LOCALLY STATIONARY PROCESS AND ITS APPLICATIONS: A BRIEF TUTORIAL

HENGRUI LUO AND KORI KHAN

1. Introduction to Locally Stationary Processes

Nonstationary data occurred in various scenarios like spatial statistics, geology and time series [2, 8]. Among other approaches that can be used for modeling a seemingly nonstationary time series dataset, 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, let us consider a typical stationary process $AR(1)$ first. Suppose the $AR(1)$ process is defined by

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

where $Z_t \sim WN(0, \sigma^2)$ as $t \in \{1, \cdots, T\}$. Then the time range $t \in \{1, \cdots, T\}$ defines a domain within which the process X_t can be described by an $AR(1)$ process. This specification is different from the usual autoregressive process where the $t \in \mathbb{Z}$ is usually assumed. Therefore locality is introduced in the model definition. We can imagine the following definition for a time series X_t on different time domain ($\phi_1 \neq \phi_2$):

$$\phi_1(B)X_t = X_t - \phi_1 X_{t-1} = Z_t t = 1, 2, \cdots, T$$

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

which is like a conjunction of two time series over the integer time domain $t \in \mathbb{Z}^+$. It is no longer stationary because $\gamma_X(1,2) = Cov(X_1, X_2) = \phi_1$ is clearly different from $\gamma_X(T+1, T+2) = Cov(X_{T+1}, X_{T+2}) = \phi_2$. However, on $[1, T]$ and $[T+1, \infty)$ X_t are separately $AR(1)$ time series and hence stationary. Therefore we can call such a time series a "locally stationary" process. In a greater generality, we could let the defining polynomial of $AR(1)$ processes vary and get an even more general locally stationary process in following form

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

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

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

Therefore a locally stationary $AR(1)$ process can be defined through a time-varying polynomial $\phi(B, t)$. An alternative perspective is to investigate the spectral representation of the $AR(1)$ process. 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 with the same calculation as above we know a general locally stationary $AR(1)$ can be defined as a process with spectral density in form of

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

We should notice that although these two ways of describing a locally stationary $AR(1)$ process are equivalent. In more generality, the spectral representation seems to be preferred.

We focus on the contrast and comparison between likelihoods of stationary and local stationary processes. For simplicity we restrict our attention to the unit interval. We write down our **locally stationary** time series model

$$X_{t,T} + \sum_{j=1}^{p} \alpha_j \left(\frac{t}{T}\right) \cdot X_{t-j,T} = \sigma \left(\frac{t}{T}\right) \cdot Z_t$$

$$Z_t \sim WN(0, 1), t = 1, \cdots, T$$

with corresponding spectral density $S_X(f)$ as defined above.

However, it is also convenient to use following local approximation by the **stationary** process around a fixed time point $u_0 = \frac{t_0}{T}$,

$$\tilde{X}_t(u_0) + \sum_{j=1}^{p} \tilde{\alpha_j}(u_0) \cdot \tilde{X_{t-j}}(u_0) = \sigma(u_0) \cdot Z_t$$

$$Z_t \sim WN(0, 1), t = 1, \cdots, T$$

2. Stationary Approximation

Intuitively speaking, this idea of analyzing is parallel to the method of using Taylor expansion around a certain point on a certain function. In Taylor expansion, we use the $f(x_0) + (x - x_0)f'(x_0) + \cdots$ to approximate $f(x)$ around $x = x_0$. In following procedure we use the derivative processes $\tilde{X}_t(u_0)$ at each time point $u_0 \in (0, 1)$ to approximate the $X_{t,T}$ and control the error term. Here we 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 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,

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

$$\tilde{L}_T(u_0, \theta) = \prod_{i=1}^{T} \frac{1}{\sqrt{2\pi\sigma(u_0)^2}} \cdot exp\left(-\frac{(\tilde{X}_i - \tilde{\alpha_1}(u_0) \cdot \tilde{X}_{i-1})^2}{2\sigma(u_0)^2}\right)$$

For the same value of $u_0 = \frac{t}{T}$, these two likelihoods are essentially the same with probability one (with respect to the product measure of $X_1, \cdots, X_T$ in the case of T samples). Therefore they converge to the same limiting likelihood function $L(u_0, \theta) = L_T(u_0, \theta)$ (because the likelihood function is actually defined for different T's).

