The Stata Journal (2022)
22, Number 3, pp. 702–712 DOI: 10.1177/1536867X221124553

The inverse hyperbolic sine transformation and retransformed marginal effects

Edward C. Norton
University of Michigan
Ann Arbor, MI
and National Bureau of Economic Research
Cambridge, MA
ecnorton@umich.edu

Abstract. In this article, I show how to calculate consistent marginal effects on the original scale of the outcome variable in Stata after estimating a linear regression with a dependent variable that has been transformed by the inverse hyperbolic sine function. The method uses a nonparametric retransformation of the error term and accounts for any scaling of the dependent variable. The inverse hyperbolic sine function is not invariant to scaling, which is known to shift marginal effects between those from an untransformed dependent variable to those of a log-transformed dependent variable, when all observations are positive.

Keywords: st0689, inverse hyperbolic sine, transformation, marginal effects

1 Introduction

The inverse hyperbolic sine function is widely used in empirical research to transform the dependent variable. The motivation to use this transformation is that it allows for nonpositive values and can reduce the influence of outliers in a right-skewed distribution. The natural logarithm transformation is often used for skewed distributions, but $\ln(y)$ is not defined when y is zero or negative. Linear regression can be used for dependent variables that span the entire real line, but the results can be greatly influenced by outliers if the distribution of y is skewed. The inverse hyperbolic sine transformation is one alternative to untransformed linear regression that can potentially solve both problems.

However, the inverse hyperbolic sine transformation creates additional challenges. The first is that in a linear regression with an inverse hyperbolic sine-transformed dependent variable, the estimated coefficients have no intrinsic meaning. It is necessary to retransform the predicted values back to the original scale to calculate quantities of interest, which are typically marginal effects. The retransformation back to the original scale of the outcome variable is not trivial (Manning 1998). In this article, I show how to estimate marginal effects on the original scale in Stata after retransforming results from a linear regression with an inverse hyperbolic sine-transformed dependent variable. I apply Duan's (1983) nonparametric smearing estimate to the inverse hyperbolic sine retransformation to get both marginal effects and predicted values on the original scale.

The second challenge is that the inverse hyperbolic sine is not invariant to scaling (Aihounton and Henningsen 2021). In contrast, linear regression with an untransformed dependent variable *is* invariant to scaling, meaning that changing the units of the dependent variable between, say, dollars, pennies, euros, pounds, and yuans will not change the final interpretation. The regression coefficients adjust predictably. Unlike linear regression with an untransformed dependent variable, an inverse hyperbolic sine-transformed dependent variable is sensitive to scaling. The marginal effects on the original scale will change if that dependent variable is rescaled (for example, from dollars to pennies). The two scaling extremes—dividing or multiplying the dependent variable by a large number before applying the inverse hyperbolic sine transformation—will reproduce marginal effects on the original scale that are either equal to marginal effects from a linear regression with an untransformed dependent variable or equal to marginal effects from a log-transformed model, when all observations are positive.

I show how to estimate marginal effects on the original scale of the outcome variable for the inverse hyperbolic sine model over a wide range of scaling factors in Stata and compare the results. Although the inverse hyperbolic sine transformation can also be used for an explanatory variable, such transformations are beyond the scope of this study (see Bellemare, Barrett, and Just [2013] for one example).

2 Inverse hyperbolic sine

The inverse hyperbolic sine function, also known as the area hyperbolic sine function (denoted `asinh()`), is the natural logarithm of y plus an additional term equal to the square root of y^2 plus 1. The inverse hyperbolic sine function is

$$\sinh^{-1}(y) = \ln\left(y + \sqrt{y^2+1}\right)$$

The inverse hyperbolic sine function has several nice properties. It passes through the origin because, when $y = 0$, then $\ln(1) = 0$. The inverse hyperbolic sine function is symmetric around 0, meaning that $\sinh^{-1}(y) = -\sinh^{-1}(-y)$. For large values of y, $\sinh^{-1}(y)$ is approximately equal to $\ln(y)$ plus a constant [$\ln(2) \approx 0.693$].

