Nonparametric Regression for Locally Stationary Time Series [Vogt, 2012]

David Veitch

University of Toronto

March 31 2022

Agenda

Motivation

Theory of locally stationary processes one way to model nonstationarity. [Dahlhaus, 1996] considered time varying spectral representations.

$$X_{t,T} = \mu\left(\frac{t}{T}\right) + \int_{-\pi}^{\pi} \exp(i\lambda t) A_{t,T}^{0}(\lambda) d\xi(\lambda) \tag{1}$$

Others considered it within a parametric context (i.e. coefficients change smoothly over time). [Dahlhaus and Rao, 2006] studied ARCH models with time-varying coefficients.

$$X_{t,N} = \sigma_{t,N} Z_t, \quad Z_t \sim \text{i.i.d} \tag{2}$$

$$\sigma_{t,N}^2 = a_0\left(\frac{t}{N}\right) + \sum_{j=1}^{\infty} a_j\left(\frac{t}{N}\right) X_{t-j,N}^2 \tag{3}$$

This paper introduces nonparametric framework as natural extension to models with time-varying coefficients.

General Idea

$$Y_{t,T} = m\left(\frac{t}{T}, X_{t,T}\right) + \epsilon_{t,T} \qquad t = 1, \ldots, T \tag{4}$$

with

$$\mathbb{E}[\epsilon_{t,T}|X_{t,T}] = 0$$
$$Y_{t,T} \in \mathbb{R}$$
$$X_{t,T} \in \mathbb{R}^d.$$

Y, X, ϵ assumed to be locally stationary, and m is a function which is allowed to change smoothly over time.

General Idea - Example

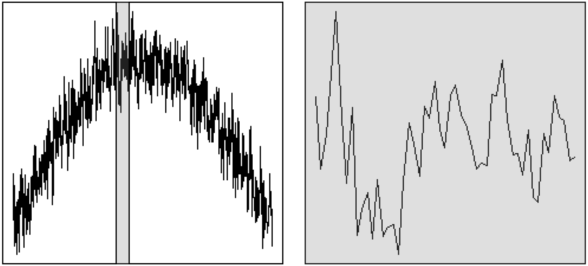


Figure: Estimate m by assuming small segments of $Y_{t,T}$ (LHS) strictly stationary (RHS $\approx$ strictly stationary).

Defining Local Stationarity

Intuition: Locally around each rescaled time point u, the process $X_{t,T}$ can be approximated by a stationary process.

Definition 2.1

The process $\{X_{t,T}\}$ is locally stationary if for each rescaled time point $u \in [0,1]$ there exists an associated process $\{X_t(u)\}$ with the following two properties

1. $\{X_t(u)\}$ is strictly stationary with density $f_{X_t(u)}$
2. $||X_{t,T} - X_t(u)|| \leq \left(\left|\frac{t}{T} - u\right| + \frac{1}{T}\right) U_{t,T}(u)$ a.s.

where $\{U_{t,T}(u)\}$ a process of positive variables with $\mathbb{E}[U_{t,T}(u)^\rho] < C$ for some $\rho > 0$, $C < \infty$ independent of u, t, T. $|| \cdot ||$ an arbitrary norm on $\mathbb{R}^d$.

Since ρ-th moments of $U_{t,T}(u)$ uniformly bounded, $U_{t,T}(u) = O_p(1)$ therefore as a result of Definition 2.1

$$||X_{t,T} - X_t(u)|| = O_p\left(\left|\frac{t}{T} - u\right| + \frac{1}{T}\right). \tag{5}$$

Author looks at three different classes of locally stationary models:

1. Locally stationary nonlinear AR models
2. Kernel estimation
3. Locally stationary additive models

Locally Stationary Nonlinear AR Models

$\{Y_{t,T} \mid t \in \mathbb{Z}\}_{T=1}^{\infty}$ is a time-varying nonlinear autoregressive process (tvNAR) if $Y_{t,T}$ evolves according to

$$Y_{t,T} = m\left(\frac{t}{T}, Y_{t-1,T}^{t-d}\right) + \sigma\left(\frac{t}{T}, Y_{t-1,T}^{t-d}\right)\epsilon_t \tag{6}$$

