Série des Documents de Travail

n° 2019-13

# Dealing with the log of zero in regression models

Christophe BELLEGO[1]
Louis-Daniel PAPE[2]



[1] CREST; ENSAE. E-mail : Christophe.Bellego@ensae.fr
[2] CREST; Ecole Polytechnique. E-mail : Louis.PAPE@ensae.fr

# DEALING WITH THE LOG OF ZERO IN REGRESSION MODELS*

Christophe BELLÉGO — *CREST - ENSAE*

Louis-Daniel PAPE — *CREST - École Polytechnique*

August 28, 2019

**Abstract**

Log-linear and log-log regressions are one of the most used statistical model. However, handling zeros in the dependent and independent variable has remained obscure despite the prevalence of the situation. In this paper, we discuss how to deal with this issue. We show that using Pseudo-Poisson Maximum Likelihood (PPML) is a good practice compared to other approximate solutions. We then introduce a new complementary solution to deal with zeros consisting in adding a positive value specific to each observation that avoids some numerical issues faced by the former.

**Keywords:** Log(0), Log of zero, Log-log, Bias, Elasticity, PPML

*We thank David Benatia, Laurent Davezies, Xavier D'Haultfoeuille, Alain Trognon, and Michael Visser for insightful discussions and comments. Any remaining errors are ours.

# 1 INTRODUCTION

Econometric specifications often involve log-transformations for three main reasons. (1) A log-log relationship leads to an elasticity $\beta$.[1] (2) The log can also linearize a theoretical model (e.g., a Cobb Douglas production function) or appear in a structural model (e.g, demand estimation using data aggregated at the market level). (3) It can be used to reduce heteroskedasticity within feasible generalized least squares procedure.

However, it often occurs that the variable taken in log contains non-positive values. For instance, a company can employ no worker, a product can have zero sales in a given market, or two countries have zero trade in a given year. Measurement errors can also generate non-positive values. In this case, the log is undefined and a fix is needed. Although this problem is quite common, the solution to be adopted is still unclear. For instance, the question "Log transformation of values that include 0 (zero) for statistical analyses?" on the forum of ResearchGate has been opened in 2014. This question has received 89 contributions and has been read about 64.000 times (at the time of April 2019), revealing the magnitude of the issue, which goes way beyond the economic field.[2]. Among these 89 contributions, there are 38 different users proposing their "personal approach" to this issue, which we classified in six categories (detailed in the next sections) in Graph 1 to show the absence of consensus. This suggests that the best practice may have failed to reach a broad audience. In addition, some methods used in practice can be misleading.

[1] Indeed,in a log-log regression such that $\log(y) = \beta \log(x) + \epsilon$, we have $\frac{\partial \log(y)}{\partial \log(x)} = \frac{\partial y}{\partial x}\frac{x}{y} = \beta$.

[2] See `https://www.researchgate.net/post/Log_transformation_of_values_that_include_0_zero_for_statistical_analyses2`. Six months earlier, in November 2018, this thread had received 77 answers and had been read 48.000 times. The 33% increase in viewership reveals an ongoing interest in this issue.