$$\text{For } y \gg 0 : \sinh^{-1}(y) \approx \ln(2y) = \ln(y) + \ln(2)$$

The derivative of the inverse hyperbolic sine function shows where that function is similar in slope to the identity and log transform functions. The derivative of $\sinh^{-1}(y)$ with respect to y is the inverse of $\sqrt{y^2+1}$.

$$\frac{d}{dy}\left\{\ln\left(y + \sqrt{y^2+1}\right)\right\} = \frac{1}{\sqrt{y^2+1}}$$

The graph of $\sinh^{-1}(y)$ against y has three distinct regions, as shown in figure 1. As y gets large, the derivative of $\sinh^{-1}(y)$ approaches $1/y$, which is the derivative of $\ln(y)$.

As y approaches 0, the derivative of $\sinh^{-1}(y)$ approaches 1, which is the slope of the untransformed line. When y is negative, $\sinh^{-1}(y)$ is equal in magnitude and opposite in sign to $\sinh^{-1}(|y|)$. Therefore, when y is large, the marginal effects for the inverse hyperbolic sine and log transformations will be nearly the same, and when y is small, the marginal effects for the inverse hyperbolic sine and identity transformations will be nearly the same (Aihounton and Henningsen 2021).

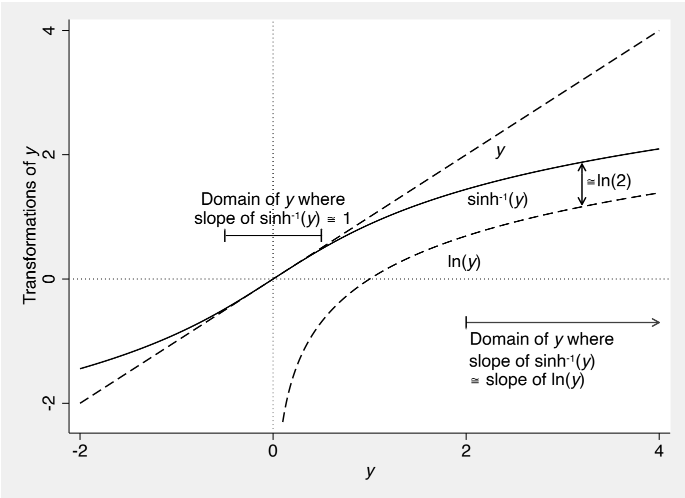

Figure 1. Three transformations of y (identity, inverse hyperbolic sine, and natural log) and the domain of y over which the slopes of the transformations are nearly identical

Suppose that you have a continuous outcome that has both positive and negative values and has a right-skewed distribution. Perhaps it is a financial variable like net income or wealth. One could estimate a linear regression with an inverse hyperbolic sine-transformed dependent variable. Let y be a continuous outcome determined by a vector of covariates $\mathbf{x}$ with a corresponding vector $\boldsymbol{\beta}$ of unknown parameters to be estimated. Let i denote individual observations, and let the error term be ϵ.

$$\sinh^{-1}(y_i) = \mathbf{x}_i'\boldsymbol{\beta} + \epsilon_i$$

If the goal is to estimate marginal effects on the original scale, then how should one proceed? One alternative is to compute elasticities. Bellemare and Wichman (2020) derive the elasticities for cases where the dependent variable y, the independent variables $\mathbf{x}$, or both y and $\mathbf{x}$ are transformed by the inverse hyperbolic sine function.

Another alternative is to fit the model using generalized method of moments (GMM). The advantage of GMM is that it avoids retransformation. Mullahy (2021) shows how to fit a GMM model with an inverse hyperbolic sine-transformed dependent variable. Unfortunately, Stata has trouble fitting inverse hyperbolic sine models with GMM when y is large—above, say, 50. For typical financial data, fitting a GMM model in Stata requires rescaling by dividing by a large number. But this returns results that are essentially the same as the untransformed model.

If the dependent variable has only positive values, then the inverse hyperbolic sine transformation is not necessary. There are well-established methods that use either the log-transformation or generalized linear models with one of several possible transformations, including the log (Manning 1998; Manning and Mullahy 2001; Deb, Norton, and Manning 2017).

