NBER WORKING PAPER SERIES

WHY TRANSFORM Y?
A CRITICAL ASSESSMENT OF DEPENDENT-VARIABLE TRANSFORMATIONS IN REGRESSION MODELS FOR SKEWED AND SOMETIMES-ZERO OUTCOMES

John Mullahy
Edward C. Norton

Working Paper 30735
http://www.nber.org/papers/w30735

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
December 2022

Thanks are owed to Partha Deb, Matt Harris, Dan Millimet, Bhavna Rai, Paul Rathouz, Jon Skinner, and participants at the 2022 Annual Health Econometrics Workshop, the Midwest Econometrics Group, the Asian Workshop on Econometrics and Health Economics, and Pennsylvania State University for helpful comments. This paper is dedicated to the memory of Will Manning, our colleague, mentor, and friend. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes
John Mullahy and Edward C. Norton
NBER Working Paper No. 30735
December 2022
JEL No. C18,C20,I10

ABSTRACT

Dependent variables that are non-negative, follow right-skewed distributions, and have large probability mass at zero arise often in empirical economics. Two classes of models that transform the dependent variable y — the natural logarithm of y plus a constant and the inverse hyperbolic sine — have been widely used in empirical work. We show that these two classes of models share several features that raise concerns about their application. The concerns are particularly prominent when dependent variables are frequently observed at zero, which in many instances is the main motivation for using them in the first place. The crux of the concern is that these models have an extra parameter that is generally not determined by theory but whose values have enormous consequences for point estimates. As these parameters go to extreme values estimated marginal effects on outcomes' natural scales approach those of either an untransformed linear regression or a normed linear probability model. Across a wide variety of simulated data, two-part models yield correct marginal effects, as do OLS on the untransformed y and Poisson regression. If researchers care about estimating marginal effects, we recommend using these simpler models that do not rely on transformations.

John Mullahy
University of Wisconsin-Madison
Dept. of Population Health Sciences
787 WARF, 610 N. Walnut Street
Madison, WI 53726
and NBER
jmullahy@facstaff.wisc.edu

Edward C. Norton
Department of Health Management and Policy
Department of Economics
University of Michigan
School of Public Health
1415 Washington Heights, M3108 SPHII
Ann Arbor, MI 48109-2029
and NBER
ecnorton@umich.edu

I think he be transform'd into a beast

Wm. Shakespeare, *As You Like It* [II,7]

Introduction

A familiar and longstanding concern in empirical economics is the specification of regression models for nonnegative outcomes y. The particular setting we consider is one where the outcomes are not only nonnegative but also where a nontrivial share of outcomes equal zero and where the zeros are true zeros, as opposed to missing, which is a different problem. For example, in health economics (the authors' field) prominent examples include healthcare use (e.g., Mullahy, 1998), spending on health services (e.g., Manning et al., 1987), and a variety of health outcome measures (e.g., Khan et al., 2008). Outcomes having similar measurement properties arise as well in labor economics, international trade, agricultural economics, and many others.

Much has been written about econometric strategies for such modeling (e.g., Deb and Norton, 2018) and much more has been written summarizing studies that have deployed such strategies in empirical analysis. The purpose of this paper is to assess the properties, merits, and shortcomings of some such methods that are used commonly in empirical work. While the methods we consider will be familiar to many readers, we speculate that some important properties of these methods may not be as familiar. Our ultimate goal is for empirical researchers considering applying such methods to be well acquainted with what particular methods can and cannot deliver with respect to answering the fundamental research questions they seek to answer.

Our particular concern is with specifications that transform the outcome variable, often done to deal with right skewness in the positives outcomes. If scalar outcomes are denoted y and an exogenous vector of covariates is denoted $\boldsymbol{x}$, then a canonical specification is $f(y) = \boldsymbol{x}'\beta + \varepsilon$, where $f(\cdot)$ is a monotonic transformation. There is, of course, no reason why transformations must be used in such settings and indeed many (including the authors) have argued in favor of specifying models that rely on untransformed outcome measures and accommodating their non-negative nature — and the strictly positive nature of their corresponding conditional means $E[y|\boldsymbol{x}]$ — using estimation strategies like log-link generalized linear models (GLMs) (Mullahy, 1998; Blough, Madden, and Hornbrook, 1999; Deb, Norton, and Manning, 2017) or Poisson regression (e.g., Wooldridge, 2010, chapter 18).

Although a transformed dependent variable need not be used to model interesting features of the conditional distributions $g(y|\boldsymbol{x})$ like conditional means, the fact remains that they *are* used often in applied work. The natural logarithm transformation of dependent variables has a long history in empirical economics. More recently, the inverse hyperbolic sine transformation has become popular. The inverse hyperbolic sine transformation has been used in labor economics (Autor et al., 2022; Gihleb et al., 2022), agricultural and environmental economics (Bellemare et al, 2013; Jayachandran et al., 2017; Yen and Jones, 1997), public economics (Kalil et al., 2022; Pence, 2006), charitable donations (Brown et al., 2015; Carroll et al., 2005), and economic history (Heblich et al., 2022).

We introduce two prominent transformation models often used in empirical research that can accommodate zero realizations of outcomes:

Natural logarithm: $f_{ln}(y, c) = \ln(y + c)$

Inverse hyperbolic sine: $f_{arcsinh}(y, k) = \operatorname{arcsinh}(ky) = \ln\left(ky + \sqrt{(ky)^2 + 1}\right)$

Much of our focus will be on the roles played by the parameters c and k. In the case of the natural log transformation, the parameter $c > 0$ guarantees that the left-hand side of $f(y) = x'\beta + \varepsilon$ is defined when $y = 0$. For the inverse hyperbolic sine transformation, although most researchers implicitly assume that $k = 1$, we explore the behavior of models that scale y multiplicatively, with $k > 0$, before applying inverse hyperbolic sine transformations (Aihounton and Henningsen, 2021; Norton, 2022).

Note that, as specified, these transformations accommodate outcomes $y \geq 0$ (the inverse hyperbolic sine also allows $y < 0$ but that is not the focus of this paper). Beyond accommodation of $y = 0$ outcomes — which our reading of the literature suggests is a key consideration when analysts use either the log or inverse hyperbolic sine transformations — such transformations will also tend to reduce the skewness of the distribution of the positive outcomes. And we emphasize that theory will typically be uninformative about which values of c and k would induce a linear specification of $E[f(y)|\boldsymbol{x}]$. Moreover, note that for either transformation, different values of the parameter p (i.e., c or k) will induce different linear specifications, i.e., $E[f(y,p)|\boldsymbol{x}] = \boldsymbol{x}'\beta(p)$. Indeed, for any particular specification $f(y,p) = \boldsymbol{x}'\beta(p) + u$ it may be more appropriate to consider $\boldsymbol{x}'\beta(p)$ as a linear predictor given a particular p rather than a conditional mean, with corresponding estimates $\boldsymbol{x}'\hat{\beta}(p)$ viewed as best linear predictors.

With all this in mind, however, one of the main goals — often *the* main goal — of empirical analysis in this area is to obtain reliable estimates of policy parameters like the marginal effects or average marginal effects of x on some functional of the conditional distribution of y, typically $E[y|\boldsymbol{x}]$. To accomplish this, estimates of the transformed models must be *retransformed* to the outcomes' natural scales, a fact that came into widespread recognition in the literature on smearing estimators developed during the RAND Health Insurance Experiment

(Duan, 1983; Manning et al., 1987) but that was actually understood much earlier (Goldberger, 1968; see also Halvorsen and Palmquist, 1980, and Thornton and Innes, 1989). How such marginal effects can be derived from estimates of $f(y) = \boldsymbol{x}' \beta + \varepsilon$, and the roles played by c and k in such retransformations, is another key focus in this paper.

Specifically, we demonstrate that as the extra parameters tend towards extreme values, the corresponding marginal effects approach those obtained from linear regression on *untransformed* outcomes (e.g., OLS) or approach a particular form of the linear probability model (LPM), which we describe below. When y is strictly positive, it has been shown that the behaviors of $\ln(y + c)$ and inverse hyperbolic sine range between the linear (y) and log-transformed ($\ln(y)$) specifications (Aihounton and Henningsen, 2021; Norton, 2022).

We show that this is *not* true if y has a mass at zero. Including zeros greatly changes the marginal effects, compared to otherwise similar regressions wherein y is strictly positive. The intuition is that these continuous transformation functions (natural log and inverse hyperbolic sine) separate the zeros from the positives in a non-continuous way and that outcomes with $y = 0$ can have enormous leverage on parameter estimates in ways that depend centrally on the extra parameters (i.e., c and k). We believe this property has not previously been recognized.

There are several important implications for applied researchers. The estimated marginal effects in transformed models with zeros are biased. The estimated marginal effects for these two transformed models lie between those from the untransformed model and those from a normed linear probability model. Including zeros in the dependent variable may result in noteworthy changes in coefficients' point estimates, depending on the values of the parameters (c and k), in part because of the high leverage of those observations. We will recommend that instead of single index models researchers working with continuous positive outcomes with a mass at zero consider alternatives (e.g., two-part models (Belotti et al., 2015)).

We recognize that the care and handling of $y = 0$ outcomes is likely of central concern and that, for whatever reasons, analysts may prefer to avoid model specifications wherein untransformed outcomes are modeled directly, as we noted above. Such care and handling is all the more challenging when analysts wish to avoid multiple-index specifications (e.g., two-part models). When $y = 0$ outcomes are prominent in a sample we show that a linear-probability model interpretation — with $y = 0$ and $y > 0$ defining the binary outcomes — may be instructive.

In what follows we provide a detailed overview of the transformation models we explore, demonstrate the aforementioned connections to the linear probability model, suggest several implications for empirical research, and provide empirical illustrations of these issues using a sample from the Medical Expenditure Panel Survey (MEPS). We also provide new insights into the recent influential paper by Bellemare and Wichman (2020).

Overview of models with transformed *Y*

We start by describing and defining two transformations, paying particular attention to how the shape changes with changes in the extra parameter (see Table 1 for a summary of the transformations). We will show graphs of the transformations as a function of the parameters and derive the derivatives as the parameters go to certain extreme values. This will provide the background information needed to understand our main points about how regression results are sensitive to these parameters when the outcome has a substantial mass at zero.

The natural logarithm transformation is widely used for outcomes that are positive, often when the goal is to compute marginal effects or elasticities. However, the natural log of 0 is not defined, so when analyzing an outcome with some zeros, authors sometimes add an arbitrary positive constant c (in most cases $c = 1$) to the outcome's natural measure so that the log transformation is defined for all sample observations. Among other settings this approach has been used in healthcare policy. For example, the Centers for Medicare and Medicaid Services used a log transformation of the ratio of the number of interns and residents to beds — plus a constant — to determine Medicare reimbursement for teaching hospitals (Rogowski and Newhouse, 1992; Dalton and Norton, 2000).

