Unitat d'Història Econòmica
UHE Working Paper 2010_01

A Critical Analysis of Dimensions and Curve Fitting Practice in Economics

Kozo Mayumi[1,*], Mario Giampietro[2] and Jesús Ramos-Martín[2, 3]

(1) Faculty of Integrated Arts and Sciences

The University of Tokushima

Minami-Josajinma 1-1

Tokushima City 770-8502, Japan

(2) Institut de Ciència i Tecnologia Ambientals

Universitat Autònoma de Barcelona

08193 Bellaterra, Spain

(3) Unitat d'Història Econòmica

Universitat Autònoma de Barcelona

08193 Bellaterra, Spain

(*) Corresponding author

e-mail: mayumi@ias.tokushima-u.ac.jp

25/02/2010

UAB
Universitat Autònoma
de Barcelona

A Critical Analysis of Dimensions and Curve Fitting Practice in Economics

Kozo Mayumi[1, *], Mario Giampietro[2], and Jesús Ramos-Martín[2, 3]

(1) Faculty of Integrated Arts and Sciences

The University of Tokushima

Minami-Josajinma 1-1

Tokushima City 770-8502, Japan

(2) Institut de Ciència i Tecnologia Ambientals

Universitat Autònoma de Barcelona

08193 Bellaterra, Spain

(3) Unitat d'Història Econòmica

Universitat Autònoma de Barcelona

08193 Bellaterra, Spain

(*) Corresponding autor,

e-mail: mayumi@ias.tokushima-u.ac.jp

Abstract

When dealing with sustainability we are concerned with the biophysical as well as the monetary aspects of economic and ecological interactions. This multidimensional approach requires that special attention is given to dimensional issues in relation to curve fitting practice in economics. Unfortunately, many empirical and theoretical studies in economics, as well as in ecological economics, apply dimensional numbers in exponential or logarithmic functions. We show that it is an analytical error to put a dimensional unit x into exponential functions (a^x) and logarithmic functions ($\log_a x$). Secondly, we investigate the conditions of data sets under which a particular logarithmic specification is superior to the usual regression specification. This analysis shows that logarithmic specification superiority in terms of least square norm is heavily dependent on the available data set. The last section deals with economists' "curve fitting fetishism". We propose that a distinction be made between curve fitting over past observations and the development of a theoretical or empirical law capable of maintaining its fitting power for any future observations. Finally we conclude this paper with several epistemological issues in relation to dimensions and curve fitting practice in economics.

Keywords: dimensions; logarithmic function; curve fitting; logarithmic specification

1. Introduction

Wassily Leontief (1982) once criticized the dismal performance of academic economists who could not "advance, in any perceptible way, a systematic understanding of the structure and the operation of a real economic system". According to Leontief, since economists have not been subjected to the harsh discipline of systematic fact-finding,

traditionally imposed on and accepted by the natural sciences, they have a predilection for deductive reasoning. This results in mathematical models full of formalism nonsense without serious empirical content.

In that letter, Leontief also touched upon the identification problem: aggregating equations in economics after they have been transformed for the purpose of curve-fitting. He used the example of a theoretical production function- a transcendental logarithmic production function, introduced by Christensen et al., 1973. It was on this occasion that he committed an analytical error concerning the issue of dimensions. This error is the main concern of this paper. What we mean by dimensions here is the elementary units (such as mass, length, time, or currency) referring to the definition of an external referent, which is required for the gathering of empirical data expressed as quantitative values assigned to the chosen proxy variable. It should be noted that dimensions in mathematics such as the Hausdorff dimension or fractal dimensions (e.g., Hurewicz and Wallman, 1948; Edgar, 1990) have nothing to do with 'dimensions' as discussed in this paper.

We believe that the vast majority of readers might wonder why dimensions matter in economics. Perhaps it is not well known that one of the first four papers published by, Georgescu-Roegen, an admirable epistemologist, in *Quarterly Journal of Economics* (Georgescu-Roegen, 1936) was concerned with dimensions in relation to the marginal utility of money. This paper by Georgescu-Roegen gave a solution to a famous controversy between A. C. Pigou and Milton Friedman[1] based on the following series where p_i stands for the price of commodity i and φ stands for utility.

$$T = \frac{p_x^2}{\varphi_{xx}} + \frac{p_y^2}{\varphi_{yy}} + \frac{p_z^2}{\varphi_{zz}} + \ldots$$

Here T is not a pure number. Rather, it has the dimension (money)^2/(utility). This is an important point, since for example, a change from US dollars to cents increases the numerical value of T by 10,000 times. "As there is no sense in speaking of a dimensional quantity as small or large, the difficulty of dimension arises at once when we pass from a mathematical constant to a quasi-constant. What do we mean, for instance, by T becoming very large as the number of commodities increases? First of all, how can we recognize whether T is large or small? By choosing appropriate units [dimensions] of measurement, T can be made to have any numerical value we please. We may try to avoid this question of dimensions by assuming the units are fixed once and for all. But this assumption does not help toward proving that the numerical value of T will increase indefinitely with the number of commodities" (Georgescu-Roegen, 1936, p. 535). Georgescu-Roegen also provided a similar discussion on the same issue in relation to Marshall's constancy of marginal utility of money (Georgescu-Roegen, 1968).

When dealing with sustainability, we are concerned with the biophysical as well as the monetary aspects of economic and ecological interactions. Therefore it is imperative to investigate the dimensional issues from a much broader perspective. This is the main reason that we discuss this issue from a slightly different angle: the issue of dimensions in relation to curve fitting in empirical analysis. Section 2 deals with two functions, the exponential function (a^x) and the logarithmic function ($\log_a x$). We show that x must be a dimensionless pure number, otherwise it does not make any sense.