Then strictly speaking, we have following theoretical result that supports our application of approximating local stationarity using global stationarity.

Theorem 1. *(Stationary approximation, [7]) Assume that $X_{t,T}$ has a representation*

$$X_{t,T} = \mu(\frac{t}{T}) + \sum_{j=-\infty}^{\infty} a_{t,T}(j) Z_{t-j}$$

with a bounded variation μ and $Z_t \sim WN(0, 1)$. Let

$$\ell(j) = \begin{cases} 1 & |j| \leq 1 \\ |j| \log^{1+\kappa}|j| & |j| > 1 \end{cases}$$

for some $\kappa > 0$ *and* $\sup_t |a_{t,T}(j)| \leq \frac{K}{\ell(j)}$ *with* K *independent of* T*. Also we need a bounded variation function* $a(\cdot, j) : (0, 1] \to \mathbb{R}$ *with*

$$\sup_u |a(u, j)| \leq \frac{K}{\ell(j)}$$
$$\sup_j \sum_{t=1}^{T} |a_{t,T}(j) - a(\frac{t}{T}, j)| \leq K$$
$$\|a(\cdot, j)\|_{TV} \leq \frac{K}{\ell(j)}$$

i.e. the coefficients that defines the time-varying local stationary time series are all bounded variation. Then the derivative stationary process for $X_{t,T}$ *around* x_0 *is*

$$\tilde{X}_t(u) \coloneqq \mu(u) + \sum_{j=-\infty}^{\infty} a(u, j) Z_{t-j}$$

The proof of the theorem is a direct check that the error term is bounded by $O\left((\frac{t}{T} - u_0) + \frac{1}{T}\right)$. The idea is to approximate the j-th time-varying coefficients $a_{t,T}(j)$ at u_0 using a fixed coefficient $a(u_0, j)$, these two $a_{t,T}(j), a(u_0, j)$ are called *closed pairs* [7]. There are multiple ways of doing this, one analysis technique is to use a partition of unity over time domain; any interpolation technique with bounded variation basis function can also be used since the time domain is discrete. The reason why this stationary approximation will not extend to continuous time series is the basic difficulty of doing a uniformly bounded function approximation on a continuous time domain.

As for consistency of the estimates from local likelihoods we introduce following result, whose proof is the classic two-step procedure, we first assert $\sup_{\theta \in \Theta} |L_T(u_0, \theta) - \tilde{L}_T(u_0, \theta)| \xrightarrow{P^T} 0$ the stationary approximation is asymptotically good; and then $\sup_{\theta \in \Theta} |L(u_0, \theta) - \tilde{L}_T(u_0, \theta)| \xrightarrow{P^T} 0$ the stationary likelihood is somehow consistent, then the MLE from local likelihood is consistent by first going to global stationary likelihood estimates. Here we again make use of the locality to assemble the results we gain from the global stationary likelihood approximation. This idea perpetuates throughout the literature about local stationary processes even beyond the scope of time series.

