2022

Dimensional Analysis and Logarithmic Transformations in Applied Econometrics

Deepankar Basu

Department of Economics, University of Massachusetts Amherst

Dimensional Analysis and Logarithmic Transformations in Applied Econometrics

Deepankar Basu*

November 25, 2022

Abstract

In economics, it is common to use dimensioned variables, e.g. earnings (measured in dollars per year), as arguments in the logarithmic function. This is conceptually problematic because a logarithmic function can only take dimensionless quantities as its argument. One way to avoid this conceptual error is to rewrite commonly used logarithmic regressions using an arbitrarily chosen reference unit so that ratios of dimensioned quantities are used in logarithmic functions. With the addition of a zero conditional mean assumption about the reference unit to the standard list of assumptions about asymptotic properties of ordinary least squares estimators, such a reformulated model can ensure consistent estimation of elasticities and semi-elasticities without relying on conceptually problematic mathematical operations.

JEL Codes: C01.

Keywords: logarithm, regression, dimensional analysis.

1 Introduction

The use of logarithmic transformations is widespread in applied econometric analysis. Thousands of papers and book chapters, including previous work by

*Department of Economics, University of Massachusetts Amherst. Email: dbasu@econs.umass.edu. I have benefited from comments by Chris Boone on a previous version of the paper. All remaining errors are of course mine.

this author, have estimated regression functions where a logarithmic transformation of some economic variable was used as a dependent or independent variable. For instance, the vast literature in applied microeconoimics that study earnings functions (Mincerian wage regressions) regress logarithm of earning (often measured by the wage rate) on measures of schooling and other covariates (Card, 2001, equation 5, p. 1132; see also the references in this paper). The equally vast literature in applied macroeconomics that investigates the variation of economic growth across countries typically regress the growth rate of per capita gross domestic product (GDP) on the logarithm of the initial level of per capita GDP, the policy variable of interest and other covariates (Rodrik, 2012, p. 138; see also the references in this paper). A large literature in applied microeconomics is devoted to estimating parameters of production functions. In this literature, the typical regression function uses the logarithm of output as the dependent variable and logarithms of the labor and capital inputs as key regressors (Levinsohn and Petrin, 2003, p. 320, 322.; see also the references to the previous literature). The use of logarithmic transformations is now so common that it is featured in all econometrics textbooks, both at the undergraduate and graduate levels, where it is typically discussed in the section on functional forms of the regression function or to introduce concepts of elasticities; for instance, see Greene (1982, p.160–61), Wooldridge (2002, p. 15–18) and Wooldridge (2016, p.171).

The widespread use of logarithmic transformations of economic variables raises a serious conceptual problem that seem to have escaped the attention of economists. On the one hand, most economic variables, e.g. earning or per capita GDP, are dimensioned quantities. They are measured in terms of relevant units, e.g. dollar per year. On the other hand, logarithmic transformations can only act on and deliver dimensionless quantities. Thus, it is not meaningful to put dimensioned quantities like earnings or per capita GDP or output or labor input, as arguments of logarithmic functions. Nor is it meaningful to use expressions involving the logarithm of units, like log-hours or log-dollars, as is used widely in the applied econometrics literature.

The fact that logarithmic transformations only act on and deliver dimensionless quantities is widely known and emphasized in the physical sciences because of the centrality of dimensional analysis (Matta et al., 2011). It has also been occasionally highlighted by economists like Mayumi and Giampietro (2010) and Shaikh (2016, p. 316), but with little impact on the practice of the mainstream of the discipline. Therefore, this paper revisits this issue.

The first contribution of this paper is to explain, using elementary ideas from mathematical analysis, why it is not meaningful to use dimensioned quantities as arguments in logarithmic functions or why the output of logarithmic transformations are themselves dimensionless. In doing so, I also point out that the argument used to derive this conclusion – that dimensioned quantities cannot be used as arguments of logarithmic or exponential functions – in Mayumi and Giampietro (2010) is faulty.

Once we accept that dimensioned quantities cannot be used in logarithmic functions, we are forced to conclude that a vast literature in applied econometrics has used conceptually meaningless quantities in its analysis. For instance, Mincerian wage regressions, estimation of production functions, estimation of cross country growth regressions, and all previous work, including this author's, which have used logarithmic transformations on dimensioned quantities have used mathematically invalid operations (that involved using logarithms of dimensioned variables).

