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

Laxman Bokati$^1$ and Vladik Kreinovich$^{1,2}$

$^1$Computational Science Program
$^2$Department of Computer Science
University of Texas at El Paso, El Paso, Texas 79968, USA
lbokati@miners.utep.edu, vladik@utep.edu
1. 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.

- The first example is that many quantities need to be described to get a good description of the state of a country’s economy.

- However, in the first approximation, Gross Domestic Product – GDP – provides a reasonably good description of this state.
2. Ubiquity of linear prediction formulas (cont-d)

• 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 $x_{t-1}, x_{t-2}, \ldots$, of this quantity.
3. Ubiquity of linear prediction formulas (cont-d)

- In mathematical terms, prediction means:
  - applying some function \( a(x_{t-1}, x_{t-2}, \ldots) \) to known values \( x_{t-i}, x_{t-2}, \ldots, \)
  - to generate the estimate for the future value \( x_t. \)

- 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 \( a(x_{t-1}, \ldots). \)

- In general, sufficiently smooth functions can be expanded in Taylor series.
4. Ubiquity of linear prediction formulas (cont-d)

- 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.

- So, in the first approximation, it is reasonable to assume that the function $a(x_{t-1}, \ldots)$ is linear, i.e., that

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

- Here $t - k$ describes the earliest value that we take into account in our prediction.
5. Ubiquity of linear prediction formulas (cont-d)

- 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 we 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 above formula, this means that:
  - if we have $x_{t-1} = \ldots = x_{t-k} = x$,
  - then the predicted value $x_t = a(x_{t-1}, x_{t-2}, \ldots, x_{t-k})$ should also be equal to the same quantity $x$. 
6. Ubiquity of linear prediction formulas (cont-d)

- In other words, we should require that \( x = w_0 + w_1 \cdot x + \ldots + w_k \cdot x \), i.e., equivalently, that
  \[
  w_0 + x \cdot (w_1 + \ldots + w_k - 1) = 0 \text{ for all } x.
  \]

- 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 \( w_1 + \ldots + w_k = 1 \).

- Since \( w_0 = 0 \), the linear dependence takes the form
  \[
  a(x_{t-1}, x_{t-2}, \ldots, x_{t-k}) = w_1 \cdot x_{t-1} + w_2 \cdot x_{t-2} + \ldots + w_k \cdot x_{t-k}.
  \]

- Now, selecting a good prediction formula means selecting appropriate values \( w_i \).
7. **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:

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

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

- While the formula has been empirically successful, there is no convincing theoretical explanation for this success.

- In this talk, we use invariance ideas to provide such a theoretical explanation.
8. How weights depend on time?

- Let us try to formulate the problem in precise terms.
- Different weights $w_i$ correspond to different time intervals:
  - the weight $w_1$ corresponds to the selected time quantum $\Delta t$,
  - 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 $w_1 + \ldots + w_k = 1$.
- To maintain this equality, we need to normalize these values by dividing each value $f(i \cdot \Delta t)$ by their sum: $w_i = \frac{f(i \cdot \Delta t)}{\sum_{j=1}^{k} f(j \cdot \Delta t)}$.
- So, a natural question is: what is an appropriate function $f(t)$?
9. 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 or $w_i$ will be multiplied by the same constant $C$ and
  - thus, the weights will remain the same:

$$\frac{\sum_{j=1}^{k} f(j \cdot \Delta t)}{\sum_{j=1}^{k} g(j \cdot \Delta t)} = \frac{\sum_{j=1}^{k} g(j \cdot \Delta t)}{\sum_{j=1}^{k} g(j \cdot \Delta t)}.$$
10. Different functions may lead to the same weights (cont-d)

- Vice versa, let us show that:
  - if two functions $f(t)$ and $g(t)$ always lead to the same weights,
  - 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 above formula implies that

\[
\frac{g(i \cdot \Delta t)}{f(i \cdot \Delta t)} = \frac{\sum_{j=1}^{k} g(j \cdot \Delta t)}{\sum_{j=1}^{k} f(j \cdot \Delta t)}.
\]

