The Log of Gravity:
Clarifying the Econometric Framework

Kenneth G. Stewart

Associate Professor
Department of Economics
University of Victoria
Victoria, British Columbia
Canada V8W 2Y2

email: kstewart@uvic.ca
phone: (250) 721-8534
fax: (250) 721-6214

July 2018

Abstract

In an intriguing and widely-cited article, Santos Silva and Tenreyro (2006) have argued that, when a linear-in-coefficients regression model is obtained by a nonlinear transformation of some initial nonlinear specification, OLS is inconsistent. I investigate and clarify the basis for this claim, and its limitations.

Keywords: loglinear regression, linearizing transformation, log-linearize, heteroskedasticity

JEL classifications: C13; C20; C51; F14

Declarations of interest: none

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

The Log of Gravity: Clarifying the Econometric Framework

July 2018

Abstract

In an intriguing and widely-cited article, Santos Silva and Tenreyro (2006) have argued that, when a linear-in-coefficients regression model is obtained by a nonlinear transformation of some initial nonlinear specification, OLS is inconsistent. I investigate and clarify the basis for this claim, and its limitations.

Keywords: loglinear regression, linearizing transformation, log-linearize, heteroskedasticity

JEL classifications: C13; C20; C51; F14

1 Introduction

Few empirical devices are as longstanding and ubiquitous to applied econometric practice as the use of functional forms that lend themselves to loglinearization. Obvious examples quickly come to mind, such as Cobb-Douglas functional forms

$$Y_i = \beta_0 X_{i1}^{\beta_1} X_{i2}^{\beta_2} \cdots X_{iK}^{\beta_K} \tag{1}$$

or the statistical earnings function of labor economics,

$$Y_i = \mathrm{e}^{\boldsymbol{x}_i'\boldsymbol{\beta}} \quad \Leftrightarrow \quad \ln Y_i = \boldsymbol{x}_i'\boldsymbol{\beta}, \tag{2}$$

where Y_i is wages and $\boldsymbol{x}_i$ a vector of worker characteristics.

In a novel and influential contribution, Santos Silva and Tenreyro (2006, henceforth SST) have argued that, when the model to be estimated is obtained as the logarithmic transformation of some primary model, OLS is inconsistent. Consequently (p. 644) "…it is not advisable to estimate $\boldsymbol{\beta}$ from the log linear model. Instead, the nonlinear model has to be estimated." They advocate applying pseudo-maximum likelihood based on the Poisson distribution, as developed in Gourieroux, Monfort, and Trognon (1984), and provide simulation evidence demonstrating its superior finite sample performance relative to alternative estimators, including nonlinear least squares (NLS). Their Monte Carlo simulations use the gravity models of international trade as the data generating process, which are an application of Cobb-Douglas functional forms. Their Poisson pseudo-maximum likelihood (PPML) estimator has been further studied in SST (2010,2011), Head and Mayer (2014), and Fally (2015) and has been implemented in both Stata and R.[1]

But the precise source of the inconsistency that PPML addresses is somewhat amorphous. In their opening paragraph SST (2006) suggest that it arises from the property of the log function that $\mathrm{E}(\ln y) \neq \ln \mathrm{E}(y)$, but later (p. 641) indicate that the issues "…extend to a broad range of economic applications where the equations under study are …transformed by a nonlinear function." Throughout the article they say that heteroskedasticity plays a role, inducing a correlation between the disturbance and the transformed regressors. Whether this heteroskedasticity is *conditional* or *unconditional* is unclear in the theoretical development, although their simulations include examples of both. Elsewhere (p. 644) they suggest that, at least in some contexts, the issue is identification: "The problem is that these parameters may not permit identification of the parameters of $\mathrm{E}[y_i|x]$."

So which is it, nonlinear variable transformations, heteroskedasticity, or identification? Or some combination? Or are these just different interpretations or manifestations of a common underlying phenomenon? Given that SST's PPML estimator has become, in the words of Fally (2015, p. 78), "A now widely-used strategy ..." for estimating gravity models, it seems worth clarifying the econometric justification for abandoning nonlinearly transformed regressions, not only in the gravity context but in other areas of application as well.

