $R(a,b) = E(a + xb - q)^2$ over a, b.

By Jensen's inequality, we have

$$R(a,b) = E\{E[(a + xb - q)^2 | x\beta, \varepsilon]\}$$
$$\geq E\{(a + E[xb|x\beta] - q)^2\}.$$

By Condition (A6), the conditional expectation in the square bracket is linear in $x\beta$, therefore the minimizer for R has the form $(a, c\beta)$, where a and c are scalars to be determined. The optimal values of a and c can be determined from the equations

$$\partial R/\partial a = 0, \ \partial R/\partial c = 0,$$

from which it follows that the optimal values are

$$c^* = \text{Cov}(q,\eta)/\text{Var}(\eta),$$
$$a^* = E(q) - c^* \bullet E(x\beta).$$

The optimal value of c given above establishes Proposition 2.2 in the general case for the OLS regression of t on x. Furthermore, under conditions (A4) and (A5), we have identity (2.6); this establishes Proposition 3.2.

REFERENCES

Bickel, P. J., and K. A. Doksum, "An Analysis of Transformations Revisited," *Journal of the American Statistical Association*, Vol. 76, 1981, pp. 293-311.

Box, G.E.P. and D. R. Cox, "An Analysis of Transformation," *Journal of the Royal Statistical Society*, Series B, Volume 26, 1964, pp. 211-252.

Box, G.E.P. and D. R. Cox, "The Analysis of Transformations Revisited, Rebutted," *Journal of the American Statistical Association*, Volume 77, 1982, pp. 209-210.

Box, G.E.P. and C. A. Fung, "Some Considerations in Estimating Data Transformations," Technical Summary Report #2609, Mathematics Research Center, University of Wisconsin, Madison, 1983.

Brillinger, D. R., "A Generalized Linear Model with Gaussian Regressor Variables," in Bickel, Doksum, and Hodges (eds.), *A Festschrift for Erich L. Lehmann,* Wadsworth Statistics/Probability Series, Wadsworth International Group, Belmont, CA, 1982, pp. 97-114.

Burrill, C. W., *Measure, Integration, and Probability*, McGraw-Hill, New York, 1972.

Carroll, R. J., and D. Ruppert, "Discussion of 'The Analysis of Transformed Data' by Hinkley and Runger," *Journal of the American Statistical Association,* Volume 79, 1984, pp. 312-313.

Duan, N., "Smearing Estimate: A Nonparametric Retransformation Method," *Journal of the American Statistical Association*, Volume 78, 1983, pp. 605-610.

Duan, N., "Scale-Sensitivity for Prediction Procedures in the Box-Cox Power Transformation Model," unpublished manuscript, June 1986.

Duan, N., and K.-C. Li, "The Ordinary Least Squares Estimation for the General-Link Linear Models, with Applications," Technical Summary Report #2880, Mathematics Research Center, University of Wisconsin-Madison, 1985.

Duan, N., W. G. Manning, C. N. Morris, and J. P. Newhouse, "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business & Economic Statistics*, Volume 1, 1983, pp. 115-126.

Hinkley, D. V. and G. Runger, "The Analysis of Transformed Data," *Journal of the American Statistical Association*, Volume 79, 1984, pp. 302-309.

Morris, Carl N., "Discussion of 'A Case Study of Bayesian/Likelihood Methods of Inference: Estimating the Total in a Finite Population Using Transformations to Normality,' by D. Rubin," *Proceedings of the Conference on Scientific Inference, Data Analysis, and Robustness*, Academy Press, New York, 1984.

Newhouse, Joseph P., et al., "Some Interim Results from a Controlled Trial of Cost Sharing in Health Insurance," *New England Journal of Medicine*, Vol. 305, 1981, pp. 1501-1507.

Rubin, D. B. "Discussion of 'The Analysis of Transformed Data' by Hinkley and Runger," *Journal of the American Statistical Association*, Volume 79, 1984, pp. 309-312.

Rubin, D. B., "A Case Study of Bayesian/Likelihood Methods of Inference: Estimating the Total in a Finite Population Using Transformations to Normality," *Proceedings of the Conference on Scientific Inference, Data Analysis, and Robustness*, Academy Press, New York, 1984.

Stein, C. M., "Estimation of the Mean of a Multivariate Normal Distribution," *Annals of Statistics*, Vol. 9, 1981, pp. 1135-1151.

Taylor, Jerome M. G., "The Retransformed Mean after a Fitted Power Transformation," *Journal of the American Statistical Association*, Vol. 81, 1986, pp. 114-118.

RAND/N-2397-SIMS/RC