Adding a constant c to y shifts the graph of $f_{ln}(y, c) = \ln(y + c)$ plotted against y to the left by c units (see Figure 1). When $c = 1$, $\ln(y + 1)$ passes through the origin at $y = 0$. As c approaches zero, the y intercept goes to $-\infty$. There are further insights from taking the derivative of $f_{ln}(y, c)$ with respect to y.

$$\frac{d\ln(y + c)}{dy} = \frac{1}{(y + c)}$$

As c grows large, the derivative with respect to y at $y = 0$ goes to zero, whereas when c goes to zero, the derivative at $y = 0$ goes to infinity.

$$\lim_{y=0,c\to 0} \frac{d\ln(y + c)}{dy} = \lim_{y=0,c\to 0} \frac{1}{(y + c)} = \infty$$

This means that as $c \to 0$ the function $f_{ln}(y, c)$ will separate the zeros from all positive values of y so long as those positive values are bounded strictly away from zero. In this case, by “separate” we mean that as $c \to 0$ the difference $\ln(y + c) - \ln(0 + c)$ grows arbitrarily large for any $y > 0$. This matters for estimating the marginal effect of a covariate x on the outcome y because in a typical regression, the coefficient for x reflects both the extensive margin for changes in y from 0 to non-zero and the intensive margin for changes in positive values of y. The coefficient is a weighted average of both the extensive and intensive margins. As the difference between $y = 0$ and $y > 0$ changes, it changes the weights between these two marginal effects. Our main point is that the estimated coefficient is much closer to describing effects on the extensive margin than previously thought (and is also normalized, which further changes the magnitude), and that interpreting this as an intensive margin is incorrect and biased.

The inverse hyperbolic sine transformation has some of those same features. The inverse hyperbolic sine function, also known as the *area hyperbolic sine function* (denoted $arcsinh(ky)$), is the natural logarithm of y plus an additional term equal to the square root of y-squared plus one. It is used by researchers who either have a nonnegative dependent variable or one that also is sometimes negative (Pence, 2006; Bellemare and Wichman, 2020; Aihounton and Henningsen, 2021; Norton, 2022). For large values of y, the inverse hyperbolic sine function is similar to the log transformation, differing by only $\ln(2)$. It also passes through the origin and is symmetric about the origin because it permits negative values of y.

Changing the parameter k is equivalent to changing the scale of y, for example, from dollars to pennies, or between euros and yen. Graphically this means stretching or compressing the graph in an east-west direction (see Figure 2). The derivative of the inverse hyperbolic sine function with respect to y declines with y, holding k constant.

$$\frac{d\text{arcsinh}(ky)}{dy} = \frac{k}{\sqrt{(ky)^2 + 1}}$$

The derivative of the inverse hyperbolic sine function, at $y = 0$ and as $k \to \infty$ goes to infinity, is equal to infinity.

$$\lim_{y=0,k\to\infty} \frac{d\text{arcsinh}(ky)}{dy} = \lim_{y=0,k\to\infty} \frac{k}{\sqrt{(ky)^2 + 1}} = \infty$$

One implication of this is that the function $f_{arcsinh}(y, k)$ will, like the log transformation, separate the zeros from the positive values of y. The issue about the regression coefficient reflecting both the extensive and intensive margins also applies to the inverse hyperbolic sine function when the dependent variable sometimes equals zero.

Finally, we briefly mention one other transformation function. The power functions were proposed by Box and Cox (1964) to transform a continuous variable to be more symmetric. We focus on a specification $f_P(y, c) = y^{\lambda}$ that is simpler than the original Box-Cox function $f_{BC}(y, \lambda) = (y^{\lambda} - 1)/\lambda$, which permitted the natural log transformation as a limiting case as $\lambda \rightarrow 0$. The most common power transformation in empirical work is the square root function, with $\lambda = 0.5$ (Ettner et al., 1998; Lindrooth, Norton, and Dickey, 2002; Veazie, Manning, and Kane, 2003). Although other power transformations with $\lambda > 0$ are possible in theory, only the square root transformation is easily retransformed back to the original scale. Therefore, we do not estimate power functions in our examples because it is not possible to transform marginal effects back to the original scale for arbitrary values of λ.

The power transformation also separates the zeros from the positives. For small positive λ the power function pulls all positive values towards 1. It both pulls large positive values towards 1 and small positive values that are less than 1 up to one, leaving zeros unchanged (see Figure 3). As λ goes to zero from above this effectively creates a binary outcome.

In summary, these transformation functions treat zeros quite differently than small positives in the sense that, for extreme values of the parameters (c, k, and λ), the transformed zeros become distant from transformations of *any* positive value. This matters in a regression, as we will show in examples, because the zeros will then exert high leverage. Changing the values of the parameters c, k, and λ will consequently change the leverage of the zeros and therefore the estimated parameters and marginal effects. In fact, in the next section we will demonstrate that in the limit (as the parameters go to certain values), the marginal effects will converge to those of a normed linear probability model.

Connection to the linear probability model

In this section we demonstrate that for a simple linear regression with a dependent variable transformed by either the log function or the inverse hyperbolic sine function, that the coefficients approach those of a linear probability model, suitable scaled. Consider a simple linear model with one covariate,

$$f(y_j, p) = \beta_0(p) + \beta_1(p)x_j + u_j \qquad \text{where } p \in \{c, k\}$$

Let there be N total observations with N_0 observations in the $y_j = 0$ subsample and N_+ observations in the $y_j > 0$ subsample, these having corresponding index sets J_0 and J_+ with $J = J_0 \cup J_+$. Let $s_0 = N_0/N$ denote the fraction of observations in the $y = 0$ subsample. Note that $f(y_j, p) = f(0, p)$ is constant for all j in J_0 and recall that $f(0, p) = 0$ always for the inverse hyperbolic sine and power transformations and when $c = 1$ for the log transformation. Define

$m_{f_0}(p) = \frac{1}{N_0}\sum_{j\in J_0} f(y_j, p)$	Subsample mean of $f(y_j, p)$ in J_0, equals $f(0, p)$
$m_{f_+}(p) = \frac{1}{N_+}\sum_{j\in J_+} f(y_j, p)$	Subsample mean of $f(y_j, p)$ in J_+
$m_{x_0}(p) = \frac{1}{N_0}\sum_{j\in J_0} x_j$	Subsample mean of x_j in J_0
$m_{x_+}(p) = \frac{1}{N_+}\sum_{j\in J_+} x_j$	Subsample mean of x_j in J_+
$m_x(p) = \frac{1}{N}\sum_{j\in J} x_j$	Sample mean of x_j
$m_{x^2}(p) = \frac{1}{N}\sum_{j\in J} x_j^2$	Sample mean of x_j^2

$m_{xf_+}(p) = \frac{1}{N_+}\sum_{j \in J_+} x_j f(y_j, p)$ Subsample mean of $x_j f(y_j, p)$ in J_+

To streamline notation, assume henceforth and without loss of generality that $m_x = 0$, i.e., that x is centered on its sample mean. After a bit of algebra, the least squares estimate of $\beta_1(p)$ can be shown to be:

$$\hat{\beta}_1(p) = \frac{1}{D}\left(s_0 m_{x_0} m_{f_0}(p) + (1 - s_0) m_{xf_+}(p)\right) \quad (1)$$

where

$$D = \det\begin{bmatrix} 1 & 0 \\ 0 & m_{x^2} \end{bmatrix} = m_{x^2} > 0$$

In those cases where $f(0, p) = 0$ equation (1) simplifies to:

$$\hat{\beta}_1(p) = \frac{1}{D}(1 - s_0) m_{xf_+}(p)$$

We can thus consider explicitly how $\hat{\beta}_1(p)$ varies with any of the parameters p. Specifically, variation in p changes two quantities in equation (1): (a) $m_{f_0}(p) = f(0, p)$ (with variation only for the log transformation); and (b) $m_{xf_+}(p)$, the average value of the product $x_j f(y_j, p)$ in the J_+ subsample. That is,

$$\frac{d\hat{\beta}_1(p)}{dp} = \frac{1}{D}\left(s_0 m_{x_0} \frac{m_{f_0}(p)}{dp} + (1 - s_0)\frac{m_{xf_+}(p)}{dp}\right)$$

Consider now a binary-outcome ("B") model where all outcomes in the J_+ subsample are assigned the value $m_{f_+}(p)$,

$$f_B(y_j, p) = \beta_{0,B}(p) + \beta_{1,B}(p) x_j + v_j$$

where

$$f_B(y_j, p) = \begin{cases} f(0, p), & y_j = 0 \\ m_{f_+}(p), & y_j > 0 \end{cases}$$

The corresponding least-squares estimate of $\hat{\beta}_{1,B}(p)$ is

$$\hat{\beta}_{1,B}(p) = \frac{1}{D}\left(s_0 m_{x_0} m_{f_0}(p) + (1 - s_0) m_{x_+} m_{f_+}(p)\right) \tag{2}$$

In those cases where $f(0, p) = 0$ equation (2) simplifies to

$$\hat{\beta}_{1,B}(p) = \frac{1}{D}(1 - s_0) m_{x_+} m_{f_+}(p)$$

Note that m_{x_0} and m_{x_+} necessarily have opposite signs (unless both are zero). For the inverse hyperbolic sine and power transformations $m_{f_0}(p) = 0$ and $m_{f_+}(p) > 0$ for the log transformation these terms can be positive, zero, or negative.

The difference between $\hat{\beta}_1(p)$ and $\hat{\beta}_{1,B}(p)$ is

$$\hat{\beta}_1(p) - \hat{\beta}_{1,B}(p) = \frac{(1 - s_0)}{D}\left(m_{xf_+}(p) - m_{x_+} m_{f_+}(p)\right)$$

$$= \frac{(1 - s_0)}{D} \operatorname{cov}_+\left(x, f(y, p)\right)$$

where $\operatorname{cov}_+\left(x, f(y, p)\right)$ is the empirical covariance in the J_+ subsample between x and $f(y, p)$.

The standard 0-1 linear probability ("LP") model arises from an affine transformation of $f_B\left(y_j, p\right)$:

$$f_{LP}\left(y_j, p\right) = \frac{f_B\left(y_j, p\right) - f(0, p)}{m_{f_+}(p) - f(0, p)} = 1\left(y_j > 0\right)$$

As such, $\hat{\beta}_1(p) - \hat{\beta}_{1,B}(p) = ((1 - s_0)/D) \operatorname{cov}_+\left(x, f(y, p)\right)$ can be rewritten as

$$T\hat{\beta}_1(p) - \hat{\beta}_{1,LP} = \frac{(1 - s_0)T}{D} \operatorname{cov}_+\left(x, f(y, p)\right)$$

where $\hat{\beta}_{1,LP}$ is the estimate from the standard linear probability model with $f_{lp}\left(y_j, p\right)$ as the dependent variable and where $T = \left(m_{f_+}(p) - f(0, p)\right)^{-1}$. For the log transformation, T goes to zero as c goes to zero because $m_{f_+}(c)$ is finite and $f(0, c)$ goes to minus infinity. For the inverse hyperbolic sine transformation, T goes to zero as k goes to infinity because $m_{f_+}(k)$ goes to plus infinity and $f(0,c)=0$. As such the rescaled $\hat{\beta}_1(p)$ and linear probability model estimate, $\hat{\beta}_{1,LP}$, differ only to the extent that the rescaled $\text{cov}_+\big(x, f(y,p)\big)$ is nonzero. This rescaled difference will tend to zero as the transformation parameters go to their extreme values.