Due to the logical consequences outlined in Section 2, it is unfortunate that many empirical and theoretical studies in economics, as well as in ecological economics, use dimensional numbers in exponential or logarithmic functions. Section 3 introduces several unfortunate examples of this analytical error. These observations are presented with the hope that they will orient future quantitative analyses of sustainability issues toward more constructive ends without introducing unwanted and unnecessary analytical errors into discussions on theoretical issues. Since many empirical studies adopt the logarithmic specification of the production function in general, Section 4 investigates conditions of data sets under which a particular logarithmic specification is superior to the usual regression specification. Using a simple example, we show that the logarithmic specification superiority, in terms of least square norm, is heavily dependent on the available data set. Section 4 also deals with economists' "curve fitting fetishism". We claim that it is essential to make a clear distinction between curve fitting over past observations and the development of a theoretical or empirical law that must be capable of fitting any future observations. Section 5 concludes this paper with several epistemological issues in relation to dimensions and curve fitting practice in economics.

2. Exponential and logarithmic functions revisited

First of all, it should be remembered that an exponential function and a logarithmic function with the same base are the inverse functions of each other.

Let's start with exponential function a^x. Since $a^x = e^{x\ln a}$, the following is true.

$$a^x = 1 + x\ln a + \frac{(x\ln a)^2}{2!} + \frac{(x\ln a)^3}{3!},,,+ \frac{(x\ln a)^n}{n!},,,, \quad (1).$$

In particular, if $a=e$ (e is the base of the natural logarithm),

$$e^x = 1 + x + \frac{x^2}{2!} + \frac{x^3}{3!},,,+ \frac{x^n}{n!},,,,, \quad (2).$$

For (2) to make sense, *each element $x^k/k!$ ($k=0,1,,,$)* must either have the same dimension or be a dimensionless pure number. If neither of these two conditions are satisfied, the summation cannot be done. For example, if x is a number with a physical dimension (e.g. 10kg), then the expression (2) is nonsensical. For this reason, it is nonsensical to use a monetary unit, for example, US dollars in a given year of reference, as x in (2). The nonsense of this operation is illustrated in Figure 1. Since a logarithmic function is the inverse function of an exponential function, it is also absurd to put a dimensional number into a logarithmic function However, it is instructive to deal with this issue directly by using an infinite expansion expression for $\log_e(1+x)$:

$$\log_e(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} +,,,+ \frac{(-1)^{n-1}x^n}{n} +,,,, \quad \text{where} \quad -1 < x < 1. \quad (3)$$
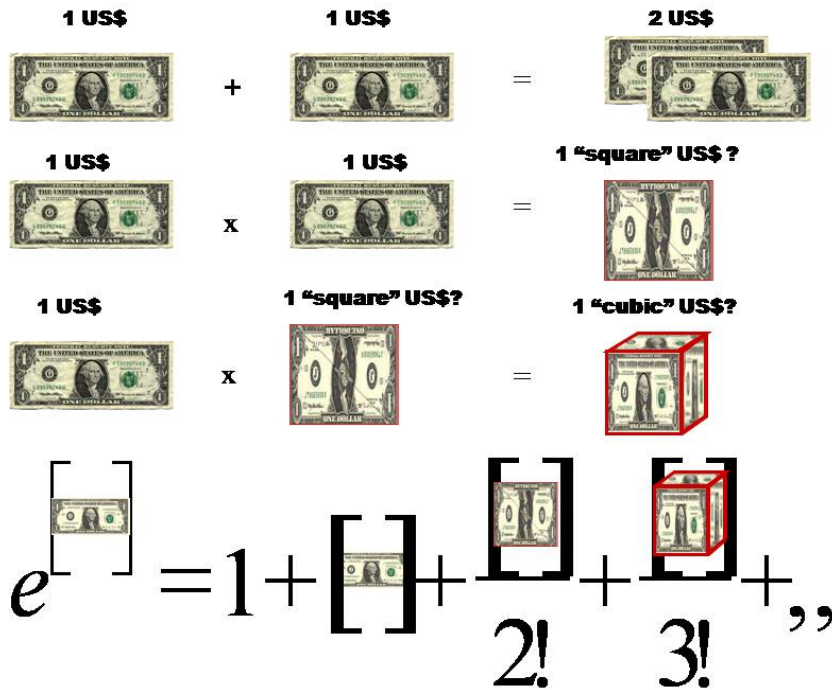
Figure 1. "Exponential Function" with US$

Suppose now we assign to x the value of 0.5 U.S. dollars. Again (3) would not make any sense. This example clarifies the point that we cannot put a dimensional unit in logarithmic functions, either. Of course it is possible to put b times c in a logarithmic function $\log(bc)$ if b and c have opposite dimensions which make their product dimensionless. For example b has the dimension of (time) and c has the dimension of 1/(time). This is the reason why we can use e^{-rt} in the present value calculation, where r has dimension of 1/(time) and t has dimension of (time).

The logarithmic function has a variety of advantages. It can conveniently represent a set made of very large and small numbers; it roughly correspond to the human perception of smell, sound and light; and it makes it possible to carry out the multiplication of ratios by simple addition and subtraction.

For example, the decibel (dB) is a logarithmic unit of measurement that expresses the magnitude of a physical quality (usually power or intensity) relative to a *reference level*. Since it expresses a ratio of two quantities with the same unit, *it is a dimensionless unit*. For example, sound pressure level (SPL) is expressed as:

$$SPL = 20\log_{10}\frac{P_1}{P_0} \quad (4)$$

where P_0 is the reference sound pressure of 2 x 10^{-5} N/m^2.

Since from relation (4), it seems that the following expression could be accepted even if a and b were numbers with the same dimension:

$$\ln a - \ln b = \ln \frac{a}{b} \quad (5)$$

But *a* and *b* must satisfy the following formula as well:

$$\ln a + \ln b = \ln ab \quad (6)$$

If *a* and *b* are numbers with the same dimension, three terms in (6) do not make any sense. To repeat, in logarithmic formulae, all the terms must be dimensionless pure numbers!