Logarithmic regressions are, of course, useful in many contexts, especially when the researcher is interested in estimating elasticities or semi-elasticities. Hence, it is desirable to come up with a method to allow the use of logarithmic transformations in applied econometric work that, at the same time, avoids using conceptually meaningless quantities. That motivates the second, constructive, contribution of this paper. I offer a way to address this problem, i.e. to rewrite the model and generate an estimable logarithmic regression function that does not use dimensioned quantities as arguments of logarithmic or exponential functions.

The basic idea behind my proposal is simple. I ask researchers to choose a reference unit and use this unit to rewrite the model in such a way that ratios of dimensioned quantities enter the logarithmic function. The use of *ratios* of dimensioned quantities as arguments in logarithmic functions is a mathematically valid operation because the ratios are dimensionless. In my proposed framework, the coefficients of interest, e.g. elasticities or semi-elasticities, can still be interpreted in the standard manner. Thus, while we avoid mathematically invalid operations like putting in a dimensioned quantity inside a logarithmic function, we retain the ability to estimate elasticities. Consistent estimation of the parameters in the reformulated model requires a zero conditional mean assumption about the reference unit. Once we add this to the list of standard orthogonality assumptions for asymptotic analysis of ordinary least squares estimators, we are ensured consistency.

The rest of this paper is organized as follows: in section 2, I explain why logarithmic functions only take and give dimensionless quantities; in section 3, I offer a simple way to rewrite common logarithmic regressions that avoid the problem of using dimensioned quantities in logarithmic functions; I discuss interpretation of the coefficients and assumptions necessary for consistent estimation; in section 4, I conclude the paper with a broader plea to use dimensional analysis in economics.

2 Logarithm and Dimensions

There are various ways to understand why logarithms, and all other transcendental functions, only act on and deliver dimensionless quantities. All of these approaches rely on noting that transcendental functions are defined in pairs, one being the inverse function of the other (Thomas and Finney, 1996, chapter 6). For instance, the exponential and logarithm function, the main ones of interest in this paper, are inverse functions of each other, over the correct domains of definition of each.

2.1 Definition of Logarithm

Let us start out by recalling that the definite integral of a continuous function, f, over a closed interval $[a, b]$,

$$\int_a^b f(t)dt, \tag{1}$$

is the *limit* of Riemann sums of the form $\sum_{i=1}^{n} f(c_i)\Delta t_i$, where the closed interval $[a, b]$ is partitioned into n subintervals $[t_{i-1}, t_i]$, with $a = x_0 < x_1 < \cdots < x_{n-1} < x_n = b$, $\Delta t_i = t_i - t_{i-1}$ and c_i is any number in the interval $[t_{i-1}, t_i]$ (Thomas and Finney, 1996, p. 313).[1]

Using this understanding of definite integrals, we now recall the definition of the natural logarithm (Binmore, 1982, chapter 13) as,

$$\ln x = \int_1^x \frac{1}{t}dt, \quad x > 0, \tag{2}$$

and note that in this case the Riemann sums are of the form $\sum_i (1/c_i)\Delta t_i$ because the function we are integrating in (2) is $f(t) = 1/t$. Since the unit of measurement of c_i and Δt_i are the same, each term in the Riemann sum is dimensionless. Hence, the integral, being the limit of Riemann sums, is also dimensionless. This establishes the fact that when the logarithmic function operates on a quantity, the *result* is a dimensionless quantity. To see why

[1] The limit that defines the integral can be more precisely stated as follows: let $S(P)$ denote the Riemann sum for some partition P of the closed interval $[a, b]$; then, the integral is supremum of the Riemann sums over all partitions of $[a, b]$, i.e. $\int_a^b f(t)dt = \sup_P S(P)$ (Binmore, 1982, p. 122).

the logarithmic function only takes dimensionless quantities as its *argument*, we need to think about its inverse.

2.2 Logarithm and Exponential as Inverse Functions

The inverse of the logarithmic function exists and is known as the exponential function. Thus, we have,

$$\ln x = y \text{ if and only if } \exp y \equiv e^y = x, \tag{3}$$

where y is a real number and $x > 0$ is a a positive real number (Binmore, 1982, chapter 14.4).