Theorem 2. *(Dahlhaus-Rao, Theorem 3.2 in* [2]*) Suppose that* Θ *is compact with* $\theta_0(u_0) \coloneqq \arg\min_{\theta \in \Theta} L(u_0, \theta) \in Int\Theta$*, the function* $L(u_0, \theta)$ *is continuous at* $\theta_0(u_0)$ *and the minimum* $\theta_0(u_0)$ *is unique. If*

$$\sup_{\theta \in \Theta} |L_T(u_0, \theta) - \tilde{L}_T(u_0, \theta)| \xrightarrow{P^T} 0$$
$$\sup_{\theta \in \Theta} |L(u_0, \theta) - \tilde{L}_T(u_0, \theta)| \xrightarrow{P^T} 0$$

then

$$\hat{\theta}_T(u_0) - \theta_0(u_0) \overset{P^T}{\to} 0$$

A very perceptual view is to view $L(\bullet, \theta)$ as a functional defined on the product sample space where θ are the *autoregressive coefficients and variance parameters* of the locally stationary process at time point $u_0 \in (0, 1)$ that we defined for. For example in locally stationary AR(1) $\theta = (\phi_1, \sigma^2)$ at time point u_0. These parameters will obviously vary as the time parameters u_0 changes.

$$(\mathcal{X}_1 \times \cdots \times \mathcal{X}_T, p_{X_1} \times \cdots \times p_{X_T}) =: \left(\mathcal{X}^T, P^T\right)$$

that maps a time point u_0 to a function $L(u_0, \theta) : \Theta \to \mathbb{R}$. With some nontrivial efforts, then $L(\bullet, \theta)$ can be regarded as a Banach valued random variable whose law of probability can be determined by $p_{X_1} \times \cdots \times p_{X_T}$ for a fixed T. The asymptotic techniques which can be applied onto the Banach valued random variables like Banach law of large numbers and Banach CLT [9]. This is a higher view on the asymptotic results we can obtained, with the least constraints on the time varying time series we defined. However, more refined techniques are available by asymptotic expansion of the time varying coefficients in literature we are not discussing in the report in details. In fact, the ergodic principle in Banach category updates most classic martingale theorems and Donsker-type result at the price of loss of rate of convergence.

The two theorems above reduces the work of doing inference on a locally stationary time series to doing inference on an approximating stationary time series, which we are familiar with. For example, we can obtain a Yule-Walker estimate for autoregessive coefficient for each fixed time point u_0 by solving the Yule-Walker equations based on the stationary approximation $\tilde{X}_t$ and then denote this Yule-Walker estimate $\alpha(u_0)$ and $\sigma^2(u_0)$ since it depends on the stationary approximation constructed at this specific point u_0. It is readily to think that $\alpha(u_0), \sigma^2(u_0)$ as a function as u_0 varies. Then a natural question to ask is that how 'close' $\alpha(u_0)$ is to the real coefficient $a_1(u_0)$, since now we are comparing two curves in the time domain. This consideration also justifies our assumption on the bounded variation assumption on the coefficients of the locally stationary time series from another angle.

As in the introduction section, we end our discussion with a result about the spectral density of the stationary approximation.

Theorem 3. *(Spectral density of stationary approximation, Theorem 4.3 in* [2]*) Assume that $X_{t,T}$ satisfies the assumptions in Theorem 1 and in addition for some $i = 1, 2, \cdots$*

$$\sup_{u\in(0,1)} \left| \frac{\partial^i \mu(u)}{\partial u^i} \right| \leq K$$

$$\sup_{u\in(0,1)} \left| \frac{\partial^i a(u,j)}{\partial u^i} \right| \leq \frac{K}{\ell(j)}$$

$$\sup_{t,T} \left| a_{t,T}(j) - a(\frac{t}{T}, j) \right| \leq \frac{K}{T \cdot \ell(j)}$$

the corresponding (true) spectral density $\tilde{S}(u, f)$ of the stationary approximation process at time point u (remember that the stationary process approximation carries out at each of these time points)

$$\tilde{X}_t(u) \coloneqq \mu(u) + \sum_{j=-\infty}^{\infty} a(u,j) Z_{t-j}$$

will approximate the spectral density $S_T(u, f)$ of the locally stationary process $X_{t,T}$ in

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

Besides stationary approximation proposed by Dahlhaus [4] whose fundamental idea is approximating the defining polynomial of the process, an alternative way of approximating the locally stationary process is to approximate the time-varying spectral density function $S_T(u, f)$ using a pre-specified collection of basis in the spectral spaces. A widely accepted basis in practice is wavelet basis proposed by Nason and his collaborators [6].

