Variable Scaling and Hypothesis Testing in the Gravity Model*

Anton Yang† Russell Hillberry‡

Yale University Purdue University

September, 2022

Abstract

Statistical inference from a Likelihood Ratio (LR) test involving the Poisson Pseudo Maximum Likelihood (PPML) estimator is sensitive to data scaling choices. We demonstrate—both analytically and empirically—that the LR test statistic is inversely proportional to the scale of the data on the dependent variable. Scaling the data affects hypothesis tests of model restrictions; it does not affect parameter estimates or t-tests of the significance of individual variables. Our finding is relevant to a large literature on the gravity model of trade, where PPML is the preferred estimator and the data are routinely scaled.

Keywords: Likelihood ratio test; Gravity model; Poisson Pseudo Maximum Likelihood; Variable scaling

JEL classification: C12; C13; C81; F14

---

*Declarations of interest: None.

†Room 112, 53 Wall Street, New Haven, CT 06511. Email: anton.yang@yale.edu. The paper was written while Yang was employed both at Yale University and Iowa State University.

‡*Corresponding author*: 692 Krannert Building, 403 W. State Street, West Lafayette, IN 47907-2053. Email: rhillber@purdue.edu.

1 Introduction

The enormous literature on the gravity model of trade has adopted the Poisson Pseudo Maximum Likelihood (PPML) function as its preferred econometric objective. In this literature the data are often scaled to facilitate numerical solution of a nonlinear estimation model.[1] In this paper we show that the scale of the data on the dependent variable affects the value of the log-likelihood function in the PPML framework, and more importantly, the conclusions from a likelihood ratio (LR) test of model restrictions. We first show these results analytically, and then offer a simple empirical example. Our insight is relevant to settings beyond the particular application of the gravity model, and may apply to other likelihood functions.

2 The PPML Estimator

Consider the estimation specification suggested by Santos Silva and Tenreyro (2006a):

$$y_i = \exp(x_i\beta) + \epsilon_i, \tag{1}$$

where y_i, $i = 1, .., n$, can be observed from data, x_i is a vector of exogenous variables, β a vector of associated parameters, and ϵ_i an error term with $E[\epsilon_i|x] = 0$. Following Gourieroux et al. (1984), choose β to maximize the log-likelihood function:

$$L(\beta) = K - \sum_i^n \exp(x_i\beta) + \sum_i^n y_i x_i \beta, \tag{2}$$

where K is a constant term.

Our analytic exercises will use this general framework, but to make our example concrete consider a simple PPML gravity model of trade:

[1] In Santos Silva and Tenreyro (2006a) the value of bilateral trade is expressed in units of one thousand U.S. dollars. Anderson and van Wincoop (2003) scale both trade flows and Gross Domestic Products so they are expressed in millions of dollars.

$$y_{ij} = \exp[\beta_0 + \beta_1 \log(x_i) + \beta_2 \log(x_j) + \beta_3 z_{ij}] + \epsilon_{ij}, \tag{3}$$

where y_{ij} are observed bilateral trade flows between locations i and j, the exporting and importing location's incomes are denoted by x_i and x_j, and z_{ij} a variable that applies bilaterally to locations i and j. The log transformation on the right hand side of (3) is standard in the gravity literature.

3 Numerical Scaling and Hypothesis Testing

In this section we show that, in the PPML framework, the scaling of data on the dependent variable affects both the likelihood statistic and the LR statistic associated with a model restriction. Statistical inference about the relevance of a model restriction thus depends on data scale.