3. Fallacies about dimensions in analytical economics: some examples

Logarithmic specification of production functions is extremely popular in empirical analysis found in economic journals. As a matter of fact, we picked up a few samples from the first two issues of *Ecological Economics* vol. 56 in 2006 to check whether the scientific community of Ecological Economics was more aware of this issue, but we found similar dimensional errors.

Before reviewing examples of the poor use of mathematical functions in relation to the issue of dimensions, we have to reveal our own error with respect to dimensions, which was made when using a logarithmic function (Pastore, et al., 2000). In that paper, after introducing the biophysical indicator of socioeconomic development BEP (Bio-Economic Pressure), we tried to check whether or not BEP is well correlated with other traditional economic indicators. BEP is the energy consumption within a particular year divided by the hours of human labor allocated to the productive sectors of the economy. The dimensions of BEP is (energy unit per year)/(work hours). In spite of this fact, we mistakenly put this quantitative assessment of BEP into a logarithmic function. Of course this operation cannot be accepted.

Morse (2006) discusses the dangers inherent in attempting to simplify using mathematical equations something as complex as development. He did so by exploring the Lynn and Vanhanen theory of deterministic development. Before presenting Morse's own critique, he presented a regression result by introducing an alternative specification into the original Lynn and Vanhanen model as follows:

$$\log_{10}(GDP/capita) = 0.77 + 0.0342(nationalIQ) \quad (7)$$

Morse states that "it has to be noted that it is possible to improve upon the original Lynn and Vanhanen regression using the logarithmic (base 10) of GDP/capita as the dependent variable" (Morse, 2006, p. 82). Morse also mentions that transforming "the GDP/capita data increases the R^2 from 54% to 70%, *suggesting a significant improvement in the model*" (italics added, Morse, 2006, p. 82). Of course, as discussed before, his transformed regression cannot be accepted.

Richmond and Kaufmann (2006) analyze the effect of fuel mix, model specification, and the level of development on the presence and size of a turning point in the relationship between income and energy use and/or carbon emissions. They state:

"Income is measured by GPD. Data for GDP are obtained from the Penn World Tables 6.1. *GDP is measured in 1996 US\$, and is converted from local currency units using PPP indices*" (italics added, Richmond and Kaufmann, 2006, p. 178).

They use one specification for the effect of income on energy use or carbon emissions as follows:

$$\ln(Y_{it}) = \alpha + \beta \ln(X_{it}) + \phi \ln(Z_{it}) + \mu_{it} \quad (8)$$

where Y_{it} is per capita energy use (BTU/Population) or per capita carbon emissions (CO_2/Population); X_{it} is per capita GDP; Z_{it} is the vector of fuel shares; μ_{it} is the regression error. This form of specification with Y_{it} and X_{it} cannot be accepted either.

Let's turn to several examples from neoclassical production and cost function theory. The first example is Arrow et al (1961). Arrow et al. tried to investigate the substitution between capital and labor within the neoclassical production theory. In Section I of that paper they used regression analysis incorporating the following variables (Arrow et al., 1961, p. 227):
V: value added in thousands of U.S. dollars
L: labor input in man-years
W: wages (total labor cost divided by L) in dollars per man-year

They statistically *tested* the following two simple relations using these three variables:

$$\frac{V}{L} = c + dW + \eta \quad (9)$$

$$\log \frac{V}{L} = \log a + b \log W + \varepsilon \quad (10)$$

Of course relation (8) cannot be used judging by the dimensions *V/L* and *W* which they used.

Pindyck (1979) looked at the role of energy in the structure of production and tried to obtain new evidence on the extent of capital, labor and energy substitution within the translog cost function theory. He used the following specification of the price index for energy:

$$\log P_E = \alpha_0 + \sum \alpha_i \log P_i + \sum_i \gamma_{ij} \log P_i \log P_j \quad (11)$$

where P_E is the price *index* for energy and P_i (i=1,2,3,4) are the fuel prices for oil, gas, coal and electricity. Again, this specification cannot be used for P_i.

Pindyck's formulation surely comes from the original formulation of transcendental logarithmic production and price frontiers investigated by Christensen et al. (1973). Let's look at this original formulation. They assumed that there are two outputs—consumption (*C*) and investment (*I*)—and two inputs—capital (*K*) and labor (*L*). The corresponding prices are q_C, q_I, q_K, q_L. They call *F* the production frontier in the following formulation:

$$\ln(F+1) = \alpha + \alpha_C \ln C + \alpha_I \ln I + \alpha_K \ln K + \alpha_L \ln L + \alpha_A \ln A + \ln C(\frac{1}{2}\beta_{CC} \ln C +$$

$$\beta_{CI} \ln I + \beta_{CK} \ln K + \beta_{CL} \ln L + \beta_{CA} \ln A) + \ln I(\frac{1}{2}\beta_{II} \ln I + \beta_{IK} \ln K + \beta_{IL} \ln L + \beta_{IA} \ln A) +$$

$$\ln K(\frac{1}{2}\beta_{KK} \ln K + \beta_{KL} \ln L + \beta_{KA} \ln A) + \ln L(\frac{1}{2}\beta_{LL} \ln L + \beta_{LA} \ln A) + \ln A(\frac{1}{2}\beta_{AA} \ln A)$$

(12), where, *according to them,* A is an index of technology.