There are corresponding consistency and asymptotic results for the spectrum approximation approach, but the point is that we can always approximating the nonstationarity from either the transition operator of the process (defining polynomial in time series) or the spectrum of the process. These two legs of dealing with nonstationarity are usually equivalent, yet sometimes one leg is easier to work with.

The example we work with is the time-varying $AR(p)$ process, one remark to make is that $AR(p)$ process has more significance than pedagogic purpose if we see these 'two legs' of approximating the locally stationary time series. 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. If we have already known the local domains $I_1, I_2, \cdots$ upon which the X_t are stationary, then we can fit a time-varying $AR(p_j)$ process on I_j for sufficiently large p_j on each domain to get a satisfying approximation. This idea of approximating the local stationary part using $AR(p)$ processes reduces our discussion of approximating locally stationary processes to approximating locally stationary $AR(p)$ processes.

3. Gaussian-Whittle Likelihood Theory

In this section, we will discuss a generalization of the Whittle likelihood method for the estimation of parametric models for Gaussian locally stationary processes. As we saw in the Gaussian processes theory, the specific form of Gaussian likelihood will allow us to derive stronger results. First, we introduce a more general definition of a Gaussian locally stationary process adapted from [3] and [5].

Definition 4. A sequence of Gaussian multivariate stochastic processes $X_{t,T} = (X^1_{t,T}, \dots, X^d_{t,T})^T, t = 1, \dots, T, T \geq 1$ is called a *Gaussian locally stationary process* with *transfer function matrix* A° and *trend function* μ if there exists a representation such that,

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

where the following conditions holds for all $a, b = 1, \dots, d$.

(1) $\xi(f) = (\xi_1(f), \cdots, \xi_d(f))$ is a complex vector valued Gaussian vector process on the spectral domain $[-\frac{1}{2}, \frac{1}{2}]$ with $\overline{\xi_a(f)} = \xi_a(-f)$, $E(\xi_a(f)) = 0$, and

$$E\left\{d\xi_a(f)d\xi_b(f)\right\} = \delta_{ab} \cdot \eta(f + \mu)df d\mu$$

where $\eta(f) = \sum_{j=-\infty}^{\infty} \delta(f + j)$ is the period 1 extension of the Dirac delta function.

(2) There exists a constant $K > 0$ and an 1-periodic matrix valued function $A : [0, 1] \times \mathbb{R} \to \mathbb{C}^{d\times d}$ with $\overline{A(u, f)} = A(u, -f)$ and

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

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

We call $S_X(u, f) = A(u, f)\overline{A(u, f)}^T$ the time-varying spectral density decomposition of the process. Local stationarity is a property of the whole triangular array $X_{t,T}$; however, here we focus on the case in which we observe a realization $X_{1,T}, \dots, X_{T,T}$ with T fixed.

We assume this realization is from a locally stationary Gaussian process with true mean function μ and transfer function A. We fit a class of locally stationary Gaussian processes indexed by parameter $\theta \in \Theta$ with mean function μ_θ and transfer function A_θ, where the parameter space Θ is compact. We will call $S_{X,\theta} = |A_\theta(u, f)|^2$ the time varying spectral density of the model process and μ_θ the model mean function.

For mean-zero univariate stationary processes, the Whittle likelihood employs the periodogram to approximate the negative Gaussian likelihood.

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

A natural inquiry is whether the result can be generalized for use in locally stationary processes. With a modification of the Toeplitz matrices occurring in the covariance structure, the approximating result by Whittle in the stationary case can be generalized into the locally stationary case. This generalization makes use of the concept of pre-periodogram, which we will now discuss.

