Supplemental Appendix to When Are Estimates Independent of Measurement Units?

Neil Thakral*   Linh T. Tô†

March 2025

Abstract

This document contains appendix material for Thakral and Tô (2025*c*).

*Thakral (Brown University): Box B, Providence, RI 02912 (email: neil_thakral@brown.edu).

†Tô (Boston University and Stanford University): Stanford Institute for Economic Policy Research, 366 Galvez St, Stanford, CA 94305 (email: linhto@stanford.edu).

Contents

List of Supplementary Appendix Figures

List of Supplementary Appendix Tables

A Discussion of other approaches in the literature

A.1 Comparison to Chen and Roth (2024) and Mullahy and Norton (2024)

This section discusses our contributions relative to existing studies along three dimensions: theoretical results on measurement-unit independence, theoretical results on measurement-unit dependence, and practical recommendations.

A.1.1 Results on measurement-unit independence

Mullahy and Norton (2024) do not present any results on measurement-unit independence, so this section focuses on comparing with the result in Chen and Roth (2024).

The main result of Chen and Roth (2024) shows that the only treatment effect estimate that is unit-independent, having a percent interpretation, is on the log of the outcome. This result is derived under the assumption of a binary treatment, and does not offer a constructive solution for data with zero- or negative-valued outcomes. Furthermore, it is already evident that logarithmic transformations uniquely give a percent interpretation from the fact that a regression coefficient represents the semi-elasticity (implied percent change in outcomes) if and only if the transformation is logarithmic (since $\beta = \frac{\beta}{yf'(y)}$ if and only if $f(y) = \log(y)$). Most importantly, this particular unit-independence criterion rules out the standard practice of estimating treatment effect in levels, for which marginal effects can be obtained on a consistent scale, and ignores the more fundamental issue of statistical inference.

Table S4 highlights how this result compares to our equivalence theorem. While the special case of the log transformation of the dependent variable directly results in semi-elasticity estimates (which goes hand in hand with the strong assumption of a constant semi-elasticity), coefficient estimates in nonlinear models (e.g., probit, logit, or models with interaction terms) generally lack a useful direct interpretation and need to be converted to semi-elasticities.

A.1.2 Results on measurement-unit dependence

Both Chen and Roth (2024) and Mullahy and Norton (2024) present theoretical results on scale dependence with non-negative dependent variables. Table S5 summarizes how our results compare with theirs.

While the results in Chen and Roth (2024) only pertain to coefficient estimates, Mullahy and Norton (2024) share our emphasis on the importance of interpreting the estimates on a consistent scale. Thus, while Chen and Roth (2024) shows that treatment effect estimates converge to zero when the scale factor approaches zero, Mullahy and Norton (2024) conclude that they approach the result from an untransformed linear regression when appropriately rescaled.

The theoretical results in Chen and Roth (2024) impose an assumption on the relationship between the independent variable and the dependent variable, namely that the treatment has an effect on the extensive margin of the outcome.[1] They do not formally consider the case of no extensive margin response but strongly suggest that having no extensive margin response eliminates the problems associated with quasi-logarithmic transformations. Our results highlight that although one may not be able to obtain any arbitrary value in this case, quasi-logarithmic transformations still do not give meaningful estimates. The case where a treatment does not affect the extensive margin of the outcome is of particular interest because it occurs with positive probability in experiments with finite samples.

The results in Mullahy and Norton (2024) and our paper do not impose any specific restrictions on the independent variables. Chen and Roth (2024), however, focus on the case of a binary treatment. While they also consider an extension where the independent variable can take a continuum of values on some set, these formal results do not cover cases such as multinomial treatments (i.e., discrete, non-binary, finite treatments).

Mullahy and Norton (2024) consider the specific quasi-linear transformations $\text{arsinh}(y)$ and $\log(y + c)$, while our paper and that of Chen and Roth (2024) consider broader classes of transformations. The results in Chen and Roth (2024) focus on what they refer to as "log-like" transformations, which they define as continuous, weakly increasing functions that are asymptotically equivalent to $\log(y)$ and defined on the domain $[0, \infty)$. We consider transformations that may be defined for negative values, as our definition of quasi-logarithmic over $\mathbb{R}_+$ considers functions over the domain $D \supseteq [0, \infty)$, and our definition of quasi-logarithmic over $\mathbb{R}$ considers functions over the domain $\mathbb{R}$, without any domain restrictions. Thus, the set of transformations we consider includes functions such as $\log(y+1)$ and $\text{arsinh}(y)$ over all values where they are defined. Our definition additionally imposes the requirement that the derivative of the transformation function is asymptotically equivalent to that of the logarithmic function, but this only becomes necessary for the results on semi-elasticities.

A.1.3 Practical recommendations

Below is a discussion of the recommendations offered by Mullahy and Norton (2024) and Chen and Roth (2024), which we summarize in Table S6.

Mullahy and Norton (2024) recommend using two-part models to separately analyze extensive and intensive margin responses. Two-part models (Cragg, 1971) provide a way of

[1]The result in Mullahy and Norton (2024) implicitly imposes this restriction as well since they do not arrive at a separate conclusion for the case of no extensive margin response.

analyzing data that contain "true zeros," representing genuine instances of zero values in the outcome variable. 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. 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.

Chen and Roth (2024) recommend estimating alternative target parameters. Since they show that it is not possible to (1) point identify any parameter that is (2) an average of (some function of) individual-level treatment effects satisfying (3) scale invariance, they propose alternatives that relax each of these conditions.

Lee (2009) bounds. If the goal of using the log transformation is to approximate the proportional effect along the intensive margin, then instead of point identifying the effect, they suggest using the Lee (2009) approach to bound the intensive margin effect. Bounding offers an alternative solution to sample selection problems (Horowitz and Manski, 2000) by focusing on intensive margin effects and analyzing "worst-case scenarios" or assuming monotonicity in the effect of treatment assignment on sample selection. Bounding approaches shift the emphasis to estimating a range of plausible impacts on outcomes for types with nonzero 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 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.

Poisson regression or quantile effects. If the goal of using the log transformation is to have a percent interpretation, then instead of estimating an average of individual-level treatment effects, they suggest using Poisson regression (Nelder and Wedderburn, 1972) to estimate the ATE in levels as a percentage, or to estimate quantile effects. We discuss Poisson regression in Section 5.2 and thus focus on quantile regressions below.

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 the median regression estimate. Finally, we emphasize that quantile regressions do not circumvent the scale-dependence issues we document that arise when using quasi-logarithmic transformations.[2]

An alternative transformation proposed by Chen and Roth (2024). If the goal of using the log transformation is to capture decreasing returns, then instead of using scale invariant transformations, they suggest using the transformation $L(y) = \log(y + c\mathbb{1}_{\{y=0\}})$ to calibrate the value placed on the intensive and extensive margins. The proposed Chen-Roth transformation is a limiting case of quasi-logarithmic transformations, specifically when the scale factor is large, and it does not avoid the problems of using quasi-logarithmic transformations. In particular, it is non-monotonic and leads to scale-dependent semi-elasticities and t-statistics. One can instead use power transformations to capture decreasing returns more flexibly and to calibrate the value placed on the intensive and extensive margins.[3]

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

[3]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.

A.2 Remarks on causality and interpretation

Since Chen and Roth (2024) cast the problem in a potential outcomes framework, we conclude this section with some discussion of 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 Supp. Appendix B 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., 2025). Logarithmic and power transformations imply percent changes that satisfy scale-invariance properties, while quasi-logarithmic 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 (2025*b*) show.

Appendix C of Chen and Roth (2024) offers some additional thoughts on interpreting the results from power transformations from the perspective of the potential outcomes framework. Specifically, they argue that a potential outcomes model of the form $Y(d,U)^k = \beta_0 + d\beta_1 + U$, where U is an individual-level random variable and the treatment d has a constant nonzero effect on the outcome, fails to accommodate situations where some units have $Y = 0$ for multiple treatment levels. First, this reasoning is premised on a constant individual-level relationship, ignoring the possibility of heterogeneous treatment effects and the focus on average effects in OLS regressions. Second, the same logic would forbid any linear model, including the standard untransformed one corresponding to the power $k = 1$. An argument hinging on an interpretation that fails to account for heterogeneity seems insufficient to justify abandoning OLS in the presence of zero-valued outcomes.

B Empirical illustration of severity of scale dependence

Our empirical application focuses on a dataset that allows us to systematically evaluate various aspects of scale dependence.

