Ultrametric Wavelet Regression of Multivariate Time Series: Application to Colombian Conflict Analysis

F. Murtagh (SFI Dublin, RHUL London)
M. Spagat (RHUL London)
J.A. Restrepo (Bogotá, Colombia)
Violence data in Colombia from 1990 to 2004

• We used 144 numerical attributes relating to killings and injuries coming from 20,288 events
• We aggregated by month to give: 204 successive months, crossed by 144 attributes
Correspondence analysis
Changepoints discussed: (i) between 1997 and 1998; and (ii) between 2001 and 2002

- **1998 changepoint:**
  - Guerrilla gains from 1996 reversed, a process that was seen to be so by 1998
  - High levels of Govt casualties in 1997-1998 due to guerrilla operations against isolated military and police bases; Govt got upper hand through airborne weaponry
  - Paramilitaries were in “upsurge period” from 1997 whereas before they were more involved in drug trafficking; paramilitary groups consolidated in 1997; initially large numbers of paramilitaries killed
  - Overall change of these disparate aspects was most intense in 1997-1998
Hierarchical clustering of the 2D factors 1, 2 correspondence analysis result

- Sequence-constrained complete link agglomerative criterion used
- Year 1 = 1990 to year 15 = 2004
- Main caesuras or changepoints are in moving from:
  - Year 8 = 1997 to year 9 = 1998; and
  - Year 12 = 2001 to year 13 = 2002
Correspondence analysis
Changepoints discussed: (I) between 1997 and 1998; and (ii) between 2001 and 2002

- 2002 changepoint:
- Peak of paramilitary activity with major setback incurred through aerial bombardment by Govt forces in 2002
- Election of Pres. Alvaro Uribe in 2002
Hierarchical interpretation of Colombian conflict violence over the years 1990 to 2004

• The hierarchy organizes change by importance
• Bigger change is associated with more distinct branches of the tree
• Hierarchy defines an ultrametric topology
• Hierarchical clustering therefore characterizes the Colombian data in ultrametric terms
• Whereas the correspondence analysis characterizes the Colombian data in metric (Euclidean) terms
How can we relate narcotics economy to the Colombian violence?

• Causal mechanisms could take place in different ways
• E.g. a declining world market for narcotics could put pressure on production, leading to violent reaction when faced with declining returns
• Or an expanding market could lead to greater violence because the spoils are greater
• Distribution is far and wide, relative to locations of production
How can we relate narcotics economy to the Colombian violence? - 2/3

• Cocaine drug comes from the coca plant
• Most initially processed cocaine on the world narcotics markets comes from Colombia
• Coca plant is very different from the cocoa bean from the cacao tree, from which chocolate is made
• Coffee, from coffee beans of the coffee plant, is the leading legal agricultural export from Colombia
• Could market changes in coffee lead to product substitution in response? This is just one among many difficulties of this type of analysis
How can we relate narcotics economy to the Colombian violence? - 3/3

• Opiates come from the opium poppy
• Like cocaine, opiates are also narcotic alkaloids
• From opium we get both (medicinal) morphine and heroin
• At times poppy production has been of some significance in Colombia, although leading opiate producer worldwide has been Afghanistan
• For supply of opiates to the US, Colombia and Mexico play a major role
• As before there are many intricate linkages between producers, distributors, consumers, the State
How can we regress the Colombian change data on these datasets?

Given the hierarchical structure - the sequence of agglomerations on the same set of years as in the graphs here - we can “fold” these datasets onto the hierarchy.

Having done this in each case, we will determine the best approximation to each narcotics dataset by hard thresholding Haar wavelet coefficients on the hierarchy.
Recall: what does hard thresholding of wavelet coefficients mean in practice?

- Hierarchy on upper left is wavelet transformed; small wavelet coefficients are set to zero ("hard thresholding"); hierarchy is re-constituted (inverse wavelet transform); result is on upper right
- Net result can also be seen in lower right
- This can be contrasted with "usual" partitioning on lower left
- We have a piecewise or clusterwise smooth of our data
- See Murtagh, JoC, The Haar wavelet transform of a dendrogram, 24, 3-32, 2007
Reprise…

• We have time series of narcotics data
• We have “folded” them over the hierarchy resulting from the Colombian violence data
• That hierarchy, resulting from the Colombian violence data, represented change over time at varying levels and scales
• For each time series of narcotics data we have thus induced a hierarchy
• Then we have checked out all possible modifications of these hierarchies, using wavelet-based hard thresholding
• Then we reconstructed the original data (hence: we applied the inverse wavelet transform; and the net overall effect is called the wavelet filtering of the original data)
The original narcotics time series are overplotted with the filtered time series. Left: 3-segment outcome; right: our selected best fit (MSE = mean square error).
Conclusions - 1/3

• Approximations to the opiate signals are tighter compared to the cocaine signals
• Due to our selection of partition, the opiates wholesale has 4 segments whereas the other signals have 3 segments in the two wholesale signals are similar in regard to years spanned, and segments in the retail signals are similar in regard to years spanned
• Narcotics prices are largely falling in this period, so it is not possible to look for links between prices and violence
• Instead we are looking for subtle effects
Conclusions - 2/3

• For both wholesale signals the ultrametric wavelet approximation is the left of, or equal to, the wholesale price signal from and including years 8 = 1997 to 12 = 2001.

• Information from the hierarchy, expressing the Colombian conflict, was lower than the wholesale price.

• The wholesale price itself was falling.

• So does this point to the Colombian conflict as having played a causative role in the wholesale price movement?
Conclusions - 3/3

• Unanticipated finding:
• We could expect the Colombian conflict to be a driver of US cocaine prices
• What we found is that the Colombian violence could also be a driver in regard to US opiate prices
• This could be a useful lead