7-2015

Why Linear (and Piecewise Linear) Models Often Successfully Describe Complex Non-Linear Economic and Financial Phenomena: A Fuzzy-Based Explanation

Hung T. Nguyen
New Mexico State University - Main Campus, hunguyen@nmsu.edu

Vladik Kreinovich
University of Texas at El Paso, vladik@utep.edu

Olga Kosheleva
University of Texas at El Paso, olgak@utep.edu

Songsak Sriboonchitta
Chiang Mai University, songsakecon@gmail.com

Comments:
Technical Report: UTEP-CS-15-59

Why Linear (and Piecewise Linear) Models Often Successfully Describe Complex Non-Linear Economic and Financial Phenomena: A Fuzzy-Based Explanation

H. T. Nguyen[1,2], V. Kreinovich[3],
O. Kosheleva[3], and S. Sriboonchitta[2]
[1]Department of Mathematical Sciences
New Mexico State University
Las Cruces, New Mexico 88003, USA
hunguyen@nmsu.edu
[2]Department of Economics
Chiang Mai University
Chiang Mai, Thailand
songsakecon@gmail.com
[3]University of Texas at El Paso
El Paso, TX 79968, USA
vladik@utep.edu, olgak@utep.edu

Abstract

Economic and financial phenomena are highly complex and non-linear. However, surprisingly, in many cases, these phenomena are accurately described by linear models – or, sometimes, by piecewise linear ones. In this paper, we show that fuzzy techniques can explain the unexpected efficiency of linear and piecewise linear models: namely, we show that a natural fuzzy-based precisiation of imprecise ("fuzzy") expert knowledge often leads to linear and piecewise linear models.

We also discuss which expert-motivated nonlinear models should be used to get a more accurate description of economic and financial phenomena.

1 Linear and Piecewise Linear Methods Are Surprisingly Efficient for Describing Nonlinear Economic Phenomena: Formulation of the Problem

It is well known that economic and financial phenomena are very *complex* and *non-linear*. However, surprisingly, many such phenomena are well described by *linear* models, such as the AutoRegressive-Moving-Average model with eXogenous inputs model (ARMAX) [3, 4]:

$$q_t = \sum_{i=1}^{p} \varphi_i \cdot q_{t-i} + \sum_{i=1}^{b} \eta_i \cdot d_{t-i} + \varepsilon_t + \sum_{i=1}^{q} \theta_i \cdot \varepsilon_{t-i},$$

where q_t is the value of an economic quantity at time t, d_t is the value of the external quantity at time t, and φ_i, η_i, and θ_i are constants. Here, ε_t are random variables $\varepsilon_t = \sigma_t \cdot z_t$, where z_t is white noise with 0 mean and standard deviation 1, and the dynamics of variances σ_t^2 is also described by a *linear* formula: namely, by the Generalized AutoRegressive Conditional Heterosckedasticity (GARCH) model [2, 3, 4]:

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{\ell} \beta_i \cdot \sigma_{t-i}^2 + \sum_{i=1}^{k} \alpha_i \cdot \varepsilon_{t-i}^2.$$

Sometimes, to get a more adequate description of the corresponding economic and financial phenomena, we need to use *piecewise linear* models, in which different linear models are used to describe different periods.

How can we explain this counter-intuitive success of linear and piecewise linear models in describing non-linear phenomena?

2 Need to Use Expert Knowledge and Fuzzy Logic

Models come from experts. To explain this phenomenon, let us recall that while the *parameters* of the models that describe real-world phenomena are tuned based on the observations, the *models themselves* come from experts.

Experts usually start with knowledge formulated in imprecise ("fuzzy") natural-language terms: for example, they can say that if the federal bank interest rate increases, more funds move into bonds, away from stocks. Once economists formulate such natural-language statements, other researchers precisiate these statements by transforming them into precise models.