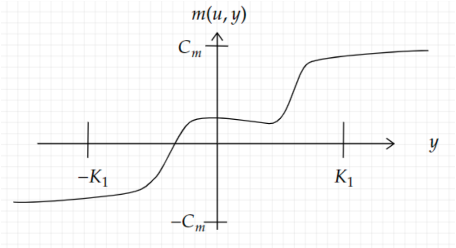
- $m(u, y), \sigma(u, y)$ smooth functions of rescaled time u
- $u \leq 0 \Rightarrow m(u, y) = m(0, y)$ and $\sigma(u, y) = \sigma(0, y)$. Similarly for $u \geq 1 \Rightarrow m(u, y) = m(1, y)$ and $\sigma(u, y) = \sigma(1, y)$.
- ϵ_t are iid mean zero

Then for all $u \in \mathbb{R}$ the associated process $\{Y_t(u) \mid t \in \mathbb{Z}\}$ is

$$Y_t(u) = m(u, Y_{t-1}^{t-d}(u)) + \sigma(u, Y_{t-1}^{t-d}(u))\epsilon_t \tag{7}$$

Some conditions sufficient to ensure tvNAR process locally statinoary and strongly mixing

- (M1) m absolutely bounded by constant $C_m < \infty$
- (M2) m Lipschitz continuous
  $|m(u, y) - m(u', y)| \leq L|u - u'| \; \forall \; y \in \mathbb{R}^d$ for some $L < \infty$
- (M3) m continuously differentiable with respect to y, and
  $\partial_j m(u, y) = \frac{\partial}{\partial y_j} m(u, y)$ have for some $K_1 < \infty$
  $\sup_{u \in \mathbb{R}, ||y||_\infty > K_1} |\partial_j m(u, y)| \leq \delta < 1$



tvNAR Process Assumptions - σ

- (Σ1) σ bounded by finite constant $C_\sigma < \infty$ from above and c_σ from below for all u, y
- (Σ2)σ Lipschitz continuous with respect to u
- (Σ3) σ continuously differentiable with respect to y, and for $\partial_j \sigma(u, y)$ we have for $K_1 < \infty$ we have $\sup_{u \in \mathbb{R}, ||y||_\infty > K_1} |\partial_j \sigma(u, y)| \leq \delta < 1$

tvNAR Process Assumptions - ϵ

- (E1) ϵ_t iid, centred $\mathbb{E}[|\epsilon_t|^{1+\eta}] < \infty$ for some positive η. ϵ has everywhere positive continuous density f_ϵ.
- (E2) f_ϵ bounded and Lipschitz $\exists\ L < \infty$ such that $f_\epsilon(z) - f_\epsilon(z')| \leq L|z - z'|\ \forall\ z, z' \in \mathbb{R}$
- (E3) d_0, d_1 constants $0 \leq d_0 \leq D_0 < \infty$, $|d_1 \leq D_1 < \infty$. f_ϵ satisfies

$$\int_{\mathbb{R}} |f_\epsilon([1 + d_0]z + d_1) - f_\epsilon(z)|dz \leq C_{D_0,D_1}(d_0 + |d_1|) \tag{8}$$

where $C_{D_0,D_1} < \infty$ only depend on bounds D_0, D_1

tvNAR Properties

Under assumptions listed we get

1. tvNAR locally stationary
2. tvNAR strongly mixing
3. Auxiliary process $\{Y_t(u)\}$ has densities that vary smoothly with u

Theorem 3.1

Let assumptions (M1)-(M3), (Σ1)-(Σ3), (E1) be fulfilled. Then

1. for each $u \in \mathbb{R}$ the process $\{Y_t(u),\ t \in \mathbb{Z}\}$ has a strictly stationary solution with ϵ_t independent of $Y_{t-k}(u)$ for $k > 0$
2. the variables $Y_{t-1}^{t-d}(u)$ have density $f_{Y_{t-1}^{t-d}(u)}$ with respect to Lebesgue measure
3. the variables $Y_{t-1,T}^{t-d}(u)$ have density $f_{Y_{t-1,T^{t-d}}(u)}$ with respect to Lebesgue measure

tvNAR Properties

Theorem 3.2

Let assumptions (M1)-(M3), (Σ1)-(Σ3), (E1) be fulfilled. Then

$$|Y_{t,T} - Y_t(u)| \leq \left(\left| \frac{t}{T} - u \right| + \frac{1}{T} \right) U_{t,T}(u) \quad a.s.$$

where variables $U_{t,T}(u)$ have property that $\mathbb{E}[(U_{t,T}(u))^\rho] < C$ for some $\rho > 0$ and $C < \infty$ independent of u, t, T.