Consider two models: a 'small' parsimonious model, and a 'large' model with more parameters. We wish to consider a hypothesis test of the form:

Hypothesis $\mathbf{H_0}$: *the 'small' model is more consistent with the data.*

Hypothesis $\mathbf{H_a}$: *the 'large' model is more consistent with the data.*

The test statistic for a LR test of $\mathbf{H_0}$ appears as:

$$\Lambda = -2[L^{H_0}(\beta) - L^{H_a}(\beta')], \tag{4}$$

where β and β' are the parameter vectors associated with the 'small' and 'large' models, respectively. Asymptotically, Λ is distributed χ^2 with k degrees of freedom, where k is the reduction in the number of parameters when moving from the 'large' to the 'small' model. Given a critical value c for this distribution, the null hypothesis will be rejected if $\Lambda > c$.[2]

[2]See Gouriéroux et al. (1982).

3.1 Implications of Scaling for Likelihood

Let the log-likelihood function for the 'small' model take the form:

$$L^{H_0}(\beta) = -\sum_i^n \exp[\beta_0 + \beta_1 \log(x_i)] + \sum_i^n y_i[\beta_0 + \beta_1 \log(x_i)], \tag{5}$$

where β_0 is a constant term, β_1 is the coefficient on $\log x_i$.[3]

In order to show that the scaling of y_i affects Λ, we introduce an arbitrary scalar $S \neq 1$ that we apply to y_i. The associated log-likelihood function is:

$$\widehat{L^{H_0}(\tilde{\beta})} = -\sum_i^n \exp[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] + \sum_i^n S y_i[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] \tag{6}$$

where $\widehat{L^{H_0}(\tilde{\beta})}$ is the log-likelihood involved with the scaling factor S, and $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are the model's parameters given the scaling. Taking the first-order conditions with respect to β terms in Equations (5) and (6) and solving the associated system of equations reveals relationships between the parameters before and after the scaling: $\beta_0 = \tilde{\beta}_0 - \log(S)$ and $\beta_1 = \tilde{\beta}_1$. Intuitively, scaling affects the size of the constant term, but other coefficients in the model are unaffected. However, the log-likelihoods before and after the scaling of y_i are not equivalent:

$$\begin{aligned}
\widehat{L^{H_0}(\tilde{\beta})} &= -\sum_i^n \exp[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] + S\sum_i^n y_i[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] \\
&= -S\sum_i^n \exp[\beta_0 + \beta_1 \log(x_i)] + S\sum_i^n y_i[\beta_0 + \log S + \beta_1 \log(x_i)] \\
&\neq -\sum_i^n \exp[\beta_0 + \beta_1 \log(x_i)] + \sum_i^n y_i[\beta_0 + \beta_1 \log(x_i)] \\
&= L^{H_0}(\beta).
\end{aligned} \tag{7}$$

The relationship only holds with equality when $S = 1$.

[3] We follow the standard PPML-gravity literature and suppress the parameter K. Equation (4) differences K from Λ anyway.

3.2 Implications of Scaling for Testing

Now shift to the large model by introducing another variable, z_i. The associated log-likelihood function is:

$$L^{H_a}(\beta') = -\sum_i^n \exp[\beta'_0 + \beta'_1 \log(x_i) + \beta'_2 \log z_i] + \sum_i^n y_i[\beta'_0 + \beta'_1 \log(x_i) + \beta'_2 \log z_i]. \tag{8}$$

The log-likelihood of the large model with scaled data appears as:

$$\widehat{L^{H_a}(\tilde{\beta}')} = -\sum_i^n \exp[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)] + S\sum_i^n y_i[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)]. \tag{9}$$

It can be shown that the parameters in the scaled and unscaled large models have the relationships: $\beta'_0 = \tilde{\beta}'_0 - \log(S)$, $\beta'_1 = \tilde{\beta}'_1$ and $\beta'_2 = \tilde{\beta}'_2$.