It is not clear how to create these indexes. However, they use the price frontier as follows:

$$\ln(P+1) = \alpha_0 + \alpha_C \ln q_C + \alpha_I \ln q_I + \alpha_K \ln q_K + \alpha_L \ln q_L + \alpha_A \ln A + \ln q_C(\frac{1}{2}\beta_{CC} \ln q_C +$$

$$\beta_{CI} \ln q_I + \beta_{CK} \ln q_K + \beta_{CL} \ln q_L + \beta_{CA} \ln A) + \ln q_I(\frac{1}{2}\beta_{II} \ln q_I + \beta_{IK} \ln q_K + \beta_{IL} \ln q_L + \beta_{IA} \ln A) + (13).$$

$$\ln q_K(\frac{1}{2}\beta_{KK} \ln q_K + \beta_{KL} \ln q_L + \beta_{KA} \ln A) + \ln q_L(\frac{1}{2}\beta_{LL} \ln q_L + \beta_{LA} \ln A) + \ln A(\frac{1}{2}\beta_{AA} \ln A)$$

Since they said that the "corresponding *prices* are q_C, q_I, q_K, q_L" (Christensen et al., 1973, p. 33, italics added), this specification cannot be used.

Of course they might argue that these variables are indexes. Unfortunately these indexes are on many occasions pseudo measures that produce inconsistent ranking orders. There is an indispensable characteristic for an index to be a pseudo measure according to Georgescu-Roegen (1971). Perhaps the following example is the best one to show this characteristic.

There are at least two indices (a type of sum based on a set of values of p_i) to explain a kind of "dispersion".

$$E = \sum_1^s p_i^2 \quad (14)$$

$$H = \sum_1^s p_i \ln p_i \quad (15)$$

where $\sum_1^s p_i = 1$.

Both of these indices reach their minimum for the microstate of lowest order, namely, $p_1 = p_2 = ,,, = p_s$ (and only for this), and their maximum for any macrostates of highest order, $p_k = 1$ for any k (and only for this).

However, *these indices produce inconsistent magnitudes of E and H depending on the values taken by p_i.* In another words, E does not rank order in the same way H does. It is possible to prove this phenomenon in general, but for the sake of simplicity we will use the following example.
Given the following two distributions:

#1$\{p_1, p_2, p_3\} = \{0.1, 0.2, 0.7\}$ and

#2$\{p_1, p_2, p_3\} = \{0.1, 0.5, 0.4\}$

the corresponding values of E and H are as follows:

$E_{\#1} = 0.54 > E_{\#2} = 0.42$, and, $H_{\#1} = -0.8018 > H_{\#2} = -0.9433$

For this example, the ranking is preserved for both E and H.

However, given the following two distributions:

#3$\{p_1, p_2, p_3\} = \{0.2, 0.2, 0.6\}$ and

#2$\{p_1, p_2, p_3\} = \{0.1, 0.5, 0.4\}$

the corresponding values of E and H are as follows:

$E_{\#3} = 0.44 > E_{\#2} = 0.42$, and, $H_{\#3} = -0.9502 < H_{\#2} = -0.9433$

So, E and H are pseudo measures. Generalizing this example we can conclude that creating an index that is not a pseudo measure is a difficult problem.

4. "Curve fitting fetishism" in economics

As suggested in the previous section when we discussed the Morse paper (2006), it is worthwhile to investigate the issue of whether or not using logarithmic functions as a dependent variable really improves the least square norm. We compare two regressions for the same data set as follows:

$$Y = \alpha_1 + \beta_1 X + u_1 \quad (16)$$

$$\ln Y = \alpha_2 + \beta_2 X + u_2 \quad (17)$$

where u_1 and u_2 are regression error terms.

The corresponding least square norms are:

$$R_1^2 = \frac{\{n(\sum_1^n X_i Y_i) - (\sum_1^n X_i)(\sum_1^n Y_i)\}^2}{\{n(\sum_1^n X_i^2) - (\sum_1^n X_i)^2\}\{n(\sum_1^n Y_i^2) - (\sum_1^n Y_i)^2\}} \quad (18)$$

$$R_2^2 = \frac{\{n(\sum_1^n X_i \ln Y_i) - (\sum_1^n X_i)(\sum_1^n \ln Y_i)\}^2}{\{n(\sum_1^n X_i^2) - (\sum_1^n X_i)^2\}\{n(\sum_1^n (\ln Y_i)^2) - (\sum_1^n \ln Y_i)^2\}} \quad (19)$$

We discuss the condition on values of data set (X_i, Y_i) for which $R_2^2 > R_1^2$ is satisfied.

First we consider the following quadratic form:

$$n(\sum_{1}^{n} X_i Y_i) - (\sum_{1}^{n} X_i)(\sum_{i}^{n} Y_i) = {}^{t}X \begin{pmatrix} (n-1) & -1,,,, & -1 & -1 \\ -1 & (n-1),,,, & -1 & -1 \\ ,,, & ,,, & ,,, & ,,, \\ -1 & ,,, & -1 & (n-1) \end{pmatrix} Y = {}^{t}XAY \quad (20)$$

where $X = {}^{t}(X_1, X_2,,, X_n)$ and $Y = {}^{t}(Y_1, Y_2,,, Y_n)$ (t signifies transpose)

Since $\det(\lambda E - A) = \lambda(\lambda - n)^{n-1}$ where E is the unit matrix, there are two eigenvalues of A, namely, O and n.

An eigenvector associated with the eigenvalue O is ${}^{t}(1,1,,,,1)$.

(n-1) eigenvectors associated with the eigenvalue n are ${}^{t}(\delta_{1k}, \delta_{2k},,,, \delta_{(n-1)k}, -1)$ where δ_{ik} is the Kroneker delta and $k = 1,2,,,,(n-1)$.

