When Are Estimates Independent of Measurement Units?

Neil Thakral
Brown University

Linh T. Tô
Boston University

July 2023*

Abstract

Data transformations often facilitate regression analysis, yet many commonly used transformations make hypothesis testing misleading because the results depend on the measurement units of the data. This paper aims to address this issue by characterizing the set of transformations where measurement units do not affect conclusions in linear regressions. The equivalence theorem establishes that desirable properties—scale-equivariant coefficient estimates, scale-invariant t-statistics, and scale-invariant semi-elasticities—arise if and only if the transformation is a logarithmic or a power function. Power transformations thus offer a natural extension of logarithmic transformations that both preserves the essential feature of obtaining unit-independent estimates for unitless quantities of interest and can handle zero or negative values. On the other hand, popular alternatives that approximate the shape of the logarithmic function at large values, such as adding a small positive constant before applying a logarithmic transformation or the inverse hyperbolic sine transformation, result in similar inferences as in an untransformed linear regression when expressing outcomes in large measurement units and imply arbitrarily large effect sizes or arbitrarily large confidence intervals when expressing outcomes in small measurement units. We demonstrate using data from a randomized experiment that such transformations reverse the sign or significance of treatment effect estimates for up to 15 out of 49 outcome variables when measurement units are changed to natural alternatives (e.g., from US dollars to local currency).

*Thakral: Department of Economics, Brown University (email: neil_thakral@brown.edu). Tô: Department of Economics, Boston University (email: linhto@bu.edu). We are grateful to Erzo F. P. Luttmer for inspiring us to write this paper and for the hospitality of the Industrial Relations Section at Princeton University. We thank Jean-Jacques Forneron, Lawrence Katz, Patrick Kline, Kevin Lang, David Lee, David McKenzie, Jesse Shapiro, and audiences at Boston University, Princeton University, University of Chicago Harris School of Public Policy, University of Michigan, and the European Labor Symposium for Early Career Economists for helpful comments and feedback. We also thank Michael Briskin for excellent research assistance.

1 Introduction

Ordinary least squares (OLS) regression analysis stands as one of the most ubiquitous and fundamental tools for empirical research. While traditionally associated with linear relationships, OLS accommodates nonlinear relationships between variables through transformations. A common transformation involves using the logarithmic function, which facilitates coefficient interpretations as percent changes, but the presence of zero- or negative-valued observations naturally creates a challenge. To address this, researchers in economics and related fields often use alternatives such as applying the logarithmic transformation after adding a small positive constant or taking the inverse hyperbolic sine (arsinh, or IHS). Although these alternatives offer practical ways of dealing with zero or negative values, they remain ad hoc approaches, often adopted without fully considering their implications.

To assess whether a regression model delivers meaningful results, a basic criterion is that applying the model to the same dataset presented in different measurement units should result in the same conclusions. For example, a meaningful model should not imply that an intervention causes a significant effect on an outcome when measured in US dollars but not in Kenyan shillings. Situations in which researchers face a choice of measurement units apply to many economic variables: measuring incidence rates such as births, disease, crime, or mortality (e.g., per 100 population or 100,000 population); measuring monetary values in different currencies (scale factors up to over 100,000 in the present); or measuring rates over time, such as trade volume per second or annually (scale factors up to about 30 million). This paper asks which transformations in linear regressions satisfy the following desirable property: that changing the measurement units of the data does not change the model's conclusions.

We formalize the notion of measurement-unit-independent transformations by introducing the concepts of *scale equivariance*, that any scaling of the data can be reversed by appropriately rescaling the estimator, and *scale invariance*, that any scaling of the data does not affect the estimator. For a simple example, consider a linear regression without any transformation. In this case, changing the scale of the outcome variable correspondingly changes the magnitudes of the regression coefficients, and rescaling these estimates results in the same magnitudes as in the original units; additionally, the change in scaling does not affect the implied percent change in outcomes. Thus the coefficient estimates are scale equivariant and the semi-elasticity estimates are scale invariant. Moreover, the standard error estimates also scale the same way as the coefficient estimate, and thus scaling does not alter their statistical significance, i.e., the t-statistics are scale invariant. The logarithmic transformation also satisfies these properties, as changing the scale of the outcome variable only modifies the constant term in a logarithmic regression.

To characterize the full set of measurement-unit-independent transformations, we establish the following equivalence theorem. Requiring any of the desirable properties—scale-equivariant coefficient estimates, scale-invariant t-statistics, and scale-invariant semi-elasticity estimates—is equivalent to requiring all of them. Moreover, the set of transformations satisfying these properties consists exactly of affine functions of a logarithmic or a power function.

We then evaluate the extent and consequences of failing to achieve scale equivariance empirically by examining the increasingly popular IHS transformation using data from a randomized controlled experiment on cash transfers in Kenya (Haushofer and Shapiro, 2016). These data serve as an excellent case study given its broad range of outcomes: assets, consumption, education, female empowerment, food security, health, income, investment, and psychological well-being. The dataset includes 98 baseline and endline measures in IHS form, which the original study uses for robustness exercises, and all of these variables have natural alternative choices of units, such as US dollars or Kenyan shillings for monetary outcomes, or µg/dL or nmol/L for stress hormone levels.

Our analysis reveals that changing measurement units to natural alternatives reverses the sign of treatment effect estimates for up to 22 percent of endline outcomes. Moreover, for up to 18 percent of endline outcomes, the natural change in units alters the conclusion of testing the null hypothesis of no treatment effect at the 5 percent significance level. In total, the natural change in units can reverse the sign or significance of the estimated treatment effect for up to 31 percent of endline outcomes, and an arbitrary choice of units can lead to a reversal in up to 82 percent of cases.

Given the extensive use of the transformations $\log(y+1)$ and $\operatorname{arsinh}(y) = \log\left(y + \sqrt{y^2+1}\right)$, we further investigate the severity of scale dependence for a class of transformations that encompasses these.[1] For such quasi-logarithmic transformations, we provide the following comprehensive characterization. First, using larger units of measurement for the outcome variable (i.e., applying smaller scale factors) leads to a regression coefficient that approaches zero, with a t-statistic and implied semi-elasticity that approach the result that would arise from an untransformed linear regression. This holds regardless of the presence of zero- or negative-valued observations. Second, the effect of using smaller units of measurement for the outcome variable (i.e., applying larger scale factors) depends on whether the independent variable correlates with the sign of the outcome variable. With a non-zero correlation, the absolute value of the regression coefficient and implied semi-elasticity approach infinity. With a zero correlation, the confidence interval becomes arbitrarily large, whenever logarithmic

[1] The number of publications in economics and finance journals that use these transformations currently exceeds 1,000 per year (Appendix Figure 1).

transformations do not apply. Third, we illustrate that the scaling factor required to achieve a given coefficient need not depend on the fraction of zero-valued observations. In particular, a given level of sensitivity to measurement units can occur even with a vanishingly small fraction of zero-valued observations. We establish an analogous set of results for quasi-logarithmic transformations of covariates.[2]

Given the potential for highly misleading empirical results when using scale-dependent transformations, we provide practical recommendations for using scale-equivariant transformations effectively and interpreting the resulting semi-elasticities. We also highlight the differences between using transformations and alternative approaches such as using generalized linear models (e.g., Poisson regressions), presenting an analysis of the relationship between wages and education using data from the Current Population Survey as a concrete example.

This paper contributes to a methodological literature on the sensitivity and interpretation of estimates involving transformed data. Several recent papers document sensitivity of empirical estimates using $\text{arsinh}(y)$ as the outcome variable to arbitrary choices regarding scaling (Bellemare and Wichman, 2020; de Brauw and Herskowitz, 2021; Delius and Sterck, 2020; Aihounton and Henningsen, 2021; Norton, 2022; Mullahy and Norton, 2022; Chen and Roth, 2023), extending prior work that yields similar conclusions when using $\log(y + 1)$ as the outcome variable (Flowerdew and Aitkin, 1982; King, 1988).[3] While these previous examples emphasize that arbitrary scale factors can cause estimates to change in magnitude, our empirical analysis establishes that a natural change of units can dramatically alter the sign and significance of treatment effect estimates. A significant paper by Bellemare and Wichman (2020) demonstrates that coefficients from regressions involving the IHS transformation do not directly have meaningful interpretations and explains how to convert them to semi-elasticities. Building on these findings, we present novel results on scale dependence of coefficient estimates, semi-elasticity estimates, and t-statistics for a broad class of transformations that include the commonly used cases of $\log(y + 1)$ and $\text{arsinh}(y)$. Our characterization of scale dependence under quasi-logarithmic transformations encompasses special cases such as binary treatments and non-negative outcomes (Chen and Roth, 2023) and offers a more comprehensive characterization than prior work describing problems that can arise in the presence of an extensive margin response of a treatment on outcomes (e.g.,

[2]A list of selected publications from 2020–2023 applying the inverse hyperbolic sine to an independent variable appears in Appendix Table 2.

[3]A review article by Santos Silva and Tenreyo (2022) points out that alternatives to Poisson regression for modeling count data that allow for "overdispersion" (e.g., see Cameron and Trivedi, 2013) also exhibit scale dependence, following results due to Bosquet and Boulhol (2014) on scale dependence for negative binomial estimators. These estimators fall within the class of generalized linear models, for which Thakral and Tô (2023) provides a complete characterization of scale dependence. Also see Bellégo, Benatia and Pape (2022) for an alternative estimator that overcomes various issues with Poisson models.

Delius and Sterck, 2020; Mullahy and Norton, 2022; Chen and Roth, 2023) or cautioning against using IHS transformations with a large fraction of zero-valued observations (e.g., Bellemare and Wichman, 2020; de Brauw and Herskowitz, 2021). Our findings contribute to this dialogue by showing how neither a small fraction of zeros nor a lack of extensive margin response addresses the fundamental problems of using the IHS and related transformations.

More importantly, our work goes beyond documenting sensitivity of a particular transformation or class of transformations by fully characterizing the set of scale-equivariant transformations.[4] This leads to the positive conclusion that the set of transformations yielding scale-invariant semi-elasticity estimates generalizes beyond the special case of scale-invariant transformations and includes transformations that apply to zero and negative values. In addition, our consideration of the stability of statistical inference (e.g., see Spitzer, 1984; Wooldridge, 1992) also constitutes a novel contribution to this line of work.

This contrasts with a large literature in statistics and economics proposing alternative families of transformations with goals such as satisfying linearity, normality, or homoskedasticity properties. The simplest and perhaps most widely adopted approach involves taking logarithms after applying additive shifts (see, e.g., Bartlett, 1947; Aitchison and Brown, 1957; Draper and Hunter, 1969; Hoyle, 1973). The two-parameter Box and Cox (1964) transformation (also see Tukey, 1957; Moore, 1957; Dolby, 1963; Turner, Monroe and Lucas, 1961) falls into this category as well, along with some of its extensions that allow for the full range of negative-valued observations, such as the modulus transformation (John and Draper, 1980) and the Yeo-Johnson power transformation (Yeo and Johnson, 2000). Another pervasive proposal motivated by the presence of zero- and negative-valued observations involves using a different functional form that also approximates the shape of the logarithm: the IHS transformation (Johnson, 1949; Burbidge, Magee and Robb, 1988). The IHS transformation and its various extensions, such as generalized power transformations (Kelmansky, Martínez and Leiva, 2013) and a hybrid of the hyperbolic sine and its inverse (Ravallion, 2017), also involve taking a logarithm after applying an additive shift, which the scale-equivariance condition precludes. Scale equivariance also rules out other related transformations that do not involve additive shift terms, such as the Manly exponential transformation (Manly, 1976) and dual power transformations (Yang, 2006).

Our complete characterization of scale-equivariant transformations provides practical guidance on obtaining meaningful estimates in empirical research. A growing set of recent papers, following Thomas et al. (2006) and Ashraf et al. (2015), uses the quartic root

[4] Our results add to a small set of papers in economics that examine scale-equivariance properties of different estimators (e.g., Koenker and Bassett Jr, 1978; Newey and Powell, 1987; He et al., 1990; Machado, 1993; Kemp, 1996; Müller, 2007; Preminger and Sakata, 2007; Sakata, 2007; Tabri, 2014; Thakral and Tô, 2023).

transformation as an alternative to the logarithmic transformation, stating that it "behaves similarly to," "closely approximates," "closely follows," or "mimics the logarithmic function reasonably well," while an even larger number of papers choose to apply the IHS transformation for similar reasons, noting that it tends toward the logarithmic function asymptotically.[5] Our results provide a formal justification for adopting power transformations such as the cubic or quartic root over the more common quasi-logarithmic transformations. We also provide clarity on the interpretation of regression coefficients when using power transformations and discuss alternative approaches such as Poisson regressions.

The paper proceeds as follows. Section 2 introduces formal definitions of scale equivariance and related concepts for desirable properties of transformations in regressions. Section 3 characterizes the family of scale-equivariant transformations for linear regressions and establishes the necessity and sufficiency of using a scale-equivariant transformation for obtaining scale-invariant estimates of semi-elasticities and t-statistics. Section 4 demonstrates the extent of scale dependence using data from a randomized experiment. Section 5 derives results for scale-dependent transformations such as $\log(y+1)$ and $\operatorname{arsinh}(y)$. Section 6 summarizes our practical recommendations and discusses comparisons between transformed regressions and alternative approaches. Section 7 concludes.

2 Equivariance properties

In this section, we define basic properties of measurement-unit-independent transformations in linear regressions. We start by introducing the concept of scale equivariance and its special case scale invariance.

Scale equivariance as a mathematical concept defines mappings or functions that preserve the structure of an output when scaling the input by a constant factor. A function $V\colon \mathbb{R} \to \mathbb{R}$ is *invariant* if $V(cy) = V(y)$ for all $c > 0$. A function V is *scale equivariant* if, for any $c > 0$, there exists $\sigma(c) > 0$ such that $V(cy) = \sigma(c)V(y)$. The special case of $\sigma(c) \equiv 1$ reduces to the definition of invariance. The concept of scale equivariance is sometimes mistakenly referred to as scale invariance.[6] Scale invariance refers to the property that a function produces the same output in response to a change in measurement units of the input. The more general concept of scale equivariance simply requires that the original output can be recovered by a consistent rescaling that may depend on the scaling factor. In other words, the effect of scaling the input can be reversed by appropriately rescaling the output.