My own view is that none of these is the most useful way of understanding the SST inconsistency result and, in turn, the motivation for their PPML estimator. Instead it is best understood as an interesting but nevertheless fairly straightforward implication of the specification of the initial nonlinear model as the conditional expectation function (CEF) of the dependent variable. Advancing this view requires that we begin by briefly reviewing the standard theory of CEF models, treatments of which can be found in, for example, Goldberger (1991) and Wooldridge (2010).

2 Background: The CEF model

Economic theory often provides little guidance about the form of the population distribution from which data are drawn, so that randomly sampled data are treated simply as i.i.d. Even so, in studying the determination of one variable Y_i in relation to others $\boldsymbol{x}_i' = [X_{i1}, X_{i2}, \ldots, X_{iK}]$, theory or the economic context do sometimes suggest a suitable functional form for the relationship. Following Davidson and MacKinnon (2004) it is convenient to denote this function by $\boldsymbol{x}_i(\boldsymbol{\beta})$ in general, which permits nonlinearities in either or both of the variables and coefficients; in the special case of linearity in the coefficients this becomes $\boldsymbol{x}_i(\boldsymbol{\beta}) = \boldsymbol{x}_i'\boldsymbol{\beta}$.

Given i.i.d. sampling, all variables have marginal distributions that are invariant across the drawings i, and so the moments and parameters of these marginal distributions are constant across the observations. For example, the *unconditional* mean and variance of Y_i

might be denoted

$$\mathrm{E}(Y_i) = \mu, \qquad \mathrm{V}(Y_i) = \mathrm{E}(Y_i - \mu)^2 = \sigma^2, \tag{3}$$

so that Y_i is *unconditionally* homoskedastic. Unconditional heteroskedasticity would imply independent but non-identically distributed drawings.

Of course, these unconditional parameters of the marginal distributions are typically not the objects of empirical inquiry. Instead the researcher is interested in studying the dependence among the variables given by $\boldsymbol{x}_i(\boldsymbol{\beta})$, which is typically taken to be a specification for the mean[2] of Y_i *conditional* on the other variables,

$$\mathrm{E}(Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i(\boldsymbol{\beta}). \tag{4}$$

This framework, in which the researcher has a priori knowledge of the conditional mean, is a common paradigm in econometrics and is that adopted by SST.

2.1 An additive CEF disturbance

Although the statistical structure of the CEF model is entirely embodied in the specification (4), it is nevertheless often stated in terms of a disturbance,

$$Y_i = \boldsymbol{x}_i(\boldsymbol{\beta}) + \varepsilon_i. \tag{5}$$

However this CEF disturbance is simply defined as the discrepancy

$$\varepsilon_i = Y_i - \mathrm{E}(Y_i|\boldsymbol{x}_i) = Y_i - \boldsymbol{x}_i(\boldsymbol{\beta}) \tag{6}$$

and its properties follow by construction from this definition. These properties are well known (e.g. Goldberger, 1991, p. 49–50; Wooldridge, 2010, sec. 2.2) but to facilitate the comparison with SST it is useful to list them explicitly.

CEF Property 1: The disturbance ε_i has conditional mean of zero.

$$\begin{aligned}\mathrm{E}(\varepsilon_i|\boldsymbol{x}_i) &= \mathrm{E}[(Y_i - \boldsymbol{x}_i(\boldsymbol{\beta}))|\boldsymbol{x}_i] \\ &= \mathrm{E}(Y_i|\boldsymbol{x}_i) - \mathrm{E}[\boldsymbol{x}_i(\boldsymbol{\beta})|\boldsymbol{x}_i] \\ &= \mathrm{E}(Y_i|\boldsymbol{x}_i) - \boldsymbol{x}_i(\boldsymbol{\beta}) \\ &= 0 \end{aligned} \tag{7}$$

By the law of iterated expectations (LIE) it follows that ε_i also has zero unconditional mean: $\mathrm{E}(\varepsilon_i) = 0$.