It is reasonable to use fuzzy logic. Precisiating imprecise expert knowledge – i.e., translating it from the imprecise natural language to precise formulas – is exactly what fuzzy logic has been invented for [5, 7, 8]. Fuzzy logic techniques has been tuned on numerous practical examples, they have many successful applications. It is therefore reasonable to use these techniques to transform economic expert knowledge into a precise model.

Expert knowledge is usually described in terms of natural-language if-then rules. We want to describe the dependence of the future values $y_1, \ldots, y_m$ of the quantities of interest on the values $x_1, \ldots, x_n$ of these and related quantities at present and in the past moments of time.

Expert knowledge about this dependence usually comes in the form of several if-then rules. Let us denote the number of such rules by K.

Some of these rules enable us to *directly describe* the corresponding predictions, i.e., they provide an explicit conclusion about the future values based on the current and past values. Other rules *do not* provide such a *direct prediction*, but describe *relations* between future values, relations that help to make correct predictions. For example, a rule may say that if in the future, there is a large increase in unemployment, then this would lead to a large decrease in stock market value.

In general, each condition or conclusion of each of the rules is based:

- either on the value of one of the quantities x_i or y_j,
- or on the value of the difference between the values of a quantity at two different moments of time or, more generally, on the difference between two quantities.

For example, a reasonable rule may say that if an interest rate in one country is much higher than the interest rate in another one, then we will have a big outflow of capital into a country with a higher interest rate.

To simplify the description of the rules, let us introduce an alternative denotation of each unknown y_j as x_{n+j}. In these terms, each of these rules $k = 1, \ldots, K$ has thus the following if-then form

$$\text{if } A_{k,1}(u_{k,1}) \text{ and } \ldots \text{and } A_{k,n_k}(u_{k,n_k}) \text{ then } B_k(v_k),$$

where:

- $A_{k,j}(u_{k,j})$ and $B_k(v_k)$ are imprecise properties like "small", "medium", etc.,
- each value $u_{k,j}$ is either one of the variables, i.e., $x_{i(k,j)}$ for some $i(k,j)$, or the difference between two variables $x_{f(k,j)} - x_{s(k,j)}$; and
- each value v_k is either one of the variables, i.e., $x_{i(k)}$ for some $i(k)$, or the difference between two variables $x_{f(k)} - x_{s(k)}$.

For example, one of the possible versions of the above rule about bonds corresponds to the case when:

- $n_k = 1$,
- $u_{k,1}$ is the increase in interest rates, i.e., the difference between the current and the previous interest rates,

- v_k is the amount of money moving into bonds, i.e., the difference between the future amount y and the current u_i investments in bonds,
- $A_{k,1}$ is "big", and
- B_k is "large".

Fuzzy logic technique: reminder. Fuzzy logic helps transform if-then rules of the above type into precise formulas. The use of fuzzy logic starts with selecting *membership functions* $\mu_{k,j}(u_{k,j})$ and $\mu_k(v_k)$ representing imprecise terms $A_{k,j}$ and B_k. Specifically, e.g., for each real number $u_{k,j}$, the value $\mu_{k,j}(u_{k,j}) \in [0,1]$ is the degree to which this number satisfies the property $A_{k,j}$ (e.g., the degree to which $u_{k,j}$ is big).

Once we know the membership functions, then, for each combination of the values

$$(x_1,\ldots,x_n,x_{n+1},\ldots,x_{n+m}) = (x_1,\ldots,x_n,y_1,\ldots,y_m),$$

we can describe the degree $\mu_{k,j}(u_{k,j})$ to which each of the conditions $A_{k,j}(u_{k,j})$ is satisfied. We then need to use these degrees to find a degree to which the entire condition

$$\text{“}A_{k,1}(u_{k,1}) \text{ and } \ldots \text{ and } A_{k,n_k}(u_{k,n_k})\text{”}$$

is satisfied. To compute this combined degree, we can use a fuzzy generalization of the "and" operation of classical logic. Such generalizations are known as *"and"-operations* or *t-norms* $f_\&(a_1,a_2,\ldots)$. Once we have selected a t-norm $f_\&(a_1,a_2,\ldots)$, then, for each combination $(y_1,\ldots,y_m)$ of future values, we can compute the degree $C_k(y_1,\ldots,y_m)$ to which the condition of the k-th rule is satisfied as

$$C_k(y_1,\ldots,y_m) = f_\&(\mu_{k,1}(u_{k,1}),\ldots,\mu_{k,n_k}(u_{k,n_k})). \quad (1)$$

Similarly, we can compute the degree $\mu_k(v_k)$ to which the conclusion $B_k(v_k)$ of the k-th rule is satisfied.