Most applied econometricians are interested in estimating conditional marginal effects, so I will show how to do that next for an inverse hyperbolic sine-transformed dependent variable.

3 Duan's smearing estimate for inverse hyperbolic sine

After estimating a linear regression with an inverse hyperbolic sine-transformed dependent variable, how should one interpret the results? In particular, how can one calculate marginal effects of covariates? The coefficients are not directly interpretable as marginal effects, as they are for an untransformed linear regression. Nor are the coefficients semielasticities, as they are for a log-transform regression. I will show how to retransform the results using the hyperbolic sine function and applying Duan's (1983) smearing estimate.

The hyperbolic sine function—the inverse of the inverse hyperbolic sine function—is half the difference of two exponential terms.

$$\sinh(x) = \frac{e^x - e^{-x}}{2} \tag{1}$$

Substitute $\mathbf{x}'\boldsymbol{\beta} + \epsilon$ for x in (1), and take expectations to derive the following expression for the expected value of y on the original scale, given $\mathbf{x}$.

$$E(y|\mathbf{x}) = E\left\{\sinh(\mathbf{x}'\boldsymbol{\beta} + \epsilon)|\mathbf{x}\right\} \tag{2}$$

Duan (1983) showed how to calculate a consistent estimate of the expected value of the outcome on the original scale after fitting a linear regression model with a transformed dependent variable. His method has been widely applied to the natural logarithm transformation (Manning 1998; Manning and Mullahy 2001). Duan's proof applies not only to log transformations but also to any smooth distribution. Specifically, the key assumption for the consistency of Duan's smearing estimate is that the retransformation function is continuously differentiable, which $\sinh(\cdot)$ is.

Following Duan (1983), instead of integrating over the unknown distribution of the error term, use the empirical cumulative distribution function by averaging (2) over the estimated residuals, and substitute the least-squares estimates of the parameters $\widehat{\boldsymbol{\beta}}$. Let the sample size be N.

$$\widehat{E}(y|\mathbf{x}) = \frac{1}{N}\sum_{i=1}^{N}\sinh\left(\mathbf{x}'\widehat{\boldsymbol{\beta}} + \widehat{\epsilon}\right)$$
$$= \frac{1}{2N}\sum_{i=1}^{N}\left(e^{\mathbf{x}'\widehat{\boldsymbol{\beta}}}e^{\widehat{\epsilon}_i} - e^{-\mathbf{x}'\widehat{\boldsymbol{\beta}}}e^{-\widehat{\epsilon}_i}\right)$$

In practice, this is done in two steps. First, let $D = E(e^{\epsilon})$. Although D is unknown because ϵ has an unknown distribution, the population mean of D can be estimated by the sample mean.

$$\widehat{D} = \frac{1}{N}\sum_{i=1}^{N} e^{\widehat{\epsilon}_i}$$

Second, substitute $\widehat{D}$ into (2), and rearrange terms to get Duan's smearing estimate for the retransformation of the inverse hyperbolic sine-transformed linear regression.

$$\widehat{E}(y|\mathbf{x}) = \frac{1}{2}\left(e^{\mathbf{x}'\widehat{\boldsymbol{\beta}}}\widehat{D} - e^{-\mathbf{x}'\widehat{\boldsymbol{\beta}}}\widehat{D}^{-1}\right) \tag{3}$$

The marginal effect of a change in a continuous variable x_1 with a corresponding coefficient β_1 is the derivative of (3) and is always positive (notice that the two terms are now added, not subtracted).

$$\frac{d\widehat{E}(y|\mathbf{x})}{dx_1} = \frac{1}{2}\left(\widehat{\beta}_1 e^{\mathbf{x}'\widehat{\boldsymbol{\beta}}}\widehat{D} + \widehat{\beta}_1^{-1} e^{-\mathbf{x}'\widehat{\boldsymbol{\beta}}}\widehat{D}^{-1}\right)$$

One advantage to Duan's approach is that it is easy to estimate in Stata, as shown in the next section. A limitation is that it assumes that the variance is homoskedastic. Manning (1998) discusses how to adjust Duan's smearing estimate when there is heteroskedasticity by group.