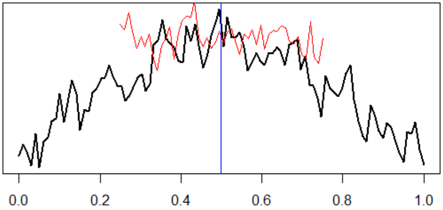


Figure: The difference between $Y_{t,T}$ (black) relative to $Y_t(u)$ increases as u moves away from 0.5 (blue).

tvNAR Properties

Theorem 3.1 + Theorem 3.2 $\Rightarrow$ tvNAR process $\{Y_{t,T}\}$ is locally stationary (Definition 2.1).

Theorem 3.3

Let $f(u, y) = f_{Y_{t-1}^{t-d}(u)}$ be the density of $Y_{t-1}^{t-d}(u)$ at $y \in \mathbb{R}^d$. If (M1)-(M3),(Σ1)-(Σ3), (E1), (E2) fulfilled then

$$|f(u, y) - f(v, y)| \leq C_y |u - v|^p$$

with some constant $0 < p < 1$ and $C_y < \infty$ continuously depending on y.

tvNAR Properties - Mixing Behaviour

We will characterize the mixing behaviour of the tvNAR process.

$$\text{Recall: Events } A, B \text{ Independent} \iff P(A \cap B) = P(A)P(B)$$
$$\iff P(A|B) = P(A)$$

Definition - α,β mixing array.

Let $(\Sigma, \mathcal{A}, P)$ be a probability space, and $\mathcal{B}, \mathcal{C}$ subfields of $\mathcal{A}$. Then

$$\alpha(\mathcal{B}, \mathcal{C}) = \sup_{B \in \mathcal{B}, C \in \mathcal{C}} |P(B \cap C) - P(B)P(C)|$$

$$\beta(\mathcal{B}, \mathcal{C}) = \mathbb{E}\left[\sup_{C \in \mathcal{C}} |P(C) - P(C|\mathcal{B})|\right]$$

$\alpha(k), \beta(k)$

For an array $\{Z_{t,T} \mid 1 \leq t \leq T\}$

$$\alpha(k) = \sup_{t,T} \sup_{1 \leq t \leq T-k} \alpha(\sigma(Z_{s,T}, 1 \leq s \leq t), \sigma(Z_{s,T}, t+k \leq s \leq T))$$

$$\beta(k) = \sup_{t,T} \sup_{1 \leq t \leq T-k} \alpha(\sigma(Z_{s,T}, 1 \leq s \leq t), \sigma(Z_{s,T}, t+k \leq s \leq T))$$

The array $\{Z_{t,T}\}$ is α mixing (strongly mixing) if $\alpha(k) \to 0$ as $k \to \infty$. Similarly β mixing if $\beta(k) \to 0$ as $k \to \infty$.

Theorem 3.4

If (M1)-(M3),(Σ1)-(Σ3), (E1)-(E3) fulfilled then the tvNAR process $\{Y_{t,T}\}$ is geometrically β mixing, that is, there exist positive constants $\gamma < 1$, $C < \infty$ such that $\beta(k) \leq C\gamma^k$.

tvNAR Properties - Mixing Behaviour

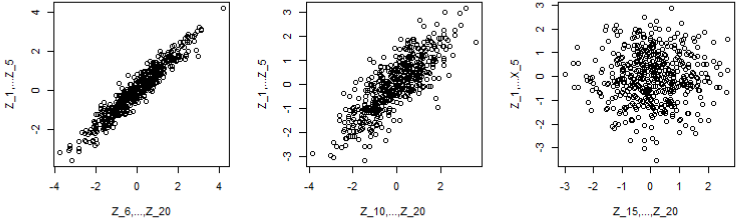


Figure: As k increases the values that $Z_1, \ldots, Z_t$ take effect the values $Z_{t+k}, \ldots, Z_T$ take less and less.

