Locally Stationary Processes

STAT 7550 Time Series

Hengrui Luo and Kori Khan

Department of Statistics
The Ohio State University

April 18, 2018

Outline

In this lecture, we are going to present you the basic notions and techniques used for locally stationary processes and carried out an example of locally AR(1) time series.

- Nonstationary data occurred in various scenarios [2, 8]
  - Spatial statistics
  - Forestry and environmental data
  - Time series and financial data
- Approaches to model nonstationary data
  - Nonlinear time series
  - Volatility models
  - Locally stationary process
    Locally stationary process method is a natural extension from the vast literature existing for stationary processes method in time series data. There are two ways of defining a locally stationary process,
    - by setting up varying coefficients at time varies in the defining equations.
    - by setting up varying spectral density form.

Definition

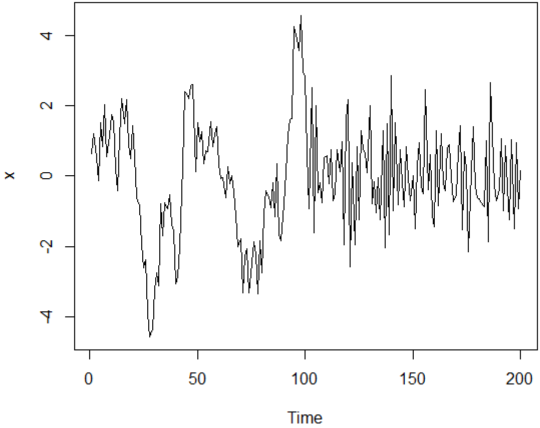
- Time varying $AR(1)$
  Suppose the $AR(1)$ process we talked about in class is defined by

  $$\phi(B)X_t = X_t - \phi_1 X_{t-1} = Z_t$$

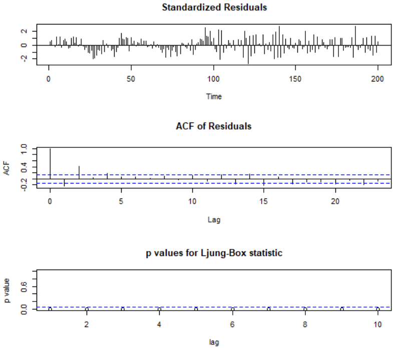
  We can imagine the following definition for a time series X_t on different time domain ($\phi_1 \neq \phi_2$):

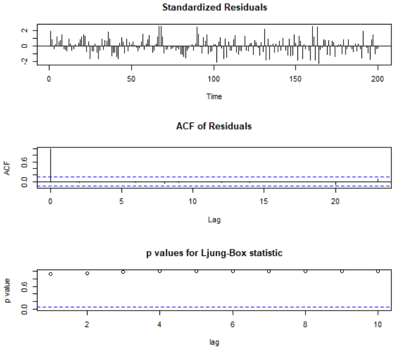
  $$\begin{aligned}\phi_1(B)X_t &= X_t - \phi_1 X_{t-1} = Z_t t = 1, 2, \cdots, T \\ \phi_2(B)X_t &= X_t - \phi_2 X_{t-1} = Z_t t = T+1, T+2, \cdots\end{aligned}$$

- Simulation example
  If we simulate from $AR(1), \phi = 0.9$ and join it with $AR(1), \phi = -0.5$ then the time series plot will look like:



- If we naively fit it to an $AR(1)$ model then the diagnostic will look like (LHS)





- From the standardized residuals we know that this is not an appropriate model and even if we fit the joined time series data to $AR(20)$ (RHS) the nonstationarity still persists. And the Ljung-Box statistics indicates that the non-stationary in the data affects the fit of the AR model.

The model that suits this kind of data

- We can either define a model from its defining equations, that is to define its transition kernel, by letting the defining polynomial of $AR(1)$ processes vary and get an even more general locally stationary process in following form

$$\phi(B)X_t = X_t - \phi_1 X_{t-1} = Z_t$$

$$\phi(B)X_t = X_t - \phi(t)X_{t-1} = Z_t, t = 1, \cdots, T$$

and our example above can be written in this form by specifying the coefficient function (as a function of time index t)

$$\phi(t) = \begin{cases} \phi_1 & t = 1, 2, \cdots, T \\ \phi_2 & t = T+1, T+2, \cdots \end{cases}$$

- Or we can define a model by its spectral density function, that is to define a piecewise spectral density function, by letting the density function have different forms for different time ranges. It is known that $AR(1)$ process $\phi(B)X_t = X_t - \phi_1 X_{t-1} = Z_t$ has a spectral density as follows, we use $S_X(\bullet)$ for spectral density function and f for frequency notation in the discussion hereafter.