B.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 (2025*a*). 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.[4] 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.

B.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.[5] Although this analysis focuses on the IHS transformation due to the presence of negative values for some of the variables, we

[4]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."

[5]When considering arbitrary units, we use the limit results established in Section 3 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.

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 S7. 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 65 out of 98 estimates in total and up to 43 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 27 estimates reverse in sign, 43 estimates reverse in statistical significance, and 60 estimates reverse in either sign or significance.[6]

Considering only the natural alternative measurement units still leads to sign or significance reversals for up to 13 baseline estimates and up to 23 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 19 hypothesis tests and reverses conclusions about whether the treatment increases or decreases outcomes at endline for up to 10 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 full sample, along with the regression estimates corresponding to the original and natural alternative units, appears in Tables S8 and S9.

The continuous treatment of the number of days spent waiting for disbursement (which restricts to a subset of households receiving lump-sum transfers) tells a similar story (Table 3). 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 36 estimates reverse in sign, 36 estimates reverse in statistical significance, and 63 estimates reverse in either sign or significance. Of these, 23 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

[6]The preferred specifications correspond to baseline controls and village fixed effects for endline estimates (column 4, Table S7) and village fixed effects for baseline estimates (difference between columns 3 and 4, Table S7).

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 3 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 (Tables S10 and S11): for example, 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.

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 3 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.

C Predicted values

Providing suitable estimators of $\mathbb{E}[y \,|\, X]$ is not obvious given the possible nonlinear transformation f. When f is nonlinear, $\mathbb{E}[y \,|\, X] \neq f^{-1}(\mathbb{E}[f(y) \,|\, X])$ from Jensen's inequality, and the difference can be substantial. As Section 4.1 notes, estimating the predicted value of the untransformed outcome variable makes it possible to compute a semi-elasticity for independent variables that are discrete.

To obtain the predicted values of the untransformed variable, we cannot simply retransform the predicted values in the transformed regression $X\hat{\beta}$ by applying f^{-1}. Instead, we need to properly account for the distribution of the error term in $\mathbb{E}[f^{-1}(X\beta+\varepsilon) \,|\, X]$ in any estimator of the predicted values. Assuming that $\varepsilon \overset{\text{i.i.d.}}{\sim} F$, one can avoid making parametric assumptions on the distribution F by using the smearing estimate that Duan (1983) proposes as an application of the bootstrap principle (Efron, 1979).[7]

Definition C.1 (Smearing estimator of the predicted value)**.** The smearing estimator of $\mathbb{E}[y \mid X]$ of the untransformed outcome variable y under transformation $f(y)$ is given by $\frac{1}{n}\sum_{i=1}^{n} f^{-1}\left(X\hat{\beta} + \hat{\varepsilon}_i\right)$. We denote this estimator as $\hat{\mathbb{E}}^f(y \mid X)$.

The smearing estimator provides a consistent nonparametric estimator of $\mathbb{E}[y \mid X]$.[8] Furthermore, compared to common parametric estimates that assume F follows a normal distribution, the smearing estimate avoids potentially large biases with departures from the parametric assumption, and it can have high relative efficiency even when the parametric assumption holds (Duan, 1983).[9]

We note that the difficulty of obtaining an unbiased estimate of $\mathbb{E}[y \mid X]$ also applies to the logarithmic transformation (with strictly positive data), which the original application by Duan (1983) focuses on. In that particular case, if a researcher is only interested in $\frac{\hat{\mathbb{E}}[y \mid x=1] - \hat{\mathbb{E}}[y \mid x=0]}{\hat{\mathbb{E}}[y \mid x=0]}$ as an estimate of the semi-elasticity for binary x (for simplicity this discussion is for the case with one independent variable x), however, then applying the smearing estimator reduces to $\exp(\hat{\beta}) - 1$, which is approximately $\hat{\beta}$ as is typically used in empirical research.[10]

[7] Duan (1983) attributes the terminology to Carl Morris's discussion of Rubin (1983).

[8] The assumption $\varepsilon \overset{\text{i.i.d.}}{\sim} F$ implicitly assumes homoskedasticity in the transformed regression. As Manning (1998) notes, the use of transformations that reduce heteroskedasticity can limit concerns about bias. Alternatively, if one specifies the form of heteroskedasticity (as is required when estimating generalized linear models) the bootstrap principle still applies: if $\varepsilon_i = g(X_i\beta)\varepsilon$, then one can use the smearing estimator substituting $\hat{\varepsilon}_i$ with $\hat{\varepsilon}_i(X) = \frac{\hat{\varepsilon}_i}{g(X_i\hat{\beta})} g(X\hat{\beta})$. See Ai and Norton (2000) for a related approach as well as Ai and Norton (2008) for an alternative that allows for heteroskedasticity of any form.

[9] When $f(y) = \log(y)$, assuming $\varepsilon \sim \mathcal{N}(0, \sigma^2)$ would imply a predicted value of $\exp(X\beta + \sigma^2/2)$.

[10] In a log-linear regression, if the object of interest is $\mathbb{E}\left[\frac{\exp((x+1)\beta+\varepsilon) - \exp(x\beta+\varepsilon)}{\exp(x\beta+\varepsilon)}\right] = \mathbb{E}[\exp(\beta) - 1]$ then an estimate of this discrete case semi-elasticity would be approximately $\exp\left(\hat{\beta} - {}^1\!/_2\widehat{\text{Var}}(\hat{\beta})\right) - 1$ as Kennedy (1981) pointed out using the relationship $\mathbb{E}\left[\exp\left(\hat{\beta}\right)\right] = \exp\left(\hat{\beta} + {}^1\!/_2\widehat{\text{Var}}(\hat{\beta})\right)$. While the ε may not cancel if f is not the logarithmic transformation, one can still apply the bootstrap principle in the smearing estimator if one chooses this definition of a discrete semi-elasticity.

D Theoretical results on scale dependence

To provide an accessible exposition, we present all arguments for one variable, noting that the same approach generalizes via an application of the Frisch-Waugh-Lovell theorem at the cost of more notation.

D.1 Scale dependence of quasi-logarithmic transformations of outcomes

Proof of "large units" case of Proposition 1. Let β_c be the coefficient in the transformed regression with scale factor c for an independent variable x_i. Since $\ell(cy_i) \to \ell(0)$ as $c \to 0$, we have $\ell(cy_i) - \frac{1}{n}\sum_{j=1}^{n} \ell(cy_j) \to 0$, and hence

$$\begin{aligned}
\hat{\beta}_c &= \frac{\operatorname{cov}(x_i, \ell(cy_i))}{\operatorname{var}(x_i)} \\
&= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\ell(cy_i) - \overline{\ell(cy)} \right)}{\operatorname{var}(x)} \\
&= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\ell(cy_i) - \frac{1}{n}\sum_{j=1}^{n} \ell(cy_j) \right)}{\operatorname{var}(x)} \\
&\to 0
\end{aligned}$$

as $c \to 0$ (we refer to the limit from the right side if ℓ is only defined for non-negative values).

Moreover, since $\ell(cy) \approx \ell(0) + cy\ell'(0)$ as $c \to 0$, we have

$$\begin{aligned}
\hat{\xi}_c = \frac{\hat{\beta}_c}{c\bar{y}\ell'(c\bar{y})} &= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\ell(cy_i) - \frac{1}{n}\sum_{j=1}^{n} \ell(cy_j) \right)}{\operatorname{var}(x)} \frac{1}{c\bar{y}\ell'(c\bar{y})} \\
&= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\frac{\ell(cy_i)}{c\ell'(c\bar{y})} - \frac{1}{n}\sum_{j=1}^{n} \frac{\ell(cy_j)}{c\ell'(c\bar{y})} \right)}{\operatorname{var}(x)} \frac{1}{\bar{y}} \\
&= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\frac{\ell(cy_i)-\ell(0)}{c\ell'(c\bar{y})} - \frac{1}{n}\sum_{j=1}^{n} \frac{\ell(cy_j)-\ell(0)}{c\ell'(c\bar{y})} \right)}{\operatorname{var}(x)} \frac{1}{\bar{y}} \\
&\to \frac{\sum_{i=1}^{N} (x_i - \overline{x}) \left(y_i - \frac{1}{n}\sum_{j=1}^{n} y_j \right)}{\operatorname{var}(x)} \frac{1}{\bar{y}}
\end{aligned}$$

