Why Threshold Models: A Theoretical Explanation

Thongchai Dumrongpokaphan[1], Vladik Kreinovich[2], and Songsak Sriboonchitta[1]

[1]Chiang Mai University, Thailand
tcd43@hotmail.com, songsakecon@gmail.com
[2]University of Texas at El Paso,
El Paso, Texas 79968, USA
vladik@utep.edu

1. Linear Models Are Often Successful in Econometrics

- In econometrics, often, linear models are efficient.
- In linear models, the values $q_{1,t}, \ldots, q_{k,t}$ of quantities $q_1, \ldots, q_k$ at time t can be predicted as linear f-s of:
  - the values of these quantities at previous moments of time $t-1$, $t-2$, ..., and
  - of the current (and past) values $e_{m,t}, e_{m,t-1}, \ldots$ of the external quantities $e_1, \ldots, e_n$:

$$q_{i,t} = a_i + \sum_{j=1}^{k} \sum_{\ell=1}^{\ell_0} a_{i,j,\ell} \cdot q_{j,t-\ell} + \sum_{m=1}^{n} \sum_{\ell=0}^{\ell_0} b_{i,m,\ell} \cdot e_{m,t-\ell}.$$

2. General Ubiquity of Linear Models in Science and Engineering

- At first glance, the ubiquity of linear models in econometrics is not surprising.
- Indeed, linear models are ubiquitous in science and engineering in general.
- Indeed, we can start with a general dependence

$$q_{i,t} = f_i\left(q_{1,t}, q_{1,t-1}, \ldots, q_{k,t-\ell_0}, e_{1,t}, e_{1,t-1}, \ldots, e_{n,t-\ell_0}\right).$$

- In science and engineering, the dependencies are usually smooth.
- Thus, we can expand the dependence in Taylor series and keep the first few terms in this expansion.
- In particular, in the first approximation, when we only keep linear terms, we get a linear model.

3. Linear Models in Econometrics Are Applicable Way Beyond the Taylor Series Explanation

- In science and engineering, linear models are effective in a small vicinity of each state, when:
  - the deviations from a given state are small
  - and we can therefore safely ignore terms which are quadratic (or of higher order) in them.
- However, in econometrics, linear models are effective even when deviations are large.
- How can we explain this unexpected efficiency?

4. Why Linear Models Are Ubiquitous in Econometrics

- A possible explanation for the ubiquity of linear models in econometrics was proposed in our 2015 paper.
- Example: predicting the country's Gross Domestic Product (GDP) $q_{1,t}$.
- To estimate the current year's GDP, we use:
  - GDP values in the past years, and
  - different characteristics that affect the GDP, such as the population size, the amount of trade, etc.
- In many cases, the corresponding description is unambiguous.
- However, in many other cases, there is an ambiguity in what to consider a country.

5. Why Linear Models Are Ubiquitous (cont-d)

- Indeed, in many cases, countries form a loose federation: European Union is a good example.
- Most of European countries have the same currency.
- There are no barriers for trade and for movement of people between different countries.
- So, from the economic viewpoint, it make sense to treat the European Union as a single country.
- On the other hand, there are still differences between individual members of the European Union.
- So it is also beneficial to view each country from the European Union on its own.
- Thus, we have two possible approaches to predicting the European Union's GDP.

6. Why Linear Models Are Ubiquitous (cont-d)

- We can treat the whole European Union as a single country, and apply the general formula to it.
- We can also apply the general formula to each country c independently, and add the predictions:

$$q_{i,t}^{(c)} = f_i\left(q_{1,t}^{(c)}, q_{1,t-1}^{(c)}, \ldots, q_{k,t-\ell_0}^{(c)}, e_{1,t}^{(c)}, e_{1,t-1}^{(c)}, \ldots, e_{n,t-\ell_0}^{(c)}\right).$$

- The overall GDP $q_{1,t}$ is the sum of GDPs of all the countries: $q_{1,t} = q_{1,t}^{(1)} + \ldots + q_{1,t}^{(C)}$.
- Similarly, the overall population, etc., can be computed as the sum of the values from individual countries:

$$e_{m,t} = e_{m,t}^{(1)} + \ldots + e_{m,t}^{(C)}.$$

- Thus, the prediction of $q_{1,t}$ based on applying the formula to the whole European Union takes the form

$$f_i\left(q_{1,t}^{(1)} + \ldots + q_{1,t}^{(C)}, \ldots, e_{n,t-\ell_0}^{(1)} + \ldots + e_{n,t-\ell_0}^{(C)}\right).$$