In other words, estimated parameters from the natural log and inverse hyperbolic sine transformations are closely related to estimated parameters from the linear probability model, but normalized by a number that also grows as the transformation parameters go to their extreme values. It may thus be worth contemplating whether estimates of standard linear probability models provide useful evidence that circumvents entirely reliance on the complex machinery of dependent-variable transformation and retransformation.

Implications (when $y \geq 0$, including mass at 0)

There are six main implications from the key result that when the continuous outcome has a mass at zero that the estimated marginal effects in a regression with a transformed dependent variable lie between the marginal effects from the untransformed model and the marginal effects of the normed linear probability model.

First, when the dependent variable has a mass at zero it is not true that the estimated marginal effects for the inverse hyperbolic sine transformation lie between the untransformed model and $\ln(y)$, as implied by (Aihounton and Henningsen, 2021). The marginal effects for the log transformation $\ln(y+c)$ will also not approach those of $\ln(y)$ when some observations have $y=0$. The intuition for what should happen when there is a mass at zero is misled by what happens in models without zeros.

Second, the presence of zeros changes the estimated coefficients. Of course, the amount of change depends on the data, the distribution of the outcome, and the model specification. But the presence of observations that, by definition, are at one extreme of the distribution of the dependent variable means that those observations will have more leverage than other observations in the middle of the distribution. Furthermore, the amount of leverage depends a lot on the values of the three parameters (c, k, and λ).

Third, the choice of those parameters has no basis in theory, at least none that we have seen. The Box-Cox power transformations (1964) were introduced to reduce skewness and reshape the distribution of the outcome towards symmetry. The inverse hyperbolic sine transformation has been hailed as a way to include zeros (and in some applications negative values) in a single transformed dependent variable that also reduces skewness. Recently, it has been pointed out that marginal effects and elasticities will vary depending on whether the outcome is scaled or not (Aihounton and Henningsen, 2021; Norton, 2022). For instance, there is no theoretical reason why financial outcomes should be expressed in one currency unit or another. Or, put another way, there is no reason to expect *a priori* that results in dollars should be preferred to results in thousands of dollars or results in Euros. In linear models with untransformed LHS variables such scaling does not change the final interpretation at all, but in the inverse hyperbolic sine transformation it does. Aihounton and Henningsen (2021) suggest ways to choose the scaling factor for an inverse hyperbolic sine model based on statistical fit. Finally, there is no theoretical reason for choosing any particular positive constant to add to the log transform model. Adding 1.0 is a common default, but has no theoretical justification.

Fourth, there is no data generating process (that we know of) that is appropriate for a single index function, has a continuous distribution on $y > 0$, and has a mass at zero (other than the Tobit model, but that is restricted to distributions that are censored, which is rarely the case with real economic data). This fact alone should be a warning that transformations of the dependent variable are unlikely to represent a model that generates marginal effects across the entire range of y. It is also hard to think of how to run simulations using data generated from a single-index transformation model, because such a model cannot generate the mass at zero. Instead, simulations need to be done using a two-part model to generate the model. Only then can we see how the bias in the marginal effects, if any, depends on things like the percentage of observations that are zero, and the parameters (c and k).

Fifth, we recommend using a two-part model whenever there is a mass at zero for a continuous outcome (Belotti et al., 2015). This would solve many of the problems described above and allow for estimation of marginal effects at both the intensive and extensive margins, at the cost of estimating twice as many parameters. Alternatively single-index specifications that do not rely on transformations (e.g., OLS, log-link GLM, Poisson regression) may also merit consideration. (Of course, with non-negative *integer-valued* outcomes (count data) there will often be natural parametric relationships between the zero and positive outcomes (e.g., Poisson and negative binomial distributions) in the sense that they do not "separate" in the manner discussed above, so-called zero-inflated count-data distributions perhaps notwithstanding.)

Sixth and specific to the inverse hyperbolic sine transformation, which also allows negative values: With outcomes that can take on both positive and negative values but has no mass at zero, it is still the case that in the limit, as k goes to ∞, the results will approximate that of a linear probability model. In this case, however, the dependent variable for that LPM is an indicator for whether the outcome is positive or negative.

Data for empirical example

To demonstrate our points, we need a data set with a dependent variable that is nonnegative, skewed, and has a mass at zero. One such variable is healthcare expenditures. The Medical Expenditure Panel Survey (MEPS) is a national survey on the use of medical care in the United States. The MEPS is publically available at meps.ahrq.gov/mepsweb. We use data on a sample of 133,659 non-elderly adults from 2008–2014 (Deb and Norton, 2018).

The dependent variable is total healthcare expenditures for a person in one year, including all of their inpatient, outpatient, and out-of-pocket spending. Expenditures can never be negative but are highly right skewed (see Figure 4 for a histogram). About 24% of the sample spends zero in a year. Eight and a half percent spend more than $10,000. The highest expenditure is over $2 million. Among those who spend at least $1, the mean is $4,780 and the median is $1,234.

For illustrative purposes, we control for just a few key covariates (see Table 2). The mean age is 40, with a range from 18 to 64. A little more than half of the persons in the sample are female. Family income has a mean of $64,000 and it is measured in thousands of dollars. In terms of health status, 23% have at least one health limitation, and the average SF12 physical and mental health scales are each about 50, with a theoretical range of zero to 100. A higher value of the physical and mental health scales indicates better health status. In the empirical examples we will be interested in comparing the estimated marginal effects.

Empirical examples

For the empirical examples, we start with results for two models that serve as benchmarks. One is the linear regression of the untransformed dependent variable and the other is the linear probability model. The dependent variable for both is based on total annual healthcare expenditures, including the zeros. For the linear probability model, the dependent variable is whether the person spent at least \$1, as opposed to spending nothing. The covariates are individual level variables for demographics, income, and health status, as described in the data section. Later we will show how using a transformed dependent variable and adjusting the three extra parameters (*c* and *k*) to more extreme values can reproduce either of these benchmark results.

The covariates for the linear regression predict individual healthcare expenditures in the expected way (see the first column of Table 3). Expenditures increase with age by about \$37 per year. Females spend more than \$600 more per year than males on average. Healthcare expenditures increase with family income (healthcare is a normal good). People who have any health limitations spend nearly \$4,000 more than those who have none, on average. Spending decreases with improvements in both physical and mental health status, as measured by the SF12. The coefficients for the linear probability model to predict the probability of any spending all have the same sign, but of course are smaller in magnitude because the marginal effects have the interpretation of a change in the probability of spending any amount compared to spending nothing. For example, the probability of positive spending is 14 percentage points greater for women than men (see the second column of Table 3).

Next, we show empirically that the log transformed and the inverse hyperbolic sine transformed outcomes yield retransformed marginal effects on the original scale that depend strongly on the values of *c* and *k*. Later we will show that they converge in the limit to the marginal effects from either the untransformed *y* or the linear probability model, but for now we just want to show that the results vary. It is necessary to convert the marginal effects back to the original scale (dollars) to be comparable across models. The coefficients are not comparable.

The retransformation for the log function uses Duan's smearing estimate (1983), specifically, the mean of the exponentiated residuals multiplied by the exponentiated linear index. The retransformation for the inverse hyperbolic sine function uses a similar retransformation, based on Duan's general function (Norton, 2022). Because the retransformation for power functions does not have a formula that can be expressed simply in terms of the linear index and the residuals (except for the square root), we do not show how marginal effects for power functions vary with λ. While it is possible to derive retransformed models for positive integer values of $\lambda > 2$ the conditional means of such models depend on higher order moments of the residuals.

The retransformed marginal effects for the log transform model depend strongly on the additive constant c. As c increases from the tiny value of one millionth to the typical value of one to the enormous value of one million, the marginal effects decrease from implausibly large values (see the left column of Table 4) to values that are quite close to those from the untransformed y model (compare the right column of Table 5 to the left column of Table 3). Of note, the estimated marginal effects when $c = 1$, the default for most researchers who use the log transformation plus a constant, are many times as large as those from the basic linear model. Because the data are from a real data set and not simulated, we do not know the true value of the marginal effects. But the enormous range of marginal effects raises a red flag that warrants further investigation.

The results for the inverse hyperbolic sine transformation show a similarly wide range of marginal effects, depending on the parameter k. However, this time, the tiny value of k (left column of Table 6) yields marginal effects close to the untransformed model, and the enormous values of k (right column of Table 7) have implausibly large marginal effects. Again, assuming that $k = 1$, which is typical for researchers who use the inverse hyperbolic sine transformation, yields marginal effects that are many times as large as the untransformed model.

Although the marginal effects for the log transformation with $c = .000001$ and the inverse hyperbolic sine transformation with $k = 1{,}000{,}000$ do not appear to have anything in common with the linear probability model, we next show that they do. First, we show that the zeros become progressively more separate from the positive values when the parameters go to the extreme, using box and whisker plots (see Figures 5 and 7). For all three transformations, the zeros become separated from the bell-curve of transformed positive values. The distance between the zeros and the mean of the transformed positives becomes large relative to the spread of the positives as shown earlier. This is seen most clearly with graphs of the power transformation. When $\lambda = .01$, the distribution is essentially a binary variable. All the positive values are pulled towards one (and if there were any outcomes in fractions of pennies, those too would get transformed to be nearly one). For the log and inverse hyperbolic sine transformations, one can imagine that as *c* gets even smaller and as *k* gets even larger that the distributions would also become essentially that of a binary variable (although not with values of 0 and 1).

If the distributions are normed as described in the theory section, then the regression of the normed transformed outcome will have marginal effects (after retransforming back to the original scale) close to those of a linear probability model. We show this in two ways. First, we run the normed regressions (see results in Tables 8, 9, and 10). We show the results for three different values of the parameters. The first has $c = k = 1$, and the marginal effects are not too far off, but certainly not identical (see Table 8). The second has $1/c = k = 1{,}000{,}000$. The marginal effects are now much closer to the linear probability model, but still all larger in absolute value by somewhere between 6% and 37% (see Table 9). The third uses the extreme value of a googol, that is, $1/c = k = 10^{100}$. Now the marginal effects for the log and inverse hyperbolic sine transformations are quite close (see Table 10).

We also show this visually by plotting the normed data against age, one representative continuous explanatory variable, and fitting the best fit regression line. This shows two things. First, that the transformed positive values ($f(y|y > 0)$) for both transformations collapse around 1, when normed by $[f(\bar{y}) - f(0)]$. Second, the slopes of the regression lines get close to the slope of the linear probability model as the parameters go to extreme values (see Figures 6 and 8).