The relationship in (3) is not restricted to defining the logarithm with base 'e', but can be defined for any other meaningful base. Thus, for a real number y, a positive real number x, and a positive number, $b > 0$ that is not equal to 1, we have,

$$\log_b x = y \text{ if and only if } b^y = x. \tag{4}$$

When $b = e$, (4) gives (3). Since natural logarithms, i.e. logarithms with base e, are most commonly used in econometrics, instead of logarithms with other bases, I will restrict my comments to the former.

Equation (3) can show why the argument of the logarithmic function must be dimensionless. It tells us that the number e, a pure number without dimensions, must be raised to the power y to give us x.[2] We have already seen above, using the argument about Riemann sums, that y is a dimensionless quantity. Since e is a pure number, e^y is therefore dimensionless. Since e^y is equal to x, the latter must be dimensionless too. This establishes, with reference to (3), that the argument of a logarithmic function must be a dimensionless quantity.

2.3 A Faulty Argument

Before I turn to drawing out the implications of the argument about logarithms, let me point out that there are some incorrect arguments that deliver the correct conclusion. Such arguments are common in online physics forms, on Wikipedia and, unfortunately, has also percolated into Mayumi and Giampietro (2010). The argument runs as follows: write the infinite series expansion of, for instance, exponential function,

$$e^x = 1 + \frac{x}{1} + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots + \frac{x^n}{n!} + \cdots ,$$

and note that for this equation to make sense, each term on the right must have the same dimension or be dimensionless. Since terms involving powers of x cannot have the same dimension as x, it follows that x must be dimensionless (Mayumi and Giampietro, 2010, p. 1605). We can write the infinite series expansion for $\ln(1 + x)$ to derive the same conclusion.

[2] The number $e = \lim_{n\to\infty} 1 + (1/1!) + (1/2!) + \cdots + (1/n!) \cdots = 2.7182818284500\ldots$

This argument delivers the correct conclusion, but is based on an incorrect argument (Matta et al., 2011, p. 69). To understand the problem, recall that the infinite series expansion written above is just the Taylor series expansion. For any continuously differentiable function, $f(x)$, the Taylor series expansion around x_0, if it exists, is given by

$$f(x_0 + h) = f(x_0) + h\frac{df(x_0)}{dx} + h^2\frac{d^2f(x_0)}{dx^2} + h^3\frac{d^3f(x_0)}{dx^3} + \cdots$$

Considering the dimension of the general term on the right hand side,

$$h^n\frac{d^nf(x_0)}{dx^n}$$

we can note that the dimension of h^n is exactly equal to the dimension of dx^n, both being small changes in x raised to the n-th power. Hence, dimensionally, h^n cancel $1/dx^n$, and we are left with $d^nf(x_0)$. This is the change in the change in ... (n times) of $f(x)$ at x_0. Hence, its dimension, as of every other term on the right hand side, is the same as the dimension of $f(x)$. "Therefore, the addition (or subtraction) of the terms in a Taylor expansion is *numerically* and *dimensionally* permissible and the equation satisfies dimensional homogeneity ... The reason for the necessity of including only dimensionless real numbers in the arguments of transcendental function is not due to the dimensional nonhomogeneity of the Taylor expansion, but rather to the lack of physical meaning of including dimensions and units in the arguments of these function." (Matta et al., 2011, p. 69–70).

2.4 Log Takes and Gives Dimensionless Quantities

Let me now return to the main argument of this paper and remind the reader of the conclusion that it is only meaningful to use dimensionless quantities as arguments of logarithmic (or exponential) transformations, and that the result of using a logarithmic (or exponential) transformation is itself a dimensionless quantity. This has important implications for applied econometric practice.

First, this implies, for instance, that it would not be meaningful to use the logarithm of earning because this variable is a dimensioned quantity measured in, e.g., dollars per year or dollars per hour. This means that the specification of earnings functions used in applied microeconometrics, where log of earnings is used as the dependent variable is problematic. It uses the logarithmic transformation of a dimensioned quantity - which is conceptually meaningless. Thus, equation (5) in Card (2001, p. 1132)

$$\log y = a_0 + \bar{b} S_i - \frac{1}{2} k_i S_i^2 + a_i + \left(b_i - \bar{b} \right) S_i,$$

is problematic because the dependent variable, $\log y$, is conceptually meaningless. Similarly, equation (6) in Levinsohn and Petrin (2003)

$$y_t = \beta_0 + \beta_l l_t + \beta_k k_t + \beta_i i_t + \omega_t + \eta_t$$

is flawed because it involves many terms, e.g. $y_t = \log Y_t, l_t = \log L_t$, that are mathematically inadmissible.