[5] A list of 36 selected references using the cubic, quartic, or quintic root transformation, mostly published since 2019, appears in Appendix Table 1.

[6] For example, see James et al. (2013).

We apply scale equivariance to estimators of transformed linear regressions. Scale equivariance formalizes the idea that scaling unit-based data, our inputs, should not alter interpretations of the basic estimators, our outputs.

2.1 Scale equivariance of estimators

Definition 1 (Scale equivariance)**.** We say that an estimator $\Theta\colon \mathbb{R}^{n(m+1)} \to \mathbb{R}$ is *scale equivariant* if, for every $c > 0$, there exists $\psi(c) > 0$ such that, for any dataset $x_1, \dots, x_m, y \in \mathbb{R}^n$, we have $\Theta(x_1, \dots, x_m, cy) = \psi(c)\Theta(x_1, \dots, x_m, y)$.

Consider the function $\theta(y \,|\, x_1, \dots, x_m)$ defined as $\Theta(x_1, \dots, x_m, y)$. *Scale invariance* refers to the special case of $\psi(c) \equiv 1$ in Definition 1, or equivalently $\theta(y)$ being homogeneous of degree 0. We use the term *exact scale equivariance* to refer to the special case $\psi(c) = c$, i.e., when $\theta(y)$ is homogeneous of degree 1. Note however, that scale equivariance is a distinct and broader concept than the two special cases of interest.

We will apply these definitions to important estimators used in empirical research including coefficient and semi-elasticity estimates, and t-statistics for hypothesis testing.

2.2 Scale equivariance of transformations

Let f be a strictly increasing twice-differentiable function, which we refer to as a *transformation*. Consider the ordinary least squares (OLS) model $f(y) = X\beta + \varepsilon$, where $\beta = \begin{bmatrix}\beta_0 & \cdots & \beta_m\end{bmatrix}'$, where X is a $n \times (m+1)$ matrix that includes a column of ones as its first column, and $\mathbb{E}[\varepsilon \,|\, X] = 0$. Denote the OLS estimator of the coefficient $\beta_j(X, f(y))$ by $\hat{\beta}_j(X, f(y))$ and its standard error estimate by $\widehat{\text{s.e.}}\left(\hat{\beta}_j(X, f(y))\right)$, for each $j \in \{0, 1, \dots, m\}$.

The ratio between the coefficient estimate $\hat{\beta}_j$ and its standard error $\widehat{\text{s.e.}}\left(\hat{\beta}_j\right)$ forms the basis for statistical hypothesis testing, and thus the first property we introduce for evaluating whether a data transformation results in meaningful estimates states that the conclusions of hypothesis tests should not depend on the measurement units of the data.

Property 1 (Scale invariance of t-statistics)**.** *A transformation f has a scale-invariant t-statistic $t_{\hat{\beta}_j(X, f(y))}$ if, for any dataset $x_1, \dots, x_m, y \in \mathbb{R}^n$, the ratio between the coefficient estimate $\hat{\beta}_j(X, f(cy))$ and the standard error estimate $\widehat{\text{s.e.}}_j\left(\hat{\beta}_j(X, f(cy))\right)$ is unchanged for all $c > 0$.*

Failure of this property can have undesirable consequences. Even if the t-statistic only varies slightly with a change in the measurement unit, simply presenting the data in one unit versus another could push a result from being statistically insignificant to statistically significant. In Section 4, we show that using the one of the most common transformations that are asymptotically equivalent to the logarithm transformation, the inverse hyperbolic sine, could even change the sign and significance of the t-statistic for the same dataset presented in US dollars versus Kenyan shillings.

Property 1 only requires that the coefficient and standard error estimates are affected "in the same way" when the outcome variable is presented in a different measurement unit. It does not specify how the coefficients themselves should change with rescaling of the data, and is therefore entirely distinct from Property 2, scale-equivariant coefficient estimate, defined below.

Property 2 (Scale equivariance of coefficient estimates)**.** *A transformation f has a scale-equivariant coefficient $\hat{\beta}_j(X, f(cy))$ if, for any dataset $x_1, \ldots, x_m, y \in \mathbb{R}^n$ and for every $c > 0$, there exists $\psi(c) > 0$ such that $\hat{\beta}_j(X, f(cy)) = \psi(c)\hat{\beta}_j(X, f(y))$.*

The above definition can be applied to any particular coefficient estimate. As a shorthand, we also say that f is a *scale-equivariant transformation* if $\hat{\beta}_j(X, f(y))$ is a scale-equivariant estimator for all $j \in \{1, \ldots, m\}$. In other words, under a scale-equivariant transformation f, changing the measurement units of the dependent variable by a factor of c is equivalent to scaling the OLS estimators (excluding the constant term corresponding to $j = 0$) by $\psi(c)$. If $\psi(c) \equiv 1$, then we say that f is a *scale-invariant transformation*. The identity transformation $f(y) = y$ in an untransformed linear regression is exactly scale equivariant with $\psi(c) = c$ for every coefficient $\hat{\beta}_j$, $j \in \{0, \ldots, m\}$. The logarithmic transformation $f(y) = \log(y)$ in a log-linear regression is scale invariant with $\psi(c) = 1$ for every coefficient $\hat{\beta}_j$, $j \in \{1, \ldots, m\}$.

Next, we consider the semi-elasticity of y with respect to x_j evaluated at a particular value y_0. In a transformed OLS regression, this is given by $\hat{\xi}_{x_j}(X, f(y))\Big|_{y=y_0} := \frac{1}{y}\frac{\partial y}{\partial x_j}\Big|_{y=y_0}$. Since $\beta_j(X, f(y)) = f'(y)\frac{\partial y}{\partial x_j}$, we alternatively write the semi-elasticity as $\hat{\xi}^f_{x_j}(y_0, \hat{\beta}) = \frac{1}{y_0}\frac{\hat{\beta}_j}{f'(y_0)}$.[7] Given the interpretation of a percent change in y with respect to changes in x_j, which is inherently unitless, a basic requirement for transformed regressions would be to preserve the semi-elasticity estimates for covariates of interest.

Property 3 (Scale invariance of semi-elasticity estimates)**.** *A transformation f has a scale-invariant semi-elasticity $\hat{\xi}_{x_j}(X, f(y))$ if, for any dataset $x_1, \ldots, x_m, y \in \mathbb{R}^n$, changing the measurement units of the dependent variable does not change the semi-elasticity estimate: $\hat{\xi}_{x_j}(X, f(y))\Big|_{y=y_0} = \hat{\xi}_{x_j}(X, f(cy))\Big|_{y=y_0}$ for all $c > 0$ and all $y_0 \neq 0$.*

While the standard semi-elasticity definition used above applies to continuous x_j, the concept extends to the case of binary x_j. To measure what the model implies about the

[7] If f is the logarithmic function, this reduces to $\hat{\beta}_j$, which is independent of y_0.

percent change in y with respect to x_j, the predicted value $\mathbb{E}[y \mid X]$ allows us to construct a natural alternative, namely, $\frac{\mathbb{E}[y \mid X, x_j=1] - \mathbb{E}[y \mid X, x_j=0]}{\mathbb{E}[y \mid X, x_j=0]}$. For any nonlinear transformation $f(y)$ of the outcome variable, including $\log(y)$, obtaining the predicted value of the untransformed outcome variable requires accounting for the fact that the error term no longer cancels in $\mathbb{E}[f^{-1}(X\beta + \varepsilon) \mid X]$, even when $\mathbb{E}[\varepsilon \mid X] = 0$. We discuss a non-parametric estimator of the predicted value in Appendix A and note that our results below involving semi-elasticities hold in the discrete case as well Thakral and Tô (2023).

3 Characterization of scale-equivariant transformations

Each of the properties from Section 2 represents a fundamental desirable feature of data transformations, as any violation would imply that the same regression and data can yield inconsistent conclusions. This section examines the relationship between these distinct properties and the restrictions they impose on the transformation f.

Recall that we define Property 1 on t-statistics, Property 2 on coefficient estimates, and Property 3 on semi-elasticity estimates with respect to individual covariates, indexed by j. These definitions account for the fact that in regression analysis, researchers often focus on a subset of covariates while regarding others as control variables. The next set of results shows that requiring any of these properties for any of the covariates imposes the same restriction on the transformation f.

Proposition 1. *If any of Property 1, Property 2, or Property 3 holds for some $j \in \{1, \ldots, m\}$, then f must be an affine function of a logarithmic or power function.*

A distinct proof accompanies each property, highlighting the nontriviality of their interconnectedness. A proof for Property 1 applying differential equation methods appears in Appendix B.3.1, and a proof for Property 2 relying on functional equation techniques appears in Appendix B.2.1. Proposition 2 below establishes an equivalence between Property 2 and Property 3, which completes the proof of Proposition 1.

Proposition 2. *A transformation f has scale equivariant coefficient $\hat{\beta}_j$ (Property 2) if and only if it has scale invariant semi-elasticity $\hat{\xi}_{x_j}$ (Property 3).*

The proof in Appendix B.4 consists of showing that Property 2 and Property 3 each hold if and only if the transformation f satisfies the condition $f(cy) = \psi(c)f(y) + h(c)$. This functional equation implies $\psi(c_1 c_2) = \psi(c_1)\psi(c_2)$ (Lemma B.3), and therefore $\psi(c) = c^k$ (Lemma B.4). When $k = 0$, we obtain a Cauchy-like functional equation $f(cy) = f(y) + h(c)$ and show that its solutions consist of an affine function of $\log(y)$ (Lemma B.5). When

$k \neq 0$, we show that h, and consequently f, must be an affine function of a power function (Appendix B.2.1).

Notice that $f(cy) = \psi(c)f(y)$ alone would define a scale-equivariant function; however, scale-equivariant transformations also contain an additive component $h(c)$. This highlights a distinction between power and logarithmic transformations. Power transformations, including the identity transformation (i.e., the untransformed linear regression), preserve the essential properties for all coefficients, including the constant term, even if the coefficient estimates themselves vary with scaling. The logarithmic transformation, on the other hand, uses the constant term to absorb any changes in measurement units but otherwise preserves the same properties for the covariates.

The next result establishes that a strong form of the converse of Proposition 1 also holds.

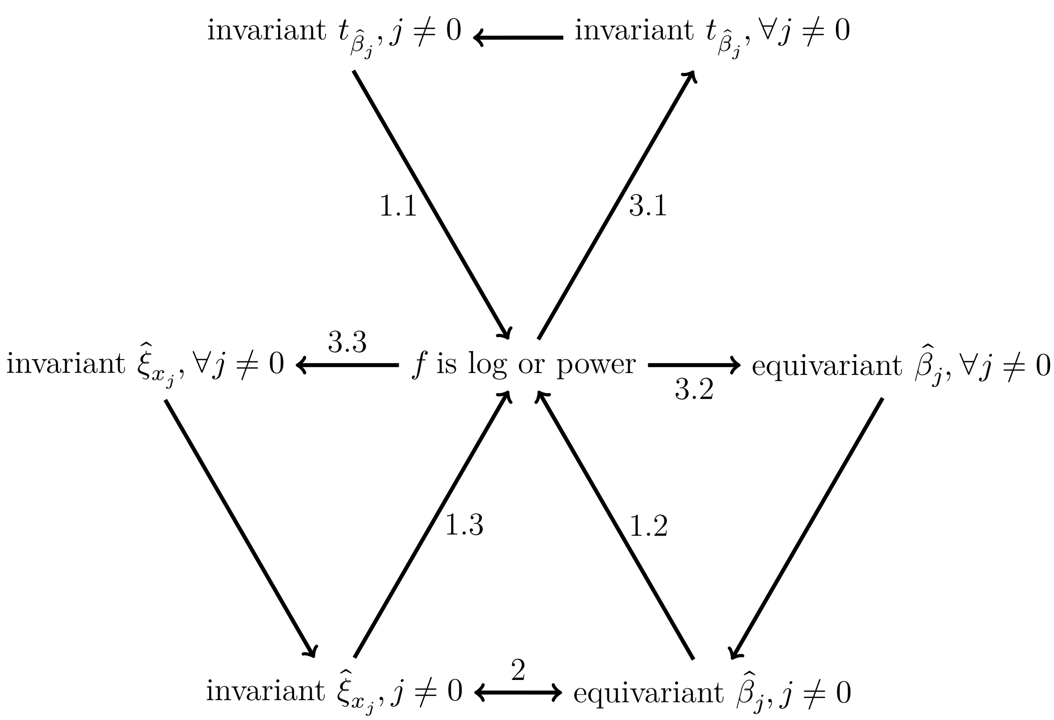
Proposition 3. *If f is an affine function of a logarithmic or power function, then all of Property 1, Property 2, and Property 3 hold for all $j \in \{1, \dots, m\}$.*

Verifying whether a given transformation exhibits scale equivariance requires less effort than deriving functional forms for the transformation from the property. All three properties follow from showing that logarithmic and power transformations result in regression coefficients that satisfy scale equivariance (Appendix B.2.2) and that the standard errors satisfy scale equivariance with the same rescaling factor $\psi(c)$ (Appendix B.3.2), given the equivalence between Property 2 and Property 3. For the case of power transformations, as noted above, all three properties hold for the constant term ($j = 0$) as well.

Combining the results from Proposition 1 and Proposition 3 immediately establishes the following result.

Corollary 1. *Requiring any of Properties 1 to 3 for any independent variable x_j ($j \neq 0$) is equivalent to requiring all three properties for all independent variables x_j ($j \neq 0$).*

The diagram below summarizes the equivalence between the properties for individual covariates.



We state the following theorem to summarize the preceding results.