The difference between the log-likelihoods using the scaled data is given by:

$$\begin{aligned}\widehat{L^{H_0}(\tilde{\beta})} - \widehat{L^{H_a}(\tilde{\beta}')} = &-\sum_i^n \exp[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] + S\sum_i^n y_i[\tilde{\beta}_0 + \tilde{\beta}_1 \log(x_i)] \\ &+ \sum_i^n \exp[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)] - S\sum_i^n y_i[\tilde{\beta}'_0 + \tilde{\beta}'_1 \log(x_i) + \tilde{\beta}'_2 \log(z_i)].\end{aligned} \tag{10}$$

Dividing both sides by S in Equation (10) and substituting the derived parametric relationships into the test statistics produces:

$$
\begin{aligned}
\frac{\widehat{\Lambda(\tilde{\beta})}}{S} = \frac{-2\left[\widehat{L^{H_0}(\tilde{\beta})} - \widehat{L^{H_a}(\tilde{\beta}')}\right]}{S} = &-2\Bigg\{-\frac{1}{S}\sum_i^n \exp[\beta_0 + \log(S) + \beta_1 \log(x_i)] \\
&+ \sum_i^n y_i[\beta_0 + \log(S) + \beta_1 \log(x_i)] \\
&+ \frac{1}{S}\sum_i^n \exp[\beta_0' + \log(S) + \beta_1' \log(x_i) + \beta_2' \log(z_i)] \\
&- \sum_i^n y_i[\beta_0' + \log S + \beta_1' \log(x_i) + \beta_2' \log(z_i)]\Bigg\}.
\end{aligned} \tag{11}
$$

The right hand side of Equation (11) is equal to the test statistic with unscaled data: $\Lambda(\beta) = -2\left[L^{H_0}(\beta) - L^{H_a}(\beta')\right]$. The lesson is that $\Lambda(\beta)$ varies inversely with the scale applied to the y variable. The critical value of the χ^2 distribution is unchanged, but the test statistic changes with S, polluting statistical inference.

4 Empirical Example

To illustrate the relevance of our insight, we offer a simple example using the data and empirical specification proposed by Santos Silva and Tenreyro (2006a). Those authors estimate a gravity model with 14 independent variables and an unreported constant term. Their specification contains two related variables: an indicator that the origin and destination countries share a colonial tie, and another indicator that the two countries share a common language. Since common historical forces often drove these two outcomes, it can be difficult for an applied researcher to know whether or not both variables should be included in a gravity regression. In Santos Silva and Tenreyro (2006a) the coefficient on the common language dummy is statistically significant, while the coefficient on the colonial tie variable is not. Our empirical example is a test of the hypothesis that a model without the colonial tie variable is equivalent to a model that includes it, thereby justifying estimation of the smaller model (without the colonial tie dummy).

Table 1: PPML estimates with different scalings of trade flows

	S=1000		S=1*		S=0.001		S=1e-6	
Model	w/ Colony	w/o Colony	w/ Colony	w/o Colony	w/ Colony	w/o Colony	w/ Colony	w/o Colony
Shipment in	USD	USD	1,000 USD	1,000 USD	mil. USD	mil. USD	bil.USD	bil.USD
Comlang	0.75**	0.76**	0.75**	0.76**	0.75**	0.76**	0.75**	0.76**
	(0.13)	(0.08)	(0.13)	(0.08)	(0.13)	(0.08)	(0.13)	(0.08)
Colony	0.03	n.a.	0.03	n.a.	0.03	n.a.	0.03	n.a.
	(0.15)	n.a.	(0.15)	n.a.	(0.15)	n.a.	(0.15)	n.a.
$L^a(\beta')$	-8.70197e11		-870246443		-888289.69		-2630.97	
$L^0(\beta)$	-8.7025e11		-870297478.5		-888340.72		-2631.02	
Λ	1.02e8**		102071**		102.07**		0.10207	
$P > \chi^2$	0.00		0.00		0.00		0.75	

Results from a replication of the full specification in Santos Silva and Tenreryo (2006a) with different scalings of the trade variable; *indicates the scale used Santos Silva and Tenreyro; **indicates statistical significance at the 5% level; standard errors in parentheses.