Theorem 3.2 Outline

1. Preliminaries
2. Triangle Inequality
3. Backward Iteration
4. Triangle Inequality Bound
5. Bounding Norm of Matrix Product

Different types of Y

$$Y_{t,T} = m\left(\frac{t}{T}, Y_{t-1,T}^{t-d}\right) + \sigma\left(\frac{t}{T}, Y_{t-1,T}^{t-d}\right)\epsilon_t \quad \text{Locally Stationary}$$

$$Y_t(u) = m\left(u, Y_{t-1,T}^{t-d}(u)\right) + \sigma\left(u, Y_{t-1,T}^{t-d}(u)\right)\epsilon_t \quad \text{Strictly Stationary}$$

$$\underline{Y}_{t,T} = Y_{t,T}^{t-d+1} = (Y_{t-d+1,T}, Y_{t-d+2,T}, \ldots, Y_{t-1,T}, Y_{t,T})$$

$$\underline{Y}_{t,T}(u) = Y_{t,T}^{t-d+1}(u) = (Y_{t-d+1,T}(u), Y_{t-d+2,T}(u), \ldots, Y_{t-1,T}(u), Y_{t,T}(u))$$

Linearization Δ Terms

By mean value theorem and Taylor's theorem

$$m(v, \underline{Y}_{t-1}(v)) - m(u, \underline{Y}_{t-1}(u)) = \Delta^m_{t,0} + \sum_{j=1}^{d} \Delta^m_{t,j}(Y_{t-j}(v) - Y_{t-j}(u))$$

$$\Delta^m_{t,0} = m(v, \underline{Y}_{t-1}(v)) - m(u, \underline{Y}_{t-1}(v))$$

$$\Delta^m_{t,j} = \Delta^m_j(u, \underline{Y}_{t-1}(u), \underline{Y}_{t-1}(v))$$

$$\Delta^m_j(u, y, y') = \int_0^1 \partial_j m(u, y + s(y' - y))ds$$

Matrix Notation

- $||\cdot||$ Euclidean norm for vectors
- Spectral norm for $d \times d$ matrices $||A||_2 = \max_{||x||=1} |Ax|$ =square root of max eigenvalue $A^T A$
- Matrices of these forms are used

$$B(z) = \begin{bmatrix} z & \cdots & z & z \\ 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{bmatrix} \tag{9}$$

Theorem 3.2 Outline - Triangle Inequality

$$|Y_{t,T} - Y_t(u)| \leq \left| Y_{t,T} - Y_t\left(\frac{t}{T}\right) \right| + \left| Y_t\left(\frac{t}{T}\right) - Y_t(u) \right|$$

Bounding both terms is similar, will focus on $\left| Y_t\left(\frac{t}{T}\right) - Y_t(u) \right|$.

Theorem 3.2 Outline - Backward Iteration

$$Y_t\left(\frac{t}{T}\right) - Y_t(u) = (\Delta^m_{t,0} + \Delta^\sigma_{t,0}\epsilon_t) + \sum_{j=1}^{d}(\Delta^m_{t,j} + \Delta^\sigma_{t,j}\epsilon_t)\left(Y_{t-j}\left(\frac{t}{T}\right) - Y_{t-j}(u)\right)$$

can rewrite in matrix form

$$\underline{Y}_t\left(\frac{t}{T}\right) - \underline{Y}_t(u) = A_t\left(\underline{Y}_{t-1}\left(\frac{t}{T}\right) - \underline{Y}_{t-1}(u)\right) + \underline{\xi}_t$$

$$A_t = \begin{bmatrix} \Delta t,1^m + \Delta^\sigma_{t,1}\epsilon_t & \cdots & \Delta t,d-1^m + \Delta^\sigma_{t,d-1}\epsilon_t & \Delta t,d^m + \Delta^\sigma_{t,d}\epsilon_t \\ 1 & & 0 & 0 \\ & \ddots & & \vdots \\ 0 & & 1 & 0 \end{bmatrix}$$

$$\underline{\xi}_t = (\Delta^m_{t,0} + \Delta^\sigma_{t,0}\epsilon_t, 0, \ldots, 0)^T$$

Iterate the aformentioned matrix form to get

$$\left|\left|\underline{Y}_t\left(\frac{t}{T}\right) - \underline{Y}_t(u)\right|\right| \leq ||\underline{\xi}_t|| + \left|\left|\sum_{r=0}^{n-1}\prod_{k=0}^{r} A_{t-k}\underline{\xi}_{t-r-1}\right|\right| + \left|\left|\prod_{k=0}^{n} A_{t-k}\left(\underline{Y}_{t-n-1}\left(\frac{t}{T}\right) - \underline{Y}_{t-n-1}(u)\right)\right|\right|$$