By utilizing the Gram-Schmidt orthogonalization process (e.g., Roman, 2008, p. 214), we can construct the following normalized orthogonal matrix P (i.e., ${}^{t}PP = E$) from the n eigenvectors obtained above:

$$P = (P_1, P_2,,,, P_n) \quad (21)$$

where

$$P_1 = {}^{t}(\frac{1}{\sqrt{2}}, 0, 0,,,,, 0, \frac{-1}{\sqrt{2}})$$

$$P_2 = {}^{t}(\frac{-1}{\sqrt{2^2+2}}, \frac{2}{\sqrt{2^2+2}}, 0, 0,,, 0, \frac{-1}{\sqrt{2^2+2}})$$

$$P_3 = {}^{t}(\frac{-1}{\sqrt{3^2+3}}, \frac{-1}{\sqrt{3^2+3}}, \frac{3}{\sqrt{3^2+3}}, 0, 0,,, 0. \frac{-1}{\sqrt{3^2+3}})$$

$$P_4 = {}^{t}(\frac{-1}{\sqrt{4^2+4}}, \frac{-1}{\sqrt{4^2+4}}, \frac{-1}{\sqrt{4^2+4}}, \frac{4}{\sqrt{4^2+4}}, 0, 0,,, 0, \frac{-1}{\sqrt{4^2+4}})$$

$$\vdots$$

$$P_{n-1} = {}^{t}(\frac{-1}{\sqrt{(n-1)^2+(n-1)}}, \frac{-1}{\sqrt{(n-1)^2+(n-1)}},,,, \frac{-1}{\sqrt{(n-1)^2+(n-1)}}, \frac{n-1}{\sqrt{(n-1)^2+(n-1)}})$$

$$P_n = {}^{t}(\frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}},,,,, \frac{1}{\sqrt{n}}, \frac{1}{\sqrt{n}})$$

Therefore, after transforming X and Y into the following forms

$$Z = {}^{t}PX \quad (21)$$
$$W = {}^{t}PY \quad (22)$$
$$T = {}^{t}P(\ln Y) \quad (23)$$

where

$Z = {}^{t}(Z_1, Z_2,,, Z_n)$, $W = {}^{t}(W_1, W_2,,, W_n)$ and for notational purpose we represent

${}^{t}(\ln Y_1, \ln Y_2,,, \ln Y_n)$ as $\ln Y$.

Then we obtain

$$n(\sum_{1}^{n} X_i^2) - (\sum_{1}^{n} X_i)^2 = n\sum_{i}^{n-1} Z_i^2 \quad (24),$$

$$n(\sum_{1}^{n} Y_i^2) - (\sum_{1}^{n} Y_i)^2 = n\sum_{1}^{n-1} W_i^2 \quad (25) \quad \text{and}$$

$$n(\sum_{1}^{n} X_i Y_i) - (\sum_{1}^{n} X_i)(\sum_{1}^{n} Y_i) = n\sum_{1}^{n-1} Z_i W_i \quad (26)$$

It should be noted that since the eigenvalue corresponding to $^t(1,1,,,,1)$ is zero, the nth term does not show up in (24), (25) and (26).

Thus we obtain

$$R_1^2 = \frac{(\sum_{1}^{n-1} Z_i W_i)^2}{n(\sum_{1}^{n-1} Z_i^2)(\sum_{1}^{n-1} W_i^2)} \quad (27)$$

In the same way, we obtain

$$R_2^2 = \frac{(\sum_{1}^{n-1} Z_i T_i)^2}{n(\sum_{1}^{n-1} Z_i^2)(\sum_{1}^{n-1} T_i^2)} \quad (28)$$

Let θ_1 be the angle between $Z^1 = {}^t(Z_1, Z_2,,,, Z_{n-1})$ and $T^1 = {}^t(T_1, T_2,,,, T_{n-1})$ and let θ_2 be the angle between $Z^1 = {}^t(Z_1, Z_2,,,, Z_{n-1})$ and $W^1 = {}^t(W_1, W_2,,,, W_{n-1})$.

These angles are measured counterclockwise from Z.

We are interested in the region (θ_1, θ_2) that satisfies the following inequality:

$$\frac{R_2^2}{R_1^2} = \frac{\{\|Z^1\|\|T^1\|\cos\theta_1\}^2 \times \|W^1\|^2\}}{\{\|Z^1\|\|W^1\|\cos\theta_2\}^2 \times \|T^1\|^2\}} > 1 \quad (29)$$

where $\|Z^1\| = \sqrt{\sum_{1}^{n-1} Z_i^2}$, $\|T^1\| = \sqrt{\sum_{1}^{n-1} T_i^2}$, and $\|W^1\| = \sqrt{\sum_{1}^{n-1} W_i^2}$.

Therefore, relation (29) becomes
$$(\cos\theta_1 + \cos\theta_2)(\cos\theta_1 - \cos\theta_2) > 0 \quad (30)$$

It is easy to check whether or not relation (30) is satisfied with a given data set on (X_i, Y_i).

Since $\cos\theta_1 = (\sum_{1}^{n-1} Z_i T_i)/(\sqrt{\sum_{1}^{n-1} Z_i^2}\sqrt{\sum_{i}^{n-1} T_i^2})$ and $\cos\theta_2 = (\sum_{1}^{n-1} Z_i W_i)/(\sqrt{\sum_{1}^{n-1} Z_i^2}\sqrt{\sum_{i}^{n-1} W_i^2})$, these two values are easily calculated from relations (21), (22) and (23).