$$= \frac{\hat{\beta}}{\bar{y}} = \hat{\xi}_u.$$

Recall that the variance of $\hat{\beta}_c$ is determined by the sum of squared residuals. Since $\hat{\beta}_c \approx \ell'(0)c\hat{\beta}_{\text{untransformed}}$, we have

$$\begin{aligned}
e_c'e_c &= \sum_i (\ell(cy_i))^2 + \sum_i \left(\hat{\beta}_c x_i\right)^2 - 2\sum_i \ell(cy_i)\hat{\beta}_c x_i \\
&\approx \left(\sum_i (\ell'(0)cy_i)^2 + \sum_i \left(\ell'(0)c\hat{\beta}_{\text{untransformed}} x_i\right)^2 - 2\sum_i \ell'(0)cy_i \cdot \ell'(0)c\hat{\beta}_{\text{untransformed}} x_i\right) \\
&= (\ell'(0)c)^2 \left(\sum_i y_i^2 + \sum_i \left(\hat{\beta}_{\text{untransformed}} x_i\right)^2 - 2\sum_i y_i \hat{\beta}_{\text{untransformed}} x_i\right) \\
&= (\ell'(0)c)^2\, e_{\text{untransformed}}' e_{\text{untransformed}},
\end{aligned}$$

and thus $\frac{\hat{\beta}_c}{\widehat{\text{s.e.}}(\hat{\beta}_c)} \to \frac{c\ell'(0)\hat{\beta}_{\text{untransformed}}}{c\ell'(0)\widehat{\text{s.e.}}(\hat{\beta}_{\text{untransformed}})} = t_{\hat{\beta}_{\text{untransformed}}}$. □

Proof of "small units" case of Proposition 1. Suppose $\ell(\cdot)$ is quasi-logarithmic over $\mathbb{R}_+$. First consider the case that $y_i = 0$ for some i. Note that

$$\begin{aligned}
\ell(cy_i) - \overline{\ell(cy)} &\approx \log(cy_i) - \overline{\log(cy)} \\
&\approx \log(1 + cy_i) - \overline{\log(1 + cy)}
\end{aligned}$$

as $c \to \infty$. This implies

$$\begin{aligned}
\hat{\beta}_c &= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\ell(cy_i) - \frac{1}{n}\sum_{j=1}^n \ell(cy_j)\right)}{\text{var}(x)} \\
&\approx \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\log(1 + cy_i) - \frac{1}{n}\sum_{j=1}^n \log(1 + cy_j)\right)}{\text{var}(x)} \\
&= \frac{\sum_{\{i:y_i>0\}} (x_i - \overline{x}) \left(\log(1 + cy_i) - \frac{1}{n}\sum_{\{j:y_j>0\}} \log(1 + cy_j)\right)}{\text{var}(x)} \\
&+ \frac{\sum_{\{i:y_i=0\}} (x_i - \overline{x}) \left(\log(1 + cy_i) - \frac{1}{n}\sum_{\{j:y_j>0\}} \log(1 + cy_j)\right)}{\text{var}(x)}
\end{aligned}$$

$$= \frac{\sum_{\{i:y_i>0\}} (x_i - \overline{x}) \Big(\log(1+cy_i) - \frac{1}{n} \sum_{\{j:y_j>0\}} \big(\log(\frac{1}{c} + y_j) + \log(c) \big) \Big)}{\mathrm{var}(x)}$$

$$+ \frac{\sum_{\{i:y_i=0\}} (x_i - \overline{x}) \Big(-\frac{1}{n} \sum_{\{j:y_j>0\}} \big(\log(\frac{1}{c} + y_j) + \log(c) \big) \Big)}{\mathrm{var}(x)}$$

$$= \frac{\sum_{\{i:y_i>0\}} (x_i - \overline{x}) \Big(\log(1+cy_i) - \frac{1}{n} \sum_{\{j:y_j>0\}} \log(\frac{1}{c} + y_j) \Big)}{\mathrm{var}(x)}$$

$$+ \frac{\sum_{\{i:y_i=0\}} (x_i - \overline{x}) \Big(-\frac{1}{n} \sum_{\{j:y_j>0\}} \log(\frac{1}{c} + y_j) \Big) - \sum_{i=1}^{n} (x_i - \overline{x}) \frac{n_{y>0}}{n} \log(c)}{\mathrm{var}(x)}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \Big(\mathbf{1}_{\{y_i>0\}} \log(1+cy_i) - \frac{1}{n} \sum_{\{j:y_j>0\}} \log(\frac{1}{c} + y_j) \Big) - \sum_{i=1}^{n} (x_i - \overline{x}) \frac{n_{y>0}}{n} \log(c)}{\mathrm{var}(x)}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \Big(\mathbf{1}_{\{y_i>0\}} \log(\frac{1}{c} + y_i) - \frac{1}{n} \sum_{\{j:y_j>0\}} \log(\frac{1}{c} + y_j) \Big)}{\mathrm{var}(x)}$$

$$+ \frac{\sum_{\{i:y_i>0\}} (x_i - \overline{x}) \log(c) - \sum_{i=1}^{n} (x_i - \overline{x}) \frac{n_{y>0}}{n} \log(c)}{\mathrm{var}(x)}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \Big(\mathbf{1}_{\{y_i>0\}} \log(\frac{1}{c} + y_i) - \frac{1}{n} \sum_{\{j:y_j>0\}} \log(\frac{1}{c} + y_j) \Big)}{\mathrm{var}(x)}$$

$$+ \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \mathbf{1}_{\{y_i>0\}} \log(c) - \sum_{i=1}^{n} (x_i - \overline{x}) \frac{n_{y>0}}{n} \log(c)}{\mathrm{var}(x)}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \Big(\mathbf{1}_{\{y_i>0\}} \log(\frac{1}{c} + y_i) - \frac{1}{n} \sum_{\{j:y_j>0\}} \log(\frac{1}{c} + y_j) \Big)}{\mathrm{var}(x)}$$

$$+ \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \Big(\mathbf{1}_{\{y_i>0\}} - \frac{n_{y>0}}{n} \Big) \log(c)}{\mathrm{var}(x)}$$

$$= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \Big(\mathbf{1}_{\{y_i>0\}} \log(\frac{1}{c} + y_i) - \frac{1}{n} \sum_{\{j:y_j>0\}} \log(\frac{1}{c} + y_j) \Big)}{\mathrm{var}(x)}$$

$$+ \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \Big(\mathbf{1}_{\{y_i>0\}} - \overline{\mathbf{1}_{\{y_i>0\}}} \Big) \log(c)}{\mathrm{var}(x)}$$

$$
\begin{aligned}
&= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\mathbf{1}_{\{y_i>0\}} \log(\frac{1}{c} + y_i) - \frac{1}{n} \sum_{\{j: y_j > 0\}} \log(\frac{1}{c} + y_j) \right)}{\text{var}(x)} \\
&+ \frac{\left(\frac{n_{y=0}}{n} \sum_{\{i: y_i > 0\}} (x_i - \overline{x}) - \frac{n_{y>0}}{n} \sum_{\{i: y_i = 0\}} (x_i - \overline{x}) \right) \log(c)}{\text{var}(x)}
\end{aligned}
$$

Thus, as long as $\frac{1}{n_{y>0}} \sum_{\{i: y_i>0\}} (x_i - \overline{x}) \neq \frac{1}{n_{y=0}} \sum_{\{i: y_i=0\}} (x_i - \overline{x})$ or equivalently $\mathbb{E}[x \mid y > 0] \neq \mathbb{E}[x \mid y = 0]$, we have $\left|\widehat{\beta}_c\right| \to \infty$ as $c \to \infty$.