7. Why Linear Models Are Ubiquitous (cont-d)

- The sum of individual predictions takes the form

$$f_i\left(q^{(1)}_{1,t},\ldots,e^{(1)}_{n,t-\ell_0}\right)+\ldots+f_i\left(q^{(C)}_{1,t},\ldots,e^{(C)}_{n,t-\ell_0}\right).$$

- We require that these two predictions return the same result:

$$f_i\left(q^{(1)}_{1,t}+\ldots+q^{(C)}_{1,t},\ldots,e^{(1)}_{n,t-\ell_0}+\ldots+e^{(C)}_{n,t-\ell_0}\right)=$$

$$f_i\left(q^{(1)}_{1,t},\ldots,e^{(1)}_{n,t-\ell_0}\right)+\ldots+f_i\left(q^{(C)}_{1,t},\ldots,e^{(C)}_{n,t-\ell_0}\right).$$

- In mathematical terms, this means that the function f_i should be *additive.*
- It also makes sense to require that very small changes in q_i and e_m lead to small changes in the predictions.
- So, the function f_i are continuous.

8. Why Linear Models Are Ubiquitous (cont-d)

- It is known that every continuous additive function is linear.
- Thus the above requirement explains the ubiquity of linear econometric models.

9. Need to Go Beyond Linear Models

- While linear models are reasonably accurate, the actual econometric processes are often non-linear.
- Thus, to get more accurate predictions, we need to go beyond linear models.
- Linear models correspond to the case when we:
  - expand the original dependence in Taylor series and
  - keep only linear terms in this expansion.
- So, to get a more accurate model, a natural idea is:
  - to take into account next order terms in the Taylor expansion,
  - i.e., quadratic terms.

10. The Above Idea Works Well in Science and Engineering, But Not in Econometrics

- Quadratic models are indeed very helpful in science and engineering.
- However, surprisingly, in econometrics, different types of models turn out to be more empirically successful.
- Namely, so-called *threshold models* in which the expression f_i is piece-wise linear.
- In this talk, explain the surprising efficiency of piecewise-linear models in econometrics.

11. Why the Name "Threshold Models"?

- When $q_{1,t} = f_1\left(q_{1,t-1}\right)$, such models can be described by:
  - listing thresholds $T_0 = 0, T_1, \ldots, T_S, T_{S+1} = \infty$ separating different linear expressions, and
  - linear expressions corresponding to each of the intervals $[0, T_1]$, $[T_1, T_2]$, ..., $[T_{S-1}, T_S]$, $[T_S, \infty)$:

$$q_{1,t} = a^{(s)} + a_1^{(s)} \cdot q_{1,t-1} \text{ when } T_s \le q_{1,t-1} \le T_{s+1}.$$

12. Linear Models: Reminder

- The ubiquity of linear models is explained if we assume that for loose federations, we get the same results:
  - whether we consider the whole federation as a single country
  - or whether we view it as several separate countries.
- A similar assumption can be made for a company consisting of several reasonable independent parts, etc.

13. Towards a More Realistic Assumption

- If we always require the above assumption, then we get exactly linear models.
- However, in practice, we encounter some non-linearities.
- This means that the above assumption is not always satisfied.
- Thus, to take into account non-linearities, we need weaken the above assumption.
- It should not matter that much if inside a loose federation, we move an area from one country to another.
- One area becomes slightly bigger and another slightly smaller – but the overall economy remains the same.

14. A More Realistic Assumption (cont-d)

- However, from the economic sense, it makes sense to expect somewhat different results:
  - from a "solid" country – in which the economics is tightly connected, and
  - from a loose federation of sub-countries, in which there is a clear separation between different regions.
- Thus, we make a weaker requirement:
  - the sum of the result of applying prediction to sub-countries should not change
  - if we slightly change the values within each sub-country – as long as the sum remains the same.

15. A More Realistic Assumption (cont-d)

- The crucial word here is "slightly"; there is a difference between:
  - a loose federation of several economies of about the same size – as in the European Union, and
  - an economic union of, say, France and Monaco, in which Monaco's economy is much smaller.
- To take this difference into account, it makes sense to divide the countries into finitely many groups by size.
- We apply the-same-prediction requirement only when changing keeps each country in its group.
- These groups should be reasonable from the topological viewpoint.