Since it not meaningful to operate the logarithmic transformation on per capita GDP because this variable is a dimensioned quantity measured in real dollars per year, the specification of growth regressions in applied macroeconomics, where log of the initial per capita GDP is used as an independent variable is, once again, problematic because it uses the logarithmic transformation of a dimensioned quantity - which is conceptually meaningless. Thus, equation (1) in Rodrik (2012, p. 138)

$$g_t = \alpha \ln y_{t0} + Z_t' \beta + \gamma s_t + \varepsilon_t$$

is inadmissible because one of the independent variables, $\ln y_{t0}$, is conceptually meaningless.

Second, it implies that it is not meaningful to use expressions like log-hours or log-points of wage or log-dollars, because logarithm of units (or of quantities measured in units) or expressions involving log(units), e.g., log(dollars), are conceptually meaningless (Matta et al., 2011, p. 68). For instance, the unit of measurement on the vertical axis in Figure 1 in Acemoglu et al. (2019) is conceptually meaningless.[3]

Finally, it implies that recent attempts to use the inverse hyperbolic sine function in place of the logarithm (Bellemare and Wichman, 2020) suffers

[3]Regressions in this paper use log per capita GDP measured in year 2000 dollars as the main outcome variable (Acemoglu et al., 2019, p. 55). By the analysis of this paper, that is conceptually problematic.

from the same problem of dimensionality as the logarithm. This is because hyperbolic functions share the same property with logarithms in that they cannot admit dimensioned quantities as arguments:

> In addition to logarithms, it is equally meaningless to include dimensioned quantities as the arguments of trigonometric or hyperbolic functions because these are defined as ratios (the sine of an angle is the ratio of the length of the opposite side to the length of the hypotenuse, the cosine is the ratio of the length of the adjacent side to the length of the hypotenuse, etc.) The hyperbolic functions, themselves defined in terms of either exponential or trigonometric functions, cannot operate on quantities to which physical dimensions are attached either. (Matta et al., 2011, p. 68)

These observations about the logarithm (and transcendental functions more generally) force us to confront another question: How can we reorient applied econometric practice so that it avoids the above problem? I want to argue that if reformulated with the use of a reference unit, logarithmic regressions can be rigorously justified. The key in this task of reformulation is to avoid using dimensioned quantities as arguments of logarithmic or exponential functions; and this can be achieved by using an arbitrary reference unit, as I now show.

3 Estimating Elasticity and Semi-Elasticity

3.1 The Reformulated Model

Suppose we have a random sample of size N for the following variables, y, x_1, x_2, and we wish to estimate the elasticity and semi-elasticity of y with respect to x_1 and x_2, respectively.[4] We start by positing the following relationship between the dependent variable y_i (measured in any units), the independent variables x_{1i} and x_{2i} (measured in whatever units are relevant), and the error term u_i (which is unit-less), for an arbitrary unit i,

$$y_i = x_{1i}^{\beta_1} e^{\beta_2 x_{2i}} e^{u_i}, \qquad y_i > 0, x_{1i} > 0, \quad i = 1, 2, \ldots, n, \tag{5}$$

where β_1 is a dimensionless constant, β_2 is a constant with a dimension that is reciprocal of x_{2i}, and u_i is a dimensionless random variable. The positivity restrictions in (5) are important as they will allow me to meaningfully use logarithmic transformations and divisions.

It is important to note three important dimensional assumptions in (5). First, the fact that β_1 is dimensionless allows me to write x_i^{β} as a meaningful quantity. For, it does not make sense to raise a quantity to the power of another quantity if the latter quantity is dimensioned (Matta et al., 2011). For instance, while it is meaningful to raise the number 10 to the power of

[4] I work with the simplest cross sectional setting where x_1 and x_2 are scalar random variables. It is relatively easy to extend to the case where x_1 and x_2 are vectors of random variables.