If $\mathbb{E}[x \mid y = 0] = \mathbb{E}[x \mid y \neq 0]$, the expression for $\widehat{\beta}_c$ becomes

$$
\begin{aligned}
\widehat{\beta}_c &= \frac{\sum_{\{i: y_i \neq 0\}} (x_i - \overline{x}) \left(\log(\frac{1}{c} + y_i) - \frac{1}{n} \sum_{\{j: y_j \neq 0\}} \log(\frac{1}{c} + y_j) \right)}{\text{var}(x)} \\
&+ \frac{\sum_{\{i: y_i = 0\}} (x_i - \overline{x}) \left(-\frac{1}{n} \sum_{\{j: y_j \neq 0\}} \log\left(\frac{1}{c} + y_j\right) \right)}{\text{var}(x)} \\
&\to \frac{\sum_{\{i: y_i \neq 0\}} (x_i - \overline{x}) \left(\log(y_i) - \frac{1}{n} \sum_{\{j: y_j \neq 0\}} \log(y_j) \right) + \sum_{\{i: y_i = 0\}} (x_i - \overline{x}) \left(-\frac{1}{n} \sum_{\{j: y_j \neq 0\}} \log(y_j) \right)}{\text{var}(x)} \\
&= \frac{\sum_{\{i: y_i \neq 0\}} (x_i - \overline{x}) \left(L(y_i) - \frac{1}{n} \sum_{\{j: y_j \neq 0\}} L(y_j) \right) + \sum_{\{i: y_i = 0\}} (x_i - \overline{x}) \left(0 - \frac{1}{n} \sum_{\{j: y_j \neq 0\}} L(y_j) \right)}{\text{var}(x)} \\
&= \frac{\sum_{\{i: y_i \neq 0\}} (x_i - \overline{x}) \left(L(y_i) - \frac{1}{n} \sum_{j=1}^{n} L(y_j) \right) + \sum_{\{i: y_i = 0\}} (x_i - \overline{x}) \left(0 - \frac{1}{n} \sum_{j=1}^{n} L(y_j) \right)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(L(y_i) - \frac{1}{n} \sum_{j=1}^{n} L(y_j) \right)}{\text{var}(x)} \\
&= \widehat{\beta}_{\text{Limit}}
\end{aligned}
$$

as $c \to \infty$.

Now consider the case that $y_i > 0$ for all i. As $c \to \infty$,

$$
\begin{aligned}
\widehat{\beta}_c &\approx \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\log(\frac{1}{c} + y_i) - \frac{1}{n} \sum_{j=1}^{n} \log(\frac{1}{c} + y_j) \right)}{\text{var}(x)} \\
&\to \frac{\sum_{i=1}^{n} (x_i - \overline{x}) \left(\log(y_i) - \frac{1}{n} \sum_{j=1}^{n} \log(y_j) \right)}{\text{var}(x)}
\end{aligned}
$$

$$= \hat{\beta}_{\text{Limit}}.$$

Now suppose $\ell(\cdot)$ is quasi-logarithmic over $\mathbb{R}_+$. In this case, we have

$$\begin{aligned}
\hat{\beta}_c &= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\ell(cy_i) - \frac{1}{n}\sum_{j=1}^n \ell(cy_j)\right)}{\text{var}(x)} \\
&\approx \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i)\log(1+c|y_i|) - \frac{1}{n}\sum_{j=1}^n \text{sgn}(y_j)\log(1+c|y_j|)\right)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i)\log(1+c|y_i|) - \frac{1}{n}\left(\sum_{\{j:y_j>0\}}\log(1+cy_j) - \sum_{\{j:y_j<0\}}\log(1-cy_j)\right)\right)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x}) (\text{sgn}(y_i)\log(1+c|y_i|))}{\text{var}(x)} \\
&- \frac{\sum_{i=1}^n (x_i - \overline{x}) \frac{1}{n}\left(\sum_{\{j:y_j>0\}}\big(\log(1+cy_j) - \log(c)\big) + n_{y>0}\log(c) - \sum_{\{j:y_j<0\}}\big(\log(1-cy_j) - \log(c)\big) - n_{y<0}\log(c)\right)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i)\log(1+c|y_i|) - \frac{1}{n}\left(\sum_{\{j:y_j>0\}}\log(\frac{1}{c}+y_j) - \sum_{\{j:y_j<0\}}\log(\frac{1}{c}-y_j)\right) - \frac{n_{y>0}\log(c) - n_{y<0}\log(c)}{n}\right)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i)\log(1+c|y_i|) - \frac{1}{n}\left(\sum_{\{j:y_j\neq 0\}}\text{sgn}(y_j)\log(\frac{1}{c}+|y_j|)\right)\right) - \sum_{i=1}^n (x_i - \overline{x}) \frac{n_{y>0} - n_{y<0}}{n}\log(c)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i)\log(\frac{1}{c}+|y_i|) - \frac{1}{n}\left(\sum_{i=1}^n \text{sgn}(y_j)\log(\frac{1}{c}+|y_j|)\right)\right)}{\text{var}(x)} \\
&+ \frac{\sum_{\{i:y_i>0\}} (x_i - \overline{x})\,\text{sgn}(y_i)\log(c) - \sum_{i=1}^n (x_i - \overline{x}) \frac{n_{y>0} - n_{y<0}}{n}\log(c)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i)\log(\frac{1}{c}+|y_i|) - \frac{1}{n}\left(\sum_{i=1}^n \text{sgn}(y_j)\log(\frac{1}{c}+|y_j|)\right)\right)}{\text{var}(x)} \\
&+ \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i) - \frac{n_{y>0} - n_{y<0}}{n}\right)\log(c)}{\text{var}(x)} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x}) \left(\text{sgn}(y_i)\log(\frac{1}{c}+|y_i|) - \frac{1}{n}\left(\sum_{i=1}^n \text{sgn}(y_j)\log(\frac{1}{c}+|y_j|)\right)\right)}{\text{var}(x)} \quad +
\end{aligned}$$

which shows that $\text{sgn}(\hat{\beta}_c)$ approaches the sign of a regression of $\text{sgn}(y_i)$ on x_i as $c \to \infty$.

Since $\ell'(y) \to \frac{1}{y}$ as $y \to \infty$ (and also as $y \to \infty$ if $\ell(\cdot)$ is quasi-logarithmic over $\mathbb{R}$), we have $\hat{\xi}_c = \frac{\hat{\beta}_c}{c\bar{y}\ell'(c\bar{y})} \to \hat{\beta}_c$ as $c \to \infty$.

First suppose $\mathrm{sgn}(y)$ is constant. This case is straightforward since $y_i > 0$ for all i implies $\ell(cy) \approx \log(c) + \log(y)$, and $y_i < 0$ for all i implies $\ell(cy) \approx -\log(|cy|) = -\log(c) - \log(|y|)$, and thus $t_{\hat{\beta}_c} = t_{\hat{\beta}_{\text{Limit}}}$. In addition, $y_i = 0$ for all i implies $\hat{\beta}_c = \hat{\beta}_{\text{Limit}} = 0$ and hence $t_{\hat{\beta}_c} = t_{\hat{\beta}_{\text{Limit}}} = 0$.

Now suppose $\mathrm{sgn}(y)$ is not constant. In this case, since $\hat{\beta}_c \approx \hat{\beta}_{\text{Limit}} + \log(c)\hat{\beta}_{\text{sign}}$, the sum of squared residuals is given by

$$\begin{aligned}
e_c'e_c &\approx \sum_i (\mathrm{sgn}(y_i)\log(1+c|y_i|))^2 + \sum_i \left(\log(c)\hat{\beta}_{\text{sign}}x_i\right)^2 - 2\sum_i \log(c)\mathrm{sgn}(y_i)\log(1+c|y_i|)\hat{\beta}_{\text{sign}}x_i \\
&\approx \sum_i \left(\mathrm{sgn}(y_i)\log(\frac{1}{c} + |y_i|) + \mathrm{sgn}(y_i)\log(c)\right)^2 \\
&+ \sum_i \left(\log(c)\hat{\beta}_{\text{sign}}x_i\right)^2 - 2\sum_i \log(c)\mathrm{sgn}(y_i)\left(\log(\frac{1}{c} + |y_i|) + \log(c)\right)\hat{\beta}_{\text{sign}}x_i \\
&\approx \sum_i (\mathrm{sgn}(y_i)\log(|y_i|) + \mathrm{sgn}(y_i)\log(c))^2 \\
&+ \sum_i \left(\log(c)\hat{\beta}_{\text{sign}}x_i\right)^2 - 2\sum_i \log(c)\mathrm{sgn}(y_i)\left(\log(|y_i|) + \log(c)\right)\hat{\beta}_{\text{sign}}x_i \\
&\approx \sum_i (\mathrm{sgn}(y_i)\log(c))^2 + \sum_i \left(\log(c)\hat{\beta}_{\text{sign}}x_i\right)^2 - 2\sum_i \mathrm{sgn}(y_i)\left(\log(c)\right)^2\hat{\beta}_{\text{sign}}x_i \\
&= (\log(c))^2\left(\sum_i (\mathrm{sgn}(y_i))^2 + \sum_i \left(\hat{\beta}_{\text{sign}}x_i\right)^2 - 2\sum_i \mathrm{sgn}(y_i)\hat{\beta}_{\text{sign}}x_i\right) \\
&= (\log(c))^2 e_{\text{sign}}'e_{\text{sign}}
\end{aligned}$$