16. A More Realistic Assumption (cont-d)

- For example, we should require that each of the corresponding domains D of possible values is:
  - contained in a closure of its interior $D \subseteq \overline{\text{Int}\,(D)}$,
  - i.e., that each point on its boundary is a limit of some interior points.
- Each domain should be strongly connected – in the sense that:
  - each two points in each interior
  - should be connected by a curve which lies fully inside this interior.
- Let us describe the resulting modified assumption in precise terms.

17. A More Realistic Assumption (cont-d)

- We assume that:
  - the set of all possible values of the input $v = (q_{1,t}, \ldots, e_{n,t-\ell_0})$ to the function f_i
  - is divided into a finite number of non-empty non-intersecting strongly connected domains $D^{(1)}, \ldots, D^{(S)}$.
- We require that each of these domains is contained in a closure of its interior $D^{(s)} \subseteq \overline{\text{Int}\left(D^{(s)}\right)}$.
- Let's assume that the following conditions are satisfied for the fours inputs $v^{(1)}$, $v^{(2)}$, $u^{(1)}$, and $u^{(2)}$:
  - the inputs $v^{(1)}$ and $u^{(1)}$ belong to the same domain,
  - the inputs $v^{(2)}$ and $u^{(2)}$ also belong to the same domain (may be different from the domain of $v^{(1)}$),
  - and we have $v^{(1)} + v^{(2)} = u^{(1)} + u^{(2)}$.

18. A More Realistic Assumption (cont-d)

- Then we should have
$$f_i\left(v^{(1)}\right) + f_i\left(v^{(2)}\right) = f_i\left(u^{(1)}\right) + f_i\left(u^{(1)}\right).$$
- Our main result is that under this assumption, the function $f_i(v)$ is piece-wise linear.
- This result explains why piece-wise linear models are indeed ubiquitous in econometrics.

19. Comment

- The functions f_i are continuous; so:
  - on the border between two domains with different linear expressions E and E',
  - the two linear expressions should attain the same value.
- Thus, the border between two domains can be described by the equation $E = E'$, i.e., $E - E' = 0$.
- Since both expressions are linear, the equation $E - E' = 0$ is also linear.
- Thus, this equation describes a (hyper-)plane in the space of all possible inputs.
- So, the zones are separated by hyper-planes.

20. Acknowledgments

- This work was supported by:
  - Chiang Mai University, Thailand,
  - Chiang Mai Center for Excellence in Econometrics, and
  - the US National Science Foundation via grant HRD-1242122 (Cyber-ShARE Center of Excellence).
- The authors are greatly thankful to Professor Hung T. Nguyen for his help and encouragement.

21. Proof: Part 1

- We want to prove that the function f_i is linear on each domain $D^{(s)}$.
- Let us first prove that this function is linear in the vicinity of each point $v^{(0)} \in \mathrm{Int}(D^{(s)})$.
- Indeed, by definition of the interior, it means that there exists a neighborhood of the point $v^{(0)}$ that fully belongs to the domain $D^{(s)}$.
- To be more precise, there exists an $\varepsilon > 0$ such that:
  - if $|d_q| \le \varepsilon$ for all components d_q of the vector d,
  - then the vector $v^{(0)} + d$ also belongs to $D^{(s)}$.

22. Proof: Part 1 (cont-d)

- Thus, because of our assumption, if for two vectors d and d', we have

$$|d_q| \le \varepsilon, \quad |d'_q| \le \Delta, \text{ and } |d_q + d'_q| \le \varepsilon \text{ for all } q, \text{ then :}$$

$$f_i\left(v^{(0)} + d\right) + f_i\left(v^{(0)} + d'\right) = f_i\left(v^{(0)}\right) + f\left(v^{(0)} + d + d'\right).$$

- Subtracting $2f_i\left(v^{(0)}\right)$ from both sides of this equality, we conclude that for the auxiliary function

$$F(v) \stackrel{\text{def}}{=} f_i\left(v^{(0)} + v\right) - f_i\left(v^{(0)}\right), \text{ we have}$$

$$F(d + d') = F(d) + F(d').$$

- Each vector $d = (d_1, d_2, \ldots)$ can be represented as

$$d = (d_1, 0, \ldots) + (0, d_2, 0, \ldots) + \ldots$$

- If $|d_q| \le \varepsilon$ for all q, then the same inequalities are satisfied for all the terms in the right-hand side.