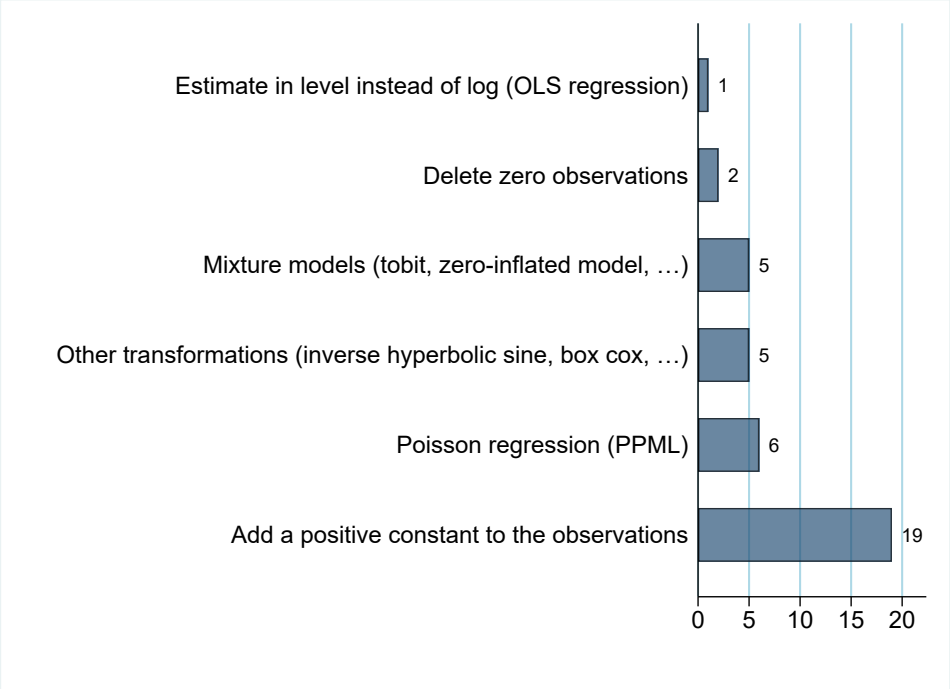


Figure 1: Number of "personal solutions" by category on the ResearchGate forum

This paper's contribution is to clarify this practical issue in a didactic way. To do so, we discuss the best practices to handle non-positive values in the independent variable and the dependent variables when one wants to estimate a log or a log-log regression. First, we review the common naive methods that may be used by practitioners and that generate biases. Second, we explain why PPML is a good solution to handle this issue without bias. Third, we propose a new solution that can easily be implemented and where the coefficient $\beta$ can directly be interpreted as a (semi-) elasticity. It consists in adding a positive value specific to each observation, in the spirit of what is commonly done in practice. This method can complement PPML proposed by Santos Silva and Tenreyro (2006) when the latter faces computational difficulties with large dependent variables. Finally, we extend the framework to log-log regressions where the independent variable can contain zeros.

## 2 Common misconceptions among practitioners

Several solutions were adopted by practitioners to work around this issue. A first solution is to delete the non-positive observations (Young and Young (1975)). However, this radical solution might introduce in turn selection bias if the occurrence of non-positive values is not random and if the number of concerned observation is high. For

instance, when analyzing the link between doctor visits and health, healthy patients are more likely to have zero appointments with their doctor during an observation period. Deleting these observations would change the scope of the study, narrowing it to the less healthy patients.

A second solution uses log-like transformations that can be applied with non-positive observations. A popular one is the inverse hyperbolic sine transformation introduced by Johnson (1949), and developed by Burbidge et al. (1988).[3] The inverse hyperbolic sine transformation writes $f(y) = \log(y + (y^2+1)^{0.5})$, which tends toward $\log(2y)$ as $y$ increases. Its similarity with the log function has led some to believe that they can be used interchangeably. However, for small $y$, this transformation can behave differently than the log function. Besides, as shown in Bellemare and Wichman (2019), the interpretation of the estimated coefficients is not trivial and the underlying elasticity is potentially biased or undefined.[4] Moreover, if the underlying model naturally writes in log, then this transformation is still an approximation that is likely to bias the estimated coefficients. For recent applications of this method, see McKenzie (2017) or Card and DellaVigna (2017).

A third solution is to add a positive constant $c$ to all observations $Y$ so that $Y + c > 0$ (see for instance MaCurdy and Pencavel (1986), Rogowski and Newhouse (1992) or Criscuolo et al. (2019)). It is not easy to find papers clearly stating that they implemented this transformation. We think that concretely, many practitioners use this solution without even mentioning it because they think that adding a very small constant is not going to be harmful. However, the choice of the constant is discretionary and likely to bias the coefficient estimates. Contrary to linear regressions, log-linear regressions are not robust to linear transformation of the dependent variable. This is due to the non-linear nature of the log function. Log transformation expands low values and squeezes high values. Therefore, adding a constant will distort the (linear) relationship between zeros and other observations in the data. The magnitude of the bias generated by the constant actually depends on the range of observations in the data. For that reason, adding the smallest possible constant is not necessarily the best worst solution. A variant of that solution consists in adding a constant value only to the problematic observations that are non-positive. Then, the model is often estimated rather by adding a dummy variable to indicate such a treatment, but this alternative

[3] An extended concave version of this transformation has recently been provided by Ravallion (2017).