Therefore, $\frac{\hat{\beta}_c}{\widehat{\text{s.e.}}(\hat{\beta}_c)} = \frac{\hat{\beta}_{\text{sign}}}{\widehat{\text{s.e.}}(\hat{\beta}_{\text{sign}})} = t_{\hat{\beta}_{\text{sign}}}$ as $c \to \infty$. In the case that $\hat{\beta}_{\text{sign}} = 0$, this implies $t_{\hat{\beta}_c} = 0$. $\square$

D.2 Scale dependence of quasi-logarithmic transformations of covariates

Proof of "large units" case of Proposition 2. Since $\ell(cx) \approx \ell(0) + cx\ell'(0)$ as $c \to 0$, we have

$$\begin{aligned}
\ell(cx_i) - \overline{\ell(cx)} &\approx cx_i\ell'(0) - \overline{cx\ell'(0)} \\
&= c\ell'(0)\left(x_i - \overline{x}\right)
\end{aligned}$$

for small values of c. Therefore, as $c \to 0$, we have

$$\begin{aligned}
\widehat{\beta}_c &= \frac{\operatorname{cov}(\ell(cx_i), y_i)}{\operatorname{var}(\ell(cx_i))} \\
&= \frac{\sum_{i=1}^n \left(\ell(cx_i) - \overline{\ell(cx)}\right)(y_i - \overline{y})}{\sum_{i=1}^n \left(\ell(cx_i) - \overline{\ell(cx)}\right)^2} \\
&\to \frac{\sum_{i=1}^n c\ell'(0)\,(x_i - \overline{x})\,(y_i - \overline{y})}{\sum_{i=1}^n \left(c\ell'(0)\,(x_i - \overline{x})\right)^2} \\
&= \frac{\sum_{i=1}^n (x_i - \overline{x})\,(y_i - \overline{y})}{c\ell'(0)\sum_{i=1}^n (x_i - \overline{x})^2} \\
&= \frac{1}{c\ell'(0)} \frac{\operatorname{cov}(x_i, y_i)}{\operatorname{var}(x_i)} \\
&= \frac{1}{c} \frac{\widehat{\beta}}{\ell'(0)},
\end{aligned}$$

which implies $|\widehat{\beta}_c| \to \infty$.

Moreover,

$$\begin{aligned}
\widehat{\xi}_c(x) &= \widehat{\beta}_c c x \ell'(cx) \\
&\to \frac{1}{c} \frac{\widehat{\beta}}{\ell'(0)} c x \ell'(cx) \\
&= \frac{\widehat{\beta}}{\ell'(0)} x \ell'(cx) \\
&\to \widehat{\beta} x \\
&= \widehat{\xi}(x)
\end{aligned}$$

as $c \to 0$.

Since $\widehat{\beta}_c \approx \frac{\widehat{\beta}_{\text{untransformed}}}{\ell'(0)c}$, the sum of squared residuals is given by

$$e_c' e_c = \sum_i y_i^2 + \sum_i \left(\widehat{\beta}_c \ell(cx_i)\right)^2 - 2\sum_i y_i \widehat{\beta}_c \ell(cx_i)$$

$$
\begin{aligned}
&\approx \sum_i y_i^2 + \sum_i \left(\frac{\hat{\beta}_{\text{untransformed}}}{\ell'(0)c} \ell'(0) c x_i \right)^2 - 2 \sum_i y_i \cdot \frac{\hat{\beta}_{\text{untransformed}}}{\ell'(0)c} \ell'(0) c x_i \\
&= \sum_i y_i^2 + \sum_i \left(\hat{\beta}_{\text{untransformed}} x_i \right)^2 - 2 \sum_i y_i \hat{\beta}_{\text{untransformed}} x_i \\
&= e'_{\text{untransformed}} e_{\text{untransformed}}
\end{aligned}
$$

and

$$
\begin{aligned}
(X_c' X_c)^{-1} &= \left(\begin{bmatrix} 1 & \cdots & 1 \\ \ell(cx_{11}) & \cdots & \ell(cx_{1n}) \end{bmatrix} \begin{bmatrix} 1 & \ell(cx_{11}) \\ \vdots & \vdots \\ 1 & \ell(cx_{1n}) \end{bmatrix} \right)^{-1} \\
&= \begin{bmatrix} n & \sum_{i=1}^n \ell(cx_{1i}) \\ \sum_{i=1}^n \ell(cx_{1i}) & \sum_{i=1}^n \left(\ell(cx_{1i})\right)^2 \end{bmatrix}^{-1} \\
&= \frac{1}{n \sum_{i=1}^n \ell(cx_{1i}) - \left(\sum_{i=1}^n \ell(cx_{1i}) \right)^2} \begin{bmatrix} \sum_{i=1}^n \left(\ell(cx_i)\right)^2 & -\sum_{i=1}^n \ell(cx_{1i}) \\ -\sum_{i=1}^n \ell(cx_{1i}) & n \end{bmatrix} \\
&\approx \frac{1}{n \sum_{i=1}^n \ell'(0) c x_{1i} - \left(\sum_{i=1}^n \ell'(0) c x_{1i} \right)^2} \begin{bmatrix} \sum_{i=1}^n \left(\ell'(0) c x_i\right)^2 & -\sum_{i=1}^n \ell'(0) c x_{1i} \\ -\sum_{i=1}^n \ell'(0) c x_{1i} & n \end{bmatrix} \\
&= \begin{bmatrix} n & \sum_{i=1}^n \ell'(0) c x_{1i} \\ \sum_{i=1}^n \ell'(0) c x_{1i} & \sum_{i=1}^n \left(\ell'(0) c x_{1i}\right)^2 \end{bmatrix}^{-1} \\
&= \left(\begin{bmatrix} 1 & \cdots & 1 \\ \ell'(0) c x_{11} & \cdots & \ell'(0) c x_{1n} \end{bmatrix} \begin{bmatrix} 1 & \ell'(0) c x_{11} \\ \vdots & \vdots \\ 1 & \ell'(0) c x_{1n} \end{bmatrix} \right)^{-1} \\
&= \frac{1}{\left(\ell'(0)c\right)^2} (X'_{\text{untransformed}} X_{\text{untransformed}})^{-1}
\end{aligned}
$$

This implies $\frac{e_c' e_c}{n-k} (X_c' X_c)^{-1} = \frac{e'_{\text{untransformed}} e_{\text{untransformed}}}{n-k} \frac{1}{\left(\ell'(0)c\right)^2} (X'_{\text{untransformed}} X_{\text{untransformed}})^{-1}$ and hence $\widehat{\text{s.e.}}(\hat{\beta}_c) = \frac{\widehat{\text{s.e.}}(\hat{\beta}_{\text{untransformed}})}{\ell'(0)c}$. Therefore, $\frac{\hat{\beta}_c}{\widehat{\text{s.e.}}(\hat{\beta}_c)} = \frac{\frac{\hat{\beta}_{\text{untransformed}}}{\ell'(0)c}}{\frac{\widehat{\text{s.e.}}(\hat{\beta}_{\text{untransformed}})}{\ell'(0)c}} \to t_{\hat{\beta}_{\text{untransformed}}}$ as $c \to 0$. □

