3.10pt

Coherence Consideration in Binary Time Series Analysis

Benjamin Kedem

Department of Mathematics & Inst. for Systems Research
University of Maryland, College Park

"Give me a place to stand and rest my lever on, and I can move the Earth",
(Archimedes, 287-212 B.C.)

NU of Lisbon, October, 2013

- An interplay between coherence and logistic regression.
- Residual coherence: a graphical spectral tool for the identification of useful interaction and lagged terms.
- Instead of direct significance testing in terms of the residual coherence, the identified covariates are tested for their significance within logistic regression.

Outline

3.10pt

Let $\{X_t\}$, $t = ... -1, 0, 1, ...,$ be a zero mean stationary input. Define a system of degree n with output Y_t by the nth degree nonlinear model:

$$\begin{aligned}
Y_t &= \sum_{k=-\infty}^{\infty} h^{(1)}_{k_1} X_{t-k_1} \\
&+ \sum_{k=-\infty}^{\infty} h^{(2)}_{k_1,k_2} X_{t-k_1} X_{t-k_2} \\
&+ \sum_{k=-\infty}^{\infty} h^{(3)}_{k_1,k_2,k_3} X_{t-k_1} X_{t-k_2} X_{t-k_3} + \cdots
\end{aligned}$$

Assume $\{X_t\}$ admits a spectral representation in terms of a process of orthogonal increments $\xi_x(\lambda)$, $\lambda \in (-\pi, \pi]$. Then Y_t is an nth degree polynomial functional,

$$Y_t = \int \cdots \int \exp[it(\lambda_1 + \cdots + \lambda_n)] H_n(\lambda_1, \cdots, \lambda_n) \xi_x(d\lambda_1) \cdots \xi_x(d\lambda_n) + \cdots + \int \exp(it\lambda_1) H_1(\lambda_1) \xi_x(d\lambda_1) + H_0 \quad (1)$$

- We shall deal with quadratic system: $n = 2$.
- For integers u_k, $k = 1, ..., n$ define the lag processes $U_k(t)$ by the centered product,

$$U_k(t) = X_t X_{t+u_k} - R_{xx}(u_k)$$

  where R_{xx} is the autocovariance of X_t.
- Under the assumption that the $U_k(t)$ are stationary, it can be shown that for sufficiently large n, Y_t admits the mean-square representation (Kimelfeld 1972,1974),

$$Y_t = G_1\left(\frac{f_{xy}(\lambda)}{f_{xx}(\lambda)} Y_t\right) + \sum_{k=1}^{n}\left[\int e^{it\lambda} B_k(\lambda)\xi_{U_k}(d\lambda) + \int e^{it\lambda} A_k(\lambda)\xi_x(d\lambda)\right].$$

where:

$$A_k(\lambda) = -\frac{B_k(\lambda) f_{xu_k}(\lambda)}{f_{xx}(\lambda)}.$$

To get the $B_k(\lambda)$, define

$$\mathbf{f}_{uu}(\lambda) = (f_{u_i u_j}(\lambda)),$$

$$\mathbf{f}_{ux}(\lambda) = (f_{u_1 x}(\lambda), ..., f_{u_n x}(\lambda))', \mathbf{f}_{uy}(\lambda) = (f_{u_1 y}(\lambda), ..., f_{u_n y}(\lambda))',$$

$$\mathbf{B}(\lambda) = (B_1(\lambda), ..., B_n(\lambda))'.$$

Then, observing that $\mathbf{f}_{ux}(\lambda)$ is the conjugate transpose of $\mathbf{f}_{xu}(\lambda)$, we have

$$\mathbf{B}(\lambda) = \left(\mathbf{f}_{uu}(\lambda) - \frac{1}{f_{xx}(\lambda)}\mathbf{f}_{ux}(\lambda)\mathbf{f}_{xu}(\lambda)\right)^{-1} \left(\mathbf{f}_{uy}(\lambda) - \frac{f_{xy}(\lambda)}{f_{xx}(\lambda)}\mathbf{f}_{ux}(\lambda)\right).$$

- The definition of A_k guarantees the orthogonality of the approximating sum and G_1.
- Assume ϵ_t is orthogonal to G_1 and to the quadratic sum, and let

$$Y_t = G(t) + \epsilon_t$$
$$\equiv G_1\left(\frac{f_{xy}(\lambda)}{f_{xx}(\lambda)}, Y_t\right) + \sum_{k=1}^{n}\left[\int e^{it\lambda}B_k(\lambda)\xi_{U_k}(d\lambda) + \int e^{it\lambda}A_k(\lambda)\xi_x(d\lambda)\right] + \epsilon_t \quad (2)$$

- Then clearly the A_k and B_k do not change, and in addition,

$$0 \leq S_2(\lambda; u_1, ..., u_n) \equiv \frac{f_{gg}(\lambda)}{f_{yy}(\lambda)} \leq 1.$$