2 (a pure number without dimensions), it is meaningless to raise it to the power of 2 dollars or 2 meters! Second, the fact that β_2 is a constant with a dimension that is reciprocal of x_{2i} means that $\beta_2 x_{2i}$ is a dimensionless number. Hence, it is meaningful to use this as an argument in the exponential function, i.e. $e^{\beta_2 x_{2i}}$ is well-defined. Third, the assumption to treat the error term as unit-less is not restrictive. We have much leeway in choosing the units of the unobserved stochastic factors that comprise the error term precisely because they are unobserved. One can think of u_i as an index of a collection of unobserved random variables, each multiplied with coefficients having suitable units to make them unit-less.

Let us choose an arbitrary reference unit and index it by r, and note that the above relationship for this unit is represented by

$$y_r = x_{1r}^{\beta_1} e^{\beta_2 x_{2r}} e^{u_r}.$$

Let us now divide the equation for unit i by the equation for the reference unit and then apply the logarithmic transformation to get the reformulated model,

$$\ln\left(\frac{y_i}{y_r}\right) = \beta_1 \ln\left(\frac{x_{1i}}{x_{1r}}\right) + \beta_2\left(x_{2i} - x_{2r}\right) + \left(u_i - u_r\right). \tag{6}$$

Note that the positivity restrictions in (5) allow us to divide by x_r and y_r in the above equation. Moreover, since y_i/y_r and x_i/x_r are dimensionless quantities (because the numerator and denominator have the same dimensions), it is meaningful to use these ratios as arguments in the logarithmic function.[5]

In deriving the *reformulated model* in (6), I have avoided using dimensioned quantities as arguments in the logarithmic or exponential functions. Note also that both sides of equation (6) are dimensionless, and we have adhered to the basic requirements of dimensional homogeneity, i.e. oranges must be added to or compared with oranges and not apples (Matta et al., 2011, p. 67).

3.2 The Estimable Equation

Rearranging and redefining terms in (6), we get the estimable equation,

$$\ln \tilde{y}_i = \alpha_0 + \alpha_1 \ln \tilde{x}_{1i} + \alpha_2 x_{2i} + \tilde{u}_i, \tag{7}$$

where $\tilde{y}_i = y_i/y_r$, $\tilde{x}_i = x_i/x_r$, $\tilde{u}_i = u_i - u_r$, and the coefficients are

$$\alpha_0 = -\beta_2 x_{2r}, \alpha_1 = \beta_1, \alpha_2 = \beta_2. \tag{8}$$

Equation (7) is the estimable regression equation that will deliver an estimate of the elasticity and semi-elasticity we are interested in. To estimate the elasticity and semi-elasticity in line with the reformulation proposed in this paper, the researcher needs to choose a reference unit, indexed by r, define new variables, $\tilde{y}_i = y_i/y_r, \tilde{x}_{1i} = x_{1i}/x_{1r}$ and estimate (7) by OLS.

[5]Sometimes, logarithmic models involves logarithms only on the right hand side. This is just a special case of the model in (7) and requires no special discussion.

3.3 Interpretation of Coefficients

The interesting feature of the reformulated model in (6) is that β_1 and β_2 have the exact same interpretations that would arise in a standard regression of this form even though in deriving it, I have managed to avoid using dimensioned quantities in logarithmic or exponential transformations.

3.3.1 Elasticity

Let us start with the elasticity, β_1. Note that

$$\beta_1 = \frac{d \ln (y_i / y_r)}{d \ln (x_{1i} / x_{1r})} = \frac{x_{1i}}{y_i} \frac{dy_i}{dx_{1i}},$$

where the last equality follows from treating y_r and x_{1r} as constants while using the chain rule for differentiating the relevant functions.[6] To see this, note that

$$\frac{d \ln (y_i / y_r)}{dx_{1i}} = \frac{d \ln (y_i / y_r)}{d (y_i / y_r)} \frac{d (y_i / y_r)}{dy_i} \frac{dy_i}{dx_{1i}} = \frac{y_r}{y_i} \frac{1}{y_r} \frac{dy_i}{dx_{1i}} = \frac{1}{y_i} \frac{dy_i}{dx_{1i}},$$

and

$$\frac{d \ln (x_{1i} / x_{1r})}{dx_{1i}} = \frac{d \ln (x_{1i} / x_{1r})}{d (x_{1i} / x_{1r})} \frac{d (x_{1i} / x_{1r})}{dx_{1i}} = \frac{x_{1r}}{x_{1i}} \frac{1}{x_{1r}} = \frac{1}{x_{1i}},$$