Proof of "small units" case of Proposition 2. First consider the case that $x_i = 0$ for some i. Note that

$$\begin{aligned}\ell(cx_i) - \overline{\ell(cx)} &\approx \log(cx) - \overline{\log(cx)} \\ &\approx \log(1+cx) - \overline{\log(1+cx)} \\ &= \log\left(\frac{1}{c} + x_i\right) - \overline{\log\left(\frac{1}{c} + x\right)}\end{aligned}$$

as $c \to \infty$, which implies

$$\begin{aligned}&\hat{\beta}_c \\ &\to \frac{\sum_{i=1}^n \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y})}{\sum_{i=1}^n \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)^2} \\ &= \frac{\sum_{\{i:x_i\neq 0\}} \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y}) + \sum_{\{i:x_i=0\}} \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y})}{\sum_{\{i:x_i\neq 0\}} \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)^2 + \sum_{\{i:x_i=0\}} \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)^2} \\ &= \frac{\sum_{\{i:x_i\neq 0\}} \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y}) + \sum_{\{i:x_i=0\}} \left(\log(\frac{1}{c}) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y})}{\sum_{\{i:x_i\neq 0\}} \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)^2 + \sum_{\{i:x_i=0\}} \left(\log(\frac{1}{c}) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)^2} \\ &= \frac{\sum\limits_{\{i:x_i\neq 0\}} \left(\log(\frac{1}{c}+x_i) - \frac{n_{x=0}}{n}\log(\frac{1}{c}) - \frac{1}{n}\sum\limits_{\{j:x_j\neq 0\}} \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y}) + \sum\limits_{\{i:x_i=0\}} \left(\frac{n_{x\neq 0}}{n}\log(\frac{1}{c}) - \frac{1}{n}\sum\limits_{\{j:x_j\neq 0\}} \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y})}{\sum\limits_{\{i:x_i\neq 0\}} \left(\log(\frac{1}{c}+x_i) - \frac{n_{x=0}}{n}\log(\frac{1}{c}) - \frac{1}{n}\sum\limits_{\{j:x_j\neq 0\}} \log(\frac{1}{c}+x_j)\right)^2 + \sum\limits_{\{i:x_i=0\}} \left(\frac{n_{x\neq 0}}{n}\log(\frac{1}{c}) - \frac{1}{n}\sum\limits_{\{j:x_j\neq 0\}} \log(\frac{1}{c}+x_j)\right)^2}\end{aligned}$$

which implies $\left|\hat{\beta}_c\right| \to 0$ as $c \to \infty$ since the denominator grows on the order of $[\log(1/c)]^2$ while the numerator grows only on the order of $\log(1/c)$.

Now consider the case that $x_i > 0$ for all i. As $c \to \infty$,

$$\begin{aligned}\hat{\beta}_c &\approx \frac{\sum_{i=1}^n \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)(y_i - \overline{y})}{\sum_{i=1}^n \left(\log(\frac{1}{c}+x_i) - \frac{1}{n}\sum_{j=1}^n \log(\frac{1}{c}+x_j)\right)^2} \\ &\to \frac{\sum_{i=1}^n \left(\log(x_i) - \frac{1}{n}\sum_{j=1}^n \log(x_j)\right)(y_i - \overline{y})}{\sum_{i=1}^n \left(\log(x_i) - \frac{1}{n}\sum_{j=1}^n \log(x_j)\right)^2} \\ &= \hat{\beta}_{\text{Limit}}.\end{aligned}$$

In both cases, since $\ell'(x) \to \frac{1}{x}$ as $x \to \infty$, we have $\hat{\xi}_c(x) = \hat{\beta}_c cx\ell'(cx) \to \hat{\beta}_c$ as $c \to \infty$.

First suppose $\text{sgn}(x)$ is constant. This case is straightforward since $x_i > 0$ for all i implies $\ell(cx) \approx \log(c) + \log(x)$, and $x_i < 0$ for all i implies $\ell(cx) \approx -\log(|cx|) = -\log(c) - \log(|x|)$, and thus $t_{\hat{\beta}_c} = t_{\hat{\beta}_{\text{Limit}}}$.

Now suppose $\text{sgn}(x)$ is not constant. Following analogous steps to the case of scale dependence for y, we obtain $e_c' e_c \approx \frac{e_{\text{sign}}' e_{\text{sign}}}{\left(\log(\frac{1}{c})\right)^2}$ and $(X_c' X_c)^{-1} \to (X_{\text{sign}}' X_{\text{sign}})^{-1}$. Therefore, $\frac{\hat{\beta}_c}{\widehat{\text{s.e.}}(\hat{\beta}_c)} = \frac{\frac{\hat{\beta}_{\text{sign}}}{\log(\frac{1}{c})}}{\frac{\widehat{\text{s.e.}}(\hat{\beta}_{\text{sign}})}{\log(\frac{1}{c})}} = t_{\hat{\beta}_{\text{sign}}}$ as $c \to \infty$. □

References

Ai, Chunrong, and Edward C Norton. 2000. "Standard errors for the retransformation problem with heteroscedasticity." *Journal of health economics*, 19(5): 697–718. 11

Ai, Chunrong, and Edward C Norton. 2008. "A semiparametric derivative estimator in log transformation models." *The Econometrics Journal*, 11(3): 538–553. 11

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. 5

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. 6

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. 6

Chen, Jiafeng, and Jonathan Roth. 2024. "Logs with zeros? Some problems and solutions." *The Quarterly Journal of Economics*, 139(2): 891–936. 1, 3, 4, 5, 6, 7

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. 4

Duan, Naihua. 1983. "Smearing estimate: A nonparametric retransformation method." *Journal of the American Statistical Association*, 78(383): 605–610. 10, 11

Efron, Bradley. 1979. "Bootstrap Methods: Another Look at the Jackknife." *The Annals of Statistics*, 1–26. 11

Gardner, John, Neil Thakral, Linh Tô, and Luther Yap. 2025. "Two-stage Differences in Differences." *Mimeo*. 7

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. 8, 9, 10

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. 6

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. 5

Kennedy, Peter. 1981. "Estimation with correctly interpreted dummy variables in semilogarithmic equations." *American Economic Review*, 71(4). 11

Lee, David S. 2009. "Training, wages, and sample selection: Estimating sharp bounds on treatment effects." *Review of Economic Studies*, 76(3): 1071–1102. 5

Lemieux, Thomas. 2006. "Postsecondary education and increasing wage inequality." *American Economic Review*, 96(2): 195–199. 6

Manning, Willard G. 1998. "The logged dependent variable, heteroscedasticity, and the retransformation problem." *Journal of health economics*, 17(3): 283–295. 11

Mullahy, John, and Edward C Norton. 2024. "Why transform y? The pitfalls of transformed regressions with a mass at zero." *Oxford Bulletin of Economics and Statistics*, 86(2): 417–447. 1, 3, 4

Nelder, John Ashworth, and Robert WM Wedderburn. 1972. "Generalized linear models." *Journal of the Royal Statistical Society: Series A (General)*, 135(3): 370–384. 5

Rubin, Donald B. 1983. "A case study of the robustness of Bayesian methods of inference: Estimating the total in a finite population using transformations to normality." In *Scientific Inference, Data Analysis, and Robustness*. 213–244. Elsevier. 11

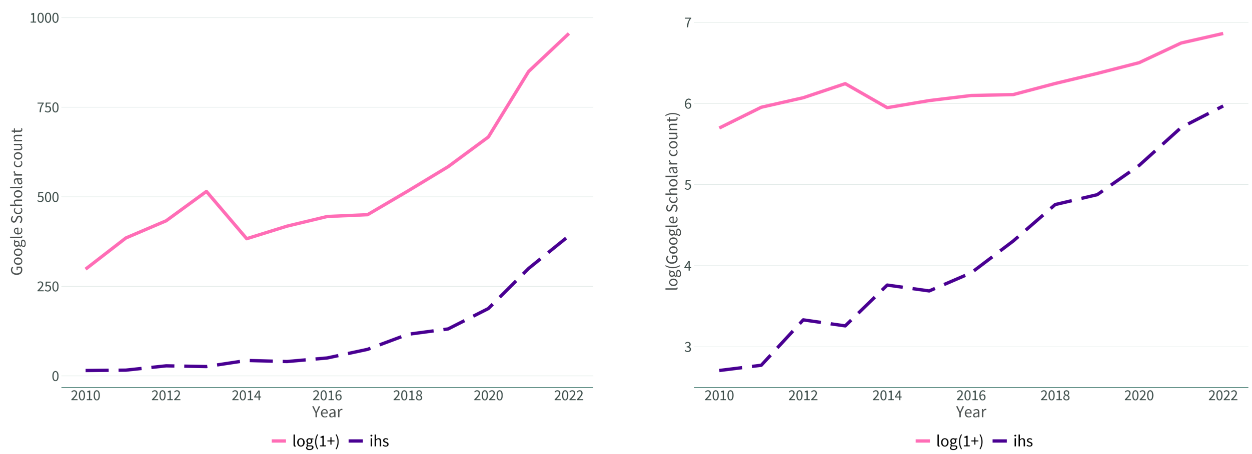
Thakral, Neil, and Linh T Tô. 2025*a*. "Excess Anticipation-Dependence in Consumption." *Mimeo*. 8

Thakral, Neil, and Linh T Tô. 2025*b*. "Scale Equivariance in Regressions." *Mimeo*. 7

Thakral, Neil, and Linh T Tô. 2025*c*. "When Are Estimates Independent of Measurement Units?" *Mimeo*.