Replace A_t matrix with

$$B_t = (1 + |\epsilon_t|)B(\Delta_t)$$

where

$$\Delta_t = \Delta 1(||\underline{\epsilon}_{t-1}||_\infty \leq K_2) + \delta 1(||\underline{\epsilon}_{t-1}||_\infty > K_2)$$

$$\Delta \geq \sup_{u,y} |\partial_j m(u, y)|$$

$$|\Delta^m_{t,j} + \Delta^\sigma_{t,j}\epsilon_t| \leq \Delta_t(1 + |\epsilon_t|)$$

Theorem 3.2 Outline - Backward Iteration

End up with

$$R_{t,n} = C(1 + ||\underline{\epsilon}_{t-n-1}||) \left|\left|\prod_{k=0}^{r} B_{t-k}\right|\right|$$

$$\left|\left|\underline{Y}_t\left(\frac{t}{T}\right) - \underline{Y}_t(u)\right|\right| \leq \left|\frac{t}{T} - u\right| \left(C(1 + |\epsilon_t|) + \sum_{r=0}^{\infty} R_{t,r}\right)$$

$$= \left|\frac{t}{T} - u\right| V_t$$

Need to show ρ-th moment of $||\prod_{k=0}^{r} B_{t-k}||$ converges exponentially fast to 0 as $n \to \infty$. If we can then $\mathbb{E}[\sum_{r=0}^{\infty} R_{t,r}]$ can be controlled.

Since matrix norms are equivalent deal with $||\cdot||_1$ column sums, specifically

$$\mathcal{B}_n = \left|\left|\prod_{k=0}^{n} B_{t-k}\right|\right|_1 \tag{10}$$

General Strategy

- Split into two cases based on a normalized sum of the lag d minimum values of ϵ
- Split product of B matricies into $B(\delta)$ (smaller norm),$B(\Delta)$ (larger norm)
- Choose δ very carefully

Theorem 3.2 Outline - Bounding Norm of Matrix Product

For example

$$\begin{bmatrix} \delta & \delta & \delta \\ 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}^3 = \begin{bmatrix} \delta^3 + 2\delta^2 + \delta & \delta^3 + 2\delta^2 & \delta^3 + \delta^2 \\ \delta^2 + \delta & 2\delta^2 + \delta & \delta^2 + \delta \\ \delta & \delta & \delta \end{bmatrix} \tag{11}$$

Therefore $||B(\delta)^d||_1 \leq C_d \delta$. Then choose a very specific δ equal to

$$\delta < [(1 + \mathbb{E}[|\epsilon_0|])^{d/(1-\kappa)} (\Delta + 1)^{\kappa d/(1-\kappa)} C_d]^{-1} \tag{12}$$

Which is not too restrictive (reminding ourselves of how δ related to m

$$\sup_{u \in \mathbb{R}, ||y||_\infty > K_1} |\partial_j m(u, y)| \leq \delta < 1 \quad (K_1 < \infty) \tag{13}$$

So as long as for some large K_1 the m function has a very small derivative, a small δ is fine.

Kernel Estimation

Here consider kernel estimation for the general model

$$Y_{t,T} = m\left(\frac{t}{T}, X_{t,T}\right) + \epsilon_{t,T} \tag{14}$$

m identified (theoretically possible to learn true values after getting an infinite number of observations) almost surely on $u \in [0,1]$.

Focus on Nadaraya-Watson (NW) estimation $\approx$ locally weighted averages.

$$\begin{aligned}
\hat{m}(u,x) &= \frac{\sum_{t=1}^{T} K_h(u - t/T) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j) Y_{t,T}}{\sum_{t=1}^{T} K_h(u - t/T) \prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)} \\
X_{t,T} &= (X_{t,T}^1, \ldots, X_{t,T}^d) \\
x &= (x^1, \ldots, x^d), \quad x \in \mathbb{R}^d
\end{aligned}$$

where K a one-dimensional kernel function

Kernel Estimation

Assumption (K2)

The array $\{X_{t,T}, Y_{t,T}$ is α-mixing. The mixing coefficients α have the property that for some $A < \infty$ and $\beta > \frac{2s-2}{s-2}$

$$\alpha(k) \leq Ak^{-\beta}$$

Thorem 4.1

Assume (K1)-(K3), C(6) and $\beta > \frac{2+s(1+(d+1))}{s-2}, \frac{\phi_T \log T}{T^\theta} h^{d+1} = o(1)$, $\theta = \frac{\beta(1-2/s)-2/s-1-(d+1)}{\beta+3-(d+1)}$ where ϕ_T slowly diverging (e.g. $\log \log T$). Finally let S be a compact subset of $\mathbb{R}^d$, and ψ the numerator of the NW estimator. Then it holds

$$\sup_{u \in [0,1],\, x \in S} \left| \hat{\psi}(u,x) - \mathbb{E}[\hat{\psi}(u,x)] \right| = O_p\left(\sqrt{\frac{\log T}{Th^{d+1}}} \right)$$