[4] In particular, he shows that if one estimates $f(y) = \alpha + \beta x + \epsilon$, then the elasticity writes $\hat{\zeta}_{yx} = \hat{\beta} x . \frac{\sqrt{y^2+1}}{y}$. This elasticity is not defined for $y = 0$, and to interpret $\beta$ as the elasticity, one needs $x = 1$ and large values of $y$.

practice generates the same kind of trouble.[5]

To understand how this solution biases the estimated coefficients, consider the following model. Let us denote $N$ the number of observations. For observation $i$, we denote the dependent variable by $y_i$ and the explanatory variables by the vector $x_i$ of size $K \times 1$. The vector of parameters is $\beta$. There is also a constant denoted $\alpha$. $\epsilon_i$ is zero-mean error term assumed to be independently distributed from one another, independent from $x_i$, and that $\mathbb{E}(\epsilon_i|x_i) = 0$. Suppose that one wants to estimate the following functional form.

$$\log(y_i) = \alpha + x_i'\beta + \epsilon_i \tag{2.1}$$

If there exists at least one $y_i = 0$, this function is not defined. The popular fix consists in adding a constant to the dependent variable, that we denote by $\Delta$. Exponentiating on both sides, and rearranging, we get

$$y_i + \Delta = exp(\alpha + x_i'\beta + \epsilon_i) + \Delta \tag{2.2}$$

So that

$$\log(y_i + \Delta) = \log(\exp(\alpha + x_i'\beta + \epsilon_i) + \Delta) \tag{2.3}$$

We can re-arrange to get[6]

$$\log(y_i + \Delta) = \alpha + x_i'\beta + \epsilon_i + \log\left(1 + \frac{\Delta}{\exp(\alpha + x_i'\beta + \epsilon_i)}\right) \tag{2.4}$$

In practice, this is equivalent to running the following regression with ordinary least squares with a new error term $\omega_i$

$$\log(y_i + \Delta) = \alpha + x_i'\beta + \omega_i \tag{2.5}$$

$$\text{with} \quad \omega_i = \epsilon_i + \log\left(1 + \frac{\Delta}{\exp(\alpha + x_i'\beta + \epsilon_i)}\right) \tag{2.6}$$

Ordinary least squares provides unbiased coefficients if $\mathbb{E}(\omega_i|x_i) = 0$. However, $\log\left(1 + \frac{\Delta}{\exp(\alpha + x_i'\beta + \epsilon_i)}\right)$ is unobserved and correlated with $x_i$. This is in contradiction with the hypothesis that $\omega_i$ is exogenous. [7]

[5] Johnson and Rausser (1971) propose a solution where the constant variable that is added to non-positive observations is treated as an additional parameter to be estimated simultaneously with other parameters. However, their method is not built to obtain unbiased values of the other parameters, it just maximizes the fit of the model.

[6] We use the following: $\log(a + b) = log\left(a\left(1 + \frac{b}{a}\right)\right)$

[7] Indeed, for the $k^{th}$ variable $\text{COV}(x_{ik}, \omega_i) = \text{COV}\left(x_{ik}, \epsilon_i + \log\left(1 + \frac{\Delta}{\exp(\alpha + x_i'\beta + \epsilon_i)}\right)\right) =$

One may nonetheless believe this bias to be negligible for small values of the constant $\Delta$. Through Monte Carlo simulation, we show this to be inexact. We draw 100 000 observations from a Poisson distribution $\mathscr{P}_{oisson}(\lambda_i)$, where the conditional moment $\mathbb{E}(y_i|x_i) = \lambda_i$, and $\lambda_i$ is parameterized as $\lambda_i = \exp(1 + x_{i,1})$. $x_1$ is drawn from a standard uniform distribution. We vary $\Delta$ on the regression $\log(y_i + \Delta) = x_i\beta + \alpha + \epsilon_i$. Figure 2 presents the relative absolute bias of the estimates as a function of the value of $\Delta$: $100 \times |\frac{\hat{\beta}-\beta}{\beta}|$. It shows that there may exist an optimal $\Delta^*$ but that it is not necessarily the smallest possible value for $\Delta$, contrary to common belief. Moreover, the size of the bias may be substantial.