E Appendix Figures and Tables

Figure S1: Google Scholar search results



(a) Level (b) Log

Note: The figures depict the flow of papers in economics, finance, or management journals using $\log(1+\cdot)$ or arsinh transformations, obtained from a Google Scholar search. The search term for papers using the $\log(1+\cdot)$ transformation is "log" and ("plus one" OR "one plus"). The search term for papers using the arsinh transformation is "inverse hyperbolic sine." Panel (a) displays the count of papers, and panel (b) displays the logarithm of the count of papers.

Table S1: Recent use of IHS transformations of x variables

Authors	Year	Journal
Bellemare and Wichman	2020	OBES
Pruitt and Turner	2020	AERI
Braguinsky, Ohyama, Okazaki, and Syverson	2021	AERI
Deryugina and Marx	2021	AERI
Goolsbee and Syverson	2021	JPubEc
Cao and Chen	2022	AER
Chan	2022	JDE
Raff, Meyer, and Walter	2022	JPubEc
DeBacker, Panousi, and Ramnath	2023	AEJ-Macro
Djourelova	2023	AER
Hernandez-Cortes and Meng	2023	JPubEc
Kammerlander and Schulze	2023	JHE
Bahar, Hauptmann, Özgüzel, and Rapoport	2024	REStat
Bursztyn, Chaney, Hassan, and Rao	2024	AER

Note: The papers listed above use an inverse hyperbolic sine transformation of an independent variable.

Table S2: Relationship between education and hourly wages—Comparison between OLS and GLM estimates

	(1)	(2)	(3)
Log transformation (OLS)	0.0980	0.0924	0.0860
Log-link quasi-Gaussian (GLM)	0.1112	0.1155	0.1047
Log-link quasi-Poisson (GLM)	0.1075	0.1056	0.0969
Log-link quasi-Exponential (GLM)	0.0990	0.0926	0.0872
Log-link quasi-inverse Gaussian (GLM)	0.0879	0.0811	0.0788

Note: This table compares the semi-elasticity estimates from the OLS regression with $\log(\text{earnings}_i)$ as the dependent variable with semi-elasticity estimates based on a generalized linear model with a log link function under the specified assumption about the error variance. The data come from the Merged Outgoing Rotation Groups of the CPS in 1991, restricting the sample to respondents who are in the labor force and not enrolled as students. The first column corresponds to the subsample of female respondents. The second column corresponds to the subsample of nonwhite respondents. The second column corresponds to the subsample of respondents with no more than 16 years of education.

Table S3: Relationship between education and hourly wages—Men and women

	(1)	(2)	(3)	(4)	(5)	(6)
Men (cubic exp.)	0.0967	0.0967	0.0968	0.0968	0.0968	0.0970
Women (cubic exp.)	0.1114	0.1104	0.1101	0.1099	0.1098	0.1094
Men (exp. FE)	0.0919	0.0948	0.0956	0.0959	0.0961	0.0968
Women (exp. FE)	0.1093	0.1095	0.1094	0.1093	0.1092	0.1089

Note: Columns 1–5 report semi-elasticity estimates from OLS regressions with the dependent variable $(\text{earnings}_i)^k$ for powers 1, $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{4}$, and $\frac{1}{5}$, while column 6 uses $\log(\text{earnings}_i)$. We compute marginal effects at the sample average of the covariates. The data come from the Merged Outgoing Rotation Groups of the CPS in 1991, restricting the sample to respondents age 16–66 who are in the labor force, not enrolled as students, and with at least seven years of completed schooling. The first and third rows correspond to the sample of 84,119 male respondents, and the second and fourth rows correspond to the sample of 80,014 female respondents. The first two columns control for a cubic polynomial in years of potential experience, defined as the number of years someone works if they start school at age 6 and start working upon finishing their years of schooling on time, and a nonwhite indicator. The last two columns instead use fixed effects for years of potential experience.

Table S4: Comparison with literature—Results on measurement-unit independence

	CR	MN	TT
Result	The only treatment effect estimate that is unit-independent, having a percent interpretation, is on the log of the outcome.	N/A	Log/power iff scale-invariant inference iff scale-invariant semi-elasticity iff scale-equivariant coefficients.
Comment	Criterion rules out standard practice of estimating treatment effect in levels, for which marginal effects can be obtained on a consistent scale.		Includes identity transformation as a special case.
Comment	It is evident that coefficient represents a semi-elasticity (implied percent change in outcomes) only under a log transformation: $\beta = \frac{\beta}{yf'(y)}$ iff log.		Coefficient estimates may lack useful interpretation; need to be converted to semi-elasticities, as with many nonlinear models (e.g., probit, logit, interaction terms).
Comment	Does not provide a constructive solution for data with zero- or negative-valued outcomes.		Only accommodates negative values when power is a ratio of odd numbers.
Comment	Assumes a binary treatment		
Comment	Does not consider statistical inference		

Table S5: Comparison with literature—Results on scale dependence

	CR	MN	TT
Quantities of interest	coefficient	marginal effect	semi-elasticity, t-statistic
Set of transformations	log-like	arsinh(y) and log($y+1$)	quasi-logarithmic
Variables	Dependent	Dependent	Dependent or independent
Restrictions...			
on X values	Binary or continuous only	N/A	N/A
on Y values	non-negative	non-negative	N/A
on relationship between X and Y	X has non zero effect on $\Pr(Y=0)$	X has non zero effect on $\Pr(Y=0)$	N/A

Table S6: Comments on recommendations in existing literature

Authors	Objective	Recommendation	Comment
CR	Bound the intensive margin effect	Bounding (Lee 2009)	• Decision to bound is distinct from choosing transformations.
CR	Estimate ATE in levels as a percentage	Poisson regression (Nelder and Wedderburn 1972)	• Requires assumption on conditional variance of error. • Percent interpretation achievable via OLS semi-elasticity; Poisson not required.
CR	Calibrate the value placed on the intensive and extensive margins	Chen-Roth transformation $\log(y + c\mathbb{1}_{\{y=0\}})$	• Non-monotonic transformation; limiting case of quasi-logarithmic transformation. • Implied percent changes and t-statistics depend on units of outcome variable. • Does not flexibly capture decreasing returns.
MN	Separately analyze extensive and intensive margin responses	Two-part models (Cragg 1971)	• Intensive margin response reflects selection bias.

Table S7: 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	7–29	7–24	0–8	0–3
Baseline control		10–30		1–7
Village FE	6–41	1–19	7–26	0–5
Baseline control & Village FE		0–18		0–6
Panel B: Significance flips (5 percent level)				
No controls	18–42	18–29	1–36	0–23
Baseline control		6–36		2–30
Village FE	19–47	13–27	11–44	4–21
Baseline control & Village FE		13–26		5–20
Panel C: Sign or significance flips				
No controls	23–48	23–34	1–39	0–25
Baseline control		16–43		3–35
Village FE	25–65	14–31	17–60	4–24
Baseline control & Village FE		13–29		5–24
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 sample consists of 2,390 individuals in the full sample and 2,272 individuals in the non-attriter sample. See Table 3 for details on the samples and specifications.