CEF Property 2: The disturbance ε_i is uncorrelated with any function of $\boldsymbol{x}_i$, say $\phi(\boldsymbol{x}_i)$:

$$\text{cov}[\phi(\boldsymbol{x}_i), \varepsilon_i] = \text{E}[\phi(\boldsymbol{x}_i)\varepsilon_i] = 0. \tag{8}$$

This property carries with it the implication of correct model specification: it implies that $\boldsymbol{x}_i(\boldsymbol{\beta})$ is the true conditional mean, so that $\boldsymbol{x}_i(\boldsymbol{\beta})$ fully accounts for the effects of the explanatory variables $\boldsymbol{x}_i$ on Y_i.

CEF Property 3: As a special case of CEF Property 2 in which the function $\phi(\boldsymbol{x}_i)$ is set to $\boldsymbol{x}_i$ itself, we have

$$\text{E}(\boldsymbol{x}_i\varepsilon_i) = \text{E}[\boldsymbol{x}_i(Y_i - \boldsymbol{x}_i(\boldsymbol{\beta}))] = \boldsymbol{0}. \tag{9}$$

These unconditional moment restrictions provide the basis for NLS estimation (or OLS if $\boldsymbol{x}_i(\boldsymbol{\beta}) = \boldsymbol{x}_i'\boldsymbol{\beta}$), the conventional estimator under these assumptions.

Although, by CEF Property 2, arbitrary functions $\phi(\boldsymbol{x}_i)$ of the explanatory variables in principle qualify as instrumental variables, they are irrelevant. Attempting to construct a GMM estimator based on an expanded instrument set $\boldsymbol{z}_i' = [\boldsymbol{x}_i', \phi(\boldsymbol{x}_i)]$ and moment conditions $\text{E}[\boldsymbol{z}_i(Y_i - \boldsymbol{x}_i(\boldsymbol{\beta}))] = \boldsymbol{0}$ merely reduces to NLS.

CEF Property 4: Our final implication is to note the lack of one. In contrast to Property 2, CEF Property 1 has no necessary implications for the higher order conditional moments of ε_i, which are of the general form $\text{E}[\phi(\varepsilon_i)|\boldsymbol{x}_i]$. Consider, for example, the conditional variance $\text{V}(\varepsilon_i|\boldsymbol{x}_i)$, which sets $\phi(\varepsilon_i) = (\varepsilon_i - \text{E}(\varepsilon_i|\boldsymbol{x}_i))^2$:

$$\text{V}(\varepsilon_i|\boldsymbol{x}_i) = \text{E}[(\varepsilon_i - \text{E}(\varepsilon_i|\boldsymbol{x}_i))^2|\boldsymbol{x}_i] = \text{E}(\varepsilon_i^2|\boldsymbol{x}_i).$$

Other than yielding the simplified final expression, by itself the condition $\text{E}(\varepsilon_i|\boldsymbol{x}_i) = 0$ has no necessary implications for this conditional moment. Despite the *unconditional* homoskedasticity (3), ε_i is *conditionally heteroskedastic*: its conditional variance varies with $\boldsymbol{x}_i$ and so is observation-specific.

Hence, although NLS is consistent, the default estimator for its variance should be one that is heteroskedasticity-consistent. Conditional homoskedasticity would be an additional assumption on the model for which there is generally no basis.

3 Implications of Loglinearization

Distilled to its essence, SST's critique of loglinearization seems to come down to the observation that, given the premise that the conditional mean $\mathrm{E}(Y_i|\boldsymbol{x}_i)$ is some nonlinear-in-coefficients function $\boldsymbol{x}_i(\boldsymbol{\beta})$, linearizing it with a nonlinear transformation abandons the CEF structure.