4 Code for marginal effects

This section shows example code to estimate marginal effects on the original scale after estimating a linear regression with an inverse hyperbolic sine-transformed dependent variable, using Duan's (1983) smearing estimate. Stata refers to the inverse hyperbolic sine function as `asinh()`. The example code assumes that the dependent variable is `y` and that there are three covariates (`x1`, `x2`, and `x3`); those would be changed by the user. In addition to marginal effects, the code also calculates predicted values of `y` by generating a new variable, `yhat_ihs`, that is also based on retransformed results with Duan's smearing estimate.

```
* Example code to fit IHS model and retransformed marginal effects
generate y_ihs = asinh(y)          // replace y with outcome
regress y_ihs x1 x2 x3, vce(robust) // replace x1-x3 with covariates
predict xbhat_ihs, xb
predict double ehat, residual
egen duan = mean(exp(ehat))
margins, dydx(*) expression(.5*(exp(xb())*duan - (1/(exp(xb())*duan))))
generate yhat_ihs = .5*(exp(xbhat_ihs)*duan - (1/(exp(xbhat_ihs)*duan)))
```

5 Code for multiple scaling factors

Next I show how to incorporate scaling into the marginal effects calculations. It is well known that inverse hyperbolic sine is sensitive to scaling because the inverse hyperbolic sine transformation is not scale invariant (Aihounton and Henningsen 2021). It is best to think of the scale factor as one additional parameter to the model. Aihounton and Henningsen (2021) discuss ways to choose the optimal scaling parameter.

The example code below is similar to the example code for the basic inverse hyperbolic sine retransformation but also allows for a scaling factor.

```
* Example code to fit scaled IHS model and retransformed marginal effects
scalar scale = .001                  // replace scaling factor
generate y_ihs_scale = asinh(scale*y)
 (code omitted)
margins, dydx(*) expression(.5*(exp(xb())*duan - (1/(exp(xb())*duan)))/scale)
generate yhat_ihs = .5*(exp(xbhat_ihs)*duan - (1/(exp(xbhat_ihs)*duan)))/scale
```

One way to compare results across different scaling factors is to fit several models and then compare the marginal effects and their standard errors using `estimates table`. The following program allows for such comparisons.

```
* Example program to fit IHS models with several scale
* factors and to compare the marginal effects
* Must have declared $y and $xvars as global variables
capture program drop ihs
program define ihs
    args scale name
    tempvar ihs_y ehat duan
    generate `ihs_y' = asinh(`scale'*$y)
    regress  `ihs_y' $xvars, vce(robust)
    predict  `ehat', residual
    egen `duan' = mean(exp(`ehat'))
    margins, dydx(*) expression(.5*(exp(xb())*`duan' ///
        - (1/(exp(xb())*`duan')))/`scale') post
    estimates store `name'
end
```

After defining the program `ihs`, one can use it to compare models with different scaling factors by specifying both the scaling factor and a name for stored results, as shown below in example code.

```
* Example use of ihs; arguments are scale and name
ihs .000001 mil_th
ihs .001    thou_th
ihs 1       one
ihs 100     hundred
estimates table mil_th thou_th one hundred, b(%7.2f) se(%7.2f)
```

The example code in this section, slightly modified, was used for the empirical example using Medical Expenditure Panel Survey (MEPS) in the next section.

6 Example using MEPS data

This empirical example predicts family income for a sample of 115,009 persons in the 2008–2014 MEPS, a national survey on the financing and use of medical care in the United States. Family income ranges from −182,078 to 556,128, has a median value of $47,439, and is right-skewed (see figure 2). Therefore, it is reasonable to consider transforming family income by the inverse hyperbolic sine function.