We download the data from Santos Silva and Tenreyro (2006b). We consider different scalings of the bilateral trade data, premultiplying it with different values of a scalar S. We estimate the model in Stata using the *ppmlhdfe* command.[4] Since our focus is on the two dummy variables and the likelihood statistics, we report only statistics related to these outcomes in Table 1. For each scaling of the data, we report results for PPML models estimated with and without the colonial tie variable. We arrange the results in increasing size of S, which corresponds to different choices of units for the value of bilateral trade. Columns 1-2 use one dollar units. Columns 3-4 measure trade in thousands of dollars. Columns 5-6 use million dollar units; columns 7-8 use one billion dollar units.

Column 3 is a replication of Santos Silva and Tenreyro (2006a), with almost exactly identical results. The coefficient on the common colony variable (0.03) is economically small and statistically insignificant, suggesting that perhaps it can be excluded from the model. Column 4 reports the common language coefficient in the model that excludes the colonial tie variable. Log-likelihoods for the two models are reported below the coefficient estimates, as is the test statistic Λ. The value of Λ for Santos Silva and Tenreyro's scaling of the data is 102,071, providing a clear rejection of the model without the colonial tie. This is surprising,

[4] Correia et al. (2020) develop this command to estimate the PPML model in the presence of high dimensional fixed effects. While our model does not contain high dimensional fixed effects, the *ppmlhdfe* command is still suitable.

as the coefficient on the *Colony* variable is not statistically different from zero in Column 3.

Columns 1-2 report estimates of the two models when the data are scaled in single dollar units. The choice of smaller units scales Λ upward by a factor of 1,000, incorrectly increasing the level of confidence in rejecting the null hypothesis. In columns 5-6, trade flows are scaled in millions of dollars; Λ is therefore scaled downward by 1,000 relative to the base case. The null hypothesis continues to be soundly rejected, but by a smaller margin than with the initial scaling. In columns 7-8, we scale the data into units of \$1 billion. In this case the computed value of Λ fails to reject the hypothesis that the two models are equivalent.

Note that the parameter estimates in all specifications are indifferent to the scale of the data. As in the mathematics above, scaling the data affects the constant term, but is otherwise unimportant for parameter estimation. Only the values of $L(\beta)$, Λ and the associated statistical inference are affected by scaling.

5 Other Scalings of the Data

We have shown that the scale of the data on the y-variable affects statistical inference around model restrictions. Does scaling of x-variables matter, and how does scaling on one side interact with scaling on the other? We investigate these questions using the same analytical methods, and verify the outcomes with the same empirical example.

We find that the scaling of one or more x-variables does not matter for the scale of either $L(\beta)$ or Λ. One can express Gross Domestic Product, for example, in thousands or in millions of dollars without affecting statistical inference. When we scale both the y and x variables by an equivalent factor S, we find that Λ varies inversely with S, as with y-scaling alone.

References

Anderson, James E., and Eric van Wincoop. 2003. "Gravity with Gravitas: A Solution to the Border Puzzle." *American Economic Review*, 93(1): 170–192.

Correia, Sergio, Paulo Guimarães, and Tom Zylkin. 2020. "Fast Poisson estimation with high-dimensional fixed effects." *The Stata Journal*, 20(1): 95–115.

Gourieroux, Christian, Alain Monfort, and Alain Trognon. 1984. "Pseudo Maximum Likelihood Methods: Theory." *Econometrica*, 52(3): 681–700.

Gouriéroux, Christian, Alberto Holly, and Alain Monfort. 1982. "Likelihood Ratio Test, Wald Test, and Kuhn-Tucker Test in Linear Models with Inequality Constraints on the Regression Parameters." *Econometrica*, 50(1): 63–80.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006a. "The Log of Gravity." *The Review of Economics and Statistics*, 88(4): 641–658.

Santos Silva, J. M. C., and Silvana Tenreyro. 2006b. "Data for 'The Log of Gravity'." London School of Economics. http://personal.lse.ac.uk/tenreyro/regressors.zip, accessed June 8, 2022.