To find the degree $D_k(y_1,\ldots,y_m)$ to which the k-th rule itself is satisfied, we can then apply a fuzzy implication operator $f_\to(a,b)$ – a generalization of the implication operation of classical 2-valued logic – to the degrees to which the condition and the conclusion are satisfied. As a result, for each combination $(y_1,\ldots,y_m)$, we get the following formula for the degree $D_k(y_1,\ldots,y_m)$ to which the k-th rule is satisfied for these values y_j:

$$D_k(y_1,\ldots,y_m) = f_\to(C_k(y_1,\ldots,y_m),\mu_k(v_k)). \quad (2)$$

We want to know to what degree $D(y_1,\ldots,y_m)$ *all* K rules are satisfied, i.e., to what degree the 1st rule is satisfied *and* the 2nd rule is satisfied, etc. To find this degree, it is natural to again use the "and"-operation:

$$D(y_1,\ldots,y_m) = f_\&(D_1(y_1,\ldots,y_m),\ldots,D_K(y_1,\ldots,y_m)). \quad (3)$$

Now, for each possible combination of future values $y_1, \ldots, y_m$, we know the degree $D(y_1, \ldots, y_m)$ to which these value y_j are consistent with the known values $x_1, \ldots, x_n$ and with the expert rules. If we need to come up with a numerical prediction, it is reasonable to select the "most possible" combination $(y_1, \ldots, y_m)$, i.e., the value for which the corresponding degree $D(y_1, \ldots, y_m)$ is the largest possible:

$$D(y_1, \ldots, y_m) \to \max_{y_1, \ldots, y_m}. \qquad (4)$$

To apply fuzzy logic technique, we need to select appropriate operations. As one can see from the above description, to apply the fuzzy logic technique, we need to select appropriate "and"-operation, an appropriate implication operation, and appropriate membership functions.

From the purely mathematical viewpoint, in each cases, there are many possible choices. Let us analyze which of these choices are most appropriate for the analysis of economic and financial data.

3 Specifics of Economic and Financial Expert Knowledge

In many applications areas, there is a very small amount of best experts. For example, among all the surgeons performing a certain kind of surgery, there are a few best ones; similarly, there are few doctors who are the best in diagnosing a certain rare disease, etc. Since there are few of these experts, it is not possible to utilize them in all relevant situations. In such cases, it is important to describe the expertise of each individual expert as accurately as possible – so that others can use this expertise.

In economics and finance, the situation is different. There is no clear small group of best experts: at any given moment of time, some financial and industrial leaders exhibit excellent results – only to be defeated by competitors. However, while we cannot point to a single expert as the best, there is no doubt that financial and economic leaders as a whole form a group with the desired expertise. In other words, in economics and finance, it is not that important to accurately describe the opinion of each *individual* expert, it is much more important to describe the opinion of the *group* of experts.

With this in mind, the best way to determine the corresponding membership functions is by polling: for each statement S like "an interest rate increase of 4% is big", we ask several (N) experts whether they believe this statement to be true, and if $N(S)$ of them agree that this statement is true, we take the ratio $\frac{N(S)}{N}$ as the degree $\mu(S)$ to which this statement is true.