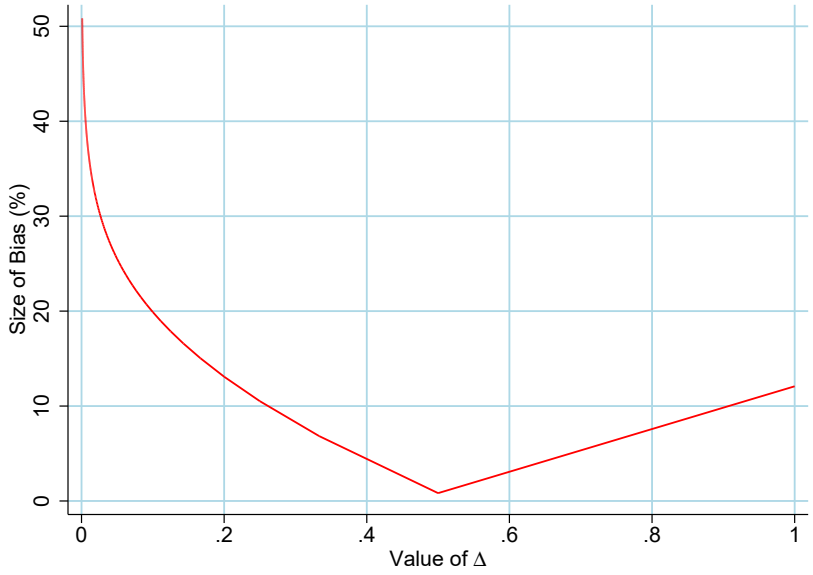


Figure 2: Bias against $\Delta$

## 3 Recommended Practice

This section presents what, we argue, should be considered a good practice. Although this is not their original goal, Santos Silva and Tenreyro (2006) proposed to use Poisson Pseudo Maximum Likelihood (PPML) as a potential solution. They consider the

$\text{COV}\left(x_{ik}, \log\left(1 + \frac{\Delta}{\exp(\alpha + x_i'\beta + \epsilon_i)}\right)\right) \neq 0$

following process.[8]

$$y_i = a_i \exp(\alpha + x_i'\beta) \quad \text{with} \quad \mathbb{E}(a_i|x_i) = 1 \tag{3.1}$$

This process is motivated by several features. First, it provides the same interpretation to $\beta$ as a semi-log model.[9] Second, this data generating process provides a logical rationalization of zero values in the dependent variable. This situation can arise when the multiplicative error term, $a_i$, is equal to zero. It corresponds to the case where $\epsilon_i \to -\infty$ in a log-model such as $\log(y_i) = \alpha + x_i'\beta + \epsilon_i$[10], which contradicts the existence of a finite conditional expectation of the error term. Third, estimating this model with PPML does not encounter the computational difficulty when $y_i = 0$. Indeed, under the assumption that $\mathbb{E}(a_i|x_i) = 1$, we have $\mathbb{E}(y_i - \exp(\alpha + x_i'\beta)|x_i) = 0$. We want to minimize the quadratic error of this moment, leading to the following first-order conditions:

$$\sum_{i=1}^{N} \big(y_i - \exp(\alpha + x_i'\beta)\big) x_i' = 0 \tag{3.2}$$

As it clearly appears, these conditions are defined even when $y_i = 0$. These first-order conditions are numerically equivalent to those of a Poisson (Gourieroux et al. (1984)). Furthermore, as originally intended in Santos Silva and Tenreyro (2006), this solution is robust to heteroscedasticity in $a_i$. This estimation can be easily implemented with standard econometric software. For instance, in Stata, the package *ppml* (Santos Silva and Tenreyro (2015)) is straightforward to use and can account for fixed effects when they are not too many of them. The package *ppmlhdfe* (Correia et al. (2019)) allows one to deal with high-dimensional fixed effects. A recent example of application can be found in Head and Mayer (2019).