Theorem (Regression Transformation Equivalence Theorem)**.** *The transformation f is scale equivariant if and only if any (and hence, all) of the following hold:*

1. *The t-statistic $t_{\hat{\beta}_j(X, f(cy))}$ is scale-invariant for some $j \in \{1, \dots, m\}$.*
2. *The t-statistic $t_{\hat{\beta}_j(X, f(cy))}$ is scale-invariant for all $j \in \{1, \dots, m\}$.*
3. *The coefficient estimate $\hat{\beta}_j(X, f(y))$ is scale-equivariant for some $j \in \{1, \dots, m\}$.*
4. *The coefficient estimate $\hat{\beta}_j(X, f(y))$ is scale-equivariant for all $j \in \{1, \dots, m\}$.*
5. *The semi-elasticity $\hat{\xi}_{x_j}(X, f(y))$ is scale-invariant for some $j \in \{1, \dots, m\}$.*
6. *The semi-elasticity $\hat{\xi}_{x_j}(X, f(y))$ is scale-invariant for all $j \in \{1, \dots, m\}$.*
7. *The transformation f satisfies the functional equation $f(cy) = \psi(c) f(y) + h(c)$ for some functions $\psi\colon \mathbb{R}_+ \to \mathbb{R}_+$ and $h\colon \mathbb{R}_+ \to \mathbb{R}$.*
8. *The transformation f is an affine function of a logarithmic or a power function: $f(y) = c_1 \log(y) + c_2$ or $f(y) = c_1 y^k + c_2$ for some constants k, c_1, c_2 satisfying $k \neq 0$ and $kc_1 > 0$.*

This establishes that logarithmic and power transformations provide the only viable options for drawing meaningful conclusions from a transformed linear regression.

Comparison of power and logarithmic transformations

Scale-equivariant transformations provide an appropriate generalization of logarithmic transformations that preserves the essential feature of obtaining unit-independent estimates for unitless quantities of interest. Logarithmic transformations, which impose the strong assumption of a constant semi-elasticity, uniquely result in regression coefficients that satisfy scale invariance. Power transformations, which relax the assumption of a constant semi-elasticity, lead to unit-dependent regression coefficients ($\hat{\beta}_c$) but still result in scale-invariant estimates of semi-elasticities ($\frac{\hat{\beta}_c}{k(cy_0)^k} = \frac{\hat{\beta}}{ky_0^k}$) and t-statistics.[8] In fact, under any power transformation, as with the special case of the identity transformation, the magnitude of the regression coefficient can achieve any arbitrarily large value given an appropriate choice of measurement units of the outcome variable; however, the standard errors scale proportionally, resulting in stable inference (i.e., scale-invariant t-statistics).

While the conclusion of the theorem holds even with strictly positive data, scale-equivariant transformations can also accommodate the presence of zero- or negative-valued data. The following observations offer more specific guidance for analyzing data with non-positive values by using power transformations defined over $[0, \infty)$ or $(-\infty, \infty)$.

Corollary 2. *Scale-equivariant transformations defined for $y = 0$ must be affine functions of y^k, where $k \geq 0$. Scale-equivariant transformations defined for $y < 0$ must be affine functions of y^{k_1/k_2}, where k_1 and k_2 are positive odd integers.*

Corollary 3. *There does not exist a scale-equivariant transformation that is strictly concave and defined at any $y < 0$.*

The impossibility of simultaneously achieving scale invariance and strict concavity over the negative domain has implications for research on measuring welfare, poverty, and inequality. Convexity in the negative domain leads to situations in which transfers from individuals with negative net wealth to those with positive net wealth can decrease measures of inequality, violating the Pigou-Dalton principle (Pigou, 1912; Dalton, 1920), as Ravallion (2017) observes. The result above notes that any concave transformation over the negative domain leads to scale dependence, raising concerns about the suitability of such measures for regression analysis.

In addition to extending beyond the strictly positive domain, power functions provide a simple and useful way of assessing robustness of regression results to the linear functional form. Relatedly, a large body of research emphasizes how concave power transformations

[8]Transformations that approximate the shape of the logarithmic function at large values, on the other hand, can result in arbitrarily large semi-elasticity estimates, as we demonstrate in Section 4.

($k < 1$) can reduce the influence of outliers in right-skewed data, following the work of Box and Cox (1964), which we discuss next.

Relationship with Box-Cox transformation

We discuss below how the equivalence theorem provides a justification for the class of functions that comprise the widely used one-parameter Box-Cox transformation. We conclude this section by noting several limitations of the Box-Cox estimation method that recent literature acknowledges, and we return to a discussion of alternatives and practical recommendations for researchers using this class of transformations in Section 6.

Box and Cox (1964) parametrize a family of transformations through the relationship that $\frac{y^\lambda-1}{\lambda}$ converges to $\log(y)$ as $\lambda \to 0$, encompassing both log and (untransformed) linear as the special cases $\lambda = 0$ and $\lambda = 1$, respectively.[9] Our equivalence theorem clarifies the desirability of log and linear models as scale-equivariant transformations and further presents a complete characterization of all such transformations. It also rules out extensions of the standard Box-Cox transformation such as the two-parameter Box-Cox transformation (Box and Cox, 1964), which attempts to accommodate zero- or negative-valued data by applying an additive shift before taking the log or power.

The original Box-Cox method—which involves assuming homoskedastic normally distributed errors and maximizing the resulting likelihood function to estimate λ—faces several limitations. First, the likelihood function requires strictly positive data. Second, since $\frac{y^\lambda-1}{\lambda}$ is bounded below for almost every λ, the normality assumption almost surely cannot hold. Third, even though estimates of λ exhibit scale invariance, the variance of $\hat{\beta}_j$ accounts for variation in $\hat{\lambda}$, resulting in t-statistics that violate scale invariance (Spitzer, 1984).[10] More recent work explores semiparametric estimation of the Box-Cox transformation, which has the potential to overcome some of the challenges above (see, e.g., Foster, Tian and Wei, 2001; Shin, 2008).

4 Empirical illustration of severity of scale dependence

The results in Section 3 imply that any transformation other than an affine function of a logarithmic or a power function exhibits scale dependence. This naturally prompts the

[9] Davidson and MacKinnon (1993) describe the Box-Cox transformation as "by far the most commonly used nonlinear transformation in statistics and econometrics" and refer to it as a "nonregression model" because of the additional parameter corresponding to the power.

[10] The likelihood-ratio test, on the other hand, satisfies scale invariance. As such, the Stata command `boxcox` only reports the likelihood-ratio test (Drukker, 2000).

question of whether commonly used transformations outside of that family can exhibit a severe degree of scale dependence in practice.

We document the proliferation of empirical research that uses the "quasi-logarithmic" transformations $\log(y+1)$ and $\text{arsinh}(y)$ in Appendix Figure 1.[11] The figure plots the number of publications in economics, finance, or management journals each year that contain "log" and the phrase "one plus" or "plus one", or papers that contain "inverse hyperbolic sine." From 2017 to 2022, the count for $\log(y+1)$ more than doubles to almost 1,000 articles, and the count for IHS more than quintuples.[12] Our empirical application focuses on a dataset that allows us to systematically evaluate various aspects of scale dependence.

4.1 Empirical setting and data

We revisit data collected from a cash transfer experiment in Kenya (Haushofer and Shapiro, 2016) and examine the extent of scale dependence for the inverse hyperbolic sine transformation. The setting and data offer several notable advantages. First, the data come from a randomized controlled trial that induces multiple independent sources of exogenous variation. This allows us to analyze specifications involving a binary treatment (receiving a transfer) as in the original work and a continuous treatment (randomly assigned waiting time) as in subsequent work by Thakral and Tô (2022). Second, the research team gathered an expansive array of over 150 distinct outcome measures spanning across different economic domains (e.g., assets, consumption, education, female empowerment, food security, health, income, investment, and psychological well-being), and the dataset contains IHS-transformed versions of 49 of these variables.[13] Third, each of the 49 variables has a natural alternative unit of measurement. While surveyed households report monetary values in Kenyan shillings, the researchers express these quantities in PPP-adjusted US dollars to facilitate interpretation (62.44 KES $\approx$ 1 USD PPP in 2012). Similarly, expressing the psychological well-being outcome measured by salivary concentration of the stress hormone cortisol in nmol/L or in µg/dL provide interchangeable representations (27.6 nmol/L $\approx$ 1 µg/dL). We analyze all 49 outcomes, leading to a total of 98 measures from the baseline and endline surveys.

[11] Section 5.1 states a formal definition of this class of transformations that apply to zero or negative values while approximating the shape of the logarithmic function at large values.

[12] These statistics likely understate the use of such transformations due to the conservative choice of search terms as well as publication lags. From a web-based poll among a convenience sample of economics researchers, Cohn, Liu and Wardlaw (2022) find that 69 percent of respondents report using such transformations.

[13] The original paper mentions applying the IHS transformation to address right-skewed variables with non-positive values while consistently referring to it as a "log specification."

4.2 Scale dependence of IHS in practice

To evaluate particularly extreme potential consequences of scale dependence in linear regressions using IHS-transformed outcome variables, we examine how often changing measurement units to either the natural alternative or to an arbitrary alternative leads treatment effect estimates to reverse in sign and in statistical significance.[14] Although this analysis focuses on the IHS transformation due to the presence of negative values for some of the variables, we note that $\log(y+1)$ tends to give similar results when it is defined since $\text{arsinh}(y) - \log(y+1)$ ranges between 0 and $\log(2) \approx 0.7$ for $y \geq 0$.

We summarize the results for the binary treatment of receiving any cash transfer in Table 1 and for the continuous treatment of the number of days spent waiting for disbursement (which restricts to a subset of households receiving lump-sum transfers) in Table 2. For each of these treatment variables, we consider four configurations of control variables, corresponding to including or excluding the baseline outcome measure and village fixed effects, and we consider two samples, corresponding to the full sample (which assigns a value of zero to missing endline responses) and the subsample of households that did not attrit. The preferred specification for analyzing endline outcomes in Haushofer and Shapiro (2016) uses the non-attriting subsample with the full set of controls, and the preferred specification for balance tests uses the same subsample with village fixed effects.

Across all specifications for the binary treatment variable, up to 52 out of 98 estimates in total and up to 40 out of 49 endline estimates can reverse in either sign or statistical significance by presenting the data in an arbitrary measurement unit. Focusing on the preferred specifications, we find that 29 estimates reverse in sign, 29 estimates reverse in statistical significance, and 54 estimates reverse in either sign or significance.[15]

Considering only the natural alternative measurement units still leads to sign or significance reversals for up to 4 baseline estimates and up to 15 endline estimates. Using IHS-transformed dependent variables, the same dataset presented in Kenyan shillings and µg/dL instead of US dollars and nmol/L reverses conclusions about statistical significance in up to 9 hypothesis tests and reverses conclusions about whether the treatment increases or decreases outcomes at endline for up to 11 out of 49 variables. A list of endline variables for which we find changes in sign or significance in the basic specification without control variables using the

[14]When considering arbitrary units, we use the limit results established in Section 5 to compare scale factors that approach zero and infinity. Furthermore, because the coefficients and t-statistics can vary non-monotonically with the scale factor, we perform a grid search for additional instances of sign or significance changes.

[15]The preferred specifications correspond to baseline controls and village fixed effects for endline estimates (column 4, Table 1) and village fixed effects for baseline estimates (difference between columns 3 and 4, Table 1).

full sample, along with the regression estimates corresponding to the original and natural alternative units, appears in Appendix Tables 5 and 6.

The continuous treatment variable tells a similar story (Table 2). Across all specifications, we generally find more reversals in sign and fewer reversals in significance compared to the binary treatment case. In the preferred specifications, we find that 38 estimates reverse in sign, 11 estimates reverse in statistical significance, and 47 estimates reverse in either sign or significance. Of these, 13 reversals in sign or significance occur with the natural alternative units.

The severity of scale dependence remains consistent across specifications. Adding baseline controls or village fixed effects does not systematically influence the severity of scale dependence for either the binary treatment or the continuous treatment. Removing attriters from the sample primarily reduces the number of sign reversals in specifications without village fixed effects but not in specifications with village fixed effects. Moreover, removing attriters leads to a similar number of reversals in significance for both natural and arbitrary alternative units and for both the binary and the continuous treatment. While this sample restriction reduces the number of zero-valued observations, we point out in Section 5 how the degree of scale dependence can remain unchanged as the fraction of zero-valued observations approaches zero. Finally, applying a stricter threshold for statistical significance does not systematically reduce the severity of scale dependence: Using arbitrary measurement units continues to reverse the significance of the binary treatment effect estimates for over half of endline outcomes in the preferred specification (Appendix Table 3).

We conclude by noting that the issues exposited above do not change the main results and conclusions of Haushofer and Shapiro (2016). The paper uses IHS transformations in a series of robustness exercises to corroborate the conclusions from untransformed linear regressions, which satisfies scale equivariance as a power transformation with an exponent of 1. With large units (such a US dollar instead of a Kenyan shilling), as Section 5 shows formally, the IHS and other related transformations such as $\log(y+1)$ give similar results to the identity transformation. Using these transformations therefore provides, at best, little or no new information about the robustness of a regression coefficient or treatment effect estimate but does not invalidate any the conclusions derived from untransformed linear regressions. To reduce the influence of observations in the right tail of the data, other power transformations such as the cubic root or quintic root can serve as alternative robustness exercises.

5 Theoretical results on scale dependence of quasi-logarithmic transformations

This section characterizes theoretically the extent of scale dependence in OLS regressions for a class of transformations that encompasses $\log(y+1)$ and $\operatorname{arsinh}(y)$.

5.1 Definitions

We start by introducing the class of *quasi-logarithmic transformations*, which exhibit the same limiting behavior as the logarithmic function for large values but apply to zero and possibly negative values.[16] For example, the transformation $\log(y+c)$ applies over the domain $(-c,\infty)$ and $\operatorname{arsinh}(y)$ applies over $\mathbb{R}$, but both functions tend toward the logarithmic function asymptotically.