Transforming the above inequality (30) into the following:

$$\sin(\theta_1 + \theta_2)\sin(\theta_1 - \theta_2) < 0 \quad (31)$$

In Figure 2 we show the region (in black) where relation (29) is satisfied.
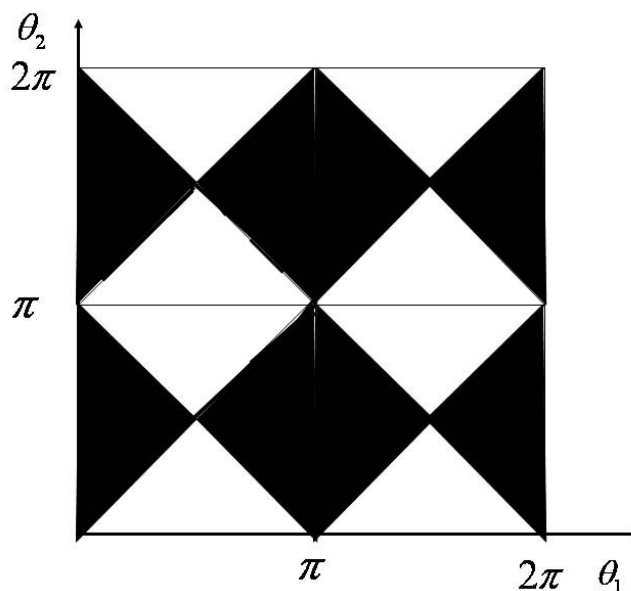

Figure 2. Regions for the two angles for $R_2{}^2 > R_1{}^2$

      The important point here is that only the relationship between the two angles measured counterclockwise from $Z^1 = {}^t(Z_1, Z_2, , , , Z_{n-1})$ characterizes the superiority of the two regressions in terms of least square norm, not necessarily the overall spatial distribution of the data. So, setting a dependent variable in logarithmic form does not necessarily improve the regression in terms of least square norm. The analysis presented above can be easily generalized and formulated for cases where logarithmic independent variables are included.

      In relation to curve fitting practice in economics, Georgescu-Roegen once aptly remarked (1966, p. 277, italics added), "econometricians seem to ignore the fact that a better fit obtained by *adding a new variable* does not mean at all that the formula is also a better law. For a formula to represent a law it is not sufficient that it should fit well the available observations: the acid test is the fit for all other observations". The present situation for econometric analyses seems to have greatly worsened due to the increasing computational power of computers and programming techniques.

In mathematics there is a famous theorem called the Weierstrass Approximation Theorem: a real-valued continuous function can be approximated *uniformly* over an

interval by a polynomial (e.g., Randolph, 1968). The polynomial has a finite length of expression in terms of variables and constants using the operations of addition, subtraction and multiplication and constant non-negative integer exponents.

For illustration purpose we construct a polynomial series that uniformly converges to a continuous function $f(x)$ where $f(x) = 0$ for $x \leq 0$ or $x \geq 1$.
Suppose the following continuous function,
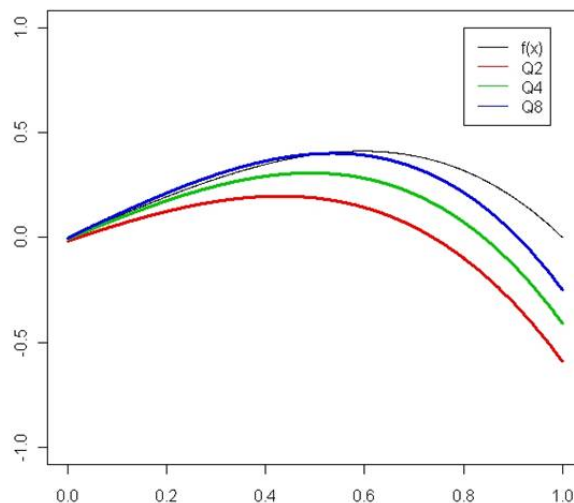$$f(x) = (1-x)e^x \sin x \quad (32) \quad \text{where } x \in [0,1]$$
The following polynomials in x converge to $f(x)$,
$$Q_2 = \frac{3\sqrt{2}}{4} \int_{-1}^{1} f(x+t)(1-t^2)^2 dt \quad (33)$$

$$Q_4 = \frac{3}{2} \int_{-1}^{1} f(x+t)(1-t^2)^4 dt \quad (34)$$

$$Q_8 = \frac{3\sqrt{2}}{2} \int_{-1}^{1} f(x+t)(1-t^2)^8 dt \quad (35)$$

Q_2, Q_4, and Q_8 are to the power of 4, 8 and 16 respectively. Figure 3 shows uniform convergence of Q_i into $f(x)$. Raising the power of polynomials corresponds to Georgescu-Roegen's sense of adding new variables.

Figure 3.   Polynomial approximation

So, it is rather easy to have a polynomial approximation that can fit perfectly well to past data using computer programming. However, the situation facing economists is

much more formidable. The "true function" *f(x)* cannot be known in advance, especially if we seriously consider the evolutionary nature of the economic process!

The resulting curve fitting - Q_2, Q_4, and Q_8 - is a series of approximations that is supposed to be a real "law". Unfortunately *f(x)* itself is simply a formal representation of the perceived behavior of a system created by a modeler. Therefore, this formal representation is based on: (i) the relevant system narrative adopted by the modeler; and (ii) data observed in the system and based on the perception of the modeler.