To exemplify the method, we simulate data with $y_i = 0$ from a Poisson distribution $\mathscr{P}_{oisson}(\lambda_i)$, where $\lambda_i$ is parameterized as $\lambda_i = \exp(-1 + 0.5x_{i,1} + 1.5x_{i,2})$. $x_1$ and $x_2$ are drawn from standard uniform distributions. We compare the estimated coefficients obtained by PPML with those obtained when we added a constant $\Delta = 0.01$, or a constant $\Delta = 1$ to $y_i$, as suggested by McCune et al. (2002) for count-data observations.

[8] Moreover, other processes than those considered in this paper can generate zero values in the dependent variable. Santos Silva et al. (2015) propose a procedure to discriminate between these competing specifications.

[9] To be clear, with $y_i = a_i \exp(\alpha + x_i'\beta)$, then $\frac{\partial y}{\partial x}\frac{1}{y} = \beta$, the model directly identifies the semi-elasticity. To obtain an elasticity, one has just to consider $y_i = a_i \exp(\alpha + \log(x_i)\beta)$, then $\frac{\partial y}{\partial x}\frac{x}{y} = \beta$.

[10] Or, equivalently, that $a_i \to 0$ if we note $\epsilon_i = \log(a_i)$

Table 1 presents the results. The coefficients obtained from PPML provides the lowest bias among all three options.

Table 1: Comparison of the PPML with misconceived tricks on simulated data

| | Real Coefficients | OLS:$\Delta = 0.01$ | | OLS:$\Delta = 1$ | | PPML | |
|---|---|---|---|---|---|---|---|
| x1 | 0.500 | 0.874*** | (0.0806) | 0.221*** | (0.0168) | 0.470*** | (0.0336) |
| x2 | 1.500 | 2.969*** | (0.0764) | 0.728*** | (0.0164) | 1.550*** | (0.0349) |
| $\alpha$ | -1.000 | -3.393*** | (0.0611) | 0.122*** | (0.0121) | -1.026*** | (0.0301) |
| Observations | 10000 | 10000 | | 10000 | | 10000 | |

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

In some cases, this method faces computational difficulties, as evidenced in Santos Silva and Tenreyro (2011). For instance, when the dependent variable has some large values, the PPML does not converge in Stata. Santos Silva and Tenreyro (2011) suggest to normalize the dependent variable to solve this issue. However, standards normalization does not always work, which could lead to misleading conclusions. We exemplify this in the next Session.

## 4 A New Complementary Solution

We provide a new complementary solution which by-passes this issue and provides robustness against computational biases generated by arbitrary normalization methods. Given that the most intuitive way to handle zeros amounts to adding a positive constant to the data points, our solution consists in adding an optimal positive value $\Delta_i$ (which varies over observations) that does not generate correlation between the error term and the regressors. This solution does not require the deletion of observations or the estimation of a supplementary parameter or the addition of a discretionary constant. It only relies on the independence between the error term and the covariates.

Consider the following data generating process such that $\mathbb{E}(\epsilon_i) = 0$ and $\epsilon_i$ independent from $x_i$,

$$\log(y_i) = \alpha + x_i'\beta + \epsilon_i \tag{4.1}$$

We now exponentiate and add a constant $\Delta_i$ to the model,

$$y_i + \Delta_i = \exp(\alpha + x_i'\beta + \epsilon_i) + \Delta_i \tag{4.2}$$

If we let $\Delta_i = \exp(x_i'\beta)$, then the previous equation can be rewritten as

$$y_i + \exp(x_i'\beta) = \exp(\alpha + x_i'\beta + \epsilon_i) + \exp(x_i'\beta) = \exp(x_i'\beta)(1 + \exp(\alpha + \epsilon_i)) \quad (4.3)$$

And since $y_i + \exp(x_i'\beta)$ is strictly non-negative for any observation, we get :

$$\log(y_i + \exp(x_i'\beta)) = x_i'\beta + \log(1 + \exp(\alpha + \epsilon_i)) \quad (4.4)$$