What to use instead of transformed models?

If the transformation models are problematic when the dependent variable is sometimes zero, what are the best alternatives? To explore this, we wanted to know the correct answer by generating data through a data generating process that allows for a mass at zero. We used the two-part model, which first models the probability that the dependent variable is zero or positive, and then, if positive, generates a continuous dependent variable. We present six representative data sets, but have also generated many more data sets and found similar conclusions. The data vary on two dimensions. First, the fraction of zeros is 5%, 30%, or 80%. Second, the distribution of the positive values is either normal or skewed (specifically, log normal).

We estimated marginal effects for five models. The first two were the natural logarithm of $y + 1$ and the inverse hyperbolic sine (with $k = 1$). We expect that these will produce biased results when the dependent variable can take on values of zero. The other three models are the two-part model (with a logit first part), untransformed y, and Poisson. The second part of the two-part model is either simple regression, for when the data generating process is normal, or GLM with a log link and gamma family distribution, for when the data generating process is skewed. During preliminary investigation we also estimated negative binomial count models, but unlike the Poisson count model those were almost never close to the two-part model. Because we know the data generating process is from a two-part model, those results are unbiased. We will check whether untransformed *y* and Poisson return similar marginal effects.

The six covariates were created so that they had different effects in the two parts of the two-part model, sometimes positive and sometimes negative. The four continuous variables (denoted x1, x2, x3, and x4 in the tables) had effects that were positive in both (x1), negative in both (x4), or alternating (x2 and x3). The two binary variables (denoted b1 and b2 in the table) had effects on only the first or the second parts in the data generating model, respectively.

The results were striking. Untransformed *y* and Poisson estimated similar marginal effects to the two-part model in all cases, as well as in all other cases that we ran (not shown) with other designs. The results for the normal distribution are in Table 11 and the results for the skewed distribution are in Table 12. There was no systematic difference in the magnitude of the standard errors across the two-part, untransformed *y*, and Poisson approaches. As expected, the two transformed models performed poorly, with estimated marginal effects quite different than the true values, even when the fraction of the sample with the dependent variable equal to zero was only 5 percent.

Bellemare and Wichman

Our findings provide some important context for the results in the well-cited paper by Bellemare and Wichman (2020; henceforth BW), which, among other things, shows how to compute elasticities when either the dependent variable or an independent variable (or both) are transformed by the inverse hyperbolic sine transformation. For example, consider the BW derivation of elasticities of the dependent variable y with respect to an independent variable x. They explain how those formulas approximate the formula for elasticities from an equivalent specification with logarithmic transformation for sufficiently large values of y and x. To better understand how large is large enough, they used simulated data and adjusted the parameter k. In their examples, multiplying y by k is equivalent to having a larger value of y. BW show that the elasticities in their simulated data become "more stable," and are virtually constant for $k > 10$ (page 54, Caveats section). Furthermore, their results show that the elasticities approach those of the linear y linear x case as $k \rightarrow 0$ (see their Table 1).

One might be tempted to conclude from the BW simulation that the inverse hyperbolic sine transformation produces elasticities that range between those of the linear transformation and the log transformation, depending on the value of k. But further investigation reveals a different conclusion. Their simulated data has no zeros (code is available at http://marcfbellemare.com/wordpress/research under Replication Materials); indeed, as BW note, the elasticities are not defined for $y = 0$. For all values of $k \geq 1$ in their data, the dependent variable is always positive. For values of $0.001 \leq k < 1$ there are exactly three negative values and no zeros out of a total sample of $N = 10{,}000$. Our paper shows the importance of zeros in calculations of marginal effects, so we reproduced the BW simulated data and calculated marginal effects for the arcsinh(y)-linear(x) specification (see Table 13). As with their results for elasticities, we find that the marginal effects approach the marginal effects for the linear(y)-linear(x) specification as $k \rightarrow 0$ and approach the marginal effects for the ln(y)-linear(x) specification as $k \rightarrow \infty$.

Our results offer additional context for the BW findings. They find stable elasticities for large y (meaning large values of k) precisely because there are neither zeros nor negative values in their data. Yet, one of the main motivations for using the inverse hyperbolic sine transformation is to apply to data with a large mass at zero. Without zeros, researchers would instead typically use a log transformation without adding a positive constant c.

There are additional insights from revisiting BW's analysis of real data taken from Dehejia and Wahba (1999), who evaluated LaLonde's (1986) experiment for the National Supported Work (NSW) demonstration. The basic design is to see whether disadvantaged workers who were randomly assigned to participate in the NSW program had higher future earnings than workers who were randomly assigned not to participate. The data set and analysis code again are found at the website http://marcfbellemare.com/wordpress/research under Replication Materials.

We analyzed the same data to demonstrate how our insights about the interpretation of results with zeros in a transformation model matters in a real data set. For simplicity, we run models in which the dependent variable, earnings in 1978, is transformed by either the inverse hyperbolic sine function or the natural log, but the independent variable of interest, earnings in 1975, is not transformed. The data set has 445 observations. In 1978 137 (30.8%) of the observations had zero earnings.

First, we show that changing the parameter k has an enormous effect on both marginal effects and elasticities this data set with a large fraction of zeros. For the inverse hyperbolic sine transformation, as k ranged from 0.000001 to 1,000,000, the marginal effect of earnings in 1975 on earnings in 1978 changed from 0.1667 to 21.2 and the earnings elasticity changed from 0.00023 to 0.357 (see Table 14). Similarly, we varied the parameter c for the natural log transformation and again both the marginal effects and elasticities changed by several orders of magnitude (see Table 15). Note that the marginal effect when k is small (0.1667) is about the same as when c is large (0.1658). We expected this result because in the limit both will approach the marginal effect for the linear untransformed dependent variable (0.1667). In contrast, when those parameters go to the other extreme, the marginal effects get large because they equal the linear probability model parameter multiplied by the mean of the transformed value of y for the subset with positive values. The parameters k and c greatly affect the estimated marginal effects and elasticities when the dependent variable is transformed and has some zero values.

Next we show how the linear probability model produces results that are closely related to those of the inverse hyperbolic sine transformation (with $k = 1$) and the natural log transformation (with $c = 1$). At first, a comparison of those three models does not seem to show similar results (see Table 16). But after norming the positive values of the transformed models by the expected value when $y > 0$ (divide by 9.242 for $arcsinh(y)$ and divide by 8.54 for $\ln(y + 1)$, see last row of Table 16), the marginal effects are quite close, as we have shown above that they must be. Visually, the normed models have slopes similar to the linear probability model, as shown in Figure 9.

Finally, we analyze the data using other models without transformations. Specifically, we estimate the two-part model, OLS with untransformed y, and Poisson. The two-part model that we estimated had a logit first part and then a generalized linear model with a log link and gamma family distribution. The two-part model has the advantage of also showing marginal effects (and elasticities) on both the extensive and intensive margins, which we also show. Given our results above using simulated data, we expect that all three of those models should give similar overall marginal effects. We find exactly that. The estimated overall marginal effect of the NSW policy variable is about \$1,720 and the estimated overall marginal effect of earnings in 1975 was about \$0.167 (see Table 17). The estimated elasticities for earnings in 1978 with respect to either NSW or earnings in 1975 were about 0.12 and 0.037, respectively. For comparison, Bellemare and Wichman (2020) estimated the elasticity of earnings in 1978 with respect to earnings in 1975 to be 0.155 (see their column 3 of their Table 3). The four-fold difference in the estimated elasticity is due, we argue, to the presence of zeros and the arbitrary choice of $k = 1$ in their inverse hyperbolic sine transformed model.

Conclusions

We studied two classes of models that transform the dependent variable y — the natural logarithm of y plus a constant and the inverse hyperbolic sine. Both are widely used for dependent variables that are non-negative, skewed, and have a large mass at zero. When these models are used to analyze data with a dependent variable that has a mass at zero, and as the models' extra parameters go to extreme values (zero or infinity), the estimated retransformed marginal effects approach those of either the untransformed linear regression or the linear probability model. The inverse hyperbolic sine was previously thought to approach the natural log model (as k gets large), but with a mass at zero we show that this is not true. We show that two-part models, which correspond to a data generating process with a mass at zero and skewed positive values, yields similar marginal effects to both OLS on the untransformed y and Poisson. If researchers care about estimating marginal effects, then we recommend using these simpler models that do not use transformations.

While this paper has suggested several promising estimation strategies models of sometimes-zero nonnegative outcomes, others not considered here may be useful foci in future research. For instance, one potentially promising line of inquiry would consider various semi-parametric approaches that allow flexible specification of the contribution of right-hand side variables to conditional means and corresponding marginal effects. Such models might avail of GLM-type link functions to accommodate the non-negative nature of the dependent variables and then specify a flexible structure of the covariates (see also Gilleskie and Mroz, 2004, for related discussion).

References

Aihounton, G.B.D., A. Henningsen. 2021. Units of measurement and the inverse hyperbolic sine transformation. *Econometrics Journal* 24(2):334–351.

Autor, D. et al. 2022. New Frontiers: The Origins and Content of New Work, 1940-2018. NBER W.P. 30389.

Bellemare, M.F., C.B. Barrett, and D.R. Just. 2013. The Welfare Impacts of Commodity Price Volatility: Evidence from Rural Ethiopia. *American Journal of Agricultural Economics* 95:877-899.

Bellemare, M.F., C.J. Wichman. 2020. Elasticities and the inverse hyperbolic sine function. *Oxford Bulletin of Economics and Statistics* 82(1):50–61.

Belotti, F., P. Deb, W.G. Manning, E.C. Norton. 2015. twopm: Two-part models. *Stata Journal* 15(1):3–20.

Blough, D. K., C. W. Madden, and M. C. Hornbrook. 1999. Modeling risk using generalized linear models. *Journal of Health Economics* 18:153–171.

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

Brown, S. et al. 2015. An Inverse Hyperbolic Sine Heteroskedastic Latent Class Panel Tobit Model: An Application to Modeling Charitable Donations. *Economic Modelling* 50:228-236.

Carroll, J., S. McCarthy, and C. Newman. 2005. An Econometric Analysis of Charitable Donations in the Republic of Ireland. *Economic and Social Review* 36:229-249.

Cox, N.J. 2011. Stata tip 96: Cube roots. *Stata Journal* 11(1):149–154.

Dalton, K., E.C. Norton. 2000. Revisiting Rogowski and Newhouse on the indirect costs of teaching: a note on functional form and retransformation in Medicare's payment formulas. *Journal of Health Economics* 19:1027–1046.

Deb, P., E.C. Norton. 2018. Modeling health care expenditures and use. *Annual Review of Public Health* 39:489–505.

Deb, P, E.C. Norton, W.G. Manning. 2017. Health Econometrics Using Stata. College Station, TX: Stata Press.

Dehajia, R.H. and S. Wahba. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association* 94:1053–1062.

Duan, N. 1983. Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association* 78(383):605–610.