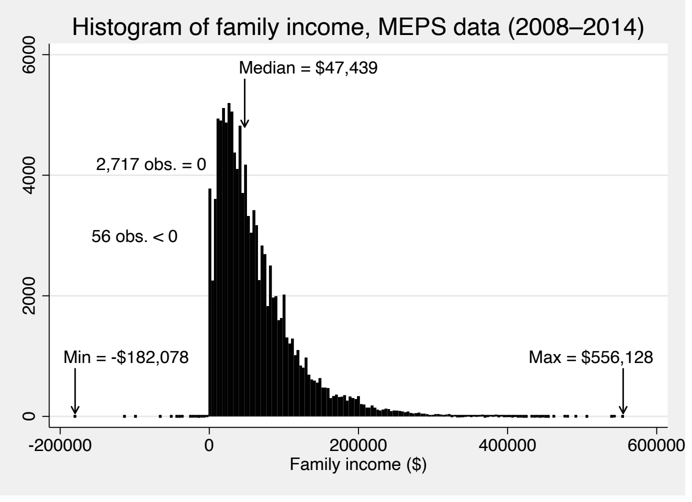

Figure 2. The histogram of family income shows a skewed distribution of positive values, with many values of zero and some negative values

The sample includes persons aged 25–65. For illustrative purposes, the simple model specification is a function of just age, gender, and the highest level of education achieved (four categorical values). The mean age is 44, more than half are women, a quarter did not complete high school, and half have a high school diploma as their highest level of education.

```
. summarize $y ihs_y ln_y age female i.education
```

Variable	Obs	Mean	Std. dev.	Min	Max
faminc	115,009	62304.58	55366.43	-182078	556128
ihs_y	115,009	11.10619	2.027033	-12.80534	13.9219
ln_y	112,236	10.6924	.9615114	1.098612	13.22875
age	115,009	44.08338	11.48955	25	65
female	115,009	.5392274	.498461	0	1
education					
No HS deg	115,009	.2642315	.4409251	0	1
HS degree	115,009	.5012303	.5000007	0	1
College deg	115,009	.1567269	.3635446	0	1
Grad. deg	115,009	.0778113	.2678756	0	1

The results compare the estimated marginal effects on the original scale for linear regressions with either an inverse hyperbolic sine-transformed dependent variable (`me_ihs`) or an untransformed dependent variable (`me_y`). The estimated coefficients of the inverse hyperbolic sine model are in the first column for completeness. The marginal effects of the inverse hyperbolic sine model are roughly half again to double the size, in absolute value, compared with those for the untransformed ordinary least-squares model. The marginal effects should be similar only if the dependent variable has been rescaled by multiplying by a tiny number, which it has not.

```
. * Results comparing IHS betas and marginal effects with untransformed OLS
. estimates table beta_ihs me_ihs me_y, b(%10.3f) se(%10.3f)
```

Variable	beta_ihs	me_ihs	me_y
age	0.009	615.570	450.733
	0.001	35.395	12.872
female			
Female=1	-0.232	-15797.450	-6239.074
	0.012	795.524	301.452
education			
HS degree	0.437	20005.268	11635.094
	0.016	680.604	311.280
College deg	1.133	76784.804	45697.617
	0.017	1327.375	528.542
Grad. deg	1.436	116905.241	71019.401
	0.019	2072.207	839.017
_cons	10.322		
	0.027		

Legend: b/se

Next are comparisons for five different rescaled inverse hyperbolic sine models along with both an untransformed model and a log-transformed model. Because $\ln(y)$ is not defined for nonpositive values of y, I dropped the 2,773 observations with nonpositive values of family income.[1] The new sample has 112,236 observations. Other than dropping the left tail of the family income distribution, the other summary statistics did not change appreciably. However, dropping the 2.5% of observations with the lowest values of the dependent variable does change the average marginal effects. The marginal effects from the two different samples cannot be compared.

The five scaling factors are 0.000000001 ("trillionth"), 0.000001 ("millionth"), 0.001 ("thousandth"), 0.1 ("tenth"), and 10 ("ten"). For this dataset and model specification, these five scaling factors give marginal effects that span from untransformed y to $\ln(y)$. The results show that scaling by one trillionth yields marginal effects and standard errors that are identical (to a few digits) to those of the untransformed model (see left two columns of the results table). Moving to the right side of the table, we see that scaling by multiplying by 10 yields marginal effects and standard errors that are identical to those of the $\ln(y)$ model.

```
. * Compare marginal effects across models
. estimates table me_y tril_th mil_th thou_th tenth ten me_lny, b(%7.1f)
> se(%7.1f)
```