Let us denote $\eta_i = \log(1 + \exp(\alpha + \epsilon_i))$, then we only need to estimate [11]

$$\log(y_i + \exp(x_i'\beta)) = x_i'\beta + \eta_i \quad (4.5)$$

This new residual term does not depend on the independent variables by virtue of the independence assumption [12]. Therefore, the parameter $\beta$ can be identified using all observations, included those where $y = 0$. Note that $\eta_i$ does not have a zero mean expected value. However, we can write $\xi_i = \eta_i - \lambda$ where $\lambda$ is a constant non-negative value such that $\mathbb{E}(\xi_i) = 0$. Thus, by independence:

$$\mathbb{E}\Big[\log(y_i + \exp(x_i'\beta)) - x_i'\beta - \lambda\Big] = 0 \quad (4.6)$$

$$\mathbb{E}\Big[x_i(\log(y_i + \exp(x_i'\beta)) - x_i'\beta - \lambda)\Big] = 0 \quad (4.7)$$

These moments suggest estimation through the non-linear generalized method of moment. We show in Appendix A that there is a unique solution to this set of empirical moments. In practice, it is very easy to implement using any standard statistical software.

For instance, with Stata, this can be executed with the following command:

```
gmm (log(y+exp({xb:x}))-{xb:}-{lambda}), instruments(x) wmatrix(robust)
```

where $y$ and $x$ respectively represent the dependent and the list of independent variables. `wmatrix(robust)` is an option to obtain robust standard errors. In `instruments`, one can also add instrumental variables.

Unfortunately, the moments imply that $\alpha$ cannot be identified directly. However, given

[11] $\eta_i$ is related to $a_i$ by the equality $\eta_i = \log(1 + \exp(\alpha + \epsilon_i)) = \log(1 + \exp(\alpha)a_i)$

[12] This estimator is not robust to heteroskedasticity in $\epsilon_i$ that may depend on a variable $x_i$. Indeed, heteroskedasticity would contradict the independence assumption.

$\hat{\lambda}$ and $\hat{\xi}$, one can recover $\hat{\eta}_i$. In turn, we have :

$$\exp(\eta_i) = 1 + \exp(\alpha + \epsilon_i) \iff \log(\exp(\eta_i) - 1) = \alpha + \epsilon_i \tag{4.8}$$

which leads to the OLS closed form solution:

$$\hat{\alpha} = \frac{\sum_{n=1}^{N} \log\left(\exp(\hat{\xi}_i + \hat{\lambda}) - 1\right)}{N} \tag{4.9}$$

We also provide in Appendix B Stata code to obtain the coefficient $\alpha$.

We illustrate this method using simulated data. We generate the following process: $y_i = \exp(x_1 + x_2)a_i$ where $a_i$ is the error term. It is a uniform on $[0, 2]$ with its value replaced with zero if the draw was below 0.4 and replaced by two if the draw was above 1.6. Table 2 displays the results. It shows that the new proposed solutions provides correct estimates, as does PPML (presented in the next Section), whereas adding $\Delta = 0.01$ or $\Delta = 1$ to all data points biases the results.

Table 2: Comparison of the best practice with the new solution on simulated data

| | Real Coefficients | OLS:$\Delta = 1$ | | PPML | | Proposed Solution | |
|---|---|---|---|---|---|---|---|
| x1 | 1.000 | 0.609*** | (0.0230) | 1.012*** | (0.0251) | 1.022*** | (0.0308) |
| x2 | 1.000 | 0.609*** | (0.0226) | 1.024*** | (0.0251) | 1.012*** | (0.0301) |
| $\alpha$ | 0 | 0.547*** | (0.0154) | -0.0192 | (0.0199) | | |
| $\lambda$ | | | | | | 0.620*** | (0.00982) |
| Observations | 10000 | 10000 | | 10000 | | 10000 | |

Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

We further illustrate the numerical issue aforementioned. We draw 100 000 observations and three variables. $x_1$ and $x_2$ are standard normal, and the error term $u$ is normal with a standard deviation of ten. Then, the dependent variable is $y = \exp(1 + x_1 + 3x_2 + u)$. By construction, the dependent variable can take large values. In Table 3, we apply PPML with two different normalization methods. In the first one, we divide $y$ by its mean value and in the second one, by its standard deviation. In both cases, PPML does not provide adequate estimates while our solution does.