Ettner, S. L., R. G. Frank, T. G. McGuire, J. P. Newhouse, E. H. Notman. 1998. Risk adjustment of mental health and substance abuse payments. *Inquiry* 35:223–239.

Gihleb, R. et al. 2022. Industrial Robots, Workers' Safety, and Health. NBER W.P. 30180.

Gilleskie, D.B. and T.A. Mroz. 2004. A Flexible Approach for Estimating the Effects of Covariates on Health Expenditures. *Journal of Health Economics* 23: 391–418.

Goldberger, A.S. 1968. The Interpretation and Estimation of Cobb-Douglas Functions. *Econometrica* 36: 464-472.

Halvorsen, R. and R. Palmquist. 1980. The interpretation of dummy variables in semilogarithmic equations. *American Economic Review* 70(3):474–475.

Heblich, S., S.J. Redding, and H.-J. Voth. 2022. Slavery and the British Industrial Revolution. NBER W.P. 30451.

Jayachandran, S. et al. 2017. Cash for Carbon: A Randomized Trial of Payments for Ecosystem Services to Reduce Deforestation. *Science* 21:267-273.

Kalil, A. et al. 2022. Welfare Reform and the Quality of Young Children's Home Environments. NBER W.P. 30407.

Khan, N., R. Kaestner, S.J. Lin. 2008. Effect of prescription drug coverage on health of the elderly. *Health Services Research* 43:1576–1597.

LaLonde, R. J. 1986. Evaluating the econometric evaluations of training programs with experimental data. *American Economic Review* 76:604–620.

Lindrooth, R. C., E. C. Norton, B. Dickey. 2002. Provider selection, bargaining, and utilization management in managed care. *Economic Inquiry* 40:348–365.

Manning, W.G. 1998. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics* 17(3):283–295.

Manning, W.G., A. Basu, J. Mullahy. 2005. Generalized modeling approaches to risk adjustment of skewed outcomes data. *Journal of Health Economics* 24(3):465–488.

Manning, W.G., J. Mullahy. 2001. Estimating log models: to transform or not to transform? *Journal of Health Economics* 20(4):461–494.

Manning, W.G., J.P. Newhouse, N. Duan, E.B. Keeler, A. Leibowitz, M.S. Marquis. 1987. "Health-insurance and the demand for medical care: Evidence from a randomized experiment." *American Economic Review* 77(3):251–277.

Mullahy, J. 1998. Much ado about two: Reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* 17: 247–281.

Mullahy, J. 2021. Inverse Hyperbolic Sine Transformations and Retransformations in Regression Models. University of Wisconsin-Madison Working Paper.

Norton, E.C. 2022. The inverse hyperbolic sine transformation and retransformed marginal effects. *Stata Journal* 22(3):702–712.

Pence, K.M. 2006. The role of wealth transformations: An application to estimating the effect of tax incentives on saving. *The B.E. Journal of Economic Analysis & Policy* 5:1–24.

Rogowski, J.A., Newhouse, J.P. 1992. Estimating the indirect costs of teaching. *Journal of Health Economics* 11:153–171.

Santos Silva, J.M.C., F. Windmeijer. 2001. Two-part multiple spell models for health care demand. *Journal of Econometrics* 104:67–89.

Thornton, R.J., J.T. Innes. 1989. Interpreting semilogarithmic regression coefficients in labor research. *Journal of Labor Research* 10:443–447.

Veazie, P.J., W.G. Manning, R.L. Kane. 2003. Improving risk adjustment for Medicare capitated reimbursement using nonlinear models. *Medical Care* 41(6):741–752.

Wooldridge, J. 2010. *Econometric Analysis of Cross Section and Panel Data*, 2nd ed. Cambridge, MA: MIT Press.

Yen, S.T. and A.M. Jones. 1997. Household Consumption of Cheese: An Inverse Hyperbolic Sine Double-Hurdle Model with Dependent Errors. *American Journal of Agricultural Economics* 79:246-251.

Table 1: Comparative information about the three transformation functions.

	Transformation of y		
	Natural Logarithm	**Inverse Hyperbolic Sine**	**Power or Box-Cox**
Formula	$\ln(y+c)$	$\text{arcsinh}(ky)$ $\ln(ky+\sqrt{ky^2+1})$	y^λ $(y^\lambda - 1)/\lambda$
Extra parameter	$c > 0$	$k > 0$	$\lambda > 0$
Domain of y	$[0,\infty)$	$(-\infty,\infty)$	$[0,\infty)$
Range of $f(y)$	$(-\infty,\infty)$	$(-\infty,\infty)$	$[0,\infty)$
Marginal effects			
$\rightarrow$ Untransformed y	$c \rightarrow \infty$	$k \rightarrow 0$	$\lambda \rightarrow 1$
$\rightarrow$ LPM	$c \rightarrow 0$	$k \rightarrow \infty$	$\lambda \rightarrow 0$

Table 2: Summary statistics for the MEPS data, 2008–2014.

	Mean	Std. Dev.	Minimum	Maximum
Outcome				
Healthcare Expenditures ($)	3,632.5	13,114.8	0	2,343,843
Covariates				
Age	40.2	13.2	18	64
Female	.534	.499	0	1
Family Income ($1,000)	60.9	55.0	-182.1	556.1
Any limitations	.191	.393	0	1
Physical health score	50.7	9.6	4.6	74.1
Mental health score	50.8	10.1	.2	78.1

N = 133,659.

Table 3: Benchmark marginal effects for untransformed *y* and for linear probability models to predict healthcare expenditures.

	Untransformed *y*	**Linear Probability Model**
Covariates		
Age	37.32	.00391
	(2.79)	(.00009)
Female	643.5	.14122
	(69.4)	(.00222)
Family Income ($1,000)	12.458	.00113
	(.648)	(.00002)
Any limitations	3888.3	.11248
	(103.8)	(.00333)
Physical health score	–245.08	–.00392
	(4.23)	(.00014)
Mental health score	–65.59	–.00314
	(3.58)	(.00011)

N = 133,659. Standard errors are in parentheses.

Table 4: Retransformed marginal effects: regression of ln(y+c) for different small values of c.

Variable	ln(y+.000001)	ln(y+.001)	ln(y+.01)	ln(y+.1)	ln(y+1)
age	3880567.09	21249.05	5329.82	1569.16	533.01
	517232.27	1244.87	233.78	52.82	13.99
female	129995982.53	689511.29	169916.33	48804.46	15981.43
	17671053.86	41343.46	7547.94	1631.75	406.59
faminc	1083879.72	5850.95	1456.14	424.09	141.76
	159909.52	394.76	72.57	15.68	3.90
anylim	122107540.36	692065.27	176771.65	53327.94	18748.31
	16989489.50	44415.43	8639.03	2016.64	549.18
pcs	-4672582.81	-27317.72	-7087.29	-2181.56	-788.15
	666559.17	1844.78	366.67	87.02	23.94
mcs	-3162728.77	-17429.72	-4386.98	-1297.73	-444.00
	461569.13	1200.99	234.84	55.49	15.47
N	133659	133659	133659	133659	133659

Legend: b/se

Table 5: Retransformed marginal effects: regression of ln(y+c) for different large values of c.

Variable	ln(y+1)	ln(y+10)	ln(y+100)	ln(y+1000)	ln(y+1000000)
age	533.01	204.66	85.40	41.34	38.13
	13.99	4.28	1.54	0.81	2.36
female	15981.43	5796.94	2231.16	1004.86	677.57
	406.59	115.57	39.08	20.09	58.79
faminc	141.76	53.02	21.24	10.12	12.22
	3.90	1.11	0.37	0.19	0.55
anylim	18748.31	7586.28	3478.27	2060.41	3810.17
	549.18	170.82	60.99	30.88	87.95
pcs	-788.15	-332.03	-162.55	-105.89	-236.20
	23.94	7.47	2.64	1.30	3.58
mcs	-444.00	-172.89	-75.08	-41.22	-63.47
	15.47	5.01	1.90	1.03	3.03
N	133659	133659	133659	133659	133659

Legend: b/se

Table 6: Retransformed marginal effects: regression of arcsinh(ky) for different small values of k.

Variable	IHS(.000001*y)	IHS(.001*y)	IHS(.01*y)	IHS(.1*y)	IHS(1*y)
age	37.96	45.77	111.46	271.75	728.22
	2.57	0.87	2.07	6.07	20.58
female	659.54	1106.10	2950.79	7834.36	22108.71
	63.96	21.72	53.18	167.32	610.01
faminc	12.46	11.04	27.88	71.06	194.78
	0.60	0.20	0.51	1.61	5.85
anylim	3900.09	2192.72	4382.03	9888.21	25315.99
	95.67	33.82	82.60	241.25	801.51
pcs	-243.47	-111.30	-200.50	-426.45	-1054.35
	3.90	1.44	3.61	10.55	34.85
mcs	-64.92	-44.28	-96.00	-228.06	-604.95
	3.30	1.11	2.52	6.97	22.38
N	133659	133659	133659	133659	133659

Legend: b/se

Table 7: Retransformed marginal effects: regression of arcsinh(ky) for different large values of k.

Variable	IHS(1*y)	IHS(10*y)	IHS(100*y)	IHS(1000*y)	IHS(1000000*y)
age	728.22	2232.20	7940.28	33339.23	7076802.30
	20.58	81.21	378.94	2139.78	1002124.08
female	22108.71	70011.07	254639.29	1086570.43	237583095.00
	610.01	2546.50	12358.86	71474.33	34280041.07
faminc	194.78	605.51	2175.00	9197.88	1978562.88
	5.85	24.48	118.70	679.68	308788.80
anylim	25315.99	75244.15	261775.27	1080851.33	222140042.63
	801.51	3071.06	13858.23	75542.75	32747451.79
pcs	-1054.35	-3057.65	-10442.00	-42492.64	-8481140.95
	34.85	131.94	584.62	3115.79	1280123.54
mcs	-604.95	-1843.07	-6528.14	-27323.12	-5765141.22
	22.38	84.11	375.90	2041.33	890776.18
N	133659	133659	133659	133659	133659

Legend: b/se

Table 8: Marginal effects for normed regressions with c=k=1.

Variable	LPM	ln(y+1)	arcsinh(1*y)
age	0.003913	0.005580	0.005438
	0.000089	0.000094	0.000093
female	0.141217	0.167319	0.165101
	0.002224	0.002343	0.002314
faminc	0.001131	0.001484	0.001455
	0.000021	0.000022	0.000022
anylim	0.112480	0.196287	0.189053
	0.003327	0.003505	0.003462
pcs	-0.003925	-0.008252	-0.007874
	0.000136	0.000143	0.000141
mcs	-0.003138	-0.004649	-0.004518
	0.000115	0.000121	0.000119
N	133659	133659	133659

Legend: b/se

Table 9: Marginal effects for normed regressions with (1/c)=k=1 million.