For example, in the case of the Cobb-Douglas functional form (1), the NLS estimator is based on the moment conditions (9) in which

$$\mathrm{E}(Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i(\boldsymbol{\beta}) = \beta_0 X_{i1}^{\beta_1} X_{i2}^{\beta_2} \cdots X_{iK}^{\beta_K}.$$

By contrast, estimation of the loglinearized model by OLS uses the moments

$$\mathrm{E}[\boldsymbol{z}_i(\ln Y_i - \ln \boldsymbol{x}_i(\boldsymbol{\beta}))] \quad \text{where} \quad \ln \boldsymbol{x}_i(\boldsymbol{\beta}) = \ln \beta_0 + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \cdots + \ln \beta_K X_{iK} \tag{10}$$

and sets the instruments to $\boldsymbol{z}_i' = [1, \ln X_{i1}, \ln X_{i2}, \ldots, \ln X_{iK}]$. SST's point is that, in relation to the initially specified nonlinear model, the moments (10) are entirely ad hoc and nothing in the theory of the CEF model implies that they should be zero. The resulting OLS estimators therefore have no desirable properties, least of all consistency. It is not so much that the OLS estimators are biased—so are the NLS estimators of the nonlinear model—but that their bias does not disappear asymptotically.

3.1 Restatement in terms of an additive ad hoc disturbance

These arguments can be restated in terms of a disturbance, which may have some interpretive value. Let $\boldsymbol{x}_i(\boldsymbol{\beta})$ be amenable to loglinearization and define a disturbance $\nu_i = \ln Y_i - \ln \boldsymbol{x}_i(\boldsymbol{\beta})$, so that the transformed model is stated as

$$\ln Y_i = \ln \boldsymbol{x}_i(\boldsymbol{\beta}) + \nu_i. \tag{11}$$

This ad hoc disturbance is not a CEF disturbance because, under the originally specified model (4), $\ln \boldsymbol{x}_i(\boldsymbol{\beta})$ is not the conditional mean of the dependent variable $\ln Y_i$:

$$\mathrm{E}(\ln Y_i|\boldsymbol{x}_i) \neq \ln \mathrm{E}(Y_i|\boldsymbol{x}_i) = \ln \boldsymbol{x}_i(\boldsymbol{\beta}). \tag{12}$$

Consistent with SST, in this respect the inconsistency of OLS applied to the loglinearized model can be interpreted as an implication of the log function that $\mathrm{E}(\ln y) \neq \ln \mathrm{E}(y)$, and the same logic would apply to linearizations using other nonlinear transformations.

Given that ν_i is not a CEF disturbance, none of the properties of a CEF disturbance need hold. Specifically:

1. Neither the conditional nor the unconditional mean of ν_i need be zero, because a derivation analogous to (7) does not go through,

$$\begin{aligned} \mathrm{E}(\nu_i|\boldsymbol{x}_i) &= \mathrm{E}[(\ln Y_i - \ln \boldsymbol{x}_i(\boldsymbol{\beta}))|\boldsymbol{x}_i] \\ &= \mathrm{E}(\ln Y_i|\boldsymbol{x}_i) - \mathrm{E}[\ln \boldsymbol{x}_i(\boldsymbol{\beta})|\boldsymbol{x}_i] \\ &= \mathrm{E}(\ln Y_i|\boldsymbol{x}_i) - \ln \boldsymbol{x}_i(\boldsymbol{\beta}) \neq 0, \end{aligned}$$

   again because of the inequality (12). This would similarly be the case for linearizations based on other nonlinear transformations. It would also be true of conditioning variables that are transformations of $\boldsymbol{x}_i$, such as $\boldsymbol{z}_i = \ln \boldsymbol{x}_i$; in general, a researcher cannot simply contrive some function $\phi(\boldsymbol{x}_i)$ to yield $\mathrm{E}[\nu_i|\phi(\boldsymbol{x}_i)] = 0$.

2. There is no reason for the ad hoc disturbance ν_i to be uncorrelated with the explanatory variables $\boldsymbol{x}_i$ or functions of them, so $\mathrm{cov}[\phi(\boldsymbol{x}_i), \nu_i] \neq 0$. As already remarked in connection with (10) there are, then, no conditions analogous to (9) on the moments

$$\mathrm{cov}[\phi(\boldsymbol{x}_i), \nu_i] = \mathrm{cov}[\phi(\boldsymbol{x}_i), \ln Y_i - \ln \boldsymbol{x}_i(\boldsymbol{\beta})] \tag{13}$$

   on which to base estimation.

