Why Exponential Almon Lag Works Well in Econometrics: An Invariance-Based Explanation

1st Laxman Bokati
Computational Science Program
University of Texas at El Paso
El Paso, Texas 79968, USA
lbokati@miners.utep.edu

2nd Vladik Kreinovich
Department of Computer Science
University of Texas at El Paso
El Paso, Texas 79968, USA
vladik@utep.edu

Abstract—In many econometric situations, we can predict future values of relevant quantities by using an empirical formula known as exponential Almon lag. While this formula is empirically successful, there have been no convincing theoretical explanation for this success. In this paper, we provide such a theoretical explanation based on general invariance ideas.

Index Terms—invariance, econometrics, expoinential Almon lag

I. Formulation of the Problem

Ubiquity of linear prediction formulas. One of the main objectives of science is to predict future events based on the past behavior of the corresponding system.

In general, to adequately describe a complex system, one needs to describe the values of many quantities characterizing this system. However, in many cases, already a single quantity provides a reasonable description of a system. Let us provide two examples.

- While many quantities need to be described to get a good description of the state of a country's economy, in the first approximation, Gross Domestic Product – GDP – provides a reasonably good description of this state.
- Similarly, to have a very good understanding of a stock market, it is desirable to have a detailed description of how the values of different stocks change with time. However, in the first approximation, a single variable – the stock market index – provides a good description of the current state of the stock market and of its dynamics.

For such descriptions, prediction means predicting the value x_t of the corresponding quantity at a future moment t based on the current and previous values x_{t-1}, x_{t-2}, ..., of this quantity. In mathematical terms, prediction means applying some algorithm $f(x_{t-1}, x_{t-2}, \ldots)$ to known values x_{t-i}, x_{t-2}, ..., to generate the estimate for the future value x_t.

The very fact that we only use a single variable to describe a system means, as we have mentioned, that we are using

a first approximation model. It is therefore reasonable to apply the same idea of the first approximation to describing possible functions $f(x_{t-1}, \ldots)$. In general, sufficiently smooth functions can be expanded in Taylor series, and by selecting appropriate terms in these series, we can get more and more accurate approximations.

- If we only keep constant and linear terms, we get the first approximation.
- If we also keep quadratic terms, we get the second-order approximation, etc.

This is a usual way to analyze physical systems in general; see, e.g., [2], [5]. So, in the first approximation, it is reasonable to assume that the function $f(x_{t-1}, \ldots)$ is linear, i.e., that

$$f(x_{t-1}, x_{t-2}, \ldots, x_{t-k}) =$$

$$w_0 + w_1 \cdot x_{t-1} + w_2 \cdot x_{t-2} + \ldots + w_k \cdot x_{t-k}, \quad (1)$$

where $t - k$ describes the earliest value that we take into account in our prediction.

One of the main ideas about prediction is that if a system remained in the same state for a long time, it will most probably remain in this state in the future. This idea is why believe in conservation laws in physics – since energy or momentum of a closed system have remained the same for a long time, we conclude that this quantity will retain the same value in the future as well.

In relation to the formula (1), this means that if we have

$$x_{t-1} = \ldots = x_{t-k} = x,$$

then the predicted value

$$x_t = f(x_{t-1}, x_{t-2}, \ldots, x_{t-k})$$

should also be equal to the same quantity x. In other words, we should require that

$$x = w_0 + w_1 \cdot x + \ldots + w_k \cdot x, \quad (2)$$

i.e., equivalently, that

$$w_0 + x \cdot (w_1 + \ldots + w_k - 1) = 0 \quad (3)$$

for all x.

In general, a linear function is always equal to 0 if both its coefficients are equal to 0, so we should have $w_0 = 0$ and

$$w_1 + \ldots + w_k = 1. \tag{4}$$

Since $w_0 = 0$, the linear dependence takes the form

$$f(x_{t-1}, x_{t-2}, \ldots, x_{t-k}) =$$

$$w_1 \cdot x_{t-1} + w_2 \cdot x_{t-2} + \ldots + w_k \cdot x_{t-k}. \tag{5}$$

Now, selecting a good prediction formula means selecting appropriate values w_i.