Definition 2 (Quasi-logarithmic over $\mathbb{R}_+$)**.** The function $\ell\colon D \to \mathbb{R}$ with $[0,\infty) \subseteq D$ is *quasi-logarithmic over* $\mathbb{R}_+$ if (i) $\lim_{y\to\infty} \frac{\ell(y)}{\log(y)} = 1$, (ii) $\ell'_+(0)$ exists, and (iii) there exists $c > 0$ such that $\ell(\cdot)$ is strictly increasing and continuously differentiable on (c,∞).

For transformations such as $\operatorname{arsinh}(y)$ that apply over $\mathbb{R}$, we add analogous conditions on the limiting behavior in the negative domain (using sgn to denote the sign function):

Definition 3 (Quasi-logarithmic over $\mathbb{R}$)**.** The function $\ell\colon \mathbb{R} \to \mathbb{R}$ is *quasi-logarithmic over* $\mathbb{R}$ if (i) $\lim_{y\to\pm\infty} \frac{\ell(y)}{\operatorname{sgn}(y)\log(|y|)} = 1$, (ii) $\ell'_+(0)$ and $\ell'_-(0)$ exist, and (iii) there exists $c, d > 0$ such that $\ell(\cdot)$ is strictly increasing and continuously differentiable on $(-\infty, d) \cup (c,\infty)$.

We distinguish between these concepts to derive results that apply for both weakly positive outcome variables as well as outcome variables with negative values.[17]

5.2 Quasi-logarithmic transformations of outcomes

The following general result, summarized in Table 3, describes how quasi-logarithmic transformations can lead to extreme levels of scale dependence. The result applies not only to weakly positive outcomes but also to negative-valued outcomes.

[16] Ravallion (2017) uses the term "log-like" to refer to this property, which we avoid due to potential confusion with other properties that share the same name; for example, see Arai (2002); Pollack (2020).

[17] While the definition of quasi-logarithmic over $\mathbb{R}$ rules out concave increasing transformations (e.g., the proposal by Ravallion 2017), Corollary 3 shows that such transformations also exhibit scale dependence. In fact, because these transformations necessarily approach $-\infty$ faster than quasi-logarithmic transformations, an extension of the proof of Proposition 4 can show that they exhibit even more severe scale dependence.

Proposition 4 (Scale dependence of quasi-logarithmic transformations of outcomes)**.** *Consider either a transformation $\ell\colon \mathbb{R} \to \mathbb{R}$ that is quasi-logarithmic over $\mathbb{R}$ or a transformation $\ell\colon [0,\infty) \to \mathbb{R}$ that is quasi-logarithmic over $\mathbb{R}_+$. For $I \in \{c,u,s,L\}$, let $\hat{\beta}_I$, $\hat{\xi}_I$, and $t_{\hat{\beta}_I} = \frac{\hat{\beta}_I}{\widehat{s.e.}(\hat{\beta}_I)}$ denote the coefficient estimate, semi-elasticity estimate, and t-statistic, respectively, corresponding to x from an OLS regression of*

$$y_I = \begin{cases} \ell(cy) & \text{if } I = c \\ y & \text{if } I = u \\ \operatorname{sgn}(y) & \text{if } I = s \\ \operatorname{sgn}(y)\log\big(|y| + \mathbb{1}_{\{y=0\}}\big) + \ell(0)\mathbb{1}_{\{y=0\}} & \text{if } I = L. \end{cases}$$

on x and Z (where $y_L = 0$ when $y = 0$ by convention).

1. *(Large units) As $c \to 0$, we have $\hat{\beta}_c \approx c\ell'(0)\hat{\beta}_u$ and thus*

$$(\hat{\beta}_c, \hat{\xi}_c, t_{\hat{\beta}_c}) \to (0, \hat{\xi}_u, t_{\hat{\beta}_u}).$$

2. *(Small units) As $c \to \infty$, we have $\hat{\beta}_c \approx \hat{\beta}_L + \log(c)\hat{\beta}_s$ and thus*

$$(\hat{\beta}_c, \hat{\xi}_c, t_{\hat{\beta}_c}) \to \begin{cases} (\pm\infty, \pm\infty, t_{\hat{\beta}_s}) & \text{if } \hat{\beta}_s \neq 0 \\ (\hat{\beta}_L, \hat{\xi}_L, 0) & \text{if } \hat{\beta}_s = 0 \text{ and } \operatorname{sgn}(y) \text{ is not constant} \\ (\hat{\beta}_L, \hat{\xi}_L, t_{\hat{\beta}_L}) & \text{if } \operatorname{sgn}(y) \text{ is constant} \end{cases}$$

To build intuition for the proof in Appendix C.1, we discuss the simple case of a weakly positive dependent variable with a single independent variable. In this case, the OLS coefficient takes the form of a ratio between the covariance of the independent variable x_i and the transformed outcome $\ell(cy_i)$, and the variance of x_i. Smaller values of c correspond to larger choices of units (e.g., millions of dollars as opposed to dollars).

As $c \to 0$, the transformed outcome approaches a constant $\ell(0)$, resulting in a covariance, and hence a coefficient, of zero. Computing the semi-elasticity requires dividing by the derivative of the transformation, $c\ell'(cy_0)$, which approaches zero at the same rate as $\ell(cy)$, thereby giving the same result as an untransformed linear regression.

As $c \to \infty$, the transformed outcome approaches $\operatorname{sgn}(y)\log(c|y|)$ for $y \neq 0$ and $\ell(0)$ for $y = 0$. In other words, $\ell(cy) \approx L(cy) = L(y) + \log(c)\operatorname{sgn}(y)$, and hence $\hat{\beta}_c \approx \hat{\beta}_L + \log(c)\hat{\beta}_s$. The result thus depends on whether y is strictly positive. If $y_i > 0$ for all i, then the coefficient estimate approaches that of a log-linear regression since $\ell(\cdot)$ is asymptotically equivalent to $\log(\cdot)$. If $y_i = 0$ for some i, then (the absolute value of) the contribution of zero-valued observations to the covariance can approach infinity. This occurs because ℓ maps zero-valued observations into a fixed value, regardless of the measurement units, but maps non-zero values into arbitrarily large quantities. Whenever the conditional mean of x differs between zero-valued and non-zero-valued outcomes (i.e., $\hat{\beta}_s \neq 0$), the difference between zero and non-zero values of y dominates the covariance term. In this case, increasing the scaling factor leads to arbitrarily large (absolute) values of the regression coefficient and implied semi-elasticity. On the other hand, if $\mathbb{E}[x \mid y = 0] = \mathbb{E}[x]$ (i.e., $\hat{\beta}_s = 0$), then the contribution of zero-valued outcomes to the covariance becomes negligible.

We expound upon this result by offering three lessons for data analysts. First, we note that the problems of using quasi-logarithmic transformations persist when $\hat{\beta}_s \to 0$. Second, we illustrate that the problems may arise even if the fraction of zero-valued observations is arbitrarily small. Third, we discuss conditions under which estimates of $\hat{\beta}_c$ change sign as c varies.

5.2.1 Extensive margin response

The intuition above suggests an important role for $\hat{\beta}_s$ in determining the limiting behavior of $(\hat{\beta}_c, \hat{\xi}_c, t_{\hat{\beta}_c})$ as $c \to \infty$.

Remark (Interpretation of $\hat{\beta}_s$). The term $\hat{\beta}_s$ captures the relationship between x and $\text{sgn}(y)$, accounting for any control variables Z. In the special case of weakly positive outcomes, $\hat{\beta}_s$ captures the relationship between x and $\mathbb{1}_{\{y>0\}}$, accounting for controls. In the special case of a binary x variable (i.e., a treatment), $\hat{\beta}_s$ captures the effect of the treatment on $\text{sgn}(y)$, accounting for controls. As a result, $\hat{\beta}_s$ captures the extensive margin response of the treatment in the special case of weakly positive outcomes with a binary treatment and no controls.

The proof of Proposition 4 involves showing that $\hat{\beta}_c \approx \hat{\beta}_L + \hat{\beta}_s \log(c)$ as $c \to \infty$. This extends beyond the scope of related results that highlight the problem of arbitrary treatment effect estimates when $\hat{\beta}_s \neq 0$ and $y \geq 0$ (Mullahy and Norton, 2022) or when a binary treatment affects the extensive margin of a weakly positive outcome (Chen and Roth, 2023). However, we note below that having a smaller, or even zero, extensive margin response does not result in meaningful estimates.

Corollary 4 (Limiting results when $c \to \infty$ for small $\hat{\beta}_s$)**.** *Suppose $\hat{\beta}_s = 0$ (e.g., no extensive margin response of a treatment). As $c \to \infty$, the coefficient estimate $\hat{\beta}_c$ approaches that of a* *regression with a non-monotonically transformed outcome variable given by*

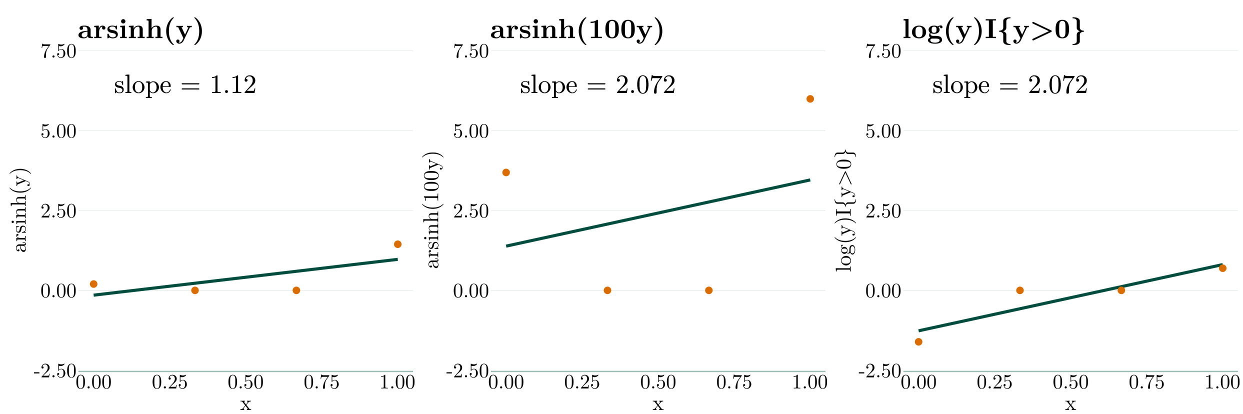
$$L(y) = \begin{cases} \log(y) & y > 0 \\ \ell(0) & y = 0. \end{cases}$$

Furthermore, if $\operatorname{sgn}(y)$ *is not constant, the* t*-statistic* $t_{\hat{\beta}_c}$ *approaches zero, or equivalently the confidence interval for* $\hat{\beta}_c$ *becomes arbitrarily wide.*

Similarly, if $\hat{\beta}_s$ *is near zero, the* t*-statistic is also near zero.*

Thus, even if a treatment has little or no extensive margin response, quasi-logarithmic transformations provide unreliable estimates. For any $\hat{\beta}_s \neq 0$, no matter how small, $\left|\hat{\beta}_c\right|$ diverges as $c \to \infty$. Moreover, $\hat{\beta}_c$ has no meaningful interpretation even if $\hat{\beta}_s = 0$ since $L(y)$ transforms the data non-monotonically, and when the outcomes do not all have the same sign (which necessitates an alternative to the logarithmic transformation), the wide confidence interval that arises renders the estimate uninformative at best.

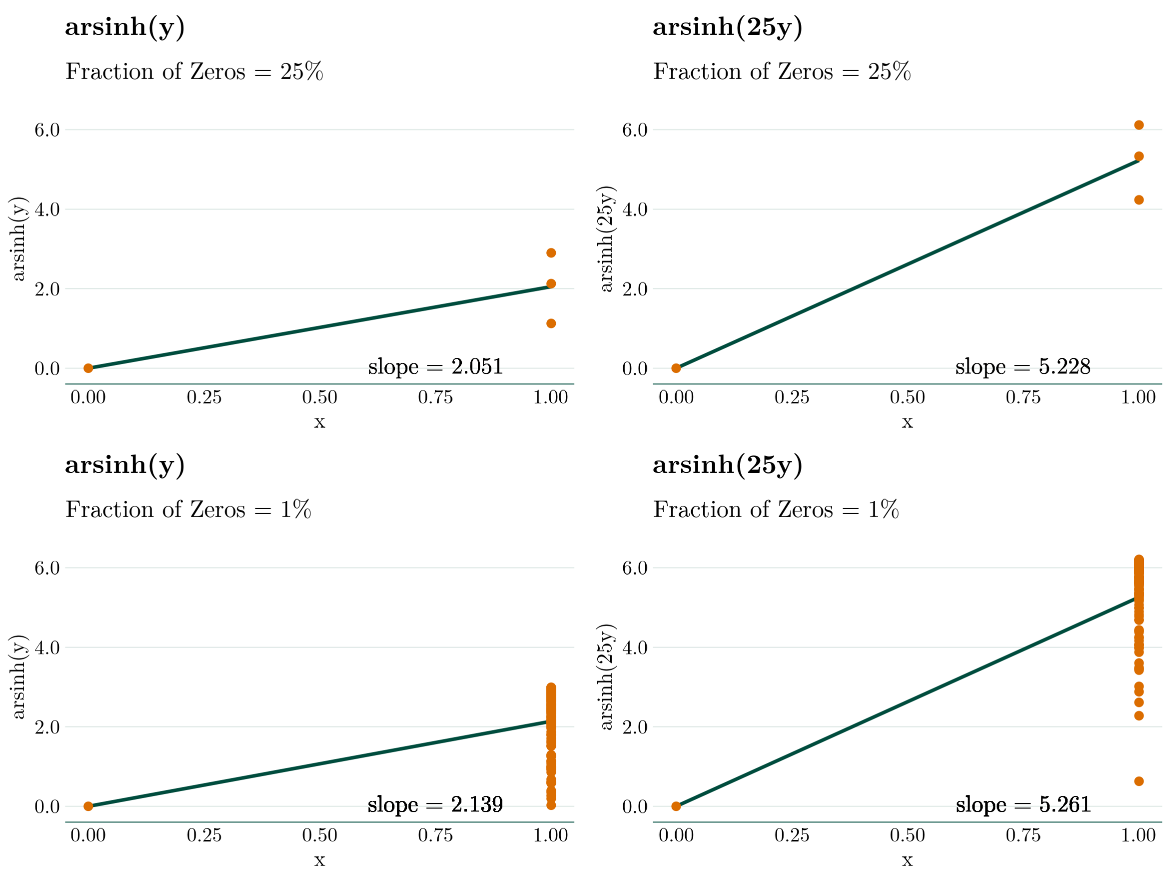
The importance of $\hat{\beta}_s$ clarifies that the distribution of x across $y = 0$ and $y > 0$ (and $y < 0$) values determines the extent of scale dependence. The four-point example below with $\{(0, 1/5), (1/3, 0), (2/3, 0), (1, 2)\}$ illustrates this. The two points with $y = 0$ share the same average x as the two points with $y > 0$. Even when the points with $y > 0$ get scaled up by using smaller units, the presence of the points with $y = 0$ does not cause instability. Nevertheless, the results do not become any more meaningful in this case. The t-statistic approaches 0, as Corollary 4 points out, leading to a completely uninformative confidence interval, and the coefficient estimate limits toward the slope that arises from using $L(y)$ as a transformation, which provides little practical value (despite $L(y)$ exhibiting scale equivariance in this case) since $L(y)$ transforms the data non-monotonically.