For locally stationary processes, the periodogram is no longer appropriate. The periodogram is a "global" measure of dependence, in some respects; giving us a measure of the covariance of lag h over the entirety of the spectral domain. In the notation of class, for univariate case we will denote this as,

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

But,for locally stationary processes, our covariance structures will now vary as function of t. So, if we instead fixed the time at t and considered the covariances at lag h centered at t, $t \leq T$, we would have "local" measure of covariance. The *pre-periodogram* is such a "local" measure of dependence, and we define it as follows for the univariate case.

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

There is a key relationship between the periodogram and pre-periodogram which will enable us to generalize Whittle's likelihood. The periodogram can be decomposed into an average of the pre-periodograms at each of the observations.

$$\begin{aligned} \hat{S}_X^{(p)} &= \sum_{h=-(T-1)}^{T-1} \left(\frac{1}{T} \sum_{t=1}^{T-|h|} X_t X_{t+|h|} \right) e^{-i2\pi fh} \\ &= \frac{1}{T} \sum_{t=1}^{T} \sum_{\substack{h \\ 1 \leq [t + \frac{1+h}{2}], [t + \frac{1-h}{2}] \leq n}} X_{[t + \frac{1+h}{2}], T} X_{[t + \frac{1-h}{2}], T} e^{-i2\pi fh} \\ &= \frac{1}{T} \sum_{t=1}^{T} J_T(\frac{t}{T}, f) \end{aligned}$$

Returning to the Whittle likelihood approximation to exact Gaussian likelihood, for the locally stationary process, we will replace $S_X(f)$, the spectral density, with the time varying spectral density $S_X(\frac{t}{T}, f)$. We will then replace the periodogram $\hat{S}_X^{(p)}(f)$ with the pre-periodogram $\tilde{J}_T(\frac{t}{T}, f)$ to obtain the generalized Whittle likelihood:

$$l_T^{GW} := \frac{1}{T}\sum_{t=1}^{T}\int_{-1/2}^{1/2}\left\{\log(2\pi S_{X,\theta}(\frac{t}{T},f)) + \frac{\tilde{J}_T(\frac{t}{T},f)}{S_{X,\theta}(\frac{t}{T},f)}\right\}df$$

More generally, for the multivariate case where we relax the assumption that $\mu = \mu_\theta = 0$ but still assume that Θ is compact, this approximation is:

$$l_T^{GW} := \frac{1}{T}\sum_{t=1}^{T}\int_{-1/2}^{1/2}\left\{\log\left((2\pi)^d \det(S_X(t/T,f))\right) + \mathrm{tr}\left[\frac{\tilde{J}_T(\frac{t}{T},f)}{S_X(\frac{t}{T},f)}\right]\right\}df$$

where

$$\tilde{J}_T(u,f)_{a,b} = \sum_{\substack{h \\ 1\le[uT+\frac{1+h}{2}],[uT+\frac{1-h}{2}]\le T}} \left\{X^a_{[uT+\frac{1+h}{2}],T} - \mu^a\left(\frac{[uT+\frac{1+h}{2}]}{T}\right)\right\} \times \left\{X^b_{[uT+\frac{1-h}{2}],T} - \mu^a\left(\frac{[uT+\frac{1-h}{2}]}{T}\right)\right\}\exp\{-i2\pi fh\}$$

At the heart of the classical likelihood approximation (i.e., in the stationary case) is a technical observation that uses the theory of Toeplitz matrices to approximate the inverse of the covariance matrix with the Toeplitz matrix of the inverse spectral density $S_X(f)$. What is driving the above generalization of the likelihood approximation for the locally stationary case is a corresponding generalization of the Toeplitz matrices theory. Since the approximation of the inverse of the covariance matrix is central to both the approximation and the asymptotic results discussed next, we will take a moment to investigate it further here.