[6] I have used a simplification while deriving the expression for β_1. The elasticity, β_1, is the partial effect of $\ln (x_i / x_r)$ on $\mathbb{E} \ln (y_i / y_r)$, the conditional expectation of $\ln (y_i / y_r)$, rather than $\ln (y_i / y_r)$. For the most part, little is lost by treating the two as the same when the relevant logarithms are well-defined. (Wooldridge, 2002, p. 17).

so that

$$\frac{d\ln(y_i/y_r)}{d\ln(x_{1i}/x_{1r})} = \frac{x_{1i}}{y_i}\frac{dy_i}{dx_{1i}}.$$

The constancy of y_r and x_{1r}, in turn, is justified by the fact that we have a random sample, so that observations on the reference unit (identified with the index r) is independent of other units (indexed by i). Thus, the model in (7) gives us the correct estimate of the elasticity that we are interested in, β_1, which is the percentage change in y for every percentage change in x_1. But, most importantly, to arrive at estimates of β_1, we do not have to use dimensioned quantities as arguments in logarithmic or exponential functions.

The elasticity, β_1, can also, of course, be written as,

$$\frac{d\ln(y_i)}{d\ln(x_{1i})},$$

but this is not admissible because in writing this expression we have to use dimensioned quantities, y_i and x_{1i}, as arguments in the logarithmic function. That is precisely why we cannot use the standard regression,

$$\ln y_i = a_0 + a_1 \ln x_{1i} + a_2 x_{2i} + u_i,$$

to estimate the elasticity of interest.[7]

[7]It is interesting that while discussing log-log regressions, Greene (1982) refers to the issue of dimensions: "This removes the units of measurement of the variables from consideration in using the [log-log] regression model." (Greene, 1982, p. 160). But there is no discussion of whether it is at all meaningful to have variables with units of measurement in the log terms.

3.3.2 Semi-Elasticity

The coefficient, β_2, in (6) has the interpretation of semi-elasticity because

$$\beta_2 = \frac{d\ln(y_i/y_r)}{dx_{2i}} = \frac{d\ln(y_i/y_r)}{d(y_i/y_r)}\frac{d(y_i/y_r)}{d(y_i)}\frac{dy_i}{dx_{2i}} = \frac{1}{y_i}\frac{dy_i}{dx_{2i}},$$

where the same argument as above can be used to justify the different steps of the differentiation on the right hand side.

3.4 Asymptotic Properties of OLS

Can the parameters of the reformulated model be consistently estimated by ordinary least squares (OLS)? The answer is in the affirmative, if the researcher is willing to add a zero conditional mean assumption about the reference unit to the list of standard orthogonality assumptions for consistency of OLS estimators. I make this explicit below.

3.4.1 Assumption and Results

Assumption 1. *The following assumptions hold.*

(a) Random sample: We have a random sample (y_i, x_{1i}, x_{2i}) *of size* N*, i.e., observations on unit* i *and unit* j *are independent for all* $i, j = 1, 2, \ldots, N$*.*

(b) Standard orthogonality assumptions: For all $i = 1, 2, \ldots, N$*,* $\mathbb{E}x_{1i}u_i = \mathbb{E}x_{2i}u_i = 0$*,* $\mathbb{E}u_i = 0$*, and the matrix of regressors has full rank.*

(c) *Zero conditional mean for reference unit: For the reference unit indexed by r,* $\mathbb{E}\left(u_r|x_{1r}\right)=0$.