Laged Coherence

- It can be shown

$$S_2(\lambda; u_1, ..., u_n) = \frac{|f_{xy}(\lambda)|^2}{f_{xx}(\lambda) f_{yy}(\lambda)} + \frac{1}{f_{yy}(\lambda)} \mathbf{B}'(-\lambda) \left(\mathbf{f}_{uu}(\lambda) - \frac{1}{f_{xx}(\lambda)} \mathbf{f}_{ux}(\lambda) \mathbf{f}_{xu}(\lambda) \right) \mathbf{B}(\lambda) \quad (3)$$

- The first term on the right hand side of (3) is the well known (squared) linear coherence which measures the degree of linear relationship between X_t and Y_t.
- The other term is due to the quadratic term corresponsing to the lag processes $U_k(t)$.
- We shall refer to $S_2(\lambda; u_1, ..., u_n)$ as *lagged coherence* (Kimelfeld 1972).

- $S_2(\lambda; u_1, ..., u_n)$ measures the validity of models of the form (2) by observing that when $S_2(\lambda; u_1, ..., u_n)$ is close to 1 for all $\lambda \in [0, \pi]$ then the signal to noise ratio is high.
- $S_2(\lambda; u_1, ..., u_n)$ may be close to one on all or part of $[0, \pi]$ due mainly to the quadeatic term in (2) represented by the sum, in which case the system is substantially quadratic.
- Similarly, $S_2(\lambda; u_1, ..., u_n)$ could be large due to the linear component, in which case the system is substantially linear.

- The quadratic contribution can also be measured by Tick's (1961) "quadratic coherency".
- Assume that X_t is Gaussian, and let C.B.S. denote the *cross bi-spectrum*, the Fourier transform of $EX_{t+t_2}X_{t+t_1}(Y_t - EY_t)$ as a function of t_1, t_2.
- $$\text{quad. coh}(\omega) = \frac{|f_{xy}(\lambda)|^2}{f_{xx}(\lambda)f_{yy}(\lambda)} + \frac{\frac{1}{2}}{f_{yy}(\lambda)} \int \frac{|\text{C.B.S.}(\omega-\lambda,\lambda)|^2}{f_{xx}(\omega-\lambda)f_{xx}(\lambda)} d\lambda$$
- $$0 < \text{quad. coh}(\omega) < 1$$

Residual Coherence

It is more convenient to define a lag process using the notation

$$X_u(t) = X_t X_{t-u} - R_{xx}(u), \quad u = 0, 1, 2, \ldots \tag{4}$$

Consider the model

$$Y_t = \sum_{k=-\infty}^{\infty} l_k X_{t-k} + \sum_{k=-\infty}^{\infty} b_k X_u(t-k) + \epsilon_t \tag{5}$$

where ϵ_t is independent noise.

- $X_t X_{t-u}$ is an "interaction" term.

The lagged coherence reduces to (Kedem-Kimelfeld 1975),

$$S_2(\lambda; u) = S_1(\lambda) + \frac{|B(\lambda)|^2}{f_{yy}(\lambda)}\left[f_{x_u x_u}(\lambda) - \frac{|f_{xx_u}(\lambda)|^2}{f_{xx}(\lambda)}\right], \quad -\pi < \lambda \leq \pi \tag{6}$$

where $u = 0, 1, 2, \ldots$, $S_1(\lambda)$ is the linear coherence as in (3),

$$S_1(\lambda) = \frac{|f_{xy}(\lambda)|^2}{f_{xx}(\lambda) f_{yy}(\lambda)}, \tag{7}$$

and

$$B(\lambda) = \frac{f_{xx}(\lambda) f_{x_u y}(\lambda) - f_{x_u x}(\lambda) f_{xy}(\lambda)}{f_{xx}(\lambda) f_{x_u x_u}(\lambda) - |f_{xx_u}(\lambda)|^2} \quad -\pi < \lambda \leq \pi.$$

Claerly, $0 \leq S_1(\lambda) \leq 1$ for all $\lambda \in (-\pi, \pi]$, and similarly

$$0 \leq S_2(\lambda; u) \leq 1, \quad -\pi < \lambda \leq \pi, u = 0, 1, 2, \ldots$$

Residual Coherence

- For a given lag u, the influence of $X_u(t)$ on Y_t can be measured by noting a significant increase in $S_2(\lambda; u)$ relative to the linear coherence $S_1(\lambda)$, for some or all $\lambda \in [0, \pi]$.
- Alternatively, as suggested recently in Khan,Katzoff,Kedem (2013), we can use the maximum *residual coherence* defined as

$$RS(u) = \max_{\lambda}\{S_2(\lambda; u) - S_1(\lambda)\}, \quad u = 0, 1, 2, \ldots \tag{8}$$

to measure the influence of the “interaction" $X_u(t)$ on Y_t. This can be done graphically.