Table S8: Scale dependence in RCT data—Sign changing variables

Label	One unit	Alternative unit
Value Of Non-Land Assets (USD)	0.2703 (0.0838) [1]	-0.1969 (0.1759) [10000]
Non-Durable Expenditure (USD)	0.0850 (0.0246) [.01]	-0.1015 (0.0670) [1]
Total Revenue, Monthly (USD)	0.1776 (0.0794) [1]	-0.0043 (0.1146) [62.440]
Value Of Non-Land Assets Excluding Roof (USD)	0.0599 (0.0807) [1]	-0.1498 (0.1210) [62.440]
Value Of Birds (USD)	0.2401 (0.0965) [1]	-0.0091 (2.3166) [1.000e+48]
Value Of Durable Goods (USD)	0.0084 (0.0734) [1]	-0.1967 (0.1145) [62.440]
Value Of Furniture (USD)	0.0412 (0.0716) [1]	-0.1375 (0.1146) [62.440]
Value Of Agricultural Tools (USD)	0.0015 (0.0066) [.01]	-0.0652 (0.0554) [1]
Food Total (USD)	0.0670 (0.0208) [.01]	-0.0909 (0.0634) [1]
Food Own Production (USD)	0.0839 (0.0600) [1]	-0.0833 (0.1010) [62.440]
Food Bought (USD)	0.0616 (0.0195) [.01]	-0.0857 (0.0627) [1]
Cereals (USD)	0.0147 (0.0082) [.01]	-0.0020 (0.0680) [1]
Fruit & Vegetables (USD)	0.0228 (0.0082) [.01]	-0.0313 (0.0535) [1]
Fats (USD)	0.0029 (0.0027) [.01]	-0.0129 (0.0458) [1]
Sugars (USD)	0.0052 (0.0036) [.01]	-0.0044 (0.0517) [1]
Other Food (USD)	0.0299 (0.0128) [.01]	-0.0592 (0.0587) [1]
Social Expenditure (USD)	0.0950 (0.0519) [1]	-0.0468 (0.0880) [62.440]
Other Expenditure (USD)	0.0063 (0.0581) [1]	-0.2020 (0.0962) [62.440]
Cortisol (No Controls)	0.0060 (0.0525) [1]	-0.0225 (0.0552) [100]
Cortisol (With Controls)	0.0214 (0.0516) [1]	-0.0266 (0.7696) [1000000]
Farm Flow Expenses, Monthly (USD)	0.1585 (0.0544) [1]	-0.1339 (0.1531) [10000]
Livestock Flow Profit, Monthly (USD)	0.1216 (0.0992) [1]	-0.0068 (1.5207) [1.000e+16]
Total Expenses, Monthly (USD)	0.3571 (0.0795) [1]	-0.0118 (0.1635) [10000]
Total Profit, Monthly (USD)	0.0085 (0.0179) [.01]	-0.1354 (0.1268) [1]

Note: The table reports endline variables for which treatment effects can change signs using the IHS transformation when there are no control variables, which corresponds to the specification presented in Column (1) of Table S7. The treatment variable is an indicator for whether a household receives an unconditional cash transfer. We report the coefficient estimate, the standard error in parentheses, and the associated scaling factor in square brackets. When a natural alternative unit results in sign flipping, we report the alternative unit (62.44 Kenyan shillings instead of 1 US dollar for monetary measures, or 0.0362 µg/dL instead of nmol/L for concentration). In other cases, we report the alternative units on the grid $1eX$ for $X \in \{-50, -48, \ldots, 348, 350\}$ that are closest to 1 that result in sign flipping. The sample consists of 2,390 individuals.

Table S9: Scale dependence in RCT data—Significance changing variables

Label	Original unit	Alternative unit
Value Of Non-Land Assets (USD)	0.2703 (0.0838) [1]	0.0606 (0.1239) [62.440]
Non-Durable Expenditure (USD)	-0.1015 (0.0670) [1]	-0.3112 (0.1090) [62.440]
Total Revenue, Monthly (USD)	0.1776 (0.0794) [1]	-0.0043 (0.1146) [62.440]
Value Of Non-Land Assets Excluding Roof (USD)	0.2398 (0.0402) [.01]	-0.4073 (0.1733) [10000]
Value Of Livestock (USD)	0.4534 (0.1122) [1]	0.4574 (0.2649) [10000]
Value Of Birds (USD)	0.2401 (0.0965) [1]	0.2312 (0.1754) [62.440]
Value Of Durable Goods (USD)	0.1615 (0.0301) [.01]	-0.4488 (0.1675) [10000]
Value Of Furniture (USD)	0.1472 (0.0263) [.01]	-0.3571 (0.1702) [10000]
Value Of Agricultural Tools (USD)	-0.0652 (0.0554) [1]	-0.2427 (0.1039) [62.440]
Value Of Radio/Tv (USD)	0.2065 (0.0828) [1]	0.5779 (0.3034) [10000]
Food Total (USD)	-0.0909 (0.0634) [1]	-0.3006 (0.1048) [62.440]
Food Own Production (USD)	0.0179 (0.0075) [.01]	-0.4907 (0.2148) [1000000]
Food Bought (USD)	-0.0857 (0.0627) [1]	-0.2909 (0.1043) [62.440]
Meat & Fish (USD)	0.2151 (0.0655) [1]	0.2430 (0.1267) [62.440]
Fruit & Vegetables (USD)	-0.0313 (0.0535) [1]	-0.2219 (0.0933) [62.440]
Dairy (USD)	0.1646 (0.0750) [1]	0.2446 (0.1693) [62.440]
Fats (USD)	-0.0129 (0.0458) [1]	-0.1943 (0.0937) [62.440]
Other Food (USD)	-0.0592 (0.0587) [1]	-0.2551 (0.0989) [62.440]
Medical Expenditure Past Month (USD)	0.3534 (0.0764) [1]	0.5445 (0.2924) [10000]
Medical Expenditure, Children (USD)	0.2294 (0.0671) [1]	0.5435 (0.2981) [10000]
Education Expenditure (USD)	0.1424 (0.0585) [1]	0.0035 (0.0040) [.01]
Social Expenditure (USD)	0.0126 (0.0034) [.01]	-0.4514 (0.1919) [1000000]
Other Expenditure (USD)	0.0063 (0.0581) [1]	-0.2020 (0.0962) [62.440]
Non-Ag Business Revenue, Monthly (USD)	0.2695 (0.1083) [1]	0.5785 (0.3066) [10000]
Non-Ag Business Investment In Durables, Monthly (USD)	0.0555 (0.0217) [1]	0.1447 (0.0860) [62.440]
Farm Revenue, Monthly (USD)	-0.0522 (0.0534) [1]	-0.2027 (0.0983) [62.440]
Farm Flow Expenses, Monthly (USD)	0.1585 (0.0544) [1]	0.0526 (0.0961) [62.440]
Total Expenses, Monthly (USD)	0.3571 (0.0795) [1]	0.2145 (0.1142) [62.440]
Total Profit, Monthly (USD)	-0.1354 (0.1268) [1]	-0.6015 (0.2750) [62.440]

Note: The table reports endline variables for which treatment effects can change in statistical significance at the 5 percent level using the IHS transformation when there are no control variables, which corresponds to the specification presented in Column (1) of Table S7. We report the coefficient estimate, the standard error in parentheses, and the associated scaling factor in square brackets. When a natural alternative unit results in sign flipping, we report the alternative unit (62.44 Kenyan shillings instead of 1 US dollar for monetary measures, or 0.0362 µg/dL instead of nmol/L for concentration). In other cases, we report the alternative units on the grid $1eX$ for $X \in \{-50, -48, \ldots, 348, 350\}$ that are closest to 1 that result in sign flipping. The sample consists of 2,390 individuals.

Table S10: Scale dependence in RCT data—Binary treatment (1 percent significance level)

	Full sample		No attriters	
	All variables (1)	Endline only (2)	All variables (3)	Endline only (4)
Panel A: Sign flips				
No controls	7–29	7–24	0–8	0–3
Baseline control		10–30		1–7
Village FE	6–41	1–19	7–26	0–5
Baseline control & Village FE		0–18		0–6
Panel B: Significance flips (1 percent level)				
No controls	11–41	11–28	3–38	3–24
Baseline control		4–27		4–29
Village FE	13–40	11–29	7–40	3–26
Baseline control & Village FE		7–29		3–27
Panel C: Sign or significance flips				
No controls	18–47	18–33	3–42	3–27
Baseline control		14–41		5–36
Village FE	19–63	12–33	14–61	3–29
Baseline control & Village FE		7–32		3–31
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 1 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 2,390 individuals in the full sample and 2,272 individuals in the non-attriter sample. See Table 3 for details on the samples and specifications.

Table S11: Scale dependence in RCT data—Continuous treatment (1 percent significance level)

	Full sample		No attriters	
	All variables	Endline only	All variables	Endline only
	(1)	(2)	(3)	(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	8–33	1–11
Baseline control & Village FE		3–14		3–14
Panel B: Significance flips (1 percent level)				
No controls	4–20	2–10	9–25	6–11
Baseline control		1–12		4–13
Village FE	5–19	2–9	4–20	3–10
Baseline control & Village FE		2–10		3–9
Panel C: Sign or significance flips				
No controls	14–67	7–34	14–55	7–24
Baseline control		2–35		9–29
Village FE	16–50	3–20	12–48	4–20
Baseline control & Village FE		5–23		6–22
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. Each cell reports the number of times the treatment effect estimate using an IHS transformation changes in sign or statistical significance at the 1 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. See Table 3 for details on the samples and specifications.