Theorem 4.2

Assume (C1)-(C6) hold and (K1)-(K3) fulfilled for both $Y_{t,T} = 1$ and $Y_{t,T} = \epsilon_{t,T}$. Let β, θ as in Theorem 4.1 and suppose $\inf_{u \in [0,1], x \in S} f(u,x) > 0$. Moreover, assume bandwidth h satisfies

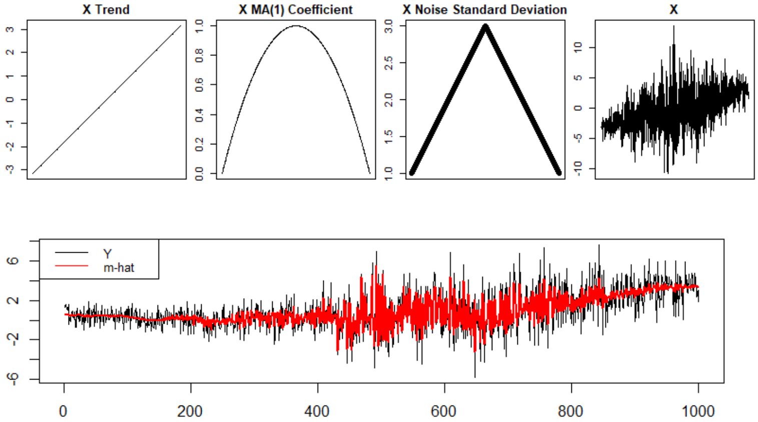
$$\frac{\phi_T \log T}{T^\theta h^{d+1}} = o(1)$$
$$\frac{1}{T^r h^{d+r}} = o(1).$$

Let $\phi_T = \log \log T$, $r = \min(\rho, 1)$ (ρ in (C1)). Let $I_h = [C_1 h, 1 - C_1 h]$. Then

$$\sup_{u \in I_h, x \in S} |\hat{m}(u,x) - m(u,x)| = O_p\left(\sqrt{\frac{\log T}{Th^{d+1}}} + \frac{1}{T^r h^d} + h^2\right)$$

Kernel Estimation - Simulation

$$Y_t = (1-t)\sin(X_t^3) + tX_t + \epsilon_t \quad \epsilon_t \sim U[0,1]$$
$$X_t = \theta X_{t-1} + \eta_t \quad \eta_t \sim \mathcal{N}(0, \sigma_t)$$



Theorem 4.1 Proof - Outline

Theorem 4.1 techniques used also in Theorem 4.2 proof.

- Preliminaries
- Truncation
- $\hat{\psi}_2$
- $\hat{\psi}_1$
- $\hat{\psi}_1$ Bound

Theorem 4.1 Proof - Preliminaries

$$\hat{\psi}(u,x) = \frac{Th^{d+1}}{\sum}\sum_{t=1}^{T} K_h\left(u - \frac{t}{T}\right)\prod_{j=1}^{d} K_h(x^j - X_{t,T}^j)W_{t,T}$$

- $W_{t,T}$ one dimensional random variables with $\mathbb{E}[|W_{t,T}|^s] \leq C > \infty$ for some $s > 2$
- $\{X_{t,T}, W_{t,T}\}$ is $\alpha-$mixing.
- For any compact set $S \subseteq \mathbb{R}^d$ where $f_{X_{t,T}}$ the density of $X_{t,T}$ we have

$$\sup_{t,T}\sup_{x\in S} \mathbb{E}[|W_{t,T}|^s|X_{t,T} = x]f_{X_{t,T}} \leq C$$

Also define

$$B = \{(u,x) \in \mathbb{R}^{d+1} \mid u \in [0,1], x \in S\}$$

$$\tau_T = \rho_T T^{1/s} \quad \rho \text{ slowly diverging}$$

Theorem 4.1 Proof - Truncation

$$\hat{\psi}(u,x) - \mathbb{E}[\hat{\psi}(u,x)] = (\hat{\psi}_1(u,x) - \mathbb{E}[\hat{\psi}_1(u,x)]) + (\hat{\psi}_2(u,x) - \mathbb{E}[\hat{\psi}_2(u,x)])$$

where

$$\hat{\psi}_1 = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\left(u - \frac{t}{T}\right) \prod_{j=1}^{d} K_h(x^j - X^j_{t,T}) W_{t,T} I(|W_{t,T}| \leq \tau_T)$$

$$\hat{\psi}_2 = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\left(u - \frac{t}{T}\right) \prod_{j=1}^{d} K_h(x^j - X^j_{t,T}) W_{t,T} I(|W_{t,T}| > \tau_T)$$