5. Conclusion

Concerning the issue of dimensions we have shown that it is an analytical fallacy that dimensional numbers may be used in exponential and logarithmic functions. However, the issue of dimensions, and of scale generally, is much more complex. In fact, after agreeing with the points made in this paper, the reader might think that as far as the same dimension is used, the summation of different terms is always possible in a meaningful way. Unfortunately this is not the case. When dealing with sustainability, we have to deal with biophysical aspects of economics, namely energy and material flows in social and ecological systems. Economists know well that it is not always possible to sum numerical values of dollars if these values refer to different years. The same applies to biophysical variables: it is not always possible to make the summation of different items in terms of energy (e.g., Joule) and weight (e.g., kg). When dealing with complex systems operating across different levels of organization and at different scales, in order to generate meaningful sums, one has always to specify a criterion of equivalence according to which we can sum "quantities" defined at different levels and in different descriptive domains. It should be noted that each metabolic system (living things, machines or social systems) has its own inherent characteristics determining the quantity and quality of energy and material transformation. Viruses consume a very specific set of molecules; the cow eats hay; cars consume energy in the form of liquid fuels; and each country consumes a mix of energy carriers with a variety of intensities, which have to be produced starting from a mix of primary energy sources (Giampietro and Mayumi, 2009). It is simply not true that all Joules are equivalent, when we account for energy flowing in a socio-economic system over a period of 30 years and across different hierarchical levels of organization. This is not true for the same reason that dollars values are not equivalent over time. . Even if combining the different energy values suffices, moving from, for example, Pica Joules (10^{-15}) to Exa Joules (10^{18}) in order to deal with different system scales, we must be aware that the quantitative assessments belonging to these different scales cannot be related to each other simply because they are expressed in a common dimensional unit (Giampietro and Mayumi, 2004; Giampietro, 2006; Mayumi and Giampietro, 2004). That is, when dealing with a representation across different scales, it is impossible to obtain the same external referent relevant for all metabolic systems, even though the dimensional unit for *the representation* is the same (e.g. Joules of energy).

When addressing the dimensions issue in relation to curve fitting practice in economics, there is an important epistemological problem. This problem regards the representation of the production process in quantitative terms. Neoclassical production functions, whether for individual firms or the aggregate economy, usually assume that any factor can always be substituted for any other factor. The implication of this

assumption is that an increase in the input of any factor always yields an increase in output. For neoclassical economists any factor is a jelly like substance, so that production is carried out everywhere in the input-output space. Such a space is assumed in the classic paper by H. S. Houthakker who formally derived the Cobb-Douglas production function based on the generalized Pareto distribution (Houthakker, 1955). As S. Islam aptly showed, the second law of thermodynamics excludes the possibility of obtaining production isoquants of the Cobb-Douglass type (Islam, 1985)[2]. However, there is more to it. Those neoclassical economists adopting the substitution assumption have not paid due attention to the essential distinction between flows and funds in the material production process (Georgescu-Roegen, 1971). This distinction leads to the heart of the issue of scale. It is the pre-analytical selection of a time horizon for the analysis, a descriptive domain associated with the choice of a given scale, that defines what is produced by an economy. On a short timescale one can decide to focus the analysis on the production of goods and services (performing an analysis of the flows). On a longer time scale, when concerned for sustainability, one can decide to focus the analysis on the processes required to produce and consume goods and services by performing an analysis of the reproduction and expansion of the funds. These two different types of analysis will provide different conclusions to the modeler. Neglecting the distinction between funds and flows results in a systematic indifference to the biophysical foundation of economic activities. It is not surprising then, that the curve fitting practice typical of aggregated production functions prevails.

Any actual material production process is limitational in the sense that within a given factory process we cannot always compensate a decrease in output due to a decrease in a fund element (e.g. capital) by an increase in a flow input (e.g. natural resources). Hence, the representation of isoquants, the concept of elasticity of substitution, and the time derivative of a function by technological improvements, lose any operational and empirical meaning (Mayumi et al., 1998). All these concepts are found in the neoclassical theory of production. However, Georescu-Roegen noticed a much more serious "analytical and conceptual fallacy" within the neoclassical treatment of the development process: "It is high time, I believe, for us to recognize that the essence of development consists of the organizational and flexible power to create new processes rather than the power to produce commodities by materially crystallized plants" (Georgescu-Roegen, 1971: 275). This power is termed as the Π-sector by Georgescu-Roegen (1971): "an economy can "take off" when and only when it has succeeded in developing a Π-sector". This issue of the Π-sector is related to the question of what is produced by the economic process. Some of those studying the functioning of socioeconomic processes seem to be confused as to what is actually produced by the economic process. According to Georgescu-Roegen, the economic process does not produce goods and services alone, but rather it produces a 'reproducible system', via an integrated process of production and consumption of goods and services. When considering the whole socioeconomic system, it is the integrated action of the productive economic sector and the sector of final consumption which have to be considered. Using Georgescu-Roegen's terminology, the economic process has the goal of reproducing and expanding the various fund elements defined simultaneously across different levels and scales. It accomplishes this task by using disposable flows. Therefore, we can conclude that an economy not only produces goods and services, *but more importantly, produces the processes required for*

producing and consuming goods and services (Giampietro and Mayumi, 2009). This neglected aspect of the economic process in conventional economics is the true reason why the curve fitting, based on dynamical system models and past data, results in continuous and inevitable failures to predict the future.

Concluding this overview of the epistemological challenges faced by those willing to generate a quantitative analysis of sustainability issues, we can say that the validation of any dynamical system model can be assured only if both "the knowledge and the definition of the modeler" and the "observed system in the model" remain stable during the given time horizon. Put in another way, the model remains valid only if the selected representation will not become either semantically or syntactically obsolete over time. Unfortunately, experience tells us that when dealing with sustainability issues, these two conditions are never respected. For example, every time econometric models failed to predict energy demand, econometricians found a ready, yet self-defeating, excuse: "history has changed the parameters" (Georgescu-Roegen, 1976). Georgescu-Roegen notes that if "history is so cunning, why persist in predicting it? What quantitative economics needs, above all, are economists such as Simon Kuznets, who would know how to pick up a small number of relevant variables, instead of relying upon the computer to juggle with scores of variables and thus losing all mental [introspective] contact with the dialectical nature of economic phenomena."