23. Proof: Part 1 (cont-d)

- Thus, we have $F(d) = F_1(d_1) + F_2(d_2) + \ldots$, where:

$$F_1(d_1) \stackrel{\text{def}}{=} F(d_1, 0, \ldots), \quad F_2(d_2) \stackrel{\text{def}}{=} F(0, d_2, 0, \ldots), \ldots$$

- For each of the functions $F_q(d_q)$, the above formula implies that $F_q(d_q + d'_q) = F_q(d_q) + F_q(d'_q)$.
- In particular, when $d_q = d'_q = 0$, we conclude that $F_q(0) = 2F_q(0)$, hence that $F_q(0) = 0$.
- Now, for $d'_q = -d_q$, this formula implies that

$$F_q(-d_q) = -F_q(d_q).$$

- So, to find the values of $F_q(d_q)$ for all d_q for which $|d_q| \le \varepsilon$, it is sufficient to consider positive d_q.
- For every natural number N, additivity implies that

$$F_q\left(\frac{1}{N} \cdot \varepsilon\right) + \ldots + F_q\left(\frac{1}{N} \cdot \varepsilon\right) (N \text{ times}) = F_q(\varepsilon).$$

24. Proof: Part 1 (cont-d)

- Thus $F_q\left(\frac{1}{N}\cdot\varepsilon\right)=\frac{1}{N}\cdot F_q(\varepsilon)$.
- Similarly, for every natural number M, we have

$$F_q\left(\frac{M}{N}\cdot\varepsilon\right)=F_q\left(\frac{1}{N}\cdot\varepsilon\right)+\ldots+F_q\left(\frac{1}{N}\cdot\varepsilon\right)\ (M\text{ times}).$$

- Thus

$$F_q\left(\frac{M}{N}\cdot\varepsilon\right)=M\cdot F_q\left(\frac{1}{N}\cdot\varepsilon\right)=M\cdot\frac{1}{N}\cdot F_q(\varepsilon)=\frac{M}{N}\cdot F_q(\varepsilon).$$

- So, for every rational number $r=\frac{M}{N}\le 1$, we have

$$F_q(r\cdot\varepsilon)=r\cdot F_q(\varepsilon).$$

- Since the function f_i is continuous, the functions F and F_q are continuous too.

25. Proof: Part 1 (cont-d)

- Thus, we can conclude that the above equality holds for all real values $r \le 1$.
- We had a formula relating r and $-r$.
- Thus, we can conclude that the same formula holds for all real values r for which $|r| \le 1$.
- Now, each d_q for which $|d_q| \le \varepsilon$ can be represented as $d_q = r \cdot \varepsilon$, where $r \stackrel{\text{def}}{=} \dfrac{d_q}{\varepsilon}$.
- Thus, the above formula takes the form $F_q(d_q) = \dfrac{d_q}{\varepsilon} \cdot F_q(\varepsilon)$, i.e., the form:

$$F_q(d_q) = a_q \cdot d_q, \text{ where } a_q \stackrel{\text{def}}{=} \frac{F_q(\varepsilon)}{\varepsilon}.$$

- Additivity implies that $F(d) = a_1 \cdot d_1 + a_2 \cdot d_2 + \ldots$

26. Proof: Part 1 (cont-d)

- By definition of the auxiliary function $F(v)$, we have

$$f_i\left(v^{(0)} + d\right) = f_i\left(v^{(0)}\right) + F(d).$$

- So for any v, if we take $d \stackrel{\text{def}}{=} v - v^{(0)}$, we would get

$$f_i(v) = f_i\left(v^{(0)}\right) + F\left(v - v^{(0)}\right).$$

- The first term is a constant, the second one is a linear function of v.
- So indeed the function $f_i(v)$ is linear in the ε-vicinity of the given point $v^{(0)}$.

27. Proof: Part 2

- To complete the proof, we need to prove that the function $f_i(v)$ is linear on the whole domain; indeed:
  - since the domain $D^{(s)}$ is strongly connected,
  - any two points are connected by a finite chain of intersecting open neighborhood.
- In each neighborhood, the function $f_i(v)$ is linear.
- When two linear function coincide in the whole open region, their coefficients are the same.
- Thus, by following the chain, we can conclude that:
  - the coefficients that describe $f_i(v)$ as a locally linear function
  - are the same for all points in the interior of the domain.
- Our result is thus proven.