Theorem 4.1 Proof - $\hat{\psi}_2$

First let $a_T = \sqrt{\log T / Th^{d+1}}$ (the bound is $O_p(a_T)$)

$$P\left(\sup_{(u,x)\in B} |\hat{\psi}_2(u,x)| > Ca_T\right) \leq P(|W_{t,T} > \tau_T \text{ for some } 1 \leq t \leq T)$$
$$\leq \underbrace{\frac{\sum_{t=1}^{T} \mathbb{E}[|W_{t,T}|^s]}{\tau_T^{-s}}}_{\text{Chebyshev}} \leq CT\tau_T^{-s} = \rho_T^{-s} \to 0$$

Next using law of total expectation

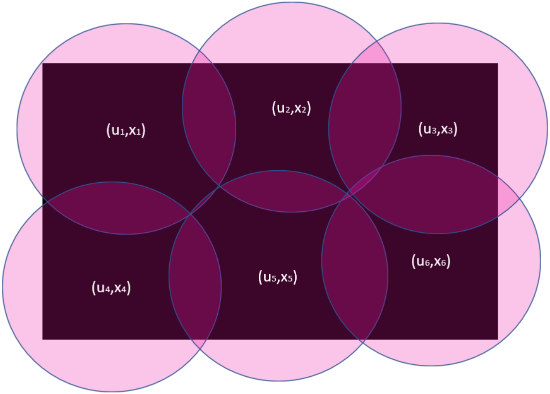
$$\mathbb{E}[|\hat{\psi}_2(u,s)] \leq \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h\left(u - \frac{t}{T}\right) \int_{\mathbb{R}^d} \prod_{j=1}^{d} K_h(x^j - w^j) \tag{15}$$
$$\times \ \mathbb{E}[|W_{t,T}|I(|W_{t,T}| > \tau_T)|X_{t,T} = w]f_{X_{t,T}}(w)dw \tag{16}$$
$$\leq \cdots \leq Ca_T \tag{17}$$

Therefore $\sup_{(u,x)\in B} |\hat{\psi}_2(u,x) - \mathbb{E}[\hat{\psi}_2(u,x)| = O_p(a_T)$

Cover the region $B = \{(u,x) \in \mathbb{R}^{d+1} \mid u \in [0,1], x \in S\}$ with $N \leq Ch^{-(d+1)} a_T^{-(d+1)}$ balls

$$B_n = \{(u,x) \in \mathbb{R}^{d+1} \mid ||(u,x) - (u_n, x_n)||_\infty \leq a_T h\} \tag{18}$$

midpoints of balls (u_n, x_n).



Want to find the difference between $\hat{\psi}_1(u,x)$ and $\hat{\psi}(u_n, x_n)$. Introduce new kernel

$$K^*(v) = C \prod_{j=0}^{d} I(|v^j| \leq 2C_1) \quad v \in \mathbb{R}^d$$

then for $(u,x) \in B_n$, $\quad T$ large

$$\left| K_h\left(u - \frac{t}{T}\right) \prod_{j=1}^{d} K_h(x^j - X^j_{t,T}) - K_h\left(u_n - \frac{t}{T}\right) \prod_{j=1}^{d} K_h(x^j_n - X^j_{t,T}) \right| \leq a_T K_h^*\left(u_n - \frac{t}{T}, x_n - X_{t,T}\right)$$

Then investigate a modified $\hat{\psi}_1$

$$\tilde{\psi}_1 = \frac{1}{Th^{d+1}} \sum_{t=1}^{T} K_h^*\left(u - \frac{t}{T}, x - X_{t,T}\right) |W_{t,T}| I(|W_{t,T} \leq \tau_T) \qquad (19)$$

Theorem 4.2 Proof - $\hat{\psi}_1$

Then analyze how $\tilde{\psi}_1$ differs from $\hat{\psi}_1$

$$\sup_{(u,x)\in B_n} \left|\hat{\psi}_1(u,x) - \mathbb{E}[\hat{\psi}_1(u,x)]\right|$$
$$\leq \left|\hat{\psi}_1(u_n,x_n) - \mathbb{E}[\hat{\psi}_1(u_n,x_n)]\right| + \left|\tilde{\psi}_1(u_n,x_n) - \mathbb{E}[\tilde{\psi}_1(u_n,x_n)]\right| + 2Ma_T$$

where M a finite constant.