   When a regression disturbance may be correlated with the regressors, the stock remedy of empirical researchers is to seek instruments $\boldsymbol{z}_i$ that are plausibly uncorrelated with the disturbance. Estimation would then be based on moment conditions of the form, in the present context,

$$\mathrm{cov}[\boldsymbol{z}_i, \nu_i] = \mathrm{cov}[\boldsymbol{z}_i, \ln Y_i - \ln \boldsymbol{x}_i(\boldsymbol{\beta})] = \mathbf{0}.$$

   The problem with this stock remedy is the availability of strong instruments. The far more appealing alternative offered by SST is to instead return to the original nonlinear model, but to estimate it by PPML instead of NLS.

3. There is no reason for the ad hoc disturbance ν_i to be conditionally homoskedastic, but then of course neither is the CEF disturbance ε_i, by CEF Property 4. It seems doubtful that it is helpful to interpret the absence of conditions on the moments (13)

as somehow being attributable to this conditional heteroskedasticity.[3] After all, the conditional heteroskedasticity of ε_i does not invalidate the moment conditions (9) or the NLS coefficient estimator that arises from them.

3.2 Restatement in terms of a multiplicative CEF disturbance

The CEF disturbance ε_i is conventionally defined so as to appear additively in the CEF model (5), but to some extent this definition is arbitrary. Formally, the discrepancy between the dependent variable Y_i and the conditional mean $\mathrm{E}(Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i(\boldsymbol{\beta})$ could just as legitimately be defined as multiplicative,

$$Y_i = \boldsymbol{x}_i(\boldsymbol{\beta})\eta_i \tag{14}$$

so that, in analogy with the definition (6), this multiplicative disturbance is implicitly defined as $\eta_i = Y_i/\boldsymbol{x}_i(\boldsymbol{\beta})$.

This is the form of the model in which SST phrase some of their analysis, drawing on the properties of η_i as they parallel those of the conventional additive CEF disturbance ε_i. In particular, in analogy with CEF Property 1 that $\mathrm{E}(\varepsilon_i|\boldsymbol{x}_i) = 0$, for $\boldsymbol{x}_i(\boldsymbol{\beta})$ to represent the conditional mean $\mathrm{E}(Y_i|\boldsymbol{x}_i)$ it must be that $\mathrm{E}(\eta_i|\boldsymbol{x}_i) = 1$ in order that

$$\mathrm{E}(Y_i|\boldsymbol{x}_i) = \mathrm{E}[\boldsymbol{x}_i(\boldsymbol{\beta})\eta_i|\boldsymbol{x}_i] = \boldsymbol{x}_i(\boldsymbol{\beta})\,\mathrm{E}(\eta_i|\boldsymbol{x}_i) = \boldsymbol{x}_i(\boldsymbol{\beta}).$$

And similarly to CEF Property 1, it follows from the LIE that the unconditional mean is also 1: $\mathrm{E}(\eta_i) = 1$.

Now consider the loglinearized multiplicative model,

$$\ln Y_i = \ln \boldsymbol{x}_i(\boldsymbol{\beta}) + \nu_i \quad \text{where} \quad \nu_i = \ln \eta_i.$$

This is the model (11) and shows that the relationship between the ad hoc disturbance ν_i and the multiplicative CEF disturbance η_i is $\nu_i = \ln \eta_i$. Some of our previous conclusions about ν_i can be understood in terms of this relationship. Specifically, even though $\ln \mathrm{E}(\eta_i|\boldsymbol{x}_i) = \ln 1 = 0$, the conditional mean of $\nu_i = \ln \eta_i$ is not zero, because the expected value of the logarithm of a random variable is not the logarithm of the expected value:

$$\mathrm{E}(\nu_i|\boldsymbol{x}_i) = \mathrm{E}(\ln \eta_i|\boldsymbol{x}_i) \neq \ln \mathrm{E}(\eta_i|\boldsymbol{x}_i) = 0.$$

This nonzero mean of the disturbance ν_i in the loglinear model (11) has the implication that, if $\ln \boldsymbol{x}_t(\boldsymbol{\beta})$ has an additive intercept, its value is not separately identified from the nonzero mean of ν_t. OLS yields an estimate of the intercept of the loglinear model only because it imposes the false identifying restriction that the disturbance ν_i has zero mean. The OLS intercept estimate cannot therefore be consistent for the true intercept, whatever that may be. In relation to the interpretations offered by SST, then, in this limited sense the inconsistency of OLS applied to a loglinearized regression can be interpreted as an identification problem.