Our objective is to find the value $\mu(S)$ as accurately as possible. It is known that in a poll, the more people we ask, the more accurate is the resulting opinion. Thus, a natural way to improve the accuracy of the poll is to ask more experts. However, there is a catch. When at first, we could only afford to poll N people, we thus selected the top leaders in the field. Now that we add N' extra experts, these experts may be too intimidated by the reputation of the original experts (like Warren Buffett) to voice their own opinions – especially if the original super-experts disagreed between themselves. With the new experts mute, we still have the same number $N(S)$ of experts who agree with the statement S – but now we have to divide it not by the original number N, but by the new number $N + N'$. As a result, instead of the original value $\mu(S) = \dfrac{N(S)}{N}$, we get a new value $\mu'(S) = \dfrac{N(S)}{N + N'}$. The values $\mu'(S)$ and $\mu(S)$ are related by a simple formula $\mu'(S) = c \cdot \mu(S)$, where $c = \dfrac{N}{N + N'}$.

Thus, for the exact same opinion, by selecting two different numbers of experts N and $N + N'$, we get two numerically different membership functions: $\mu(S)$ and $c \cdot \mu(S)$. These two membership functions represent the same expert opinion and are, thus, equivalent in some reasonable sense. It is therefore reasonable to select "and"-operations which are consistent with this equivalence.

4 Selecting an Appropriate "And"-Operation

An "and"-operation $f_\&(a, a')$ transform the degrees of belief a and a' in statements S and S' into a degree of belief in a combined statement $S \,\&\, S'$. Consistency means that if we simply re-scale each degree, i.e., replace a with an equivalent degree $c \cdot a$ and replace a' with an equivalent degree $c' \cdot a'$, for some constants c and c', then he resulting degrees $f_\&(c \cdot a, c' \cdot a')$ should also be equivalent to the original degrees, i.e., we should have $f_\&(c \cdot a, c' \cdot a') = C \cdot f_\&(a, a')$ for some constant C depending on c and c'. Thus, we arrive at the following definition.

Definition 1. *We say that a t-norm $f_\&(a, b)$ is* consistent with polling *if for every c and c' there exists a value $C(c, c')$ for which, for all a and a', we have*

$$f_\&(c \cdot a, c' \cdot a') = C(c, c') \cdot f_\&(a, a'). \tag{3}$$

It turns out that this requirement uniquely determines the "and"-operation:

Proposition 1. *The only t-norm which is consistent with polling is the product $f_\&(a, a') = a \cdot a'$.*

Proof. Let us first consider the case when $c' = 1$. In this case, the formula (3) takes the form

$$f_\&(c \cdot a, a') = C(c) \cdot f_\&(a, a'). \tag{4}$$

Similarly, for any c', we have

$$f_\&(c' \cdot c \cdot a, a') = C(c') \cdot f_\&(c \cdot a, a') = C(c') \cdot C(c) \cdot f_\&(a, a'). \tag{5}$$

One the other hand, substituting $c' \cdot c$ into the formula (4), we get

$$f_\&(c' \cdot c \cdot a, a') = C(c' \cdot c) \cdot f_\&(a, a'). \quad (6)$$

By comparing the right-hand sides of the formulas (5) and (6), we conclude that for every c and c', we have

$$C(c' \cdot c) = C(c') \cdot C(c). \quad (7)$$

A t-norm is increasing in each of the variables, so from (4), we can conclude that the function $C(c)$ is increasing. It is known (see, e.g., [1]) that every increasing solution to the functional equation (7) has the form $C(c) = c^\alpha$ for some $\alpha > 0$. Substituting this expression into the formula (4), we get

$$f_\&(c \cdot a, a') = c^\alpha \cdot f_\&(a, a'). \quad (8)$$

In particular, for $a = 1$, we get

$$f_\&(c, a') = c^\alpha \cdot f_\&(1, a') = c^\alpha \cdot a'. \quad (9)$$

The t-norm is symmetric, so we have $c^\alpha \cdot a' = c \cdot (a')^\alpha$, hence $\alpha = 1$ and $f_\&(c, a') = c \cdot a'$. The proposition is proven.