Empirical fact: exponential Almon formula works well in econometrics. Several formulas for w_i have been tried. It turned out that in many econometric applications, the following weights work the best [3]:

$$w_i = \frac{\exp(a_0 + a_1 \cdot i + a_2 \cdot i^2 + \ldots + a_n \cdot i^n)}{\sum\limits_{j=1}^{k} \exp(a_0 + a_1 \cdot j + a_2 \cdot j^2 + \ldots + a_n \cdot j^n)}. \tag{6}$$

This formula is known as *exponential Almon lag* – since it is an exponential version of a formula previously proposed by Almon.

Why this formula: what we do in this paper. While the formula (6) has been empirically successful, there is no convincing theoretical explanation for this success.

In this paper, we use invariance ideas to provide such a theoretical explanation.

II. REFORMULATING THE PROBLEM IN PRECISE TERMS AND RESULTING EXPLANATION

How weights depend on time? Different weights w_i correspond to different time intervals:

- the weight w_1 corresponds to the selected time quantum Δt,
- the weight w_2 corresponds to the time interval $2 \cdot \Delta t$, and,
- in general, the weight w_i corresponds to the time interval $i \cdot \Delta t$.

At first glance, it may look like we can simply assume that the weight is a function of time: $w_i = f(i \cdot \Delta t)$. However, this assumption will not lead to the equality (4). To maintain this equality, we need to normalize these values by dividing each value $f(i \cdot \Delta t)$ by their sum:

$$w_i = \frac{f(i \cdot \Delta t)}{\sum\limits_{j=1}^{k} f(j \cdot \Delta t)}. \tag{7}$$

So, natural question is: what is an appropriate function $f(t)$?

Different functions $f(t)$ may lead to the same weights w_i. In different situations, we may have different weights – and thus, different functions $f(t)$. However, even when the weights are the same, we may have different functions $f(t)$.

For example, if we take a function $g(t) = C \cdot f(t)$ for some constant C, then both numerator and denominator of the right-hand side of the formula (7) will be multiplied by the same constant C and thus, the weights will remain the same:

$$\frac{f(i \cdot \Delta t)}{\sum\limits_{j=1}^{k} f(j \cdot \Delta t)} = \frac{g(i \cdot \Delta t)}{\sum\limits_{j=1}^{k} g(j \cdot \Delta t)}. \tag{8}$$

Vice versa, let us show that if two functions $f(t)$ and $g(t)$ lead to the same weights, i.e., if we have (8) for all i, then the functions $f(t)$ and $g(t)$ differ only by a multiplicative constant: $g(t) = C \cdot f(t)$ for some constant C.

Indeed, the formula (8) implies that

$$\frac{g(i \cdot \Delta t)}{f(i \cdot \Delta t)} = \frac{\sum\limits_{j=1}^{k} g(j \cdot \Delta t)}{\sum\limits_{j=1}^{k} f(j \cdot \Delta t)}. \tag{9}$$

The right-hand side of the formula (9) is the same for all time intervals $i \cdot \Delta t$, i.e., is a constant. If we denote this constant by C, we get the desired equality $g(t) = C \cdot f(t)$.

The desired dependence, in general, depends on many factors. The exact form of the dependence $f(t)$ depends on many different factors, i.e., on the values of many quantities $q_1, \ldots, q_m$ that characterize these factors. In general, we have

$$f(t) = F(q_1(t), \ldots, q_m(t)). \tag{10}$$

Possibility to re-scale numerical values. A numerical value of a physical quantity depends on the selection of the measuring unit. For example, the same height can be described as 1.7 m and as 170 cm.