Residual Coherence Applied to Clipped Binary Series

Example:

- $X_t = 0.3X_{t-1} + \epsilon_t$ where ϵ_t is standard logistic noise.
- Consider an autoregression plus a past interaction covariate $X_{t-1}X_{t-2}$,

$$Z_t = 0.8Z_{t-1} + 1.5X_{t-1}X_{t-2} + \eta_t, \quad t = 1, ..., 156 \tag{9}$$

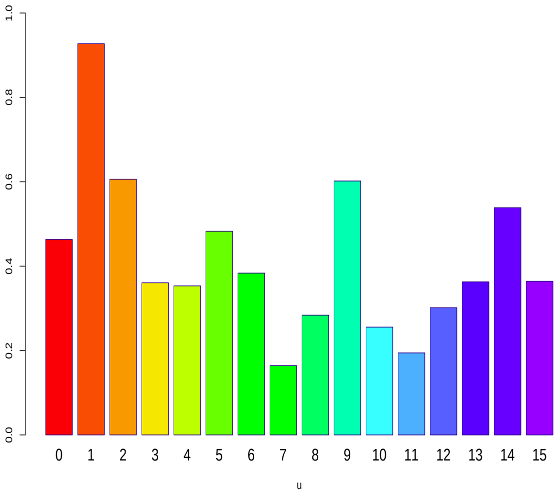
  where η is again standard logistic noise.
- Except for a constant $X_{t-1}X_{t-2}$ is a lag process with $u = 1$, and we would expect the residual coherence obtained from (X_t, Z_t) to peak at $u = 1$.

Clipping Z_t at level 5 we obtain a binary time series

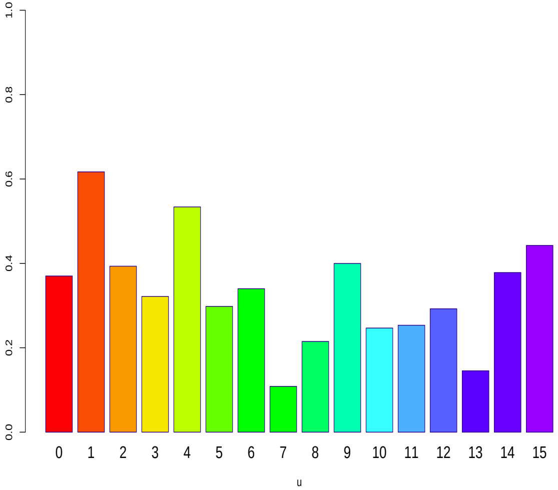
$$Y_t = \begin{cases} 1, & Z_t \geq 5 \\ 0, & Z_t < 5 \end{cases}, \quad t = 1, ..., 156.$$

The corresponding bar-plot obtained from (X_t, Y_t) again is maximized at $u = 1$ as expected, since in general clipping operations retain to a degree useful spectral information from the original baseline series which in the present case is Z_t; see Kedem (1980). Very similar bar-plots are obtained when ϵ_t and η_t are both Gaussian.

Maximum residual coherence $RS(u)$ obtained from (X_t, Z_t). $RS(u)$ peaks at $u = 1$.



Maximum residual coherence $RS(u)$ obtained from $((X_t, Y_t)$. $RS(u)$ peaks at $u = 1$.



Interactions in Logistic Regression

- The residual coherence introduced earlier can identify or point to potentially useful interactions in logistic regression.
- It should be considered as a suggestive device only.
- Significance tests can be used to ascertain that an identified interaction term is indeed a useful covariate.
- Thus, the importance of the identified interactions is determined from within logistic regression and not by coherence testing. In this way we avoid the inclusion of covariates judged useful on account of coherence but not significant on account of logistic regression analysis.

Illustration

- Consider the logistic regression model

$$\text{logit}(\pi_t(\boldsymbol{\beta})) = \beta_0 + \beta_1 Y_{t-1} + \beta_2 X_{t-1} + \beta_3 X_{t-2} + \beta_4 X_{t-1} X_{t-2} \qquad (10)$$

- The binary time series Y_t is obtained by clipping the above Z_t at level 5.
- Recall that an application of residual coherence points to $X_{t-1}X_{t-2}$ as a potential covariate.
- This is supported by the R output in the following table which points to the great significance of this (shifted) interaction term $X_{t-1}X_{t-2}$. However, X_{t-1} and X_{t-2} are not significant.

Table: Logistic regression results for model (10) showing the significance of the interaction $X_{t-1}X_{t-2}$.

