Riesz Estimator: Motivation and Implementation

Slawomir Tur

September 2024

Riesz Estimators

- **Intuitive definition**: Riesz estimators model relationships that are linear but change across segments or thresholds - **continuous piecewise linear functions**
- **Why not MARS?** (Most) economists are still apprehensive about machine learning (ML as 'black box' inference), we know less about splines than we do about piecewise linear functions, and MARS do come with some bias
- **Why not threshold regression?** Riesz algorithms work just as well with thresholds on many predictors as they do with thresholds on one

Why are they popular as objects of inquiry?

- They capture structural breaks, or 'kinks', in data where typical linear regression fails
- They naturally occur in many theoretical and practical problems:
  - The neural net activation function **ReLU** is a piecewise linear function – $ReLU(ax + b) = 0 \vee (ax + b)$
  - In finance, option payoff curves are represented by piecewise linear functions
  - In microeconomics, such functions can be found in the **revealed preference** literature

Just enough maths to help you understand

- You might have noticed that I used $\vee$ in the definition of ReLU, instead of max
- This is counterintuitive notation (blame the order theorists):
  - The down arrow, $\vee$, is the *supremum* (maximum)
  - The up arrow, $\wedge$, is the *infimum* (minimum)
- We will be working with pointwise maxima and minima, e.g. for two 4-element vectors X, Y:

X	Y	$X \wedge Y$	$X \vee Y$
1	2	1	2
3	5	3	5
5	2	2	5
6	5	5	6

Just enough maths to help you understand

- Ovchinnikov (2002); Aliprantis, Harris, and Tourky (2006): Every piecewise linear function can be represented with the right combination of maxima and minima of its linear **components**
- If a piecewise linear function has a simple enough form (more on that later), there is only a finite number of such combinations
- If we manage to extract the linear functions from each region, we can search over those combinations and get a good estimate of the piecewise function

Worked example

Let $f : [0, 2] \to \mathbb{R}$, where:

$$f(x) = \begin{cases} 2 + 0.5x & x \in [0, \frac{2}{3}] \cup [\frac{4}{3}, 2] \\ 1 + 2x & x \in [\frac{2}{3}, 1] \\ 4 - x & x \in [1, \frac{4}{3}] \end{cases}$$

Worked example

1. Enumerate the linear components of the function, and determine the set of intersecting functions E:
   - $f_1 = 2 + 0.5x, f_2 = 1 + 2x, f_3 = 4 - x$
   - All of the functions intersect, so $E = \{(1,2),(1,3),(2,3)\}$
2. Solve for the points at which each pair intersects:
   - $H_{(1,2)} = [f_1 = f_2] = [2 + 0.5x = 1 + 2x] = [x = \frac{2}{3}] = \{x = \frac{2}{3}\}$
   - $H_{(1,3)} = [f_1 = f_3] = [2 + 0.5x = 4 - x] = [x = \frac{4}{3}] = \{x = \frac{4}{3}\}$
   - $H_{(2,3)} = [f_2 = f_3] = [1 + 2x = 4 - x] = [x = 1] = \{x = 1\}$

Worked example

These points are boundaries between the cells

Algorithm Overview

- We don't have access to the linear components, nor the precise points at which they intersect – we're working with finite data, after all
- **Two approaches**:
  - **RIESZVAR(i)**: the 'analytical' method
  - **RIESZVAR(ii)**: piecewise quadratic optimisation
- **Goal**: Find the optimal combination of linear segments to minimize overall error

RIESZVAR(i) – Estimation Procedure

- **Data Splitting**: Split the ordered predictor data into k regions
- **Linear Regression**: Simple as that, run a regression on each region and extract the coefficients
- **Sperner Families**:
- **Searching**: Search across the generated combinations for k, and choose the combination that minimises the SSE

Data Splitting

Sperner Families (last bit of maths, I promise)

- Two sets X, Y are **incomparable** whenever $X \not\subseteq Y$ and $Y \not\subseteq X$
- A family of incomparable subsets of a set is a **Sperner family**
- e.g. $\{f_1, f_2\}$ and $\{f_2, f_3\}$ are both contained in $\{f_1, f_2, f_3\}$, but neither is in the other; thus, they form a Sperner family
- Remember the last step of the worked example, where we removed every superfluous set?

Sperner Families

- For each family, we take the minimum within the sets, and the maximum across the minima
- Limiting the number of elements a family can have makes the process computationally

Curse of dimensionality

- The number of Sperner families rises with k, but it remains finite
- We know this number for k-sets up to 9 elements – it's the **Dedekind numbers** sequence (A007153 on OEIS):

$$1, 4, 18, 166, 7579, 7828352, 2414682040996,$$
$$56130437228687557907786,$$
$$286386577668298411128469151667598498812364$$

- See the problem?

RIESZVAR(ii) – Numerical Estimation

- **Method**: Uses optimization to tune parameters without manually defining regions.
- **Pros and Cons**: Faster for simple forms but lacks the flexibility of RIESZVAR(i).

Challenges and Conclusions

- **Challenges**:
  - High computational cost for Sperner family calculations.
  - Balancing model complexity: too many segments may overfit, too few may miss detail.
- **Conclusion**: Riesz estimators are effective for segmented data but need optimization for broader use.

Future Directions

- **Potential Improvements**:
  - Optimizing Sperner calculations using clustering or machine learning methods.
  - Implementing adaptive methods to detect regions in larger datasets.
- **Goal**: Make Riesz estimators practical for large-scale data applications.