In general:

- if we replace the original measuring unit by a new one which is $\lambda > 0$ times smaller,
- then all the numerical values of the corresponding quantity q get multiplied by λ:

$$q \mapsto \lambda \cdot q.$$

Scale-invariance. In many cases, there is no preferred measuring unit. This means that the physical dependence should not depend in which units we use for measuring each of the quantities q_a. We should get the same formula no matter what re-scaling $q_i \mapsto \lambda_a \cdot q_a$ we apply to the numerical values of each of these quantities. In other words, the weights w_i as described by the formula (7) should not change if:

- instead of the original values $q_a(t)$,
- we use re-scaled values $\lambda_a \cdot q_a(t)$,

i.e., equivalently, if we use a re-scaled function

$$g(t) = F(\lambda_1 \cdot q_1(t), \ldots, \lambda_m \cdot q_m(t)). \tag{11}$$

We have already shown that the only possibility for two functions $f(t)$ and $g(t)$ to lead to the same weights w_i is

when $g(t) = C \cdot f(t)$ for some constant C, i.e., in our case, when

$$F(\lambda_1 \cdot q_1, \ldots, \lambda_m \cdot q_m) =$$
$$C(\lambda_1, \ldots, \lambda_m) \cdot F(q_1, \ldots, q_m) \quad (12)$$

for some constant C that, in general, depends on the re-scaling parameters λ_i.

What can we conclude from this scale-invariance. It is known (see, e.g., [1]) that all continuous solutions of the functional equation (12) have the form

$$F(q_1, \ldots, q_m) = A \cdot q_1^{a_1} \cdot \ldots \cdot q_m^{a_m}. \quad (13)$$

Vice versa, it is easy to show that all functions of the type (13) are scale-invariant, i.e., satisfy the equation (12), for

$$C(\lambda_1, \ldots, \lambda_m) = \lambda_1^{a_1} \cdot \ldots \cdot \lambda_m^{a_m}. \quad (14)$$

In different situations, we may have different dependence on the quantities q_i, i.e., we may have different values A and a_i. It is therefore reasonable to consider the class of *all* the functions of the type (13).

This class of functions is closed under multiplication and raising to a power. This closeness can be described in easier-to-process terms if we take the logarithm of both sides and consider the function

$$L(q_1(t), \ldots, q_m(t)) \overset{\text{def}}{=} \ln(F(q_1(t), \ldots, q_m(t)) \quad (15)$$

for which

$$F(q_1(t), \ldots, q_m(t)) = \exp(L(q_1(t), \ldots, q_m(t)). \quad (16)$$

For this logarithm, the equality (13) takes the form

$$L(q_1(t), \ldots, q_m(t)) = a + a_1 \cdot Q_1(t) + \ldots + a_m \cdot Q_m(t), \quad (17)$$

where we denoted $a \overset{\text{def}}{=} \ln(A)$ and

$$Q_a(t) \overset{\text{def}}{=} \ln(q_a(t)). \quad (18)$$

Formula (17) describes all possible linear combinations of functions 1 and $Q_a(t)$. Thus, the logarithmic functions corresponding to all possible expressions (13) form a finite-dimensional linear space.

Scale-invariance with respect to time. Similarly to the fact that there is usually no preferred measuring unit for measuring the quantities q_a, there is usually also no preferred unit for measuring time intervals.

Thus, it is reasonable to assume that:
- if we change the measuring unit for time interval

$$t \mapsto \mu \cdot t,$$

- then we should get the same family of functions.

In other words, it is reasonable to assume that the linear space (17) is *scale-invariant* in the sense that:
- with every function $L(t)$,
- this space should also contain functions $L(\mu \cdot t)$ corresponding to different values μ.

Final explanation. It is also reasonable to require that all the functions (17) are analytical, i.e., that they can be expanded in Taylor series; this is, as we have mentioned, a usual assumption in the analysis of physical systems; see, e.g., [2], [5].