Variable	me_y	tril_th	mil_th	thou_th	tenth	ten
age	452.2	452.2	450.2	450.0	448.5	448.4
	13.0	13.0	12.9	15.3	15.8	15.8
female						
Female=1	-5978.9	-5978.9	-5964.8	-8866.0	-8972.8	-8975.8
	304.1	304.1	301.7	342.2	350.9	351.2
education						
HS degree	11222.0	11222.0	11209.9	13478.2	13461.1	13458.9
	315.7	315.7	313.7	314.9	322.3	322.5
College deg	44788.2	44788.2	44634.1	53359.8	53445.9	53443.5
	530.7	530.7	526.5	627.9	647.0	647.7
Grad. deg	69924.0	69924.0	69595.4	82569.2	82848.8	82852.6
	839.5	839.5	831.4	1039.9	1065.1	1065.8

Legend: b/se

1. This is done to allow only for direct comparisons across the different models including $\ln(y)$; one of the motivations for inverse hyperbolic sine is the ability to include zero and negative values of y.

Variable	me_lny
age	448.4
	15.8
female	
Female=1	-8975.8
	351.2
education	
HS degree	13458.9
	322.5
College deg	53443.5
	647.7
Grad. deg	82852.6
	1065.8

Legend: b/se

The marginal effects for age do not change appreciably with changes in scaling. However, the other marginal effects change considerably with different scaling. For a real research article, instead of this illustration, it would be important to compare the fit of the various models as a function of the scaling parameter and make an informed choice about which model specification is best (Aihounton and Henningsen 2021). Any comparison, however, of log likelihoods of models with different dependent variables would need to add the Jacobian term relevant for the transformation to its respective log likelihood (Bellemare and Wichman 2020).

The code and data are available from the author upon request.

7 Conclusions

The inverse hyperbolic sine transformation is gaining popularity because it is easy to estimate with linear regression and allows for a skewed dependent variable that takes on zero and negative values.

In practice, it can be hard to estimate the marginal effects on the original scale. The estimated coefficients are not marginal effects, nor are they semielasticities like $\ln(y)$. As has been known for years, if your dependent variable has been transformed but you want to interpret on the original scale, then you must retransform the results using Duan's smearing estimate. Also note that the inverse hyperbolic sine function is not scale invariant. Scaling from low values (multiply by tiny positive number) to high values (multiply by large number) changes the marginal effects at the extremes from ordinary least squares to $\ln(y)$.

In this article, I showed how to retransform in Stata the inverse hyperbolic sine model results and calculate marginal effects and their confidence intervals on the original scale. I also showed how to compare results across multiple scaling factors.

8 Acknowledgment

I thank John Mullahy for helpful discussions and comments on an earlier version.

9 References

Aihounton, G. B. D., and A. Henningsen. 2021. Units of measurement and the inverse hyperbolic sine transformation. *Econometrics Journal* 24: 334–351. https://doi.org/10.1093/ectj/utaa032.

Bellemare, M. F., C. B. Barrett, and D. R. Just. 2013. The welfare impacts of commodity price volatility: Evidence from rural Ethiopia. *American Journal of Agricultural Economics* 95: 877–899. https://doi.org/10.1093/ajae/aat018.

Bellemare, M. F., and C. J. Wichman. 2020. Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics* 82: 50–61. https://doi.org/10.1111/obes.12325.

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

Duan, N. 1983. Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association* 78: 605–610. https://doi.org/10.2307/2288126.

Manning, W. G. 1998. The logged dependent variable, heteroscedasticity, and the retransformation problem. *Journal of Health Economics* 17: 283–295. https://doi.org/10.1016/S0167-6296(98)00025-3.

Manning, W. G., and J. Mullahy. 2001. Estimating log models: To transform or not to transform? *Journal of Health Economics* 20: 461–494. https://doi.org/10.1016/S0167-6296(01)00086-8.

Mullahy, J. 2021. Inverse hyperbolic sine transformations and retransformations in regression models. Working paper, University of Wisconsin–Madison.

About the author

Edward C. Norton is a professor in the Department of Health Management and Policy and in the Department of Economics at the University of Michigan, and he is a research associate at the National Bureau of Economic Research.