5.2.2 Fraction of zeros

While the presence of zero-valued outcome data plays a vital role in shaping scale dependence, we emphasize that the fraction of observations with $y = 0$ does not necessarily provide any indication regarding the extent of scale dependence. This contrasts with the rule of thumb by Bellemare and Wichman (2020) of having at most 30 percent zero-valued outcome data when using the IHS transformation.

Consider the example below, in which only one single data point at $(0, 0)$ has a zero-valued outcome. The remaining points take the form $(1, y_i)$, where y_is are drawn uniformly from 0.01 to 10. Regardless of how many data points have a strictly positive y value, the severity of scale dependence of $\hat{\beta}_c$ remains unchanged since only the mean y-value at $x = 0$ and $x = 1$ determine the slope of the best-fit line. Thus, even as the fraction of zero-valued outcomes approaches 0, the changes in the coefficient or semi-elasticity estimates from simply rescaling the data remain just as large.



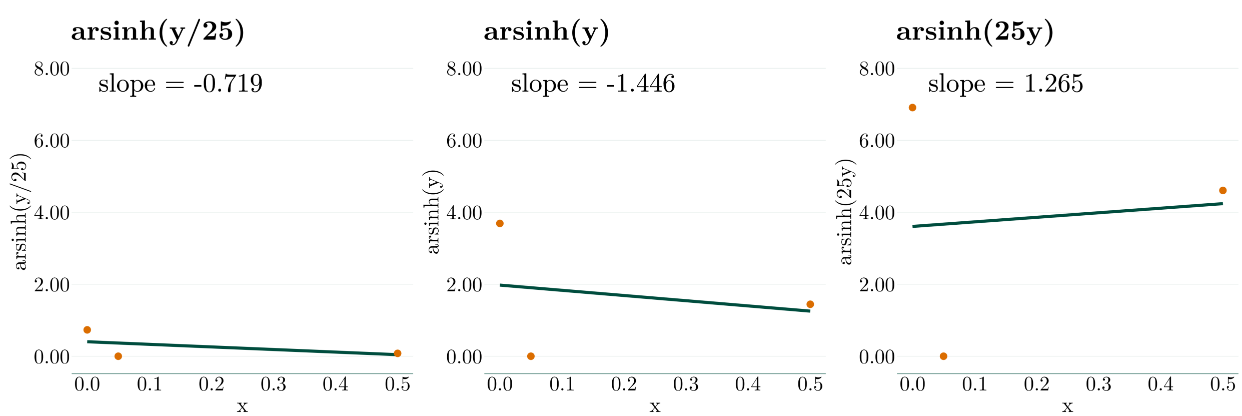
5.2.3 Sign reversals

Inspecting the limiting cases, $c \to 0$ and $c \to \infty$, leads to the following sufficient condition for changes in the sign of $\widehat{\beta}_c$.

Corollary 5. *If $\widehat{\beta}_u \widehat{\beta}_s < 0$, then different choices of units can change the sign of the coefficient and semi-elasticity estimates.*

This result simply states that a sign reversal occurs if the limiting case of large units (which leads to the same sign as an untransformed regression) and the limiting case of small units (which leads to the same sign as a regression with $\text{sgn}(y)$ as the outcome variable) have different signs.

However, sign flips remain a possibility even if $\widehat{\beta}_u$ and $\widehat{\beta}_s$ have the same sign because the estimates $\widehat{\beta}_c$ can vary non-monotonically with the scale factor c. We highlight this using a three-point example, $\{(0.05, 0), (0.5, 2), (0, 20)\}$, depicted below. The best-fit line through $(x, \text{arsinh}(y))$, corresponding to $\widehat{\beta}_c$ at $c = 1$, has a negative slope, while the slope of $\ell(cy)$ on x clearly approaches zero as $c \to 0$. To see the non-monotonicity, note that $\widehat{\beta}_c$ becomes positive and increases without bound as $c \to \infty$: The transformation ℓ always maps the data point with $y = 0$ to the same value regardless of the scale factor, whereas the two data points with $y > 0$ move relatively closer to each other and further from the zero-valued data point, mapping to increasingly larger values as c increases.



5.3 Quasi-logarithmic transformations of covariates

Applying the intuition from the previous section suggests that using quasi-logarithmic transformations on covariates instead of outcome variables would lead to similar forms of scale dependence. This finding holds importance considering the papers in economics that apply the IHS transformation to an independent variable (see Appendix Table 2 for a list of selected publications) and challenges the applicability of the guidance from Bellemare and

Wichman (2020) that "applying the inverse hyperbolic sine transformation to an explanatory variable of interest appears somewhat harmless."

Proposition 5 (Scale dependence of quasi-logarithmic transformations of covariates)**.** *Consider either a transformation $\ell\colon \mathbb{R} \to \mathbb{R}$ that is quasi-logarithmic over $\mathbb{R}$ or a transformation $\ell\colon [0,\infty) \to \mathbb{R}$ that is quasi-logarithmic over $\mathbb{R}_+$. For $I \in \{c,u,s,L\}$, let $\hat{\beta}_I$, $\hat{\xi}_I$, and $t_{\hat{\beta}_I} = \frac{\hat{\beta}_I}{\widehat{s.e.}(\hat{\beta}_I)}$ denote the coefficient estimate, semi-elasticity estimate, and t-statistic, respectively, corresponding to x_I from an OLS regression of y on*

$$x_I = \begin{cases} \ell(cx) & \text{if } I = c \\ x & \text{if } I = u \\ \operatorname{sgn}(x) & \text{if } I = s \\ \operatorname{sgn}(x)\log\big(|x| + \mathbb{1}_{\{x=0\}}\big) + \ell(0)\mathbb{1}_{\{x=0\}} & \text{if } I = L. \end{cases}$$

and Z (where $x_L = 0$ when $x = 0$ by convention).

1. *(Large units) As $c \to 0$,*

$$(\hat{\beta}_c, \hat{\xi}_c, t_{\hat{\beta}_c}) \to (\pm\infty, \hat{\xi}_u, t_{\hat{\beta}_u}).$$

2. *(Small units) As $c \to \infty$,*

$$(\hat{\beta}_c, \hat{\xi}_c, t_{\hat{\beta}_c}) \to \begin{cases} (0, 0, t_{\hat{\beta}_s}) & \text{if } \operatorname{sgn}(y) \text{ is not constant} \\ (\hat{\beta}_L, \hat{\xi}_L, t_{\hat{\beta}_L}) & \text{if } \operatorname{sgn}(y) \text{ is constant} \end{cases}$$

These results, summarized in Table 4, mirror the conclusions of Proposition 4, and the proof in Appendix C.2 follows similar logic. The case of large units for an independent variable leads to arbitrarily large rather than arbitrarily small coefficient estimates, but the semi-elasticity and t-statistic still approach the results of an untransformed linear regression. Similarly, with small units for an independent variable, the coefficient and semi-elasticity estimates approach zero rather than diverge, but the t-statistic still approaches the result of replacing the quasi-logarithmic transformation with the sgn function when this function is non-constant (which is precisely when alternatives to the logarithmic transformation are necessary).

6 Practical recommendations

This section discusses practical considerations for using data transformations effectively. We then compare the parameters that dependent-variable transformations allow researchers to estimate with the parameters that result from alternative approaches such as Poisson regressions or quantile regressions. We also discuss methods that apply when the data contain zero-valued observations that the research question does not interpret as "true zeros." We illustrate these points using data on the relationship between earnings and education. We conclude this section with a discussion of causality and interpretation.

6.1 Using power transformations

Using a range of transformations such as the powers 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, and $\frac{1}{5}$ provides a useful strategy for assessing sensitivity to departures from linearity and reducing the influence of outliers in right-skewed data (with the identity, cubic root, and quintic root applying for negative values). Box and Cox (1964), despite proposing a framework for estimating the power λ in their proposed family of transformations, point out that most often the primary focus lies in studying the effect of the explanatory variable on the outcome. In such cases, aligning with common intuitions regarding applied economic research, they argue, "we shall need to fix one, or possibly a small number, of λ's and go ahead with the detailed estimation and interpretation of the factor effects on this particular transformed scale. We shall choose λ partly in the light of the information provided by the data and partly from general considerations of simplicity, ease of interpretation, etc."

Various recent publications use power transformations to assess robustness of regression results to powers ranging from $\frac{1}{3}$ to $\frac{1}{5}$ (see Appendix Table 1), often noting the similarity between these transformations and the logarithmic transformation. These powers provide distinct ways of reducing the influence of outliers in right-skewed data while demonstrating robustness to functional form misspecification, without giving as much influence to values near zero as the logarithmic transformation does. Furthermore, as noted earlier, the cubic root and quintic root also accommodate negative-valued data.

In contrast to the constant semi-elasticity imposed by the logarithmic transformation, the power transformation y^k results in a semi-elasticity estimate of $\frac{\hat{\beta}}{k(X_0\hat{\beta})}$ at any X_0, which may provide a better fit for data in which proportional changes in the outcome require increasingly larger changes in x_j. This is the case, for instance, when a constant elasticity rather than semi-elasticity characterizes the underlying relationship.

Researchers may also choose to carry out statistical tests to confirm the suitability of a given transformation for specific purposes. To gauge the validity of a linear relationship between the transformed outcome variable and the independent variables, one can use the Ramsey (1969) Regression Equation Specification Error Test (RESET) or the Andrews (1971) test.[18] Additionally, researchers applying transformations to satisfy normality or homoskedasticity assumptions, respectively, may refer to the tests developed by Shapiro and Wilk (1965) or Breusch and Pagan (1979); however, as Davidson and MacKinnon (1993) note, "in any given case there may exist no transformation of the dependent variable that yields symmetric and homoskedastic residuals." These tests notwithstanding, graphical analyses of the data can provide useful diagnostics of model specifications as well as influential observations and outliers. Our package `tregs` provides these tools for carrying out regressions with transformed outcome variables.[19]

6.2 Using other moments of the data for transformations

Given the challenges that can arise when transforming unit-based data, we now turn to the question of whether manipulating the data to remove the units offers a viable alternative to using scale-equivariant transformations. Using a moment of the data to transform individual observations, such as using $\log(\frac{y}{\bar{y}} + 1)$ or $\log(y + y_q)$ where y_q is a quantile of the data, provides a way of removing units.

At first glance, using moments of the data to adjust observations to take on positive values may appear to resolve unit dependency since scaling all observations by a factor c results in the moments of the data being scaled accordingly. However, several problems arise when using such transformations. First, these transformations muddle the interpretation of coefficient estimates, which no longer map cleanly to semi-elasticity estimates and predicted values of the conditional mean of y.[20] Second, the absence of a principled way of choosing a data moment can lead to arbitrary changes in estimates, following the logic of the results in the previous sections. Third, even with a fixed choice of a moment-dependent transformation, computing

[18] The RESET follows from the intuition that, in a correctly specified model, additionally controlling for nonlinear functions of explanatory variables should not significantly affect the outcome. Thus, one can conduct an F-test of the null hypothesis that powers of the predicted values of the (transformed) dependent variable have a null effect on the outcome. In the context of power transformations, Andrews (1971) proposes a closely related approach leading to an exact test of the null hypothesis that the power equals a particular value.

[19] The package computes the coefficients and semi-elasticities in power-transformed regressions and provides various post-estimation tests as discussed above. The package additionally calculates the predicted value of the untransformed outcome, taking into account the error term structure using the non-parametric approach of Duan (1983) as discussed in Appendix A, and provides a visualization of the relationship in the data, both with the transformed data and with the untransformed data.

[20] Computing the semi-elasticity requires solving for $\frac{\partial y}{\partial x_j}$, which depends on how the moments (e.g., y_q or $\bar{y}$) vary with x_j. For example, using the transformation $\log(y + y_q)$ gives the relationship $\beta = \frac{1}{y+y_q}\left(\frac{\partial y}{\partial x_j} + \frac{\partial y_q}{\partial x_j}\right)$. The $y + y_q$ component creates similar challenges in isolating the conditional mean of y.

standard errors requires taking into account sampling variation in the data moment, resulting in t-statistics that violate scale invariance.