5 Selecting an Appropriate Implication Operation

An implication is naturally related to an "and"-operation. Namely, an implication $A \to B$ means that if we add it to A, we get B. If we get a statement weaker than the implication to A, then we not necessarily get B. Thus, implication can be defined as the supremum of all such "below-implication" values, i.e., as

$$f_\to(a, b) = \max\{c : f_\&(a, c) \le b\}.$$

For the case when the "and"-operation is a product, we get

$$f_\to(a, b) = \frac{b}{a} \text{ when } a > b, \text{ else } f_\to(a, b) = 1. \quad (10)$$

This is the implication operation that we will use in this paper.

6 Selecting Appropriate Membership Functions

Idea. As we have mentioned earlier, one of the main features of expert knowledge in economics and finance is that, in contrast to many other areas of knowledge, here we need to combine the opinions of several experts. It is therefore reasonable to select a family of membership functions in such a way that not only the opinion of each expert can be described by a membership function from this family, but also that the "and"-combination of these opinions should be described by a function from this same family.

Since we have decided to express "and" as a product, this means that our family F of membership functions should be closed under multiplication:

$$\text{if } \mu_1(x) \in F \text{ and } \mu_2(x) \in F, \text{ then } \mu_1(x) \cdot \mu_2(x) \in F.$$

Analysis of this idea. To analyze the situation, it is convenient to use the fact that the logarithm of the product is equal to the sum of the logarithms. Thus, if instead of the original family of functions, we consider their logarithms $f(x) \stackrel{\text{def}}{=} \ln(\mu(x))$, we can conclude that the family L of all such logarithms should be closed under addition.

In the computer, at any given moment of time, we can only represent finitely many parameters. Thus, it is reasonable to conclude that the linear space generated by the set L can also described by finitely many parameters, i.e., is *finite-dimensional.*

Scale-invariance. The numerical values of the economics- and finance-related quantities x_i and y depend on the choice of the unit. For example, if, instead of dollars, we start measuring these quantities in euros, we get different numerical values. In general, if we replace a measuring unit with another unit which is λ times larger, then the original numerical value x of the corresponding quantity is replaced by a new value $x' = \dfrac{x}{\lambda}$.

Under this re-scaling, each original membership function $\mu(x)$ takes the form $\mu'(x') = \mu(x) = \mu(\lambda \cdot x')$. It is reasonable to require that this re-scaling transform membership functions from the selected family F into functions from the same family, i.e., that the family F (and thus, the family L of the logarithms of functions $\mu \in F$) be invariant with respect to this re-scaling.

Definition 1. *Let n be an arbitrary integer. We say that a finite-dimensional linear space L of analytical functions is* scale-invariant *if for every function $f \in L$ and for every $\lambda > 0$, the function $f_\lambda(x) \stackrel{\text{def}}{=} f(\lambda \cdot x)$ also belongs to the family L.*

Proposition 2. [6] *For every scale-invariant finite-dimensional linear space L of analytical functions, every element $f \in L$ is a polynomial.*

Proof of Proposition 2. Let L be a scale-invariant finite-dimensional linear space F of analytical functions, and let $f(x)$ be a function from this family L.

By definition, an analytical function $f(x)$ is an infinite sum of monomials $m(x)$ of the type $a_k \cdot x^k$:

$$f(x) = a_0 + a_1 \cdot x + a_2 \cdot x^2 + \ldots$$

If we multiply x by λ, then the value of this monomial is multiplied by λ^k:

$$f(\lambda \cdot x) = a_0 + \lambda \cdot a_1 \cdot x + \lambda^2 \cdot a_2 \cdot x^2 + \ldots$$