4 Can loglinearization be saved?

Is, then, the age-old practice among applied researchers of estimating loglinearized regressions entirely discredited? Not necessarily: it depends on the researcher's initial specification of the CEF. The above chain of reasoning hinges entirely on the premise that the nonlinear model is the CEF, so that loglinearization results in a regression for which the disturbance is not a CEF disturbance.

For example, suppose that in the Cobb-Douglas case, instead of specifying

$$\mathrm{E}(Y_i|\boldsymbol{x}_i) = \beta_0 X_{i1}^{\beta_1} X_{i2}^{\beta_2} \cdots X_{iK}^{\beta_K}, \tag{15}$$

the researcher interprets the variable to be explained as $\ln Y_i$ and specifies its conditional mean to be

$$\mathrm{E}(\ln Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i(\boldsymbol{\beta}) = \beta_0 + \beta_1 \ln X_{i1} + \beta_2 \ln X_{i2} + \cdots + \ln \beta_K X_{iK}.$$

Then the disturbance defined as $\varepsilon_i = \ln Y_i - \mathrm{E}(\ln Y_i|\boldsymbol{x}_i) = \ln Y_i - \boldsymbol{x}_i(\boldsymbol{\beta})$ is a CEF disturbance. Along with the other properties of CEF disturbances, this yields moment conditions of the form (8). Setting the instruments to be $\boldsymbol{z}_i' = \phi(\boldsymbol{x}_i) = [1, \ln \boldsymbol{x}_i']$, these conditions are

$$\mathrm{E}[\boldsymbol{z}_i \varepsilon_i] = \mathrm{E}[\boldsymbol{z}_i(\ln Y_i - \boldsymbol{x}_i(\boldsymbol{\beta}))] = \boldsymbol{0},$$

which in turn yield the OLS estimator of the loglinearized equation.

Even so, in the case of gravity equations, and perhaps even Cobb-Douglas functions generally, it may well be that the nonlinear CEF (15) is the compelling specification. But this may not be true of other functional forms amenable to nonlinear transformation. Consider the statistical earnings function (2) of labor economics, which is routinely estimated in the log-lin form

$$\ln Y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \varepsilon_i. \tag{16}$$

The underlying basis for this specification is that wages Y_i are positive and have an empirical distribution that is skewed to the right. This distribution is much better approximated by lognormality than normality, so that $\ln Y_i$ is well-approximated by normality. Consequently specifying the dependent variable in log form is likely to yield a disturbance that is close to normally distributed, improving the likelihood that the actual finite sample distributions of classical t and F statistics will be well-approximated by their namesake distributions. (Of course, a convenient byproduct of the log-lin specification is that the $\boldsymbol{\beta}$ coefficients have interpretations as semielasticities. But this is an incidental consequence of the choice of functional form, not the justification for that choice.)

On this reasoning, were wages being studied in isolation their approximate lognormality suggests the model $\ln Y_i = \mu + \varepsilon_i$, $\varepsilon_i \sim$ n.i.d.$(0, \sigma^2)$, where μ is the unconditional mean $\mathrm{E}(\ln Y_i)$. That is, the natural object of interest is the mean of the *transformed* dependent variable, not that of the untransformed variable. By the same logic, when it is the relationship of wages to worker characteristics that is being studied, the natural object of interest is the conditional mean $\mathrm{E}(\ln Y_i|\boldsymbol{x}_i) = \boldsymbol{x}_i'\boldsymbol{\beta}$, not $\mathrm{E}(Y_i|\boldsymbol{x}_i)$. Hence it is the log-lin specification (16) that constructs $\varepsilon_i = \ln Y_i - \mathrm{E}(\ln Y_i|\boldsymbol{x}_i) = \ln Y_i - \boldsymbol{x}_i'\boldsymbol{\beta}$ to be a CEF disturbance, in turn yielding the moment conditions $\mathrm{E}(\boldsymbol{x}_i\varepsilon_i) = \mathrm{E}[\boldsymbol{x}_i(\ln Y_i - \boldsymbol{x}_i'\boldsymbol{\beta})] = \mathbf{0}$.