Variable	LPM	ln(y+.000001)	arcsinh(1000000*y)
age	0.003913	0.004481	0.004463
	0.000089	0.000089	0.000089
female	0.141217	0.150114	0.149829
	0.002224	0.002207	0.002207
faminc	0.001131	0.001252	0.001248
	0.000021	0.000021	0.000021
anylim	0.112480	0.141005	0.140090
	0.003327	0.003301	0.003301
pcs	-0.003925	-0.005396	-0.005349
	0.000136	0.000134	0.000134
mcs	-0.003138	-0.003652	-0.003636
	0.000115	0.000114	0.000114
N	133659	133659	133659

Legend: b/se

Table 10: Marginal effects for normed regressions with (1/c)=k=1 googol.

Variable	LPM	ln(y+(1/googol))	arcsinh(googol*y)
age	0.003913	0.003963	0.003963
	0.000089	0.000089	0.000089
female	0.141217	0.142001	0.141999
	0.002224	0.002220	0.002220
faminc	0.001131	0.001142	0.001142
	0.000021	0.000021	0.000021
anylim	0.112480	0.114993	0.114986
	0.003327	0.003321	0.003321
pcs	-0.003925	-0.004054	-0.004054
	0.000136	0.000135	0.000135
mcs	-0.003138	-0.003184	-0.003184
	0.000115	0.000114	0.000114
N	133659	133659	133659

Legend: b/se

Table 11 Comparison of marginal effects for $\ln(y+1)$, arcsinh(y), two-part model, untransformed y, and Poisson.

Simulated data with **5% zeros** and normal errors.

Variable	lny_N05	arcsinh_N05	tpm_N05	y_N05	pois_N05
x1	22794.98	26965.38	1508.43	1555.66	1554.79
	1356.27	1699.02	36.74	52.23	53.90
x2	10743.52	12763.15	413.94	396.95	396.24
	930.36	1134.70	35.72	44.26	45.01
x3	-11414.67	-13558.14	-414.67	-436.31	-436.13
	947.46	1158.21	33.52	44.60	45.41
x4	-17365.27	-20557.20	-1086.08	-1092.29	-1092.84
	1129.47	1401.69	36.63	47.47	48.73
b1	27426.57	32513.14	1601.47	1444.14	1446.03
	1801.23	2231.24	84.22	78.37	80.00
b2	-22.75	-80.74	390.96	332.42	332.66
	1474.15	1747.47	62.55	78.62	80.14

Simulated data with **30% zeros** and normal errors.

Variable	lny_N30	arcsinh_N30	tpm_N30	y_N30	pois_N30
x1	7341884.28	1.76e+07	4232.02	4214.54	4194.13
	704612.47	1852269.07	51.90	64.14	88.11
x2	3809680.60	9141674.87	1926.63	1923.06	1893.36
	381791.00	995846.72	57.96	67.34	80.59
x3	-3.79e+06	-9.09e+06	-1899.13	-1917.21	-1901.67
	375451.59	979349.64	57.89	66.67	79.48
x4	-5.80e+06	-1.39e+07	-3197.63	-3250.80	-3249.98
	554232.44	1454941.87	54.14	66.05	83.49
b1	9713447.38	2.33e+07	4976.94	5222.13	5290.84
	900980.47	2364976.80	106.08	133.80	159.32
b2	-261280.55	-634303.76	113.65	94.76	97.39
	251509.57	604623.88	117.80	134.15	154.63

Simulated data with **80% zeros** and normal errors.

Variable	lny_N80	arcsinh_N80	tpm_N80	y_N80	pois_N80
x1	18087.41	27172.74	3491.21	3423.92	3300.69
	2130.18	3519.80	49.93	68.46	133.79
x2	8479.43	12743.05	1556.93	1530.38	1433.02
	1038.49	1703.63	51.97	62.82	86.15
x3	-9053.16	-13605.04	-1695.59	-1637.84	-1568.73
	1097.02	1801.04	52.06	64.25	87.36
x4	-13101.94	-19684.10	-2544.21	-2462.50	-2437.30
	1549.04	2554.56	53.47	65.42	106.67
b1	22530.79	33853.54	4266.95	4171.33	4561.25
	2644.61	4355.07	104.75	127.87	198.01
b2	1341.55	2011.74	259.31	316.04	386.11
	699.52	1055.15	107.36	127.32	155.77

Table 12 Comparison of marginal effects for $\ln(y + 1)$, arcsinh(y), two-part model, untransformed y, and Poisson.

Simulated data with **5% zeros** and skewed errors.

Variable	lny_S05	arcsinh_S05	tpmglm_S05	y_S05	pois_S05
x1	120150.13	141347.29	19603.99	20747.17	20568.75
	6729.81	8596.97	233.27	371.10	306.12
x2	21511.36	27263.40	-7429.04	-8037.55	-7902.06
	3343.97	4099.24	213.14	318.41	278.78
x3	-24399.46	-30697.40	7204.99	7542.14	7468.07
	3434.48	4223.61	211.11	300.23	263.93
x4	-81813.08	-96788.09	-10375.75	-10747.63	-10761.42
	5106.32	6474.81	206.61	300.88	259.95
b1	99993.97	120119.47	3076.86	2803.69	2562.25
	7358.55	9271.95	410.19	526.22	480.84
b2	34478.04	38630.76	16531.46	16727.65	16964.43
	5623.67	6751.60	408.60	526.16	492.82

Simulated data with **30% zeros** and skewed errors.

Variable	lny_S30	arcsinh_S30	tpmglm_S30	y_S30	pois_S30
x1	8.11e+07	2.02e+08	23653.63	24548.67	24238.24
	8803576.09	2.38e+07	251.01	402.39	431.75
x2	3.46e+07	8.74e+07	-1687.74	-2176.61	-1981.72
	3866675.47	1.05e+07	236.07	348.22	320.42
x3	-3.45e+07	-8.72e+07	1452.35	1777.15	1718.85
	3845608.89	1.05e+07	234.33	333.49	312.51
x4	-6.18e+07	-1.55e+08	-14563.04	-14993.41	-14943.83
	6667594.90	1.81e+07	225.47	333.02	343.90
b1	9.73e+07	2.44e+08	12189.45	12910.89	12840.68
	1.02e+07	2.79e+07	423.15	600.11	615.15
b2	4009100.95	8974948.50	12305.20	12377.68	12664.17
	2507230.87	6280977.34	465.47	599.96	617.27

Simulated data with **80% zeros** and skewed errors.

Variable	lny_S80	arcsinh_S80	tpmglm_S80	y_S80	pois_S80
x1	109383.77	169802.81	15540.02	15661.79	15008.64
	15188.42	25615.11	286.88	469.17	738.89
x2	47552.36	74176.13	2624.62	2096.57	2191.74
	6693.55	11295.78	192.04	297.07	305.75
x3	-50968.93	-79485.02	-3080.90	-2586.01	-2482.16
	7117.41	12015.84	187.38	301.62	337.58
x4	-78342.62	-121701.33	-10311.90	-10026.02	-9751.79
	10860.38	18311.01	231.30	343.59	490.17
b1	131587.54	204719.20	13168.77	13044.92	13744.06
	18002.34	30386.64	410.51	581.20	803.25
b2	11152.68	17020.65	4629.75	5121.81	5660.42
	4300.68	6733.36	420.18	577.99	703.09

Table 13
Marginal effects corresponding to third column of Table 1, Bellemare and Wichman (2020).

	Empirical specifications		
	Linear-linear		
	0.480715		
Values of k:	*arcsinh-linear*	N with $y = 0$ (%)	N with $y < 0$ (%)
.001	0.480714	0 (0%)	3 (.003%)
.01	0.480622	0 (0%)	3 (.003%)
.1	0.484385	0 (0%)	3 (.003%)
1	0.505784	0 (0%)	0 (0%)
5	0.506556	0 (0%)	0 (0%)
10	0.506581	0 (0%)	0 (0%)
50	0.506589	0 (0%)	0 (0%)
100	0.506589	0 (0%)	0 (0%)
1000	0.506589	0 (0%)	0 (0%)
10,000	0.506589	0 (0%)	0 (0%)
	ln(y)-linear		
	0.506589		

Notes: This table presents marginal effects estimates from different empirical specifications based on a simulated data set where $x \sim N(10, 2)$, $e \sim N(0, 2)$, and $ky = 5 + 0.5kx + e$ for various values of k. The simulation was done in Stata with set seed 12345, following Bellemare and Wichman (2020).

Table 14
Extension of results of column (3) in Table 3, Bellemare and Wichman (2020).

	Dependent variable = arcsinh($k \times Earnings_{78}$)				
	$k = .000001$	$k = .001$	$k = 1$	$k = 1000$	$k = 1{,}000{,}000$
Estimated parameters					
NSW	0.001750	0.307	1.056	1.801	2.55
	(0.000632)	(0.127)	(0.416)	(0.717)	(1.02)
E_{75}	0.00000017	0.0000374	0.0001122	0.000186	0.000259
	(0.00000010)	(0.0000199)	(0.0000651)	(0.000112)	(0.000160)
Constant	0.004344	1.4779	5.803	10.173	14.544
	(0.000426)	(0.0857)	(0.280)	(0.483)	(0.688)
Obs.	445	445	445	445	445
Adj. R^2	0.0197	0.0171	0.0173	0.0164	0.0160
Estimated marginal effects and elasticities					
ME(E_{75})	0.1667	0.208	0.824	3.19	21.2
	(0.0990)	(0.114)	(0.645)	(5.06)	(69.6)
$\xi(E_{78}, E_{75})$	0.000230	0.0514	0.1546	0.256	0.357
	(0.000136)	(0.0274)	(0.0896)	(0.155)	(0.220)

Table 15
Extension of results of column (5) in Table 3, Bellemare and Wichman (2020).

	Dependent variable = ln $(Earnings_{78} + c)$				
	$c = .000001$	$c = .001$	$c = 1$	$c = 1000$	$c = 1{,}000{,}000$
Estimated parameters					
NSW	2.471	1.726	0.982	0.253	0.001726
	(0.990)	(0.687)	(0.386)	(0.102)	(0.000623)
E_{75}	0.000252	0.000178	0.000105	0.0000301	0.000000165
	(0.000155)	(0.000108)	(0.0000604)	(0.0000160)	(0.000000098)
Constant	0.290	2.827	5.365	8.0762	13.819831
	(0.668)	(0.463)	(0.260)	(0.0691)	(0.000420)
Obs.	445	445	445	445	445
Adj. R^2	0.0160	0.0165	0.0174	0.0178	0.0197
Estimated marginal effects and elasticities					
ME(E_{75})	17.0	2.73	0.7300	0.190	0.1658
	(52.7)	(3.99)	(0.542)	(0.103)	(0.0981)
$\xi(E_{78}, E_{75})$	0.347	0.246	0.1444	0.0415	0.000227
	(0.213)	(0.148)	(0.0831)	(0.0221)	(0.000134)

Table 16
Comparison of a linear probability model (LPM) with inverse hyperbolic sine and natural logarithm, with original zeros set to zero, and normed by the mean of the positive values.