It is known – see, e.g., [4] (and it not that difficult to prove) that every function from a sca;e-invariant finite-dimensional linear space of analytical functions is a polynomial. (For reader's convenience, in the following auxiliary section, we describe how this can be proven.) Thus, every logarithmic function $L(q_1(t), \ldots, a_m(t))$ is a polynomial. So, the function $f(t) = F(q_1(t), \ldots, q_m(t))$ is equal to e to the polynomial-of-t power. Thus, that due to the formula (7), the weights w_i have the desired form (6).

So, we have indeed explained the empirical success of the formula (6) – it turns out to be the only formula that satisfies the natural invariance conditions.

III. AUXILIARY SECTION: HOW TO PROVE THE RESULT THAT WE CITED

Let us consider any function $L(t)$ from the scale-invariant finite-dimensional linear space S of analytical functions. Since this function is analytical, it has the form

$$L(t) = c_0 + c_1 \cdot t + c_2 \cdot t^2 + \ldots \quad (19)$$

Some of the coefficients c_i may be equal to 0, so let us keep only non-zero terms in the Taylor expansion (19):

$$L(t) = c_{k_1} \cdot t^{k_1} + c_{k_2} \cdot t^{k_2} + \ldots, \quad (20)$$

where $k_1 < k_2 < \ldots$

Let us prove that the space S contains all the power functions t^{k_1}, t^{k_2}, etc. corresponding to all non-zero coefficients c_{k_i}. Since all power functions are linearly independent, and the space S is finite-dimensional, this would imply that the expansion (20) contains only finitely many terms and is, thus, a polynomial.

Let us first prove that the space S contains the function t^{k_1}. Indeed, since the space S is scale-invariant, with the function $L(t)$, it also contains, for every μ, the function

$$L(\mu \cdot t) = c_{k_1} \cdot \mu^{k_1} \cdot t^{k_1} + c_{k_2} \cdot \mu^{k_2} \cdot t^{k_2} + \ldots \quad (21)$$

Since S is a linear space, it also contains the function

$$a_{k_1}^{-1} \cdot \mu^{-k_1} \cdot L(\mu \cdot t) = t^{k_1} + \frac{c_{k_2}}{c_{k_1}} \cdot \mu^{k_2 - k_1} \cdot t^{k_2} + \ldots \quad (22)$$

Any finite-dimensional space is closed in the topological sense (in the sense that it contains all its limits). In the limit $\mu \to 0$, the function (22) tends to t^{k_1}. Thus, this function is indeed contained in the space S.

Since a linear space S contains the functions (20) and t^{k_1}, it also contains their linear combination

$$L(t) - c_{k_1} \cdot t^{k_1} = c_{k_2} \cdot t^{k_2} + \ldots \quad (23)$$

Thus, similarly, we can prove that the function t^{k_2} is also contained in the space S, etc.

The statement is thus proven.

References

[1] J. Aczél and J. Dhombres, *Functional Equations in Several Variables*, Cambridge University Press, 2008.

[2] R. Feynman, R. Leighton, and M. Sands, *The Feynman Lectures on Physics*, Addison Wesley, Boston, Massachusetts, 2005.

[3] E. Ghysels, A. Sinko, and R. Valkanov, "MIDAS Regressions: Further Results and New Directions", *Econometric Reviews*, 2007, Vol. 26, No. 1, pp. 53–90.

[4] A. Jaimes, C. Tweedie, T. Magoc, V. Kreinovich, and M. Ceberio, "Selecting the best location for a meteorological tower: a case study of multi-objective constraint optimization", In: M. Ceberio and V. Kreinovich (eds.), *Constraint Programming and Decision Making*, Springer Verlag, Berlin, Heidelberg, 2014, pp. 61–66.

[5] K. S. Thorne and R. D. Blandford, *Modern Classical Physics: Optics, Fluids, Plasmas, Elasticity, Relativity, and Statistical Physics*, Princeton University Press, Princeton, New Jersey, 2017.