Ideally want the bound to be $O_p(a_T)$. Can then investigate

$$\hat{Q}_T = N \max_{1\leq n\leq N} P(|\hat{\psi}_1(u_n,x_n) - \mathbb{E}[\hat{\psi}_1(u_n,x_n)]| > Ma_T)$$
$$\tilde{Q}_T = N \max_{1\leq n\leq N} P(|\tilde{\psi}_1(u_n,x_n) - \mathbb{E}[\tilde{\psi}_1(u_n,x_n)]| > Ma_T)$$

Theorem 4.1 Proof - $\hat{\psi}_1$ Bound

We can bound $\hat{Q}_T$ and $\tilde{Q}_T$ (which in this case is rewritten $\approx |\sum Z_{t,T}|$) with the simple bound

Locally Stationary Additive Models

Assume that regression function can be split up into additive components. For $x \in [0,1]^d$ have

$$\mathbb{E}[Y_{T,t} | X_{t,T} = x] = m_0\left(\frac{t}{T}\right) + \sum_{j=1}^{d} m_j\left(\frac{t}{T}, x^j\right)$$

Condition imposed that $\int m_j(u, x^j) p_j(u, x^j) dx^j = 0$ for all j, u where $p(u, x)$ the density of the strictly stationary process $\{X_t(u)\}$.

Locally Stationary Additive Models - Estimation

Utilize a smooth backfitting technique

$$\hat{p}(u,x) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\left(u, \frac{t}{T}\right) \prod_{j=1}^{d} K_h(x^j, X^j_{t,T})$$

$$\hat{m}(u,x) = \frac{1}{T_{[0,1]^d}} \sum_{t=1}^{T} I(X_{t,T} \in [0,1]^d) K_h\left(u, \frac{t}{T}\right) \prod_{j=1}^{d} K_h(x^j, X^j_{t,T}) Y_{t,T} / \hat{p}(u,x)$$

To determine minimizers for each u, $\tilde{m}_0(u), \tilde{m}_1(u,\cdot), \ldots, \tilde{m}_d(u,\cdot)$ minimizing

$$\int \left(\hat{m}(u,w) - g_0 - \sum_{j=1}^{d} g_j(w^j) \right)^2 \hat{p}(u,w) dw$$

Becomes $\approx$ weighted least squares problem.

Locally Stationary Additive Models - Result

Theorem 5.1

Let $I_h = [2C_1 h, 1 - 2C_1 h]$. Then under (Add1) and (Add2)

$$\sup_{u, x^j \in I_h} |\tilde{m}_j(u, x^j) - m_j(u, x^j)| = O_p\left(\sqrt{\frac{\log T}{Th^2}} + h^2\right)$$

Extensions

- Bandwidth selection (plug-in methods)
- Forecasting
  - Previous theorems valid for $u \in [Ch, 1 - Ch]$. Ideally get in $(1 - Ch, 1]$
  - Boundary-corrected kernels/one-sided kernels

R. Dahlhaus. On the kullback-leibler information divergence of locally stationary processes. *Stochastic Processes and their Applications*, 62 (1):139–168, 1996. ISSN 0304-4149. doi: https://doi.org/10.1016/0304-4149(95)00090-9. URL https://www.sciencedirect.com/science/article/pii/0304414995000909.

Rainer Dahlhaus and Suhasini Subba Rao. Statistical inference for time-varying ARCH processes. *The Annals of Statistics*, 34(3):1075 – 1114, 2006. doi: 10.1214/009053606000000227. URL https://doi.org/10.1214/009053606000000227.

Eckhard Liebscher. Strong convergence of sums of -mixing random variables with applications to density estimation. *Stochastic Processes and their Applications*, 65(1):69–80, 1996. ISSN 0304-4149. doi: https://doi.org/10.1016/S0304-4149(96)00096-8. URL https://www.sciencedirect.com/science/article/pii/S0304414996000968.

Michael Vogt. Nonparametric regression for locally stationary time series. *The Annals of Statistics*, 40(5):2601–2633, 2012.