The epistemological challenge associated with evolving systems is due to the mismatch between these two facts: (1) the information space used by any formal system of inference (mathematical model) must be closed, finite and discrete, otherwise it would not be possible to run such a model in finite time; (2) the information space for describing any evolving system is open and always expanding. By "information space" we mean the formal representation of the evolving system expressed in terms of the epistemological categories required to characterize its behavior. This implies that no matter how good a given model is, the simulated behavior always depends on the validity of the initial choice of typologies used in the representation. Unfortunately for modelers, individual realizations belonging to given typologies tend to evolve in time, "becoming" something else (Prigogine, 1978). Thus, the validity of any model of an evolving system is bound to expire due to two plausible reasons:

(a) semantic obsolescence - the set of relevant attributes for the observed system must change in time, since the concerns justifying the modeling will naturally evolve with the advancement of knowledge. Thus, the qualities monitored and the priority given to various criteria of performance, will sooner or later cease to reflect the modeler's perception of relevance to the goals and problem structure (e.g. outdating of the narratives of neoclassical economics theory).

(b) syntactic obsolescence - the set of relevant attributes for the observed system remains the same for the concerned modeler, but the model can no longer provide accurate prediction of the values taken by key indicators, since the observed system has become something else (outdating of the validity of the curve fitting parameters). The model is no longer able to simulate the movements of the system within its original state space.

Note
1. Georgescu-Roegen states on this controversy: The "verdict was against Friedman. As we all know, if you disagree with him however little, Milton Friedman would clobber you: "you are *totally* wrong" So I felt immensely gratified when Milton introduced me before a lecture at the University of Chicago as the only economist who had proved him wrong. Of course, my lecture, on Brazilian monetary inflation, "was *totally* wrong"" (Georgescu-Roegen, 1988. p. 27).
2. To be fair to Cobb and Douglas we should note the following: when carefully reading Cobb and Douglas' classic paper (1928), one remains awed by their meticulous attitude. They devoted almost half of their paper to the task of how to create indices for capital and labor. They were very careful about avoiding the generation of pseudo measures with the inconsistent ranking order of capital and labor indices.

References

  Arrow, K. J., Chenery, H.B., Minhas, B.S., and Solow, R.M. 1961. Capital-labor substitution and economic efficiency. *The Review of Economics and Statistics* 43(3), 225-250.
  Christensen, L. R., Jorgenson, D. W. and Lau, L. J. 1973. Transcendental logarithmic production frontiers. *The Review of Economics and Statistics* 55 (1), 28-45.
  Cobb, C. W. and Douglas, P. H. 1928. A theory of production. *The American Economic Review* 18 (1) Supplement, 139-165.
  Edgar, G. A. 1990. Measure, Topology, and Fractal Geometry. Springer, New York.
  Georgescu-Roegen, N. 1936. Marginal utility of money and elasticities of demand. *Quarterly Journal of Economics* 50, 532-539.
  Georgescu-Roegen, N. 1968. Revisiting Marshall's constancy of marginal utility of money. *Southern Economic Journal* 35, 176-181.
  Georgescu-Roegen, N. 1966. Further thoughts on Corrado Gini's Delusioni dell'econometria. *Merton* 25, 265-279.
  Georgescu-Roegen, N. 1971. *The Entropy Law and the Economic Process*. Harvard University Press, Cambridge Mass.
  Georgescu-Roegen, N. 1976. *Energy and Economic Myths*. Pergamon Press, New York.

Georgescu-Roegen, N. 1988. An emigrant from a developing country: Autobiographical Notes-I. *Banca Nationale del Lavoro Quarterly Review* 164, 3-31.

Giampietro, M. 2006 Theoretical and practical considerations on the meaning and usefulness of traditional energy analysis (Comments on "The Energetic Metabolism of the European Union and the United States" by Haberl and Colleagues) *Journal of Industrial Ecology* Vol.10(4): 173-185

Giampietro, M. and Mayumi, K. 2004. Complex systems and energy. In: C. Cleveland (Ed.), *Encyclopedia of Energy.* Elsevier, San Diego, Vol. 1, pp. 617-631.

Giampietro, M. and Mayumi, K. 2009. *The Biofuel Delusion: The Fallacy of Large-Scale Agro-Biofuel Prodcution*. The Earthscan, London.

Houthakker, H.S. 1955. The Pareto distribution and the Cobb-Douglass production function in activity analysis. *The Review of Economic Studies* 23, 27-31.

Hurewicz, W. and Wallman, H. 1948. *Dimension Theory*. Princeton University Press. Princeton.

Islam, S. 1985. Effect of an essential input on isoquants and substitution elasticities. *Energy Economics* 7,194-196.

Leontief, W. 1982. Academic economists. *Science* 217, 104-107.

Mayumi, K. and Giampietro, M. 2004. *Entropy in ecological economics*. In: J. Proops and P. Safonov (Eds.), *Modeling in Ecological Economics.* Edward Elgar, Cheltenham (UK), pp. 80-101.

Mayumi, K., Giampietro, M. and Gowdy, J. 1998. Georgescu-Roegen/Daly versus Solow/Stiglitz Revisited. *Ecological Economics* 27, 115-117.

Morse, S. 2006. Making development simple. The genetic deterministic hypothesis for economic development. *Ecological Economics* 56, 79-88.

Pastore, G. , Giampietro, M. and Mayumi, K. 2000. Societal metabolism and multiple-scale integrated assessment: empirical validation and examples of application. *Population and Environment* 22 (2), 211-254.

Pindyck, R. S. 1979. Interfuel substitution and the industrial demand for energy: an international comparison. The Review of Economics and Statistics 61(2), 169-179.

Prigogine, I. 1978. *From Being to Becoming* San Francisco: W.H. Freeman and company.

Randolph, J. F. 1968. *Basic Real and Abstract Analysis*. Academic Press, New York.

Richmond, A.K. and Kaufmann, R. K. 2006. Is there a turning point in the relationship between income and energy use and/or carbon emissions? Ecological Economics 56, 176-189.

Roman, S. 2008. *Advanced Linear Algebra*. Third Edition. Springer, New York.