Proposition 1. *If assumption 1 holds then the OLS estimators of the parameters in (7) converge in probability to the corresponding parameters in (6).*

Proof. The crucial condition for consistency of OLS estimators of the parameters in (6) requires

$$\mathbb{E}\left(\tilde{x}_{1i}\tilde{u}_i\right)=0, \quad \mathbb{E}\left(x_{2i}\tilde{u}_i\right)=0. \tag{9}$$

Let us start with the second condition,

$$\mathbb{E}\left(x_{2i}\tilde{u}_i\right)=\mathbb{E}x_{2i}\left(u_i-u_r\right)=\mathbb{E}x_{2i}u_i-\mathbb{E}x_{2i}u_r=\mathbb{E}x_{2i}u_i-\mathbb{E}x_{2i}\mathbb{E}u_r.$$

Considering the two terms on the far right, we see that assumption 1(b) will ensure the first term is zero and, since assumption 1(a) allows us to write the expectation of the product, $\mathbb{E}x_{2i}u_r$, as the product of expectations, $\mathbb{E}x_{2i}\mathbb{E}u_r$, assumption 1(b) will then ensure that the second term is also zero.

Turning to the first condition, we have,

$$\mathbb{E}\left(\tilde{x}_{1i}\tilde{u}_i\right)=\mathbb{E}\left[\frac{x_{1i}}{x_{1r}}\left(u_i-u_r\right)\right]=\mathbb{E}\left(x_{1i}u_i\right)\mathbb{E}\left(\frac{1}{x_{1r}}\right)-\mathbb{E}x_{1i}\mathbb{E}\left(\frac{u_r}{x_{1r}}\right),$$

where I have used assumption 1(a) to write expectations of products as products of expectations. Assumption 1(b) shows that the first term on the far right hand side is zero, and writing the last term on the right hand side above as

$$\mathbb{E}\left(\frac{u_r}{x_{1r}}\right) = \mathbb{E}\left(\frac{1}{x_{1r}}\mathbb{E}\left[u_r|x_{1r}\right]\right),$$

where the outer expectation on the right hand side is with respect to the distribution of x_{1r}, and using assumption 1(c) allows us to conclude that this term is zero.[8]

The conclusion now follows from an application of Theorem 4.1 in Wooldridge (2002, p. 53). □

The implication of this theorem is that if we add the zero conditional mean assumption (assumption 1(c)) to the standard orthogonality assumptions used for asymptotic analysis of OLS estimators, we are ensured consistent estimates of the elasticity and semi-elasticity in the reformulated model (7). Thus, we are able to avoid using conceptually meaningless mathematical operations, e.g. using dimensioned quantities as arguments in the logarithmic function, and also to derive consistent estimates of the elasticity and semi-elasticity. An additional homoskedasticity assumption (Wooldridge, 2002, p. 54) will deliver asymptotic normality of the OLS estimators. Moreover, such assumptions can be significantly weakened, allowing for heteroskedasticity, serial correlation (if the time dimension is present in the data set) and clustering. These considerations are less important than consistency

[8] This is precisely where we need the zero conditional mean assumption. Zero correlation between x_{1r} and u_r will not suffice because we have a nonlinear function of x_{1r} in the expectation.

(Wooldridge, 2002, p. 56), and hence, in this paper, I focus on the latter.

3.4.2 The Trade Off

The proposal of this paper would allow researchers to avoid using conceptually questionable mathematical operations while estimating useful quantities like elasticities and semi-elasticities. Of course, the proposal is not costless. The zero conditional mean assumption is the cost we have to bear to allow consistent parameter estimates in the reformulated model (7). The zero conditional mean assumption is necessary because zero covariance does not carry over to nonlinear functions, and a crucial component of the proposal in this paper involves division by x_{1r}, giving rise to a nonlinear function of x_{1r}.

The zero conditional mean assumption amounts to assuming that *all* functions of the regressor which needs to be log-transformed is uncorrelated with the error term for the reference unit. One can go a step further and make this assumption for all units. This would be tantamount to assuming that the model is correctly specified (Wooldridge, 2002, p. 18). This is of course stronger than the standard orthogonality assumptions used in the asymptotic analysis of OLS estimators. Therefore, researchers face a trade-off.

On the one hand, they can ignore the fact that using dimensioned quantities as arguments in logarithmic, or other transcendental, functions is conceptually problematic. If they do so, they can continue estimating logarithmic regressions in the standard way as, for example,

$$\ln y = a_0 + a_1 \ln x + u,$$

where y and x are dimensioned quantities. This method will give them numerically correct estimates of elasticity, but to do so, they will need to use conceptually meaningless terms involving logarithms of dimensioned quantities, like ln(y dollar per hour).