Some of the coefficients a_i may be zeros – if the original expansion has no monomials of the corresponding order. Let k_0 be the first index for which $a_{k_0} \neq 0$. Then,

$$f(x) = a_{k_0} \cdot x^{k_0} + a_{k_0+1} \cdot x^{k_0+1} + \ldots$$

Since the family L is scale-invariant, it also contains the function $f_\lambda(x) = f(\lambda \cdot x)$. At this re-scaling, each term $a_k \cdot x^k$ is multiplied by λ^k; thus, we get

$$f_\lambda(x) = \lambda^{k_0} \cdot a_{k_0} \cdot x^{k_0} + \lambda^{k_0+1} \cdot a_{k_0+1} \cdot x^{k_0+1} + \ldots$$

Since L is a linear space, it also contains a function

$$\lambda^{-k_0} \cdot f_\lambda(x) = a_{k_0} \cdot x^{k_0} + \lambda \cdot a_{k_0+1} \cdot x^{k_0+1} + \ldots$$

Since L is finite-dimensional, it is closed under turning to a limit. In the limit $\lambda \to 0$, we conclude that the term $a_{k_0} \cdot x^{k_0}$ also belongs to the family L.

Since L is a linear space, this means that the difference

$$f(x) - a_{k_0} \cdot x^{k_0} = a_{k_0+1} \cdot x^{k_0+1} + \ldots$$

also belongs to L. If we denote, by k_1, the first index $k_1 > k_0$ for which the term $a_{k_1} \neq 0$, then we can similarly conclude that the corresponding term $a_{k_1} \cdot x^{k_1}$ also belongs to the family L, etc.

We can therefore conclude that for every index k for which term $a_k \neq 0$, the corresponding term $a_k \cdot x^k$ also belongs to the family L.

Monomials of different total order are linearly independent. Thus, if there were infinitely many non-zero coefficients $a_k \neq 0$, we would have infinitely many linearly independent function in the family L – which contradicts to our assumption that the family L is a finite-dimensional linear space.

So, in the expansion of the function $f(x)$, there are only finitely many non-zero terms. Hence, the function $f(x)$ is a sum of finitely many monomials – i.e., a polynomial.

The proposition is proven.

Conclusion about membership functions. So, we conclude that the logarithms of the membership functions are polynomials, and thus, each membership function has the form $\mu(x) = \exp(P(x))$ for an appropriate polynomial $P(x)$.

Simplest possible membership functions are Gaussian. We need to make sure that $\mu(x) \in [0, 1]$ for all x; this excludes linear functions $P(x) = a + b \cdot x$, since for them $\exp(P(x))$ tends to ∞ either when $x \to \infty$ (for $b > 0$) or when $x \to -\infty$ (for $b < 0$). Thus, the simplest possible membership functions of the type $\mu(x) = \exp(P(x))$ are the functions corresponding to quadratic polynomials $P(x)$.

Each quadratic polynomial $P(x)$ can be represented as $C \cdot (x - a)^2 + b$ for some C, a, and b. Thus,

$$\mu(x) = \exp(P(x)) = \exp(b) \cdot \exp\left(C \cdot (x - a)^2\right).$$

The requirement that $\mu(x) \in [0,1]$ for all x implies that $C < 0$, i.e., that $C = -c$ for some $c > 0$. Also, we have argued earlier that the degrees of belief (and thus, membership functions) can only be defined modulo a multiplicative constant. So, we can safely conclude that

$$\mu(x) = \exp\left(-c \cdot (x-a)^2\right),$$

i.e., that all the membership functions that we consider are Gaussian.

7 Justification of Piecewise Linear Dependence

Derivation. Let us show that when the "and"-operation is a product, the implication operation comes from the product, and all membership functions are Gaussian, the standard fuzzy logic procedure (1)–(4) leads to piecewise linear dependencies. This will complete our justification.

