A Selection Criterion for Piecewise Stationary Long-Memory Models

Li Song & Pascal Bondon
2012 Statistical Signal Processing Workshop

Introduction
We consider the problem of estimating multiple structural breaks in a long-memory FARIMA signal. The number of break points as well as their locations, the orders and the parameters of each regime are assumed to be unknown. A selection criterion based on the minimum description length (MDL) principle is proposed and is compared favorably with two existing criteria by means of Monte Carlo simulations.

Piecewise FARIMA model description
A piecewise FARIMA time series \( \{Y_t\} \), \( t = 1, \ldots, n \) is modeled by

\[
Y_t = X_{\gamma t} + \epsilon_t, \quad \gamma_1 \leq t < \gamma_2,
\]

where \( \{X_t\}, t \in \mathbb{Z} \), is the FARIMA \( \{p, d, q\} \) process defined by the difference equation

\[
\phi(B)Y_t = \theta(B)(1 - B^{-\alpha})^{\lambda} \epsilon_t,
\]

where \( m \): unknown BP number.

\( \gamma_j \) : BP between the \( j \)th and \( j + 1 \)th FARIMA regime; \( j = 1, \ldots, m, \gamma_0 = 1 \) and \( \gamma_{m+1} = n + 1 \).

\( \{\epsilon_t\} \): sequence of iid zero-mean Gaussian random variables with unit variance \( \epsilon_t \sim N(0,1) \), \( t \in \mathbb{Z} \), \( j = 1, \ldots, m, \gamma_0 \).

\( B \): backward operator \( BX = X_t - X_{t-1} \) with real coefficients have no common zeros and neither \( \phi(x) \) nor \( \theta(x) \) has zeros in the closed unit disk \( z \in C, |z| \leq 1 \).

\( \lambda \): Process \( (1 - B^{-\alpha})^{\lambda} \epsilon_t \) is defined by

\[
(1 - B^{-\alpha})^{\lambda} \epsilon_t = \sum_{k=0}^{\lambda} \binom{\lambda}{k} (-1)^{\lambda - k} B^{-k \alpha} \epsilon_t,
\]

The parameters of the \( j \)th regime are \( \epsilon_j = (d_j, \phi_j, \theta_j, \gamma_j, \alpha_j) \) and \( \epsilon_{j+1} \) is constant for each interval \( (\gamma_j, \gamma_{j+1}) \). The piecewise FARIMA process \( \{Y_t\} \) is characterized by the BP number \( m \), the BP locations \( \gamma_0 \ldots \gamma_m \) and the parameters \( \epsilon_1 \ldots \epsilon_{m+1} \).

Model selection using MDL
Fitting model (1)-(2) to the data \( y = \{y_1, \ldots, y_n\} \) consists in finding the "best" vector \( \gamma = (m, \gamma_0, \gamma_1, \ldots, \gamma_m) \). Let \( L(\gamma) \) denote the code length of an object. The "best" model is the one minimizing \( L(\gamma) \). We adopt the two-part description length method used by Rissanen; see e.g. Lee (2001). Then

\[
L(\gamma) = L(n) + L(y|\gamma),
\]

where \( L(n) \) is the length of the code of the number of observations in the \( \gamma \)th piece in (1). Then we select the best model (1)-(2) for \( y \) as the one that minimizes \( L(\gamma) \).

Conclusions
• Criterion 1: works better than both 0 and \( \kappa \) in terms of percentage of selection of the true break points, precisionness of estimation of the break fractions, and precision of selection of the true orders.

Application to River Nile data
The raw data is the series of yearly minimal water levels of the Nile River for the years 622-1284 AD \( (n = 663) \). The Nile River data is modeled using a two regimes piecewise FARIMA model with a BP at 722 AD by the three criteria.

References