On the other hand, they can decide to stop using dimensioned quantities as arguments in logarithmic, or other transcendental functions, because such operations are conceptually problematic. If they choose to take this route, then they can use the reformulated model (7) suggested in this paper. With the addition of a zero conditional mean assumption, they are guaranteed consistent estimates. The zero conditional mean assumption (assumption 1(c)) is of course more stringent than the standard orthogonality assumptions used to derive consistency of OLS estimators.

4 Conclusion

Logarithmic transformations are used widely in applied economics to estimate elasticities and semi-elasticities. In most application, logarithmic transformations are applied to economic variables that are measured in some units (e.g. earnings, measured in dollars per year; per capita GDP, measured in real dollars per year). This is problematic because the logarithmic function, like other transcendental functions, can only act on and deliver dimensionless quantities. To avoid using dimensioned quantities in logarithmic (or exponential) functions, in this paper I have offered a simple and constructive way to re-write logarithmic regressions by using an arbitrary reference unit and using ratios of dimensioned quantities. If we add a zero conditional mean assumption about the reference unit to the standard assumptions used in the asymptotic analysis of OLS estimators, we can ensure that the reformulated model will provide consistent estimates of elasticity and semi-elasticity. By adopting this approach, economists can avoid using meaningless quantities in their empirical analyses and yet derive meaningful estimates of important magnitudes like elasticities and semi-elasticities.[9]

This paper makes the broader case that we, as economists, need to pay more attention to the dimensions of variables we use in our theoretical and empirical analyses (Mayumi and Giampietro, 2010). Dimensional analysis, which is common in the physical sciences, should be adopted in economics. The only economics book that I am aware of that seriously discussed dimensions of variables and carried out some dimensional analysis is Foley et al. (2019, section 2.3). While they are not sufficient, correct dimensions of equations and variables are necessary for economically meaningful work (Matta

[9]An alternative methodology is suggested in Mayumi and Giampietro (2010) that allows a regression without the use of logarithms to have a higher R-squared than a corresponding regression where some or all variables are in logs. If the motivation for using logs is to estimate elasticities and not to ensure higher R-squared, the algorithm in Mayumi and Giampietro (2010, section 4) would have limited use.

et al., 2011). Paying attention to dimensions of variables can often help in identifying inadvertent errors in arguments or analyses.

References

Acemoglu, D., Naidu, S., Restrepo, P., and Robinson, J. A. (2019). Democracy does cause growth. *Journal of Political Economy*, 1(127):47–100.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. *Oxford Bulletin of Economics and Statistics*, 1(82):50–611.

Binmore, K. G. (1982). *Mathematical Analysis: A straightforward approach.* Cambridge University Press, Cambride, UK, 2nd edition.

Card, D. (2001). Estimating the return to schooling: Progress on some persistent econometric problems. *Econometrica*, 5(69):1127–1160.

Foley, D. K., Michl, T. R., and Tavani, D. (2019). *Growth and Distribution.* Harvard University Press, 2nd edition.

Greene, W. H. (1982). *Econometric Analysis.* Prentice Hall, Boston, MA, 7th edition.

Levinsohn, J. and Petrin, A. (2003). Production functions using inputs to control for unobservables. *Review of Economic Studies*, 2(70):317–341.

Matta, C. F., Massa, L., Gubskaya, A. V., and Knoll, E. (2011). Can one take the logarithm or the sine of a dimensioned quantity or a unit? Dimensional analysis involving transcendental functions. *Journal of Chemical Education*, 1(88):1127–1160.

Mayumi, K. and Giampietro, M. (2010). Dimensions and logarithmic function in economics: A short critical analysis. *Ecological Economics*, (69):1604–1609.

Rodrik, D. (2012). Why we learn nothing from regressing economic growth on policies. *Seoul Journal of Economics*, 2(25):137–151.

Shaikh, A. (2016). *Capitalism: Competition, Conflict, Crises*. Oxford University Press, New York, NY.

Thomas, G. B. and Finney, R. L. (1996). *Calculus and Analytic Geometry*. Addison-Wesley Publishing Company, Reading, MA, 9th edition.

Wooldridge, J. M. (2002). *Econometric Analysis of Cross Section and Panel Data*. MIT Press, Cambridge, MA.

Wooldridge, J. M. (2016). *Introductory Econometrics: A Modern Approach*. Cenage Learning, Boston, MA, 6th edition.