Indeed, we assume that all membership functions are Gaussian, i.e., that

$$\mu_{k,j}(u_{k,j}) = \exp\left(-c_{k,j} \cdot (u_{k,j} - a_{k,j})^2\right)$$

and

$$\mu_k(v_k) = \exp\left(-c_k \cdot (v_k - a_k)^2\right)$$

for all k and j. Since the "and"-operation is the product, we get

$$C_k(y_1, \ldots, y_m) = f_\&(\mu_{k,1}(u_{k,1}), \ldots, \mu_{k,n}(u_{k,n_k})) =$$

$$\mu_{k,1}(u_{k,1}) \cdot \ldots \cdot \mu_{k,n}(u_{k,n_k}) =$$

$$\exp\left(-c_{k,1} \cdot (u_{k,1} - a_{k,1})^2\right) \cdot \ldots \cdot \exp\left(-c_{k,n_k} \cdot (u_{k,n_k} - a_{k,n_k})^2\right) =$$

$$\exp\left(-\sum_{j=1}^{n_k} c_{k,j} \cdot (u_{k,j} - a_{k,j})^2\right).$$

We can now use the formula (10) for the implication to find the degree $D_k(y_1, \ldots, y_m)$. Specifically, when the degree $C_k(y_1, \ldots, y_m)$ is smaller than or equal to the degree

$$\mu_k(v_k) = \exp\left(-c_k \cdot (v_k - a_k)^2\right),$$

i.e., equivalently, when

$$\sum_{j=1}^{n_k} c_{k,j} \cdot (u_{k,j} - a_{k,j})^2 \ge c_k \cdot (v_k - a_k)^2, \tag{11}$$

then $D_k(y_1, \ldots, y_m) = 1$.

On the other hand, when $C_k(y_1, \ldots, y_m) > \mu_k(v_k)$, i.e., equivalently, when

$$\sum_{j=1}^{n_k} c_{k,j} \cdot (u_{k,j} - a_{k,j})^2 < c_k \cdot (v_k - a_k)^2, \tag{12}$$

then

$$D_k(y_1,\ldots,y_m)=\frac{c_k(v_k)}{C_k(y_1,\ldots,y_m)}=$$

$$\exp\left(-\left(c_k\cdot(v_k-a_k)^2-\sum_{j=1}^{n_k}c_{k,j}\cdot(u_{k,j}-a_{k,j})^2\right)\right). \tag{13}$$

The resulting expression (3) for the maximized degree thus has the form

$$D(y_1,\ldots,y_m)=D_1(y_1,\ldots,y_m)\cdot\ldots\cdot D_K(y_1,\ldots,y_m)=$$

$$\prod_{k\in\mathcal{K}}\exp\left(-\left(c_k\cdot(v_k-a_k)^2-\sum_{j=1}^{n_k}c_{k,j}\cdot(u_{k,j}-a_{k,j})^2\right)\right), \tag{14}$$

where $\mathcal{K}$ is the set of all the indices k for which the inequality (12) is satisfied.

For each combination of inputs $(x_1,\ldots,x_n)$, the maximum in (4) is attained for the values $(y_1,\ldots,y_m)$ characterized by some set $\mathcal{K}\subseteq\{1,\ldots,K\}$.

A finite set $\{1,\ldots,K\}$ has finitely many possible subsets $\mathcal{K}$. We can therefore divide the set of all possible combinations of inputs $(x_1,\ldots,x_n)$ into finitely many regions corresponding to different subsets $\mathcal{K}$. In each of these regions, the predicted values $y_1,\ldots,y_m$ can be determined by maximizing the corresponding expression (14). Since the exponent of the sum is equal to the product of the exponents, we can conclude that

$$D(y_1,\ldots,y_m)=$$

$$\exp\left(-\sum_{k\in\mathcal{K}}\left(c_k\cdot(v_k-a_k)^2-\sum_{j=1}^{n_k}c_{k,j}\cdot(u_{k,j}-a_{k,j})^2\right)\right). \tag{15}$$

The function $\exp(-z)$ is monotonically decreasing, so maximizing the expression (15) is equivalent to minimizing the expression

$$E\stackrel{\text{def}}{=}\sum_{k\in\mathcal{K}}\left(c_k\cdot(v_k-a_k)^2-\sum_{j=1}^{n_k}c_{k,j}\cdot(u_{k,j}-a_{k,j})^2\right). \tag{16}$$