$$S_X(f) = \frac{\sigma^2}{|1 - \phi_1 e^{-i2\pi f}|^2}, |f| < \frac{1}{2}$$

and our example above can be written in this form by specifying the parameter in the density function (as a function of time index t)

$$S_X(f) = \frac{\sigma^2}{|1 - \phi(t) e^{-i2\pi f}|^2}, |f| < \frac{1}{2}$$

Stationary approximation

- For simplicity we assume that within each time domain $t = 1, \cdots, T$ and reparameterize the time index as $\frac{t}{T} \in (0,1)$.
- Intuitively speaking, this idea of analyzing is parallel to the method of using Taylor expansion around a certain point on a certain function.
- In following procedure of *stationary approximation*, we use the stationary derivative processes $\tilde{X}_t(u_0)$ at each time point $u_0 \in (0,1)$ to approximate the $X_{t,T}$.
- We want to construct a stationary time series $\tilde{X}_t$ and try to approximate $X_{t,T}$ around the time point u_0.
  To see this idea better, we write down the likelihoods of these two different models for $p = 1$. i.e.

$$X_{t,T} + \alpha_1\left(\frac{t}{T}\right) \cdot X_{t-1,T} = \sigma\left(\frac{t}{T}\right) \cdot Z_t$$
$$\tilde{X}_t(u_0) + \tilde{\alpha}_1(u_0) \cdot \tilde{X_{t-1}}(u_0) = \sigma(u_0) \cdot Z_t$$
$$Z_t \sim WN(0,1), t = 1, \cdots, T$$

- The locally stationary likelihood can be written as $L_T(u_0, \theta)$ and the approximating stationary likelihood can be written as $\tilde{L}_T(u_0, \theta)$. Assume that X_0 is already known for simplicity. $X_{T,i} = \tilde{X}_i = X_i$ in the following likelihoods,

$$L_T(\frac{t}{T}, \theta) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma\left(\frac{t}{T}\right)^2}} \cdot exp\left(-\frac{(X_{T,i} - \alpha_1\left(\frac{t}{T}\right) \cdot X_{T,i-1})^2}{2\sigma\left(\frac{t}{T}\right)^2}\right)$$

$$\tilde{L}_T(u_0, \theta) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma(u_0)^2}} \cdot exp\left(-\frac{(\tilde{X}_i - \tilde{\alpha}_1(u_0) \cdot \tilde{X}_{i-1})^2}{2\sigma(u_0)^2}\right)$$

How good is our stationary approximation?

- In our $AR(1)$ example, our $X_{t,T}$ has a representation

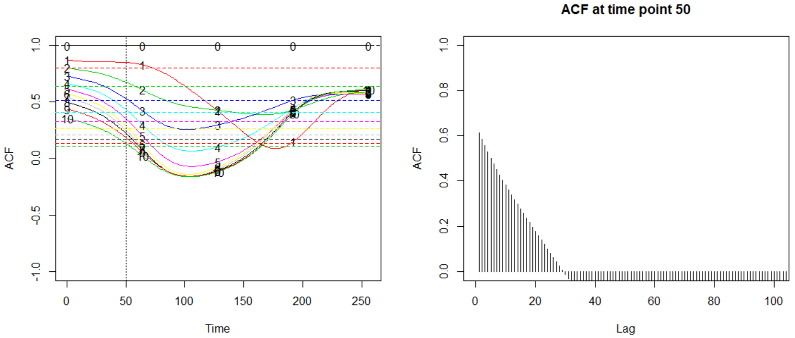
$$X_{t,T} = \mu(\frac{t}{T}) + a_1(t)Z_{t-1}, t = 1, \cdots, T$$

Then under some regularity condition (see our report or [2, 7])the derivative stationary process for $X_{t,T}$ around u is

$$\tilde{X}_t(u) := \mu(u) + a_1(u)Z_{t-1}$$

with estimated mean μ and the error term is bounded by $O\left((\frac{t}{T} - u_0) + \frac{1}{T}\right)$. Correspondingly, the spectral density near u is also a good approximation in sense that $\tilde{S}(u,f)$ will approximate the spectral density $S_T(u,f)$ of the locally stationary process $X_{t,T}$ around u up to an error term of $o(1)$

$$\int_{-\frac{1}{2}}^{\frac{1}{2}} |S_T(u,f) - \tilde{S}(u,f)|^2 du = o(1)$$



The ACF are varying as the time (horizontal axis) varies. According to our locally stationary processes model for two-segment AR data we discussed above, we can use wavelet basis (so sample size has to be 2^n)to estimating the spectral density of the stationary approximation processes in `costat` package [6]. Different colors of curves means the different lags global ACF $C_t(h)$; different numbers denotes the lag. The dashed lines is the AR ACF if fitted to AR(1) (0.8 here).
For example, we extract the locally stationary approximating stationary process at time 50, i.e. $C_{t=50}(h)$ as below (RHS).