	$\hat{\beta}$	SE	p-value
Intercept	-2.18220	0.38383	1.31e-08
Y_{t-1}	3.66744	0.54258	1.39e-11
X_{t-1}	0.03879	0.15594	0.804
X_{t-2}	-0.22056	0.16708	0.187
$X_{t-1}X_{t-2}$	0.57688	0.12933	8.18e-06

Application to LA Mortality

- Shumway et al (1988) analyzed filtered weekly mortality data in Los Angeles County from January 1, 1970 to December 31, 1979, using a regression model in terms of temperature (quadratic) and a log-pollution covariate, plus autoregressive noise.
- The data were reanalyzed in Kedem and Fokianos (2002) by Poisson regression, where it was found that total mortality Z_t, $t = 1, ..., 508$, depends on itself, temperature T_t, and log carbon momoxide $C_t = \log(CO_t)$.
- It is interesting to see what further insight might be gained using logistic regression.
- Define: $X_t = T_t/10$

Define the clipped series,

$$Y_t = \begin{cases} 1, & Z_t \geq 180 \\ 0, & Z_t < 180 \end{cases}, \quad t = 1, ..., 508.$$

This gives the binary time series

111101001000101000010000000000001000000110101111111111111111000000000000000000100001000 10
10100011111111111110000000000001000000001000000000000010011011111111101000000010000000000
0000000000000000000000001001100000000100000000000000000000000000000000000000000011111111 1
00000000000000000000000000000000000000000000000000011111000000000000000000000000000000000
00000000010000000000000000000000000000000000000000000000000001110110000000000000000000
00000000000001000000000011111001000000000000001010000000000000000000000000

Residual coherence obtained from (X_t, Y_t) suggesting (past of) $X_t X_{t-k}$, $k = 0, 2, 4$ as possible interaction covariates for logistic regression.

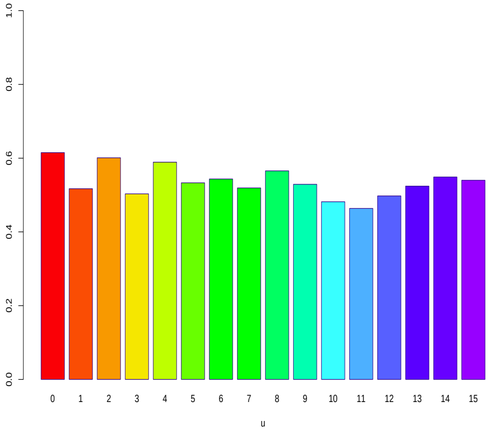


Table: Logistic regression models for clipped mortality data showing that showing the interactions $X_{t-1}X_{t-3}$ is a useful covariate.

	Model	AIC
1.	$Y_{t-1} + Y_{t-2}$	343.97
2.	$Y_{t-1} + C_{t-1}$	325.84
3.	$Y_{t-1} + Y_{t-2} + C_{t-1}$	306.95
4.	$Y_{t-1} + C_{t-1} + X_{t-1}X_{t-5}$	302.71
5.	$Y_{t-1} + Y_{t-2} + C_{t-1} + X_{t-3}$	298.94
6.	$Y_{t-1} + Y_{t-2} + C_{t-1} + X_{t-1}^2$	298.47
7.	$Y_{t-1} + Y_{t-2} + C_{t-1} + X_{t-1}$	298.35
8.	$Y_{t-1} + Y_{t-2} + C_{t-1} + X_{t-1} + X_{t-2} + X_{t-3}$	297.18
9.	$Y_{t-1} + Y_{t-2} + C_{t-1} + X_{t-1} + X_{t-3}$	297.03
10.	$Y_{t-1} + Y_{t-2} + C_{t-1} + X_{t-1}X_{t-3} + X_{t-1}X_{t-5}$	295.85
11.	$Y_{t-1} + Y_{t-2} + C_{t-1} + X_{t-1}X_{t-3}$	295.08

The table shows that the AIC is minimized at the model

$$\text{logit}(\pi_t(\boldsymbol{\beta})) = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + \beta_3 C_{t-1} + \beta_4 X_{t-1} X_{t-3} \quad (11)$$

containing the interaction $X_{t-1}X_{t-3}$ (past of $X_t X_{t-2}$) which appears more useful than its factors.

Table: Logistic regression results for model (11) showing the significance of the interactions covariate $X_{t-1}X_{t-3}$.

	$\hat{\beta}$	SE	p-value
Intercept	-3.544	1.315	7×10^{-3}
Y_{t-1}	1.662	0.351	2.12×10^{-6}
Y_{t-2}	1.148	0.361	10^{-3}
C_{t-1}	2.031	0.401	4.05×10^{-7}
$X_{t-1}X_{t-3}$	-0.059	0.017	5×10^{-4}

- The estimates are given in Table 3. Apparently $X_{t-1}X_{t-3}$ is quite significant, but not its factors.
- Model (11) can be judged further from the plots of the estimated autocorrelation and cumulative periodogram of the residuals shown in Figure 29.

Use the figure in the paper.