To minimize this expression, we can differentiating it with respect to each of the unknowns $y_1,\ldots,y_m$ and equate each of these derivatives to 0. Each of the expressions $u_{k,j}$ and v_k is linear in terms of x_i and y_j; thus, each equation $\dfrac{\partial E}{\partial y_j}=0$ is linear in terms of x_i and y_j. Thus, to find m unknown, we have a system of m linear equations $y_1,\ldots,y_m$ that linearly include x_i in the right-hand sides, i.e., equations

$$\sum_{j'=1}^{m}A_{j,j'}\cdot y_{j'}=B_j+\sum_{i=1}^{n}C_{j,i}\cdot x_i,\quad j=1,\ldots,m, \tag{17}$$

for some constants $A_{j,j'}$, B_j, and $C_{j,i}$. In matrix form, we can rewrite this as $Ay = B + Cx$, hence $y = A^{-1}(B + Cx) = A^{-1}B + A^{-1}Cx$.

Conclusion. Thus, *on each of finitely many regions, we get linear dependence of the predicted quantities y_j on the inputs $x_1, \ldots, x_m$*, i.e., we indeed get a piecewise linear dependence.

8 Discussion: Linear and Piecewise Linear Models, What Next?

Our justification of piecewise linear models is based on the selecting, among all membership functions which are consistent with our assumptions, the simplest ones – which turned out to be Gaussian.

If it turns out that in some situations, the resulting piecewise linear models are not sufficient accurate, then a natural idea is to use the next simplest class of corresponding membership functions. In general, we have membership functions of the type $\mu(x) = \exp(P(x))$ for some polynomial $P(x)$. In our analysis, we selected the simplest case when these polynomials are quadratic (linear polynomials are not possible since then we will not have $\mu(P(x)) \in [0, 1]$ for all x).

To get a more adequate description, we therefore need to consider polynomials of higher order. Cubic polynomials are not possible (for the same reason as linear ones), so the next simplest case is the case of fourth order polynomials. For second order polynomials, our analysis led us to a system of m equations each of which is linear in terms of the inputs x_i and the predicted values y_j. For fourth order polynomials, a similar analysis will lead to a system of m equations each of which is *cubic* in terms of the inputs x_i and the predicted values y_j.

Since the analysis of expert knowledge naturally leads to such cubic systems, it may be a good idea, in situations when we seek better prediction accuracy, to start adding cubic terms to the known piecewise linear models.

Acknowledgments

We acknowledge the partial support of the Center of Excellence in Econometrics, Faculty of Economics, Chiang Mai University, Thailand. This work was also supported in part by the National Science Foundation grants HRD-0734825 and HRD-1242122 (Cyber-ShARE Center of Excellence) and DUE-0926721.

References

[1] J. Aczel, *Lectures on Functional Equations and Their Applications*, Dover Publ., New York, 2006.

[2] T. Bollerslev, "Generalized Autoregressive Conditional Heteroskedasticitiy", *Journal of Econometrics*, 1986, Vol. 31, No. 3, pp. 307–327.

[3] P. J. Brockwell and R. A. Davis, *Time Series: Theories and Methods*, Springer Verlag, New York, 2009.

[4] W. Enders, *Applied Econometric Time Series*, Wiley, New York, 2014.

[5] G. Klir and B. Yuan, *Fuzzy Sets and Fuzzy Logic*, Prentice Hall, Upper Saddle River, New Jersey, 1995.

[6] O. Kosheleva and M. Ceberio, "Why polynomial formulas in soft computing, decision making, etc.?", Proceedings of the World Congress on Computational Intelligence WCCI'2010, Barcelona, Spain, July 18–23, 2010, pp. 3310–3314.

[7] H. T. Nguyen and E. A. Walker, *A First Course in Fuzzy Logic*, Chapman and Hall/CRC, Boca Raton, Florida, 2006.

[8] L. A. Zadeh, "Fuzzy sets", *Information and Control*, 1965, Vol. 8, pp. 338–353.