- Can we do better approximation?
- By rational approximation of spectral density (for example, Theorem 4.4.3 in [1]), any stationary process can be approximated by an $AR(p)$ for p sufficiently large.
- Yes, we can approximate the locally stationary processes using an additive sum of derivative stationary processes. But at the same time we should be parsimonious!
- Where else in statistics do we apply the idea of additivity ?
  - Additive modeling
  - Iterative least square methods
  - Levinson-Durbin algorithm

Generalization of Whittle Likelihood

Recall from Gaussian (mean zero) univariate stationary case, the Whittle Likelihood was:

$$\tilde{l}_T(\theta) = -2l_T(\theta) \approx \int_{-1/2}^{1/2} \left\{ \log(2\pi S_X(f)) + \frac{\hat{S}_X^{(p)}(f)}{S_X(f)} \right\} df$$

Some Uses for this:

- Under a number of assumptions, estimates asymptotically equivalent to MLE
- Computationally nicer

Can this be generalized?

A sequence of Gaussian stochastic processes $X_{t,T} = (t = 1, \ldots, T, T \geq 1)$ is called a *Gaussian locally stationary process* with *transfer function* A° and *trend function* μ if there exists a representation such that,

$$X_{t,T} = \mu(\frac{t}{T}) + \int_{-1/2}^{1/2} \exp(i2\pi ft)\ A^{\circ}_{t,T}(f)\ d\xi(f)$$

1. $\xi(f)$ a complex valued Gaussian process on $[-\frac{1}{2}, \frac{1}{2}]$ with $\overline{\xi(f)} = \xi_a(-f)$, $E(\xi(f)) = 0$ [1]
2. There exists a constant $K > 0$ and an 1-periodic function $A : [0,1] \times \mathbb{R} \to \mathbb{C}$ with $\overline{A(u,f)} = A(u,-f)$ and

$$\sup_{t,f} |A^{\circ}_{t,T}(f) - A(\frac{t}{T}, f)| \leq \frac{K}{T}$$

for all $T \in \mathbb{N}$. $A(u,f)$ and $\mu(u)$ are assumed to be continuous in u.

[1] Some additional technical conditions

For reference:

$$X_{t,T} = \mu(\frac{t}{T}) + \int_{-1/2}^{1/2} \exp(i2\pi ft)\ A^{\circ}_{t,T}(f)\ d\xi(f)$$

$$\sup_{t,f} |A^{\circ}_{t,T}(f) - A(\frac{t}{T}, f)| \leq \frac{K}{T}$$

Some Definitions:

1. $S_X(u,f) = |A(u,f)|^2$ is called the time-varying spectral density of the function
2. $S_{X,\theta}(u,f)$ is called the time varying spectral density of the model process [2]
3. μ is the trend function of the process
4. μ_θ is the trend function of the model process
5. We'll assume $\mu_\theta = \mu = 0$

[2] $\theta \in \Theta$, Θ compact

Going back:

$$\tilde{l}_T(\theta) = -2l_T(\theta) \approx \int_{-1/2}^{1/2} \left\{ \log(2\pi S_X(f)) + \frac{\hat{S}_X^{(p)}(f)}{S_X(f)} \right\} df$$

Periodogram doesn't seem well suited anymore:

$$\hat{S}_X^{(p)}(f) = \sum_{h=-(T-1)}^{T-1} \left(\frac{1}{T} \sum_{t=1}^{T-|h|} X_t X_{t+|h|} \right) e^{-i2\pi fh}$$

So instead think about local behavior, the pre-periodogram:

$$J_T(u,f) = \sum_{\substack{h \\ 1 \leq [uT + \frac{1+h}{2}], [uT + \frac{1-h}{2}] \leq T}} X_{[uT+\frac{1+h}{2}],T} X_{[uT+\frac{1-h}{2}],T}$$

Notice, that at time t, the pre-periodogram:

$$J_T(t/T, f) = \sum_{\substack{h \\ 1 \leq [t+\frac{1+h}{2}], [t+\frac{1-h}{2}] \leq T}} X_{[t+\frac{1+h}{2}],T} X_{[t+\frac{1-h}{2}],T}$$

is using $X_{[t+\frac{1+h}{2}],T} X_{[t+\frac{1-h}{2}],T}$ as a "local" estimator of the covariance at lag h at time t.