5 Conclusion

Although the inconsistency result of Santos Silva and Tenreyro has other interpretations, I have argued that it is most naturally and comprehensively understood in terms of the specification of the conditional expectation function of the dependent variable. In applications where the nonlinear function is the compelling specification for the CEF, linearization by a logarithmic or other nonlinear transformation has the implication that OLS estimation of the linearized regression is inconsistent. The reason is simple: the disturbance of the linearized regression is not a CEF disturbance and so does not have the standard properties of CEF disturbances, in particular absence of correlation with the regressors. SST's use of gravity models of international trade to illustrate this situation is one such compelling specification.

But this conclusion hinges on the initial specification of the CEF. There may be nonlinear functional forms for which their linearized version can reasonably be taken to be the conditional mean of the similarly-transformed dependent variable. In such cases the longstanding practice of OLS estimation of the linearized regression is innocuous, because the regression disturbance is a CEF disturbance. (Although as usual inference should be done with the appropriate robust covariance matrix estimator. For example, in the case of cross-sectional sampling a CEF disturbance is conditionally heteroskedastic.) I have suggested the log-lin statistical earnings functions of labor economics as an example.

Notes

[1]For the Stata implementation visit SST's Log of Gravity webpage:

http://personal.lse.ac.uk/tenreyro/lgw.html

For the R implementation see the R package gravity v0.6 which offers a PPML routine as one of its estimation options.

[2]A much less common alternative is that it is taken to be a specification for the conditional median, which leads to quantile regression.

[3]The interpretation in terms of heteroskedasticity also seems to puzzle Head and Mayer (2014, p. 172) who remark "SST frame the problem in terms of heteroskedasticity but this begs the question of which error is not homoskedastic."

References

Davidson, R., and J.G. MacKinnon (2004) *Econometric Theory and Methods* (Oxford).

Fally, T. (2015) "Structural gravity and fixed effects," *Journal of International Economics* 97, 76–85.

Goldberger, A.S. (1991) *A Course in Econometrics* (Harvard).

Gourieroux, C., Monfort, A., and A. Trognon (1984) "Pseudo maximum likelihood methods: Applications to Poisson models," *Econometrica* 52, 701–720.

Head, K., and T. Mayer (2014) "Gravity equations: workhorse, toolkit, and cookbook," Chap. 3 of *Handbook of International Economics,* vol. 4 (Elsevier).

Santos Silva, J.M.C., and S. Tenreyro (2006) "The log of gravity," *Review of Economics and Statistics* 88, 641–658.

Santos Silva, J.M.C., and S. Tenreyro (2010) "On the existence of the maximum likelihood estimates in Poisson regression," *Economics Letters* 107, 310–312.

Santos Silva, J.M.C., and S. Tenreyro (2011) "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator," *Economics Letters* 112, 220–222.

Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data,* 2nd ed. (MIT Press).

Implications for estimating marginal effects

When economists estimate a functional relationship between variables, they are often interested in using the results to characterize the marginal effect on the dependent variable Y_i of each of the explanatory variables $\boldsymbol{x}_i$. Precisely what is meant by "marginal effect" must be defined in the context of a model and, even given a model, there may be more than one useful definition. Both the questions of model choice and definition of marginal effect are distinct from actually estimating that marginal effect, however defined in the context of whatever model.

When the model being estimated is a CEF (4) this marginal effect of the kth explanatory variable is

$$\frac{\partial\,\mathrm{E}(Y_i|\boldsymbol{x}_i)}{\partial x_{ik}} = \frac{\partial \boldsymbol{x}_i(\boldsymbol{\beta})}{\partial x_{ik}}.$$

Because this is the marginal effect on the conditional mean, it is interpreted as the marginal effect on an "average" or "typical" or "representative" value of Y_i.[4]