Similar considerations arise when analyzing unitless data, such as counts (e.g., number of fatalities or bankruptcies). First, choosing to present counts in tens or millions skews estimates in the direction of the "large units" case of scale dependence. Second, using an arbitrary scale factor is analogous to using an arbitrary additive shifter since $\log(cy+1) = \log(c)+\log(y+\frac{1}{c})$ for $c > 0$; thus, the consequences of lacking a principled way to choose among quasi-logarithmic transformations with different additive shifters mirrors those outlined in Proposition 4. While $\text{arsinh}(y)$ may superficially appear to be a more disciplined transformation, it also involves an arbitrary additive shifter and only differs from $\log(y+1)$ by at most $\log(2)$. Finally, count data may exhibit specific error term structures (e.g., if the variance of $\mathbb{E}[y \mid x]$ is proportional to its mean as in a Poisson distribution with overdispersion) that make generalized linear models a more appropriate choice.

6.3 Using generalized linear models such as Poisson regressions

Generalized linear models (GLMs) provide an alternative way to apply transformations in regression analysis that additionally requires specifying the structure of the errors. Rather than directly applying a transformation to the dependent variable y, a GLM involves applying the transformation to the conditional expectation of the outcome $\mathbb{E}[y \mid X]$. In other words, GLMs explicitly express the transformed conditional expectation as a linear function of X, i.e., $f(\mathbb{E}[y \mid X]) = X\beta$, where f is called the *link function*, instead of $\mathbb{E}[f(y) \mid X] = X\beta$.[21] Furthermore, a GLM requires imposing a restriction on the conditional variance, which relates to the assumed distribution of y. For example, the Poisson regression, used for analyzing count data assumed to follow a Poisson distribution, specifies $\text{Var}[y \mid X]$ as equal to $\mathbb{E}[y \mid X]$, and the more general quasi-Poisson regression specifies $\text{Var}[y \mid X]$ as proportional to $\mathbb{E}[y \mid X]$.[22] The Poisson regression with a log link function provides a common example of

[21]These models impose distinct structures on the error term compared to their OLS counterparts that involve transforming the dependent variable, and thus neither regression framework nests the other. For example, consider the log-link GLM given by $y = \exp(X\beta) + \nu$. Since $\log(y) = X\beta + \log\left(\frac{y}{\exp(X\beta)}\right)$, this model implies $\log(y) = X\beta + \log\left(1 + \frac{\nu}{X\beta}\right)$. Since $\log\left(1 + \frac{\nu}{X\beta}\right)$ generally depends on X, this can differ substantially from the log-linear OLS regression $\log(y) = X\beta + \varepsilon$. See Santos Silva and Tenreyro (2006) for further discussion of this point in the trade literature.

[22]Although the Poisson distribution describes discrete data, GLMs in general and quasi-Poisson regressions in particular can accommodate continuous data. The Poisson likelihood function depends only on the conditional mean of the outcome, which equals the conditional variance of the outcome. The quasi-Poisson (also called pseudo-Poisson) regression assumes proportionality between the conditional mean and conditional variance without further specifying the distribution, resulting in the same quasi-maximum likelihood estimation (QMLE) problem, also referred to as the Poisson Pseudo-Maximum Likelihood (PPML) estimator. The quasi-likelihood behaves like a log-likelihood function but does not correspond to a particular probability

a GLM, though Poisson regressions may use other link functions as well.[23]

Using GLMs requires careful consideration of issues related to modeling assumptions and interpretation of parameter estimates. First, assumptions about the error variance change the GLM point estimates (not just the confidence intervals), even for a given link function such as log, as the example in Section 6.6 shows.[24] Second, GLMs result in parameters that have a different interpretation from their transformed OLS counterparts. In a log-transformed OLS regression, for instance, the coefficient estimate β_j has a structural interpretation as a constant percent change in outcomes in response to a unit change in x_j—i.e., a constant semi-elasticity $\frac{1}{y}\frac{\partial y}{\partial x_j}$. In a log-link GLM, on the other hand, the corresponding parameter estimate does not have this structural interpretation, even if the model is correctly specified. The log-link GLM instead results in a constant value for $\frac{1}{\mathbb{E}[y\,|\,X]}\frac{\partial \mathbb{E}[y\,|\,X]}{\partial x_j}$, a different definition of semi-elasticity, which may have the opposite sign from the structural relationship between y and X.[25]

6.4 Using quantile regressions

Quantile regressions model the conditional quantile of the outcome variable as a linear function of predictors. As this object differs from the conditional mean, the preferability of this method ultimately depends on the underlying model or parameter of interest. Analyzing conditional quantiles often complements rather than substitutes analysis of the conditional mean. For instance, Lemieux (2006) estimates a model of human capital with heterogeneous returns based on conditional means but also provides descriptive evidence from quantile regressions.

While quantile regressions can certainly provide useful additional information, we note three potential concerns that may arise when working with quantile regressions (QRs). First, as Angrist et al. (2002) note, "near-discreteness causes QR estimates to behave poorly and invalidates standard asymptotic theory for QR, since a regularity assumption for QR is continuity of the dependent variable." Second, as Hausman (2001) notes, classical measurement error in the outcome variable biases quantile regression estimates toward

distribution. See Wooldridge (2010) for a textbook treatment of these issues.

[23]Common choices for the link function include $\log(\mathbb{E}[y\,|\,X])$ and $(\mathbb{E}[y\,|\,X])^p$. Many of the insights from our investigation of properties of transformations in OLS also apply in the case of GLMs; Thakral and Tô (2023) characterize restrictions that scale-equivariance properties impose on the link function and the variance function in GLMs.

[24]This relates to the mean-variance confound that Drake et al. (2022) discuss.

[25]To illustrate this point, consider the following simple example based on the structural model of Lemieux (2006), which features an additive heteroskeastic measurement error component in the relationship between log wages and education. Suppose $\log(y_i) = \frac{x_i}{5} + \epsilon(x_i)$, where $\epsilon(x_i) \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, 2x_i)$, specifies the true data-generating process. In 10,000 simulations with 1,000 data points each, a log-link quasi-Poisson regression of y on x results in highly inconsistent estimates, achieving the wrong sign in over one-third of cases.

the median regression estimate. Finally, we emphasize that quantile regressions do not circumvent the scale-dependence issues we document that arise when using quasi-logarithimic transformations.[26]

6.5 Using alternative approaches to account for zeros

A variety of methods make progress by imposing assumptions about the interpretation of zero-valued data. This includes two-part models (Cragg, 1971), selection models (Heckman, 1979), and bounding approaches (Horowitz and Manski, 2000; Lee, 2009). Two-part models consist of a binary choice component to determine the probability of observing a positive outcome (as opposed to zero), followed by a regression model (such as OLS or GLM) conditional on positive outcomes. Selection models specify a latent variable that represents having a positive outcome observed. Bounding offers an alternative solution to sample selection problems by focusing on intensive margin effects and analyzing "worst-case scenarios" or assuming monotonicity in the effect of treatment assignment on sample selection.

Two-part models provide a way of analyzing data that contain "true zeros," representing genuine instances of zero values in the outcome variable. However, as Angrist (2001) shows, only the first part (extensive margin) yields a causal interpretation, while the second part (intensive margin) may reflect selection bias. Power transformations offer an alternative approach that accommodates true zeros while preserving the ability to establish causal interpretations.[27]

In selection models, on the other hand, zeros conventionally represent censored values of a latent outcome rather than the outcome process itself. These models focus on estimating causal impacts on the latent outcome, relying on an exclusion restriction for identification. Selection approaches thus face limitations when considering outcome variables that contain true zeros or negative values, such as firm profit, categories of expenditure, trade flows, or even income.

Bounding approaches shift the emphasis to estimating a range of plausible impacts on outcomes for types with non-zero outcomes. Lee (2009) motivates the focus on such intensive margin effects by the premise that zero values reflect sample selection. Ultimately, however, this approach offers a way of estimating a range of effects applicable to any particular subgroup, without requiring that non-membership in that subgroup censors outcomes. For

[26]See Campos et al. (2017) for an example in which a quantile regression uses the inverse hyperbolic sine of profit as the dependent variable.

[27]Furthermore, the choice of a power when using power transformations provides a way of specifying how to separately value changes along the intensive vs. extensive margins. When extensive margin responses involve discrete changes in outcomes, a power of k corresponds to valuing an initial marginal increase along the intensive margin k times as much as an initial unit increase along the extensive margin.

instance, one could use Lee bounds to obtain a range of estimates of the effect of a subsidy on profits among non-attriting firms or among profitable firms. The former constitutes a situation in which zeros may represent sample selection, while zeros in the latter may reflect true zeros. In either case, bounding provides a useful option for obtaining estimates that apply to the relevant subgroup. As the value of bounding lies in its ability to deliver a range of estimates for specific subgroups of the population, the decision of using bounding remains separate from that of using data transformations. In some cases, such as when analyzing highly right-skewed data, combining the use of bounding approaches with transformed outcome variables may prove valuable.

6.6 Example: Relationship between earnings and education

This section illustrates some of our suggestions regarding data transformations using data from the Current Population Survey (CPS) to analyze the relationship between earnings and years of education in 1991, the most recent year for which the CPS data do not coarsen years of education. The primary sample consists of respondents in the labor force and not enrolled as students. We additionally report results for two subsamples: those who at least attempted grade 9, and nonwhite males. We use weekly earnings as the outcome variable of interest, filling in missing values with zeros. In all specifications, we include fixed effects for age, and we report marginal effects at 39 years of age and 13 years of education, approximately the sample averages.

Power transformations

Due to the presence of zero-valued data, the logarithmic transformation does not apply, so we estimate effects under power transformations ranging from the identity transformation to the fifth root. In general, the power-transformed regression $y^k = X\beta + \varepsilon$ results in a straightforward formula for the semi-elasticity or implied percent change in untransformed outcomes at $X = X_0$: $\frac{\hat{\beta}}{k(X_0\hat{\beta})}$. Within each sample, we obtain relatively stable semi-elasticity estimates as the transformation varies (Table 5 Panel A).

The untransformed linear regression (row 1) yields a coefficient estimate equivalent to 35.89 dollars per week for the primary sample. At 13 years of education and 39 years of age, the regression model predicts earnings equivalent to 392.85 dollars per week. Taking the ratio between the coefficient and the predicted value, the model implies a $\frac{35.89}{392.85} = 9.14$ percent increase in earnings in response to a unit increase in years of education. Using any of the earnings measures as the outcome variable would lead to the same semi-elasticity estimate of 0.0914 (e.g., a coefficient of 1,866.18 dollars per year relative to a predicted value of

20,428.38 dollars per year).

The square root transformation (row 2) yields a coefficient estimate equivalent to 0.7890 $\sqrt{\text{dollars per week}}$ for the primary sample. At 13 years of education and 39 years of age, the regression model predicts square root earnings equivalent to 16.89 $\sqrt{\text{dollars per week}}$. Taking the ratio between the coefficient and the predicted value, the model implies a $\frac{0.7890}{16.89}$ = 4.67 percent increase in square root earnings in response to a unit increase in years of education. This represents a $\frac{0.7890}{\frac{1}{2}\cdot 16.89} = 2 \cdot 4.67 = 9.34$ percent increase in earnings.[28] Using any of the earnings measures as the outcome variable would lead to the same semi-elasticity estimate of 0.0914 (e.g., a coefficient of 5.6896 $\sqrt{\text{dollars / year}}$ relative to a predicted value of 121.77 $\sqrt{\text{dollars / year}}$).

The logic above for interpreting the estimates generalizes to any power transformation. The coefficient estimates come attached with units that may lack inherent importance, but these estimates simply summarize the model's predicted nonlinear effects on (untransformed) outcomes. These models allow for a non-constant implied percent changes in (untransformed) outcomes, in contrast to logarithmic transformations. We obtain semi-elasticity estimates of 9.14–10.10 percent in the full sample, 11.34–12.07 percent in the at-least-high-school sample, and 8.73–9.29 percent in the nonwhite-male sample, all evaluated at 39 years of age and 13 years of education. We note that the estimates need not vary monotonically with the choice of power.

Log-link GLMs including quasi-Poisson regression

We next estimate models of the form $\log(\mathbb{E}[y \,|\, X]) = X\beta$, or equivalently, $y = \exp(X\beta) + \varepsilon$. The estimates have a straightforward interpretation as a constant semi-elasticity, with the caveat discussed previously that this definition of semi-elasticity does not describe the structural relationship between y and X (which may even have the opposite sign).

Although the log-link GLM appears simple, estimating the model requires choosing an assumption regarding the error variance. Quasi-Poisson regressions, for example, assume proportionality between $\mathrm{Var}[y \,|\, X]$ and $\mathbb{E}[y \,|\, X]$, while quasi-exponential regressions assume proportionality with $(\mathbb{E}[y \,|\, X])^2$ and quasi-Gaussian regressions assume $\mathrm{Var}[y \,|\, X]$ is constant.

[28] The predicted value at 13 years of education corresponds to 16.89^2 dollars per week, and the predicted value at $13 + \Delta$ years of education corresponds to $(16.89 + 0.7890\Delta)^2$ dollars per week. This implies a percent change of

$$\frac{(16.89 + 0.7890\Delta)^2 - 16.89^2}{16.89^2} = \frac{2 \cdot 0.7890}{16.89}\Delta + \frac{0.7890^2}{16.89^2}\Delta^2$$

in response to an increase in years of education of Δ, or equivalently a semi-elasticity of $2 \cdot \frac{0.7890}{16.89} + \frac{0.7890^2}{16.89^2}\Delta$. This approach applies for discrete changes in independent variables, but for continuous variables (or as a first-order approximation) we consider $\Delta \to 0$.

Even with a fixed link function such as log, the specification of the error variance family affects the point estimates in these models. In our data, we find that the semi-elasticity estimates increase by 17–23 percent when changing the error variance assumption from the exponential family to the Gaussian family in all three samples, as Table 5 Panel B shows. We obtain reduced-form semi-elasticity estimates of 9.04–10.61 percent in the full sample, 10.69–10.91 percent in the at-least-high-school sample, and 8.32–10.22 percent in the nonwhite-male sample. The estimates assuming the exponential family (row 3) most closely match the magnitudes of the untransformed specification (Table 5 Panel A).