Table 3: Comparison of PPML and our solution when $y$ takes on large values

| | Real Coefficients | Normalized PPML (mean) | | Normalized PPML (s.d.) | | Proposed Solution | |
|---|---|---|---|---|---|---|---|
| x1 | 1.000 | 0.408 | (0.517) | 0.408 | (0.517) | 1.032*** | (0.0364) |
| x2 | 3.000 | 0.929*** | (0.224) | 0.929*** | (0.224) | 2.966*** | (0.0358) |
| $\alpha$ | 1.000 | -0.521 | (0.758) | -5.706*** | (0.758) | | |
| $\lambda$ | | | | | | 4.558*** | (0.0193) |
| Observations | 100000 | 100000 | | 100000 | | 100000 | |

Robust standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## 5 EXTENSION FOR INDEPENDENT VARIABLES

In many econometric applications, the main parameter of interest is an elasticity of $y_i$ with respect to $x_i$. This can be recovered by a log-log regression. In practice, it is common to have both dependent and independent variables that are equal to zero for some observations. Taking the log-transform of either of these variables is impossible. We recommend to use the following approach. Let the parameter value $\beta$ be a function of the data $x_i$, so that the $\beta$ takes the value $\beta_{(x_i>0)}$ if $x_i > 0$ and the value $\beta_{(x_i=0)}$ if $x_i = 0$. Note :

$$\log(x_i)\beta_{(x_i)} = \mathbb{1}(x_i > 0)\log(x_i)\beta_{(x_i>0)} + (1 - \mathbb{1}(x_i > 0))\beta_{(x_i=0)}$$

All we need to do now is to estimate as in the previous section :

$$\log(y_i) = \alpha + \mathbb{1}(x_i > 0)\log(x_i)\beta_{(x_i>0)} + (1 - \mathbb{1}(x_i > 0))\,\beta_{(x_i=0)} + \epsilon_i \tag{5.1}$$

This is akin to the semi-parametric approach consisting in providing a flexible formulation of $\beta$ such that it can vary with $x_i$.

To illustrate this, we draw 10 000 observations from a process $y_i = \exp(1.5\log(x_i)+\epsilon_i)$ where $\epsilon_i \sim \chi^2_1$ and $x_i$ is drawn from the standard uniform. We assume that we only have access to censored version of $x_i$ is equal to zero when $x_i < 0.1$. Table 4 presents the estimated coefficients with different methods.

Table 4: Example of the use of $\beta_{(x_i=0)}$ to account for zero values in the regressors

| | OLS | | PPML | | Proposed Solution | |
|---|---|---|---|---|---|---|
| $\beta_{(x_i>0)}$ | 1.511*** | (0.0197) | 1.465*** | (0.114) | 1.500*** | (0.0204) |
| $\beta_{(x_i=0)}$ | -4.978*** | (0.0395) | -4.892*** | (0.123) | -4.655*** | (0.0467) |
| $\alpha$ | 0.000752 | (0.0188) | 0.890*** | (0.0931) | | |
| $\lambda$ | | | | | 0.793*** | (0.00981) |
| Observations | 10000 | | 10000 | | 10000 | |

Robust Standard errors in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

## 6 Conclusion

In this paper, we discuss the best practice to deal with zeros in log-linear and log-log regressions. We first present some common misconceptions encountered in practice that may generate arbitrary sources of bias. We then explain why the PPML estimator provides a good solution. We also introduce a new solution to deal with zeros in the spirit of adding a positive constant to the dependent variable, which is particularly useful when PPML faces numerical issues. Finally, we discuss the case where the independent variable in a log-log regression also contains zeros. The conclusions can be applied in many fields in and outside of economics.

## References


BELLEMARE, M. F. AND C. J. WICHMAN (2019): "Elasticities and the Inverse Hyperbolic Sine Transformation," Working paper.

BURBIDGE, J. B., L. MAGEE, AND A. L. ROBB (1988): "Alternative Transformations to Handle Extreme Values of the Dependent Variable," *Journal of the American Statistical Association*, 83, 123–127.

CARD, D. AND S. DELLAVIGNA (2017): "What do Editors Maximize? Evidence from Four Leading Economics Journals," Working Paper 23282, National Bureau of Economic Research.