For $\underset{\sim}{X} = (X_{1,T},\ldots,X_{T,T})^T$, we can express the covariance matrix $\Sigma_\theta = \Sigma_T(A_\theta,A_\theta)$ to be the $T\times T$ block matrices with corresponding (j,k) block for $j,k = 1,\ldots,T$:

$$\Sigma_T(A,A)_{j,k} = \int_{-1/2}^{1/2}\exp\{i2\pi f(j-k)\}A_{\theta,j,T}(f)A_{\theta,k,T}(-f)^T df$$

We also introduce the $U_t(\phi)$ the $T\times T$ block matrix with (j,k) block:

$$U_T(\phi)_{j,k} = \int_{-1/2}^{1/2}\exp\{i2\pi f(j-k)\}\phi\left(\frac{1}{T}\left\lfloor\frac{j+k}{2}\right\rfloor,f\right)df$$

where $\phi(u, f)$ is a $d \times d$ matrix and the integral is calculated entry-wise in the matrix. In a result analogous to the classical case, we see that an approximation of Σ_θ^{-1} can be done with $U_T(c \cdot S_X^{-1}(u, f))$.

Theorem 5. *(Proposition 2.4 in* [5]*) Suppose the matrices A and ϕ fulfill the smoothness conditions denoted in Assumptions A.3(i)-(iii) in the Appendix of* [5] *with existing and bounded derivatives $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial f}A(u, f)_{a,b}$ and eigenvalues of $\phi(u, f)$ which are bounded from below uniformly in u and f. Then for all $\epsilon > 0$,*

$$\frac{1}{T}|\Sigma_T(A, A)^{-1} - U_T(\{c A\overline{A}^T\}^{-1})|^2 = O(T^{-1+\epsilon})$$

and

$$\frac{1}{T}|U_T(\phi)^{-1} - U_T(\{c\phi\}^{-1})|^2 = O(T^{-1+\epsilon})$$

This approximation in terms allows a generalization of the Szegö identity (**supplement the Szegö identity for univariate stationary case**).

Theorem 6. *(Generalization of Szegö Identity to multivariate locally stationary processes, Proposition 2.5 in* [5]*) Suppose A fulfills Assumptions A.3(i) in* [5]*, with bounded derivatives $\frac{\partial^2}{\partial u^2}\frac{\partial}{\partial f}A(u, f)_{a,b}$. Then for $S_X(u, f) = A(u, f)A(u, -f)^T$ for each $\epsilon > 0$*

$$\frac{1}{T}\log\det\Sigma_T(A, A) = \int_0^1 \int_{-1/2}^{1/2} \log[(2\pi)^d \det S_X(u, f)]df\,du + O(T^{-1+\epsilon})$$

If $A = A_\theta$ depends on a parameter θ and fulfills the smoothness conditions of Assumption 2.6(iii),(iv), then $O(T^{-1+\epsilon})$ is uniform in θ.

For the following discussion, we make a series of technical assumptions regarding regularity conditions on the true mean function vector μ, the transfer function matrix A, and the covariance matrix Σ_T; as well as on the corresponding model functions μ_θ, A_θ, and Σ_θ. We let θ_0 denote the true parameter and state following result. Returning for a moment to the univariate stationary case (under a number of assumptions), we know that the estimate, $\hat{\theta}$, found by minimizing the $\tilde{l}_T$ is asymptotically equivalent to MLE estimate. Happily, a corresponding result is possible here. For the following result, we let $\hat{\theta}^{GW} = \arg\min_{\theta\in\Theta} l_T^{GW}$ and $\tilde{\theta}_T$ denote the MLE of θ.

Since Whittle likelihood is a reasonable approximation to the exact Gaussian likelihood, intuitively it is easy to imagine that the consistency and asymptotic normality results of estimation based on the generalized Whittle likelihood will follows from those results for exact Gaussian likelihood.