	LPM	$\text{arcsinh}(1 \times y)$	$\ln(y+1)$
Estimated parameters			
NSW	.1078	1.056	.982
	(.0442)	(.416)	(.386)
E_{75}	.0000106	.0001122	.0001049
	(.00000691)	(.0000651)	(.0000604)
Constant	.633	5.803	5.3648
	(.0298)	(.280)	(.260)
Obs.	445	445	445
Adj. R^2	0.0148	0.0173	0.0174
Estimated normed marginal effects			
$E[asinh(y)\|y>0]$		9.242	
$E[ln(y)\|y>0]$			8.549
Normed ME(E_{75})	.0000106	.0000121	.0000123
	(.00000691)	(.00000704)	(.00000706)

Table 17
Marginal effects and elasticities from a two-part, untransformed y, and Poisson models.

	Two-Part Model			
	Logit	GLM (log link)	Untransformed y	Poisson
Estimated parameters				
NSW	.520	.17090	1750	.325
	(.216)	(.0980)	(669)	(.117)
E_{75}	.0000594	.0000137	.167	.0000266
	(.0000392)	(.0000148)	(.107)	(.0000149)
Constant	.533	8.8399	4344	8.3865
	(.137)	(.0694)	(356)	(.0776)
Obs.	445	308	445	445
Estimated marginal effects				
ME(NSW)	.1087	1309.1		
	(.0441)	(759.1)		
Overall ME(NSW)	1720.8		1750	1721.5
	(620.3)		(669)	(649.2)
ME(E_{75})	.0000124	.105		
	(.00000812)	(.114)		
Overall ME(E_{75})	.1657		.167	.1409
	(.0993)		(.107)	(.0795)
Estimated elasticities				
$\xi(E_{78}, NSW)$	.0526	.0777		
	(.0169)	(.0446)		
Overall $\xi(E_{78}, NSW)$	.1237		.1152	.1350
	(.0441)		(.0354)	(.0488)
$\xi(E_{78}, E_{75})$	.01852	.0211		
	(.00891)	(.0228)		
Overall $\xi(E_{78}, E_{75})$	.0374		.0363	.0366
	(.0222)		(.0197)	(.0205)

Notes: The second part of the two-part model is estimated with generalized linear models (GLM) with a log link and a gamma distribution.

Figure 1: Natural logarithm transformation as a function of *c*.

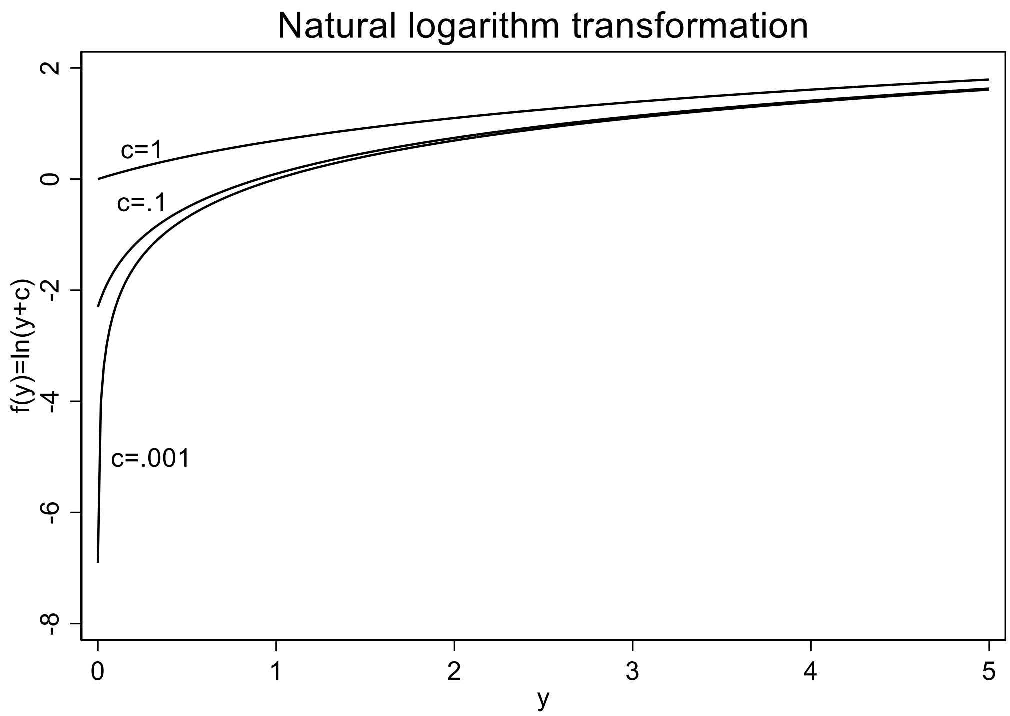

Figure 2: Inverse hyperbolic sine transformation as a function of *k*.

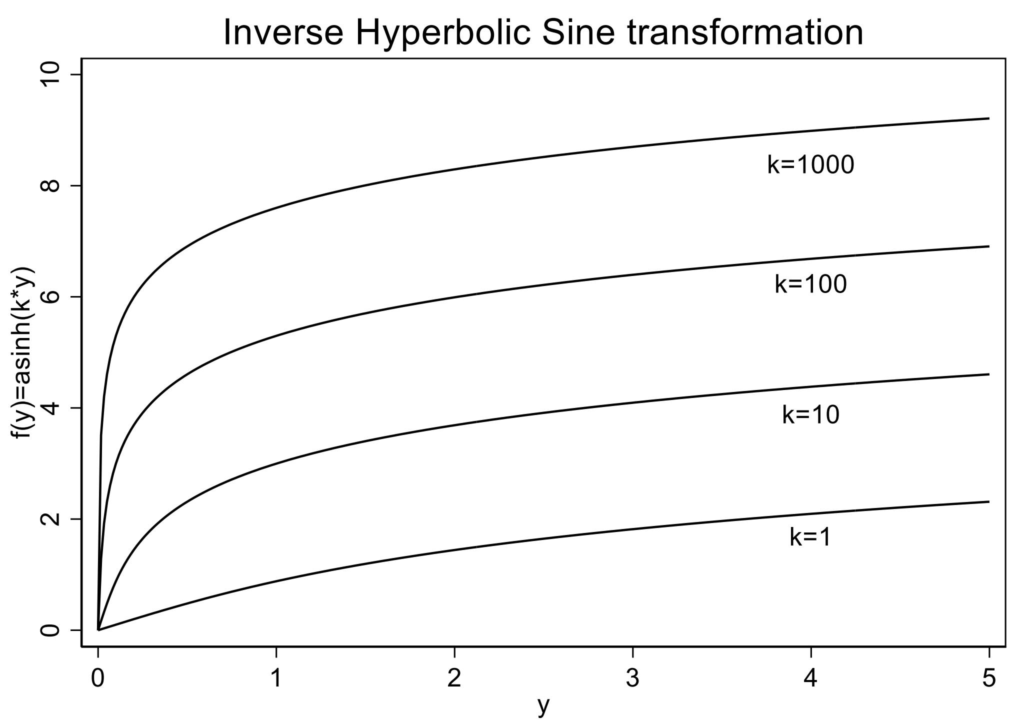

Figure 3: Power transformation as a function of λ.

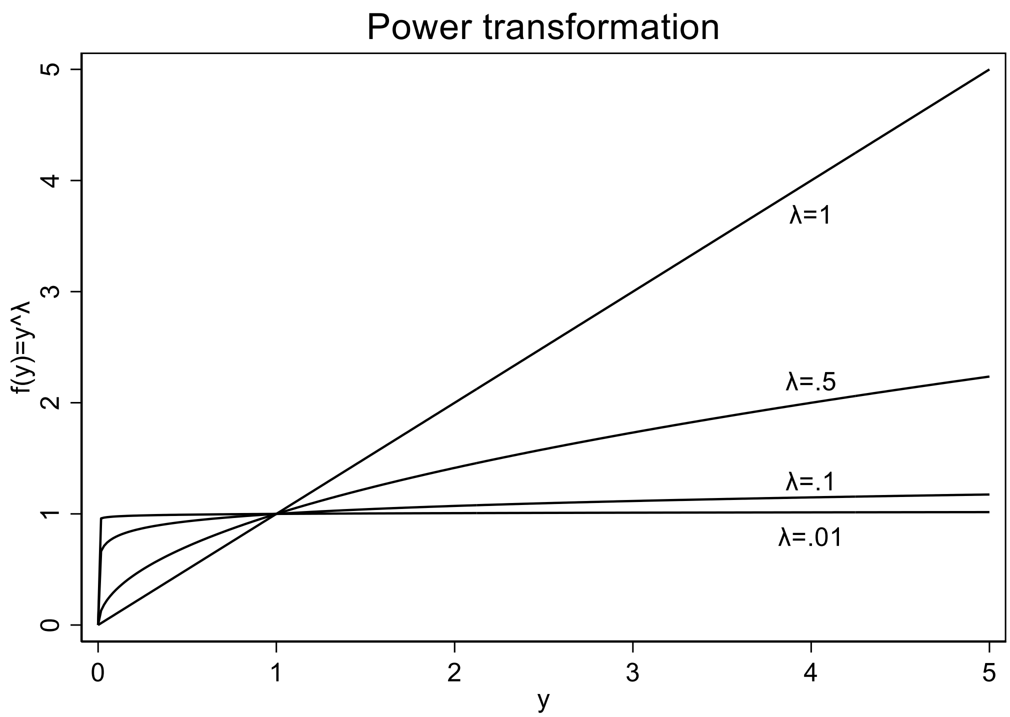

Figure 4: Histogram of total health care expenditures, MEPS data.