CORREIA, S., P. GUIMARÃES, AND T. ZYLKIN (2019): "ppmlhdfe: Fast Poisson Estimation with High-Dimensional Fixed Effects," *arXiv e-prints*.

CRISCUOLO, C., R. MARTIN, H. G. OVERMAN, AND J. VAN REENEN (2019): "Some Causal Effects of an Industrial Policy," *American Economic Review*, 109, 48–85.

GOURIEROUX, C., A. MONFORT, AND A. TROGNON (1984): "Pseudo Maximum Likelihood Methods: Applications to Poisson Models," *Econometrica*, 52, 701–20.

HEAD, K. AND T. MAYER (2019): "Brands in Motion, How Frictions Shape Multinational Production," *American Economic Review*, forthcoming.

JOHNSON, N. L. (1949): "Systems of Frequency Curves Generated by Methods of Translation," *Biometrika*, 36, 149–176.

JOHNSON, S. R. AND G. C. RAUSSER (1971): "Effects of Misspecifications of Log-Linear Functions When Sample Values Are Zero or Negative," *American Journal of Agricultural Economics*, 53, 120–124.

MACURDY, T. E. AND J. H. PENCAVEL (1986): "Testing between Competing Models of Wage and Employment Determination in Unionized Markets," *Journal of Political Economy*, 94, S3–S39.

MCCUNE, B., J. B. GRACE, AND D. L. URBAN (2002): *Analysis of Ecological Communities*, MjM Software Design.

MCKENZIE, D. (2017): "Identifying and Spurring High-Growth Entrepreneurship: Experimental Evidence from a Business Plan Competition," *American Economic Review*, 107, 2278–2307.

RAVALLION, M. (2017): "A concave log-like transformation allowing non-positive values," *Economics Letters*, 161, 130 – 132.

ROGOWSKI, J. A. AND J. P. NEWHOUSE (1992): "Estimating the indirect costs of teaching," *Journal of Health Economics*, 11, 153 – 171.

SANTOS SILVA, J. AND S. TENREYRO (2006): "The Log of Gravity," *The Review of Economics and Statistics*, 88, 641–658.

——— (2011): "poisson: Some convergence issues," *The Stata Journal*, 11, 207 – 212.

——— (2015): "PPML: Stata module to perform Poisson pseudo-maximum likelihood estimation," *Statistical Software Components*.

SANTOS SILVA, J., S. TENREYRO, AND W. F. (2015): "Testing Competing Models for Non-negative Data with Many Zeros," *Journal of Econometric Methods*, 4, 1–18.

YOUNG, K. H. AND L. Y. YOUNG (1975): "Estimation of Regressions Involving Logarithmic Transformation of Zero Values in the Dependent Variable," *The American Statistician*, 29, 118–120.

## A Proof of Uniqueness

We want to minimize the following quadratic loss:

$$\left(\hat{\beta},\hat{\lambda}\right)=\underset{(\beta,\lambda)}{\arg\min}\left\{\frac{1}{N}\sum_{i=1}^{N}\left[x_i(\log(y_i+\exp(x_i'\beta))-x_i'\beta-\lambda)\right]\right\}'\left\{\frac{1}{N}\sum_{i=1}^{N}\left[x_i(\log(y_i+\exp(x_i'\beta))-x_i'\beta-\lambda)\right]\right\} \tag{A.1}$$

Which has as first order condition :

$$0=(1/N)\sum_{i}^{N}\left[x_i(\log(y_i+\exp(x_i'\hat{\beta}))-x_i'\hat{\beta}-\hat{\lambda})\right] \tag{A.2}$$

With second order condition :

$$-(1/N)\sum_{i}^{N}x_i'x_i\left[\frac{y_i}{y_i+\exp(x_i'\beta)}\right] \tag{A.3}$$

which is negative definite, as required.

## B Calculating $\alpha$

Stata code to obtain the constant $\alpha$.

```
gmm ( log(y+exp({xb:x} ))-{xb:}-{lambda} ), instruments(x) wmatrix(robust)
gen lambda = [lambda]_cons
predict xi, res
gen eta_i = lambda + xi
gen alpha_epsilon = log(exp(eta_i)-1)
reg alpha_epsilon
```