Theorem 7. *(Asymptotic Equivalence to MLE, Adapted from Theorem 3.1 in* [5]*) Suppose that Assumption 2.6 holds and that* $\mu = \mu_{\theta_0}$ *and* $A_{\theta_0}(u, f) = A(u, f)$ *for some* $\theta_0 \in \Theta$ *(i.e., the model is correctly specified), 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

$$\Gamma_{ij} = \frac{1}{2}\int_0^1 \int_{-1/2}^{1/2} tr\{S_{X,\theta_0}(\nabla_i S_{X,\theta_0}^{-1})S_{X,\theta_0}(\nabla_i S_{X,\theta_0}^{-1})\}df du + \int_0^1 (\nabla_i \mu_{\theta_0}(u))^T S_{X,\theta_0}^{-1}(u,0)(\nabla_j \mu_{\theta_0}(u))du$$

Where both estimates are asymptotically efficient.

4. Simulation Studies

In this section, we simulate realizations of a stationary $AR(1)$ process as well as for a time-varying $AR(1)$ process to explore the differences between the two. We consider the implications of naively using methods for stationary processes to fit the time varying $AR(1)$ process.

We first generate a realization from the stationary $AR(1)$ process ($n = 256$)

$$X_t = .8X_{t-1} + Z_t$$
$$Z_t \overset{iid}{\sim} N(0, 1)$$

Next, we consider a simple form of a realization from a "time varying" $AR(1)$ model:

$$X_t = .9X_{t-1} + Z_t \ t \leq 128$$
$$X_t = -.5X_{t-1} + Z_t \ 128 < t \leq 256$$
$$Z_t \overset{iid}{\sim} N(0, 1)$$

Below, we plot the realizations, and we note that in this case we can see what appears to be a change in behavior for the locally stationary process at about $t = 128$.

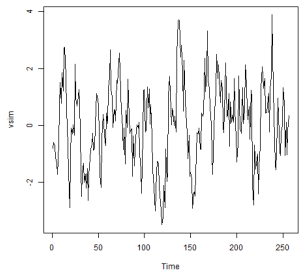

FIGURE 4.1. Stationary

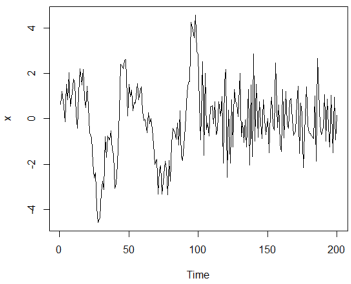

FIGURE 4.2. Time Varying

Now, let us consider if we assumed that both are stationary processes, and we fit an $AR(1)$. If we consider the residuals for the fitted models, we see that the residuals from the time-varying $AR(1)$ model show trends (particularly for $t < 128$) suggesting the model is a poor fit.