Lastly, we explore how the magnitudes of the estimates compare between log-link GLMs and the log-transformed OLS regression. For this analysis, to maintain consistent samples, we remove all zero-valued observations from the data. In our data, across all three samples, the quasi-exponential regression estimate aligns most closely with the corresponding log-transformation estimate (Table 5 Panel C). We emphasize once again, however, that these estimates correspond to different parameters. Nevertheless, this example highlights that the quasi-Poisson assumption of proportionality between $\mathrm{Var}[y \mid X]$ and $\mathbb{E}[y \mid X]$ does not necessarily provide the best alternative to using log transformations in OLS.

6.7 On causality and interpretation

Our theoretical results on scale-equivariant and scale-dependent transformations characterize general properties of OLS estimators that apply regardless of whether the estimates have a causal interpretation. Causal interpretations for a regression coefficient in a standard OLS framework arise under the conditional mean independence assumption. Yet causality alone does not guarantee meaningful model conclusions, as the empirical example in Section 4 shows. In that example, stating that the cash transfer experiment *causes* a significant increase in the inverse hyperbolic sine of total entrepreneurial revenue in US dollars and, at the same time, *causes* an insignificant decrease in the inverse hyperbolic sine of total entrepreneurial revenue in Kenyan shillings constitutes a valid interpretation of the results. While the estimated coefficients, standard errors, and conditional mean independence assumption may support the technical correctness of both statements, these interpretations highlight the misleading nature of transformations that violate scale equivariance.

The logic of our results extends more generally to a variety of methods for causal inference. This includes two-stage regression approaches such as instrumental variables estimators and differences-in-differences estimators (e.g., Gardner et al., 2023). Logarithmic and power transformations imply percent changes that satisfy scale-invariance properties, while quasi-logarithimic transformations can lead to severe scale dependence. In addition, our characterization of the family of scale-equivariant transformations in OLS (i.e., models of the form $\mathbb{E}[f(y) \mid X] = X\beta$) also applies to link functions in GLMs (i.e., models of the form $f(\mathbb{E}[y \mid X]) = X\beta$), as Thakral and Tô (2023) show.

7 Conclusion

Data transformations play a fundamental and widespread role in analyzing relationships between variables, estimating causal effects, and testing statistical hypotheses. They serve a useful purpose in developing models that provide better descriptions of the underlying data and in assessing the robustness of conclusions. Yet transformations do not come without potential pitfalls: those that are not grounded in sound statistical principles can introduce distortions and lead to erroneous inferences.

This paper proposes a simple set of criteria for evaluating whether a transformation delivers meaningful conclusions, based on the idea that a regression model applied to the same data with different units should result in unit-independent estimates for unitless quantities of interest such as semi-elasticities and t-statistics. Scale equivariance emerges as an essential property of transformations for deriving meaningful conclusions in linear regressions with unit-based data. Our regression transformation equivalence theorem unequivocally supports the use of logarithmic and power transformations, implying that power transformations provide the only viable option for analyzing data that contain zero or negative values.[29] With such data, the increasingly favored quasi-logarithimic transformations that approach $\log(y)$ at large values, such as $\log(y+1)$ and $\operatorname{arsinh}(y)$, suffer from extreme levels of scale dependence.[30] We show theoretically that using these transformations provides no new information at best and highly misleading conclusions at worst. When expressing outcomes in large measurement units (e.g., millions of people), these transformations result in similar inferences as in an untransformed linear regression. When expressing outcomes in small measurement units (e.g., number of people), these transformations imply arbitrarily large effect sizes or arbitrarily large confidence intervals. While these limiting results raise serious potential concerns about the use of quasi-logarithmic transformations, whether natural changes in units have significant practical implications constitutes an entirely separate question. We address this question using data from randomized controlled experiment with 49 distinct outcome variables. We show empirically that changing units from US dollars to Kenyan shillings changes the sign

[29] We provide a package `tregs` that facilitates conducting regression analysis using power transformations by estimating semi-elasticities, computing predicted values of the untransformed outcome, and conducting specification tests.

[30] Our package `wreckitreg` computes the necessary scaling factor for achieving any specified coefficient estimate or achievable semi-elasticity using these transformations.

and significance of treatment effect estimates for up to 31 percent of outcome variables.

On a broader level, this paper delineates a set of properties to formalize the notion of unit-independent model conclusions based on the concept of scale equivariance. Our approach provides a foundation for future research to explore the extent of scale dependence and criteria for unit-independent model conclusions in a wide range of statistical frameworks.

References

Aihounton, Ghislain BD, and Arne Henningsen. 2021. “Units of measurement and the inverse hyperbolic sine transformation.” *The Econometrics Journal*, 24(2): 334–351. 3

Aitchison, John, and James AC Brown. 1957. *The Lognormal Distribution with Special Reference to the Uses in Economics.* Cambridge University Press. 4

Andrews, DF. 1971. “A note on the selection of data transformations.” *Biometrika*, 58(2): 249–254. 24

Angrist, Joshua D. 2001. “Estimation of limited dependent variable models with dummy endogenous regressors: simple strategies for empirical practice.” *Journal of business & economic statistics*, 19(1): 2–28. 27

Angrist, Joshua, Eric Bettinger, Erik Bloom, Elizabeth King, and Michael Kremer. 2002. “Vouchers for private schooling in Colombia: Evidence from a randomized natural experiment.” *American economic review*, 92(5): 1535–1558. 26

Arai, Toshiyasu. 2002. “On the slowly well orderedness of ε_o.” *Mathematical Logic Quarterly: Mathematical Logic Quarterly*, 48(1): 125–130. 16

Ashraf, Nava, Diego Aycinena, Claudia Martínez A, and Dean Yang. 2015. “Savings in transnational households: a field experiment among migrants from El Salvador.” *Review of Economics and Statistics*, 97(2): 332–351. 4

Bartlett, Maurice S. 1947. “The use of transformations.” *Biometrics*, 3(1): 39–52. 4

Bellégo, Christophe, David Benatia, and Louis Pape. 2022. “Dealing with logs and zeros in regression models.” *arXiv preprint arXiv:2203.11820*. 3

Bellemare, Marc F, and Casey J Wichman. 2020. “Elasticities and the inverse hyperbolic sine transformation.” *Oxford Bulletin of Economics and Statistics*, 82(1): 50–61. 3, 4, 20, 21

Bosquet, Clément, and Hervé Boulhol. 2014. “Applying the GLM variance assumption to overcome the scale-dependence of the negative binomial QGPML estimator.” *Econometric Reviews*, 33(7): 772–784. 3

Box, George EP, and David R Cox. 1964. “An analysis of transformations.” *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2): 211–243. 4, 12, 23

Breusch, Trevor S, and Adrian R Pagan. 1979. “A simple test for heteroscedasticity and random coefficient variation.” *Econometrica: Journal of the econometric society*, 1287–1294. 24

Burbidge, John B, Lonnie Magee, and A Leslie Robb. 1988. “Alternative transformations to handle extreme values of the dependent variable.” *Journal of the American statistical Association*, 83(401): 123–127. 4

Cameron, A Colin, and Pravin K Trivedi. 2013. *Regression analysis of count data.* Vol. 53, Cambridge university press. 3

Campos, Francisco, Michael Frese, Markus Goldstein, Leonardo Iacovone, Hillary C Johnson, David McKenzie, and Mona Mensmann. 2017. "Teaching personal initiative beats traditional training in boosting small business in West Africa." *Science*, 357(6357): 1287–1290. 27

Chen, Jiafeng, and Jonathan Roth. 2023. "Log-like? ATEs defined with zero outcomes are (arbitrarily) scale-dependent." *arXiv preprint arXiv:2212.06080.* 3, 4, 18

Cohn, Jonathan B, Zack Liu, and Malcolm I Wardlaw. 2022. "Count (and count-like) data in finance." *Journal of Financial Economics*, 146(2): 529–551. 13

Cragg, John G. 1971. "Some statistical models for limited dependent variables with application to the demand for durable goods." *Econometrica: journal of the Econometric Society*, 829–844. 27

Dalton, Hugh. 1920. "The measurement of the inequality of incomes." *The Economic Journal*, 30(119): 348–361. 11

Davidson, Russell, and James G MacKinnon. 1993. *Estimation and inference in econometrics.* Vol. 63, Oxford New York. 12, 24

de Brauw, Alan, and Sylvan Herskowitz. 2021. "Income variability, evolving diets, and elasticity estimation of demand for processed foods in Nigeria." *American journal of agricultural economics*, 103(4): 1294–1313. 3, 4

Delius, Antonia, and Olivier Sterck. 2020. "Cash transfers and micro-enterprise performance: Theory and quasi-experimental evidence from Kenya." *Available at SSRN 3591146.* 3, 4

Dolby, James L. 1963. "A Quick Method for Choosing a Transformation." *Technometrics*, 5(3): 317–325. 4

Drake, Marshall, Fernando Payró, Neil Thakral, and Linh T Tô. 2022. "Bayesian adaptive choice experiments." Mimeo. 26

Draper, Norman R, and William G Hunter. 1969. "Transformations: Some examples revisited." *Technometrics*, 11(1): 23–40. 4

Drukker, David M. 2000. "Box-Cox regression models." *Stata Technical Bulletin*, 9(54). 12

Duan, Naihua. 1983. "Smearing estimate: A nonparametric retransformation method." *Journal of the American Statistical Association*, 78(383): 605–610. 24

Flowerdew, Robin, and Murray Aitkin. 1982. "A method of fitting the gravity model based on the Poisson distribution." *Journal of regional science*, 22(2): 191–202. 3

Foster, AM, L Tian, and LJ Wei. 2001. "Estimation for the Box-Cox transformation model without assuming parametric error distribution." *Journal of the American Statistical Association*, 96(455): 1097–1101. 12

Gardner, John, Neil Thakral, Linh Tô, and Luther Yap. 2023. "Two-stage Differences in Differences." *Mimeo.* 30

Haushofer, Johannes, and Jeremy Shapiro. 2016. "The short-term impact of unconditional cash transfers to the poor: experimental evidence from Kenya." *The Quarterly Journal of Economics*, 131(4): 1973–2042. 2, 13, 14, 15

Hausman, Jerry. 2001. "Mismeasured variables in econometric analysis: problems from the right and problems from the left." *Journal of Economic perspectives*, 15(4): 57–67. 26

Heckman, James J. 1979. "Sample selection bias as a specification error." *Econometrica: Journal of the econometric society*, 153–161. 27

He, Xuming, Jana Jurečková, Roger Koenker, and Stephen Portnoy. 1990. "Tail behavior of regression estimators and their breakdown points." *Econometrica: Journal of the Econometric Society*, 1195–1214. 4

Horowitz, Joel L, and Charles F Manski. 2000. "Nonparametric analysis of randomized experiments with missing covariate and outcome data." *Journal of the American statistical Association*, 95(449): 77–84. 27

Hoyle, Mike H. 1973. "Transformations: An introduction and a bibliography." *International Statistical Review/Revue Internationale de Statistique*, 203–223. 4

James, Gareth, Daniela Witten, Trevor Hastie, and Robert Tibshirani. 2013. *An introduction to statistical learning.* Vol. 112, Springer. 5

John, JA, and Norman R Draper. 1980. "An alternative family of transformations." *Journal of the Royal Statistical Society Series C: Applied Statistics*, 29(2): 190–197. 4

Johnson, Norman L. 1949. "Systems of frequency curves generated by methods of translation." *Biometrika*, 36(1/2): 149–176. 4

Kelmansky, Diana M, Elena J Martínez, and Víctor Leiva. 2013. "A new variance stabilizing transformation for gene expression data analysis." *Statistical Applications in Genetics and Molecular Biology*, 12(6): 653–666. 4

Kemp, Gordon CR. 1996. "Scale equivariance and the Box-Cox transformation." *Economics Letters*, 51(1): 1–6. 4

King, Gary. 1988. "Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential Poisson regression model." *American Journal of Political Science*, 838–863. 3

Koenker, Roger, and Gilbert Bassett Jr. 1978. "Regression quantiles." *Econometrica: Journal of the Econometric Society*, 33–50. 4

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *Review of Economic Studies*, 76(3): 1071–1102. 27, 28

Lemieux, Thomas. 2006. "Postsecondary education and increasing wage inequality." *American Economic Review*, 96(2): 195–199. 26

Machado, Jose AF. 1993. "Robust model selection and M-estimation." *Econometric Theory*, 9(3): 478–493. 4

Manly, Bryan FJ. 1976. "Exponential data transformations." *Journal of the Royal Statistical Society: Series D (The Statistician)*, 25(1): 37–42. 4

Moore, PG. 1957. "Transformations to normality using fractional powers of the variable." *Journal of the American Statistical Association*, 52(278): 237–246. 4

Mullahy, John, and Edward C Norton. 2022. "Why Transform Y? A Critical Assessment of Dependent-Variable Transformations in Regression Models for Skewed and Sometimes-Zero Outcomes." National Bureau of Economic Research. 3, 4, 18

Müller, Ulrich K. 2007. "A theory of robust long-run variance estimation." *Journal of Econometrics*, 141(2): 1331–1352. 4

Newey, Whitney K, and James L Powell. 1987. "Asymmetric least squares estimation and testing." *Econometrica: Journal of the Econometric Society*, 819–847. 4

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

Pigou, Arthur Cecil. 1912. *Wealth and welfare.* Macmillan and Company, limited. 11

Pollack, Paul. 2020. "A generalization of the Hardy–Ramanujan inequality and applications." *Journal of Number Theory*, 210: 171–182. 16

Preminger, Arie, and Shinichi Sakata. 2007. "A model selection method for S-estimation." *The Econometrics Journal*, 10(2): 294–319. 4

Ramsey, James Bernard. 1969. "Tests for specification errors in classical linear least-squares regression analysis." *Journal of the Royal Statistical Society: Series B (Methodological)*, 31(2): 350–371. 24