- The right-hand side of this formula 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.
- So $f(t)$ depends on the values of the quantities $q_1, \ldots, q_m$ that characterize these factors:

$$f(t) = F(q_1(t), \ldots, q_m(t)).$$
12. 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 $\lambda$:
    \[ q \mapsto \lambda \cdot q. \]
13. 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 $q_a \rightarrow \lambda_a \cdot q_a$ we apply to the numerical values of each of these quantities.
- In other words, the weights $w_i$ as described by the above formula 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
    \[ g(t) = F(\lambda_1 \cdot q_1(t), \ldots, \lambda_m \cdot q_m(t)) \].
- 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 $C$. 
14. Scale-invariance (cont-d)

- In our case, we conclude that there is a constant $C$ – which may depend on the re-scaling parameters $\lambda_i$ – for which:

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

- It is known that all continuous solutions of the above functional equation have the form $F(q_1, \ldots, q_m) = A \cdot q_1^{a_1} \cdot \ldots \cdot q_m^{a_m}$.

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

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

- 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 this type.

- This class of functions is closed under multiplication and raising to a power.
15. Scale-invariance (cont-d)

- This closeness can be described in easier-to-process terms if we take the logarithm of both sides and consider the function

  \[ L(q_1(t), \ldots, q_m(t)) \overset{\text{def}}{=} \ln(F(q_1(t), \ldots, q_m(t))). \]

- For this function, \( F(q_1(t), \ldots, q_m(t)) = \exp(L(q_1(t), \ldots, q_m(t))). \)

- For this logarithm, the above equality takes the form

  \[ L(q_1(t), \ldots, q_m(t)) = a + a_1 \cdot Q_1(t) + \ldots + a_m \cdot Q_m(t). \]

- Here we denoted \( a \overset{\text{def}}{=} \ln(A) \) and \( Q_a(t) \overset{\text{def}}{=} \ln(q_a(t)). \)

- The above formula for \( L \) describes all possible linear combinations of functions 1 and \( Q_a(t). \)

- Thus, the logarithmic functions corresponding to all possible expressions of this type form a finite-dimensional linear space.
16. Scale-invariance with respect to time

- We have already mentioned that there is usually no preferred measuring unit for measuring the quantities $q_a$.

- Similarly, 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 of $L$-functions 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 $\mu$. 
17. Final explanation

- It is also reasonable to require that all the functions $L$ 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.
- It is known that every function from a scale-invariant finite-dimensional linear space of analytical functions is a polynomial.
- 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, the weights $w_i$ have the desired form.
- So, we have indeed explained the empirical success of the exponential Almon lag formula.
- Namely, it turns out to be the only formula that satisfies the natural invariance conditions.
18. 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

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

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

$$L(t) = c_{k_1} \cdot t^{k_1} + c_{k_2} \cdot t^{k_2} + \ldots, \text{ where } k_1 < k_2 < \ldots$$

- 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:

  - 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 $\mu$, the function

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

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

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

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 above function tends to $t^{k_1}$.

Thus, this function is indeed contained in the space $S$.

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

$$L(t) - c_{k_1} \cdot t^{k_1} = c_{k_2} \cdot t^{k_2} + \ldots$$
Thus, similarly, we can prove that the function $t^{k_2}$ is also contained in the space $S$, etc.

The statement is thus proven.
21. Acknowledgments

• This work was supported in part by the National Science Foundation grants:
  – 1623190 (A Model of Change for Preparing a New Generation for Professional Practice in Computer Science), and
  – HRD-1834620 and HRD-2034030 (CAHSI Includes).

• It was also supported by the AT&T Fellowship in Information Technology.

• It was also supported by the program of the development of the Scientific-Educational Mathematical Center of Volga Federal District No. 075-02-2020-1478.