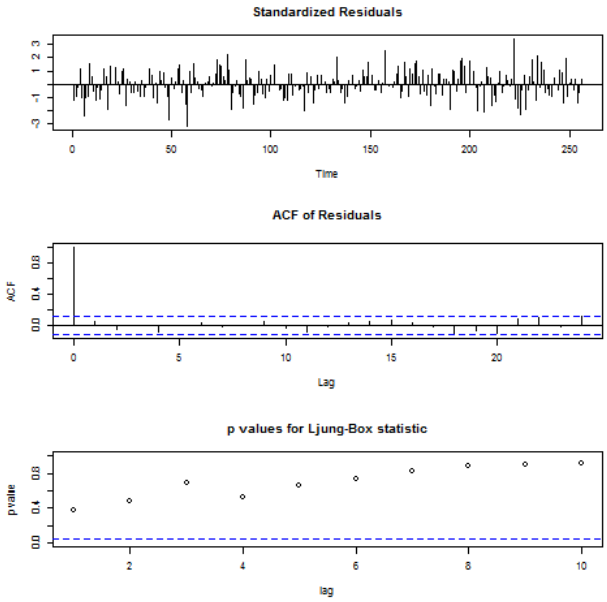

FIGURE 4.3. Stationary

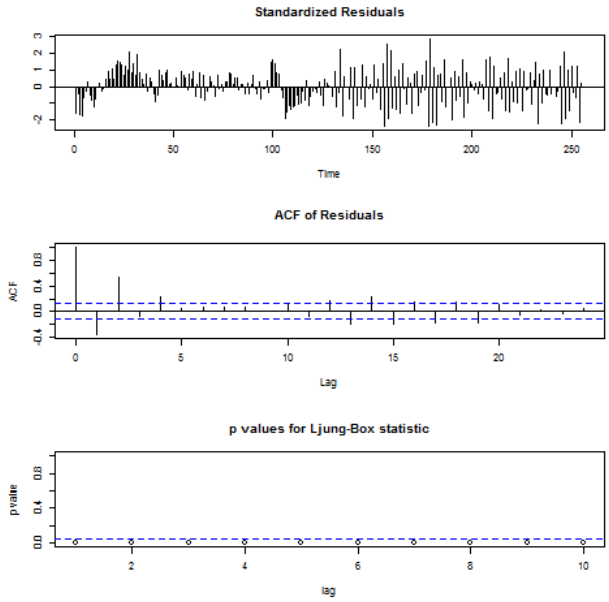

FIGURE 4.4. Time Varying

Now, we estimate the localized autocovariance using a wavelet filter for both realizations. We then plot the localized autocorrelation curves at lags 1,2, and 3 and compare them to the theoretical values we would expect from the fitted models. Clearly, for the time-varying models these observed and theoretical values are vastly different.

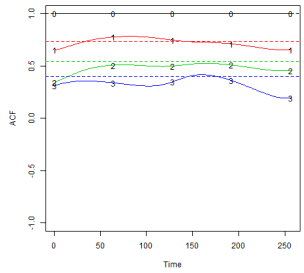

FIGURE 4.5. Stationary

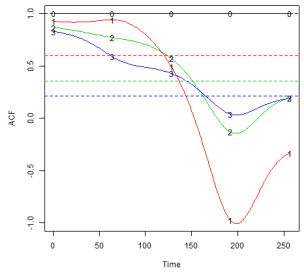

FIGURE 4.6. Time Varying

5. Discussion and Future Works

A more general theory can be built using Banach space theory by extending the idea of using stationary derivative processes approximating locally stationary process, the benefit is unification of the ergodicity behaviors (mainly consistency and central limit theorems) as mentioned. The cost is obvious that very limited possibilities of nonlinearity they can handle with certain complexity of the processes [10]. Personally I think the work in that direction is pretty much exhausted although some innovations can be done by relaxation to locally convex processes.

References

[1] Brockwell, Peter J., and Richard A. Davis. Time series: theory and methods. Springer Science & Business Media, 2013. 2

[2] Dahlhaus, Rainer. "Locally stationary processes." Handbook of statistics. Vol. 30. Elsevier, 2012. 351-413. 1, 2, 3

[3] Dahlhaus, Rainer. "Maximum Likelihood Estimation and Model Selection for Local Stationary Processes. " Nonparametric Statistics. 6 (1996): 171-191. 3

[4] Dahlhaus, Rainer. "Fitting time series models to nonstationary processes." The annals of Statistics 25.1 (1997): 1-37. 2

[5] Dahlhaus, Rainer. "A Likelihood Approximation for Locally Stationary Processes. " The annals of statistics. 38.6 (2000): 1762-1794. 3, 5, 6, 7

[6] 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. 2

[7] Cardinali, Alessandro, and Guy P. Nason. "Costationarity of locally stationary time series." Journal of Time Series Econometrics 2.2 (2010). 1, 2

[8] Paciorek, Christopher Joseph. Nonstationary Gaussian processes for regression and spatial modeling Diss. PhD thesis, Carnegie Mellon University, Pittsburgh, Pennsylvania, 2003. 1

[9] Ledoux, Michel, and Michel Talagrand. Probability in Banach Spaces: isoperimetry and processes. Springer Science & Business Media, 2013. 2

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