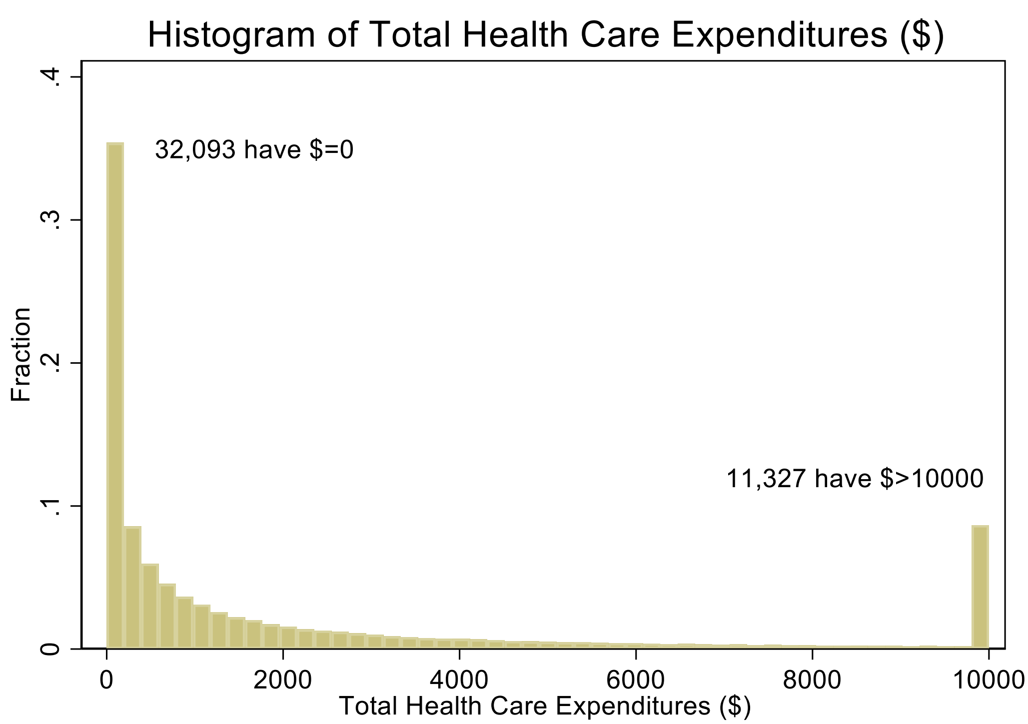

Figure 5: Horizontal box plot of distribution of $\ln(y + c)$ for different values of c.

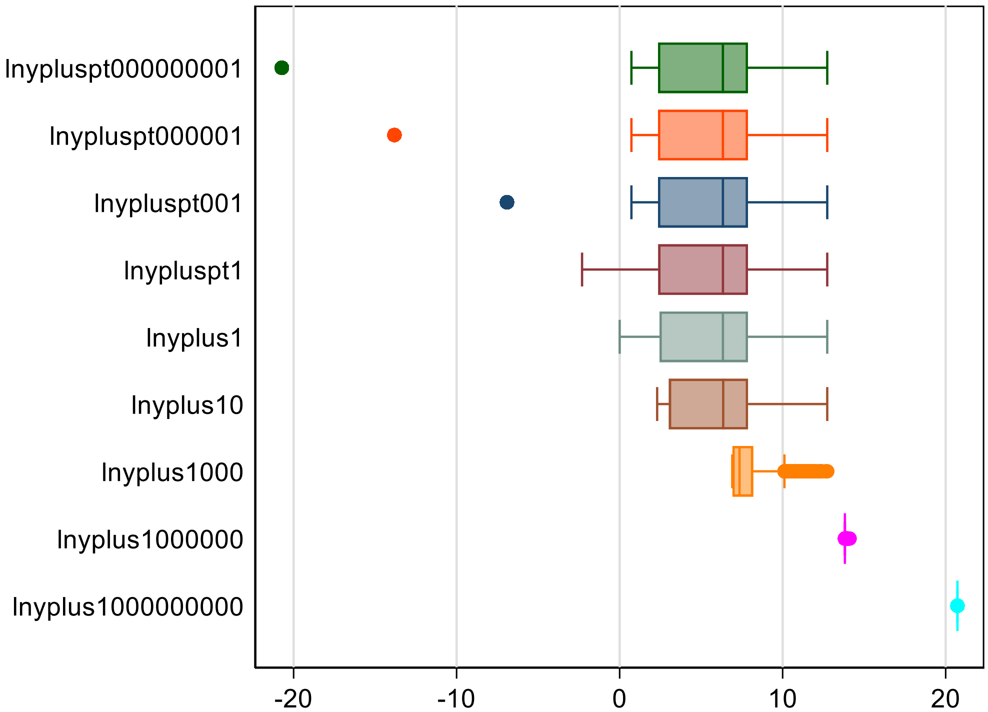

Figure 6: Regressions of normed $\ln(y + c)$ on age, for different values of c.

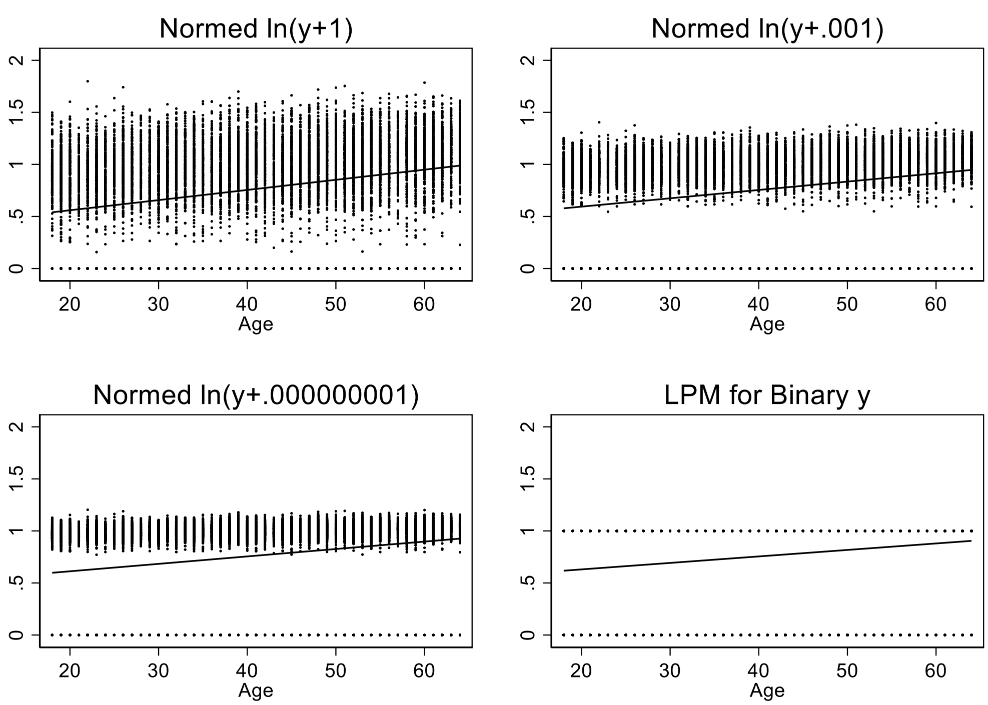

Figure 7: Horizontal box plot of distribution of $\text{arcsinh}(ky)$ for different values of k.

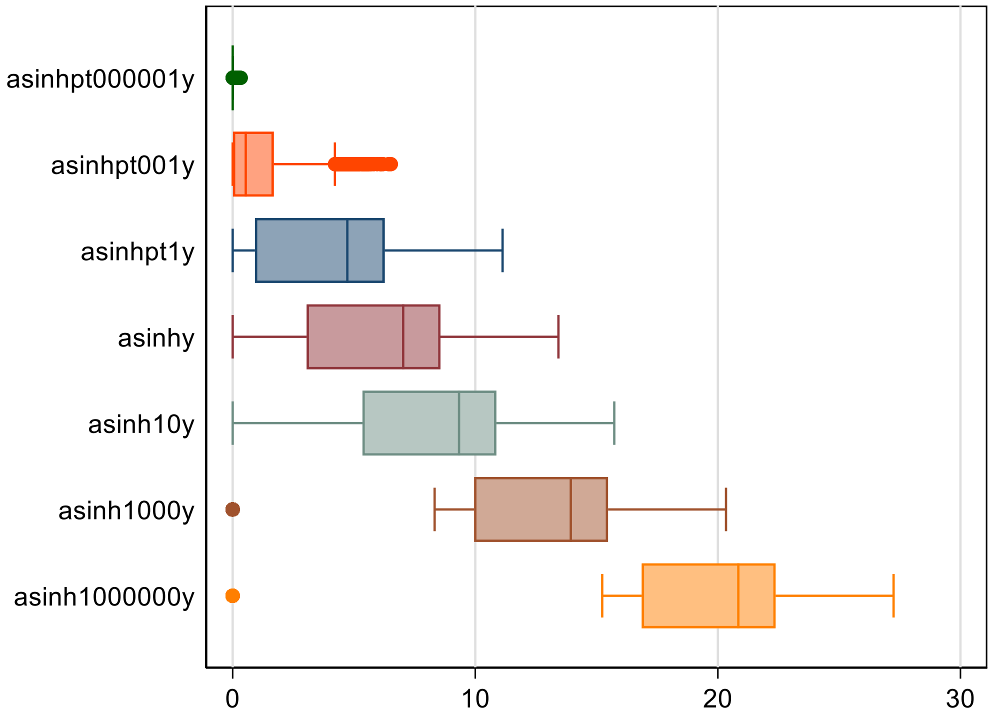

Figure 8: Regressions of normed $\text{arcsinh}(ky)$ for different values of k.

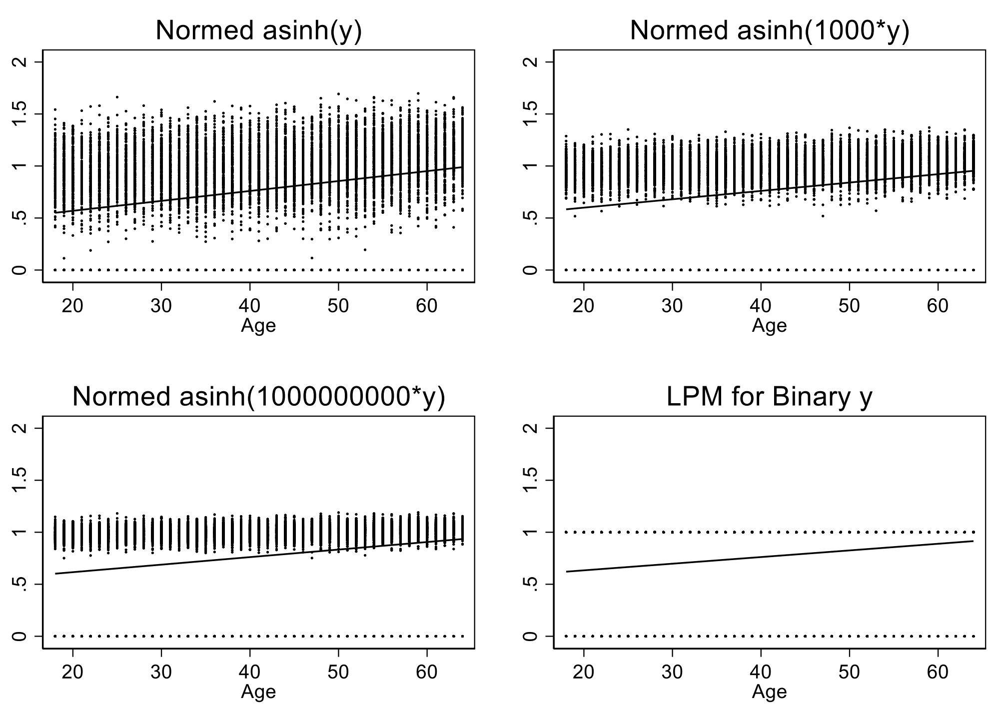

Figure 9
Comparison of the normed natural logarithm and the normed inverse hyperbolic sine to the linear probability, with original zeros set to zero. The slopes are all nearly the same (see Table 16).