$$t + \frac{1+h}{2} - (t + \frac{1-h}{2}) = h$$

You can show that

$$\hat{S}_X^{(p)} = \frac{1}{T} \sum_{t=1}^{T} J_T(\frac{t}{T}, f)$$

Now, to obtain the generalization for the Whittle likelihood for a locally stationary process, we will:

1. Replace $\hat{S}_X^{(p)}$ by the average of the pre-periodograms
2. Replace $S_X(f)$, the spectral density, with the time varying spectral density $S_X(\frac{t}{T}, f)$.

$$I_T^{GW} := \frac{1}{T}\sum_{t=1}^{T}\int_{-1/2}^{1/2}\left\{\log(2\pi S_X(\frac{t}{T}, f)) + \frac{J_T(\frac{t}{T}, f)}{S_X(\frac{t}{T}, f)}\right\} df$$

Notice what happens in the stationary case

Asymptotics for Locally Stationary Processes

- For the stationary case, we're assuming more and more "future" observations become available
- What does an asymptotic argument mean for a locally stationary process?
- One approach is to rescale the interval of observation, similar to non-parametric regression

Example: [3]
Suppose we observe:

- $X_t = g(t)X_{t-1} + \varepsilon_t,\ \varepsilon_t \stackrel{iid}{\sim} N(0,\sigma^2), t = 1,\ldots,T$
- $g(t) = a + bt + ct^2,\ |g(t)| < 1\ t \in [1,T]$

We would then define:

$$X_{t,T} = g(t/T)X_{t-1,T} + \varepsilon_t$$

$$X_{t,T} = g(t/T)X_{t-1,T} + \varepsilon_t$$

- But, what does this mean as T goes to infinity?
- It does NOT mean:
    - We are sampling more and more from a continuous time process
    - We are sampling further and further into the future
- Instead, we have more observations for each value of g.

Note: in the stationary case, where g is constant, the asymptotics will be the same as for the classical asymptotics for stationary processes

- Let $\hat{\theta}^{GW} = \arg\min_{\theta \in \Theta} l_T^{GW}$
- Let $\tilde{\theta}_T$ denote the MLE of θ.
- Let θ_0 be the "true" value
- If the model is not mis-specified (i.e.,$\mu = \mu_{\theta_0}$ and $A_{\theta_0}(u, f) = A(u, f)$ for some $\theta_0 \in \Theta$) [3], then

$$\sqrt{T}(\hat{\theta}_T^{GW} - \theta_0) \xrightarrow{d} N(0, \Gamma^{-1})$$
$$\sqrt{T}(\tilde{\theta}_T - \theta_0) \xrightarrow{d} N(0, \Gamma^{-1})$$

where

$$\Gamma = \frac{1}{2} \int_0^1 \int_{-1/2}^{1/2} \nabla \log(S_{X,\theta_0}) \nabla \log(S_{X,\theta_0})^T dfdu$$

[3] Assuming a number of technical conditions

For Further Reading I

- Brockwell, Peter J., and Richard A. Davis. Time series: theory and methods. Springer Science & Business Media, 2013.
- Dahlhaus, Rainer. "Locally stationary processes." Handbook of statistics. Vol. 30. Elsevier, 2012. 351-413.
- Dahlhaus, Rainer. "Maximum Likelihood Estimation and Model Selection for Local Stationary Processes. " Nonparametric Statistics. 6 (1996): 171-191.
- Dahlhaus, Rainer. "Fitting time series models to nonstationary processes." The annals of Statistics 25.1 (1997): 1-37.
- Dahlhaus, Rainer. "A Likelihood Approximation for Locally Stationary Processes. " The annals of statistics. 38.6 (2000): 1762-1794.

For Further Reading II

- Nason, Guy P., Rainer Von Sachs, and Gerald Kroisandt. "Wavelet processes and adaptive estimation of the evolutionary wavelet spectrum." Journal of the Royal Statistical Society: Series B (Statistical Methodology) 62.2 (2000): 271-292.
- Cardinali, Alessandro, and Guy P. Nason. "Costationarity of locally stationary time series." Journal of Time Series Econometrics 2.2 (2010).
- Paciorek, Christopher Joseph. Nonstationary Gaussian processes for regression and spatial modeling Diss. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2003.
- Ledoux, Michel, and Michel Talagrand. Probability in Banach Spaces: isoperimetry and processes. Springer Science & Business Media, 2013.

For Further Reading III

\Dahlhaus, Rainer, Stefan Richter, and Wei Biao Wu. "Towards a general theory for non-linear locally stationary processes." arXiv preprint arXiv:1704.02860 (2017).