Ravallion, Martin. 2017. "A concave log-like transformation allowing non-positive values." *Economics Letters*, 161: 130–132. 4, 11, 16

Sakata, Shinichi. 2007. "Instrumental variable estimation based on conditional median restriction." *Journal of Econometrics*, 141(2): 350–382. 4

Santos Silva, J. M. C., and Silvana Tenreyo. 2022. "The Log of Gravity at 15." *Portuguese Economic Journal.* 3

Santos Silva, J. M. C., and Silvana Tenreyro. 2006. "The log of gravity." *The Review of Economics and statistics*, 88(4): 641–658. 25

Shapiro, Samuel Sanford, and Martin B Wilk. 1965. "An analysis of variance test for normality (complete samples)." *Biometrika*, 52(3/4): 591–611. 24

Shin, Youngki. 2008. "Semiparametric estimation of the Box–Cox transformation model." *The Econometrics Journal*, 11(3): 517–537. 12

Spitzer, John J. 1984. "Variance estimates in models with the Box-Cox transformation: implications for estimation and hypothesis testing." *The Review of Economics and Statistics*, 645–652. 4, 12

Tabri, Rami Victor. 2014. "Testing for normality in linear regression models using regression and scale equivariant estimators." *Economics Letters*, 122(2): 192–196. 4

Thakral, Neil, and Linh Tô. 2023. "Scale Equivariance in Regressions." *Mimeo.* 3, 4, 8, 26, 31

Thakral, Neil, and Linh T Tô. 2022. "Anticipation and Consumption." *Mimeo.* 13

Thomas, Duncan, Elizabeth Frankenberg, Jed Friedman, Jean-Pierre Habicht, Mohammed Hakimi, Nicholas Ingwersen, Nathan Jones, Christopher McKelvey, Gretel Pelto, Bondan Sikoki, et al. 2006. "Causal effect of health on labor market outcomes: Experimental evidence." 4

Tukey, John W. 1957. "On the comparative anatomy of transformations." *The Annals of Mathematical Statistics*, 602–632. 4

Turner, Malcolm E, Robert J Monroe, and Henry L Lucas. 1961. "Generalized asymptotic regression and non-linear path analysis." *Biometrics*, 17(1): 120–143. 4

Wooldridge, Jeffrey M. 1992. "Some alternatives to the Box-Cox regression model." *International economic review*, 935–955. 4

Wooldridge, Jeffrey M. 2010. *Econometric analysis of cross section and panel data.* MIT press. 26

Yang, Zhenlin. 2006. "A modified family of power transformations." *Economics Letters*, 92(1): 14–19. 4

Yeo, In-Kwon, and Richard A Johnson. 2000. "A new family of power transformations to improve normality or symmetry." *Biometrika*, 87(4): 954–959. 4

Table 1: Scale dependence in RCT data: Binary treatment

	Full sample		No attriters	
	All variables (1)	Endline only (2)	All variables (3)	Endline only (4)
Panel A: Sign flips				
No controls	11–28	11–23	4–11	4–6
Baseline control		4–21		6–12
Village FE	5–25	1–5	4–28	0–6
Baseline control & Village FE		2–7		0–7
Panel B: Significance flips (5 percent level)				
No controls	5–39	4–26	5–35	3–22
Baseline control		4–27		8–31
Village FE	9–29	9–25	5–27	5–23
Baseline control & Village FE		9–25		6–25
Panel C: Sign or significance flips				
No controls	16–48	15–35	9–40	7–27
Baseline control		8–39		14–40
Village FE	14–52	10–30	9–51	5–27
Baseline control & Village FE		11–31		6–30
Number of outcome variables	98	49	98	49

Note: The treatment variable is an indicator for whether a household receives an unconditional cash transfer. Each cell reports the number of times the treatment effect estimate using an IHS transformation changes in sign or statistical significance at the 5 percent level as a result of changing measurement units to the natural alternative (first number) or to an arbitrary alternative (second number). The natural alternative unit refers to Kenyan shillings instead of US dollars for monetary measures, and to µg/dL instead of nmol/L for concentration. Panel A reports the number of outcome variables for which the treatment effect estimate changes from positive to negative, or vice versa. Panel B reports the number of outcome variables for which the treatment effect estimate changes between statistically significant and statistically insignificant, or changes between a statistically significant positive effect and a statistically significant negative effect. Panel C reports the number of outcome variables for which either sign or statistical significance of the treatment effect estimate changes. The first two columns present results estimated on the full sample of households, whereas the last two columns restrict to the subsample of households that responded to the endline survey. Columns (1) and (3) present results using IHS-transformed baseline and endline measures of the 49 outcome variables. Columns (2) and (4) report results for only endline measures. Each of the four rows in each panel corresponds to a different set of control variables: no control variables; controlling for the baseline outcome measure; controlling for village fixed effects; and controlling for both the baseline outcome measure and village fixed effects. The sample consists of 2,880 individuals in the full sample and 2,744 individuals in the non-attriter sample. Standard errors are clustered at the household level.

Table 2: Scale dependence in RCT data: Continuous treatment

	Full sample		No attriters	
	All variables (1)	Endline only (2)	All variables (3)	Endline only (4)
Panel A: Sign flips				
No controls	10–53	5–25	5–37	1–13
Baseline control		1–25		5–17
Village FE	11–36	1–13	8–35	1–13
Baseline control & Village FE		3–14		3–16
Panel B: Significance flips (5 percent level)				
No controls	5–17	3–7	5–20	3–9
Baseline control		1–7		1–10
Village FE	3–13	2–8	3–14	2–9
Baseline control & Village FE		1–7		2–6
Panel C: Sign or significance flips				
No controls	15–65	8–32	10–51	4–22
Baseline control		2–31		6–26
Village FE	14–48	3–21	11–47	3–21
Baseline control & Village FE		4–20		5–21
Number of outcome variables	98	49	98	49

Note: The treatment variable is the number of days between baseline survey and payment date of a cash transfer of roughly USD 404 PPP for households in the small lump-sum cash transfer treatment arm during the nine months after baseline. Each cell reports the number of times the treatment effect estimate using an IHS transformation changes in sign or statistical significance at the 5 percent level as a result of changing measurement units to the natural alternative (first number) or to an arbitrary alternative (second number). The sample consists of 298 individuals in the full sample and 284 individuals in the non-attriter sample. Standard errors are clustered at the household level. See Table 1 for details on the samples and specifications.

Table 3: Scale dependence of quasi-logarithimic transformations for outcome variables

Panel A: Definitions

Notation	Correspond to regression with
$\hat{\beta}_c$, $\hat{\xi}_c$, $t_{\hat{\beta}_c}$	outcome variable $\ell(cy)$
$\hat{\xi}_u$, $t_{\hat{\beta}_u}$	outcome variable y
$\hat{\beta}_s$, $t_{\hat{\beta}_s}$	outcome variable $\mathrm{sgn}(y)$
$\hat{\beta}_L$, $\hat{\xi}_L$, $t_{\hat{\beta}_L}$	outcome variable $L(y)$: $y = 0 \rightarrow L(y) = \ell(0)$ $y \neq 0 \rightarrow L(y) = \mathrm{sgn}(y)\log(\|y\|)$

Panel B: Transforming y with quasi-logarithmic function $\ell(y)$

Estimate	Large units, $c \rightarrow 0$	Small units, $c \rightarrow \infty$	Condition
$\hat{\beta}_c$	$\hat{\beta}_c \rightarrow 0$	$\left\|\hat{\beta}_c\right\| \rightarrow \infty$ $\hat{\beta}_c \rightarrow \hat{\beta}_L$	$\hat{\beta}_s \neq 0$ $\hat{\beta}_s = 0$
$\hat{\xi}_c$	$\hat{\xi}_c \rightarrow \hat{\xi}_u$	$\left\|\hat{\xi}_c\right\| \rightarrow \infty$ $\hat{\xi}_c \rightarrow \hat{\xi}_L$	$\hat{\beta}_s \neq 0$ $\hat{\beta}_s = 0$
$t_{\hat{\beta}_c}$	$t_{\hat{\beta}_c} \rightarrow t_{\hat{\beta}_u}$	$t_{\hat{\beta}_c} \rightarrow t_{\hat{\beta}_s}$ $t_{\hat{\beta}_c} \rightarrow 0$ $t_{\hat{\beta}_c} \rightarrow t_{\hat{\beta}_L}$	$\hat{\beta}_s \neq 0$ $\hat{\beta}_s = 0$ & $\mathrm{sgn}(y)$ is not constant $\hat{\beta}_s = 0$ & $\mathrm{sgn}(y)$ is constant

Panel C: Interpretation of $\hat{\beta}_s = 0$ when $y \geq 0$, no controls

Case	Interpretation
Any x Binary x	$\mathbb{E}[x \mid y = 0] = \mathbb{E}[x \mid y > 0]$ no extensive margin response

Note: The table summarizes the results in Proposition 4, which describe the scale dependence of the coefficient estimate $\hat{\beta}_c$, semi-elasticity estimate $\hat{\xi}_c$, and t-statistic $t_{\hat{\beta}_c}$ from an OLS regression of $\ell(y)$ on x and Z, where $\ell(\cdot)$ is a quasi-logarithmic transformation over $\mathbb{R}_+$ or $\mathbb{R}$.

Table 4: Scale dependence of quasi-logarithimic transformations for covariates

Panel A: Definitions

Notation	Correspond to regression with
$\widehat{\beta}_c$, $\widehat{\xi}_c$, $t_{\widehat{\beta}_c}$	covariate $\ell(cx)$
$\widehat{\xi}_u$, $t_{\widehat{\beta}_u}$	covariate x
$t_{\widehat{\beta}_s}$	covariate $\text{sgn}(x)$
$\widehat{\beta}_L$, $\widehat{\xi}_L$, $t_{\widehat{\beta}_L}$	covariate $L(x)$: $x = 0 \to L(x) = \ell(0)$ $x \neq 0 \to L(x) = \text{sgn}(x)\log(\lvert x\rvert)$

Panel B: Transforming x with quasi-logarithmic function $\ell(x)$

Estimate	Large units, $c \to 0$	Small units, $c \to \infty$	Condition
$\widehat{\beta}_c$	$\lvert\widehat{\beta}_c\rvert \to \infty$	$\widehat{\beta}_c \to 0$ $\widehat{\beta}_c \to \widehat{\beta}_L$	$\text{sgn}(x)$ is not constant $\text{sgn}(x)$ is constant
$\widehat{\xi}_c$	$\widehat{\xi}_c \to \widehat{\xi}_u$	$\widehat{\xi}_c \to 0$ $\widehat{\xi}_c \to \widehat{\xi}_L$	$\text{sgn}(x)$ is not constant $\text{sgn}(x)$ is constant
$t_{\widehat{\beta}_c}$	$t_{\widehat{\beta}_c} \to t_{\widehat{\beta}_u}$	$t_{\widehat{\beta}_c} \to t_{\widehat{\beta}_s}$ $t_{\widehat{\beta}_c} \to t_{\widehat{\beta}_L}$	$\text{sgn}(x)$ is not constant $\text{sgn}(x)$ is constant

Note: The table summarizes the results in Proposition 5, which describe the scale dependence of the coefficient estimate $\widehat{\beta}_c$, semi-elasticity estimate $\widehat{\xi}_c$, and t-statistic $t_{\widehat{\beta}_c}$ from an OLS regression of y on $\ell(x)$ and Z, where $\ell(\cdot)$ is a quasi-logarithmic transformation over $\mathbb{R}_+$ or $\mathbb{R}$.

Table 5: Relationship between earnings and education

	(1)	(2)	(3)
Panel A: Semi-elasticities from power transformations			
Power 1	0.0914	0.1150	0.0887
Power 1/2	0.0935	0.1134	0.0873
Power 1/3	0.0959	0.1152	0.0887
Power 1/4	0.0985	0.1178	0.0907
Power 1/5	0.1010	0.1207	0.0929
Panel B: Semi-elasticities from log-link GLMs			
Log-link quasi-Gaussian (GLM)	0.1061	0.1091	0.1022
Log-link quasi-Poisson (GLM)	0.1007	0.1084	0.0943
Log-link quasi-Exponential (GLM)	0.0904	0.1069	0.0832
Panel C: Semi-elasticities from log-transformed OLS vs. log-link GLM (restricted to data with strictly-positive outcome variable)			
Log transformation (OLS)	0.0898	0.1033	0.0809
Log-link quasi-Gaussian (GLM)	0.1069	0.1103	0.1048
Log-link quasi-Poisson (GLM)	0.0997	0.1076	0.0942
Log-link quasi-Exponential (GLM)	0.0884	0.1037	0.0810

Note: The data come from the Merged Outgoing Rotation Groups of the CPS in 1991 with the sample restrictions described in Section 6.6. The first column corresponds to our primary sample of 211,802 respondents (170,488 with nonzero earnings) in the labor force and not enrolled as students. The second column corresponds to the subsample of 202,982 respondents (163,726 with nonzero earnings) who at least attempted grade 9. The third column corresponds to the subsample of 13,733 nonwhite respondents (10,818 with nonzero earnings). We report marginal effects at 39 years of age and 13 years of education, approximately the sample averages. In Panel A, each row reports semi-elasticity estimates based on the OLS regression $\text{earnings}_i^k = \beta\text{education}_i + \gamma_{a(i)} + \varepsilon_i$, where $\gamma_{a(i)}$ denotes age fixed effects and k corresponds to the specified power. In Panel B, each row reports semi-elasticity estimates based on the GLM $\text{earnings}_i = \exp\Big(\beta\text{education}_i + \gamma_{a(i)}\Big) + \varepsilon_i$ under the specified assumption about the error variance. In Panel C, we compare the GLM estimates with estimates from the OLS regression $\ln\big(\text{earnings}_i\big) = \beta\text{education}_i + \gamma_{a(i)} + \varepsilon_i$ with the GLM estimates, restricting to nonzero earnings to ensure the sample does not differ from that of the log-transformed OLS specification.