Nonparametric methods for dependent data: the example of the Stochastic Volatility model.

Fabienne Comte[1]

Joint works with C. Lacour[2]**, V. Genon-Catalot**[1] **and Y. Rozenholc**[1]**.**

[1] MAP5, UMR 8145, Université Paris Descartes.

[2] Laboratoire de Probabilités et Statistique, Université Paris Sud-Orsay.

Continuous time model

$$\begin{cases} d\log(S_t) = \sqrt{V_t}dW_t, \\ dV_t = b_c(V_t)dt + \sigma_c(V_t)dB_t \end{cases}$$

where (W_t, B_t) is a 2-dim standard Brownian Motion,

and V_t is a positive diffusion process.

Observations: $(Z_{i\delta})_{1\leq i\leq n}$ for $Z_t = \log(S_t)$ and $k\delta = \Delta$, $n = kN$.

Assumptions: Diffusion in stationary regime.

but **non independent underlying sequence** $\Rightarrow$ geometrically β-mixing r.v.'s

Aim: Estimate b_c (and σ_c^2), without observing V but only $Z_t = \log(S_t)$ and provide **risk bounds**.

Ideas of the estimation strategy:

1) The **realized quadratic variation** associated with $(Z_{\ell\delta})_{ik+1\leq\ell<(i+1)k}$:

$$\hat{\bar{V}}_i = \frac{1}{k\delta}\sum_{j=0}^{k-1}\left(Z_{(ik+j+1)\delta} - Z_{(ik+j)\delta}\right)^2.$$

provides an approximation of the integrated volatility ($\Delta = k\delta$)

$$\bar{V}_i = \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} V_s ds. \tag{1}$$

2) If one observes (Y_i, X_i) with $\mathbf{Y_i = f(X_i)} + \varepsilon_\mathbf{i}$ where $\varepsilon_i =$ noise, then **nonparametric mean square contrasts** $\to$ good estimation of f.

Find the regression equation.

Suppose we observe directly the $(V_{i\Delta})$, then, we can write:

$$\frac{V_{(i+1)\Delta} - V_{i\Delta}}{\Delta} = \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} dV_s = \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} b_c(V_s)ds + \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} \sigma_c(V_s)dW_s$$

$$= \mathbf{b_c(V_{i\Delta})} + \underbrace{\frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} \sigma_c(V_s)dW_s}_{\textbf{noise}} + \underbrace{\frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} [b_c(V_s) - b_c(V_{i\Delta})]ds}_{\textbf{Residual term}} .$$

This regression of the $\frac{V_{(i+1)\Delta} - V_{i\Delta}}{\Delta}$ on the $V_{i\Delta}$ allows to estimate b_c (see Comte *et al.* (2007)).

Mixing sequences – Martingale properties – Δ, δ small.

Suppose we observe the $(\bar{V}_i)$

$$\bar{V}_i = \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} V_s ds,$$

then, we can write

$$\begin{aligned}\bar{\mathbf{V}}_{\mathbf{i}} &= \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta} V_s ds = \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}\left(V_{i\Delta} + \int_{i\Delta}^{s} dV_u\right) ds \\ &= \mathbf{V}_{i\Delta} + \frac{1}{\Delta}\int_{i\Delta}^{(i+1)\Delta}[(i+1)\Delta - u]dV_u.\end{aligned}$$

So we have

$$\begin{aligned}\frac{\bar{\mathbf{V}}_{\mathbf{i+1}} - \bar{\mathbf{V}}_{\mathbf{i}}}{\boldsymbol{\Delta}} &= \frac{\mathbf{V}_{\mathbf{(i+1)\Delta}} - \mathbf{V}_{\mathbf{i\Delta}}}{\boldsymbol{\Delta}} + \frac{1}{\Delta^2}\left[\int_{(i+1)\Delta}^{(i+2)\Delta}((i+2)\Delta - u)dV_u\right. \\ &\qquad \left. + \int_{i\Delta}^{(i+1)\Delta}(u - (i+1)\Delta)dV_u\right].\end{aligned}$$

$$\psi_{i\Delta}(u) = (u - i\Delta)\mathbb{I}_{[i\Delta,(i+1)\Delta[}(u) + [(i+2)\Delta - u]\mathbb{I}_{[(i+1)\Delta,(i+2)\Delta[}(u)$$

$$\frac{\bar{\mathbf{V}}_{\mathbf{i+1}} - \bar{\mathbf{V}}_{\mathbf{i}}}{\boldsymbol{\Delta}} = \mathbf{b}(\mathbf{V}_{\mathbf{i}\boldsymbol{\Delta}}) + \underbrace{\frac{1}{\Delta^2}\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(u)\sigma_c(V_u)dW_u}_{\text{noise}}$$
$$+ \underbrace{\frac{1}{\Delta^2}\int_{i\Delta}^{(i+2)\Delta} \psi_{i\Delta}(u)[b_c(V_u) - b_c(V_{i\Delta})]du}_{\text{residual}} .$$

Recall now

$$\hat{\bar{V}}_i = \frac{1}{k\delta}\sum_{j=0}^{k-1}\left(Z_{(ik+j+1)\delta} - Z_{(ik+j)\delta}\right)^2$$

is an **approximation of** $\bar{V}_i$ ($\Delta = k\delta$).

Last step: quadratic variations ($\hat{\bar{V}}_i$) built **using our effective observations** ($k\delta = \Delta$):

$$\hat{\bar{V}}_i = \bar{V}_i + u_{i,k},$$

where

$$u_{i,k} = \frac{1}{\Delta} \sum_{j=0}^{k-1} \left[\left(\int_{(ik+j)\delta}^{(ik+j+1)\delta} \sqrt{V_s} dB_s \right)^2 - \int_{(ik+j)\delta}^{(ik+j+1)\delta} V_s ds \right].$$

This yields

$$\mathbf{H_i} = \frac{\hat{\bar{\mathbf{V}}}_{\mathbf{i+1}} - \hat{\bar{\mathbf{V}}}_{\mathbf{i}}}{\boldsymbol{\Delta}} = \frac{\bar{V}_{i+1} - \bar{V}_i}{\Delta} + \frac{u_{i+1,k} - u_{i,k}}{\Delta}.$$

Finally, we obtain the development,

$$\mathbf{H_{i+1} = b_c(\hat{\tilde{V}}_i) + Z_{i+1} + R(i+1)}, \tag{2}$$

where Z_{i+1} is a noise term (with **martingale properties**):

$$Z_{i+1} = \frac{1}{\Delta^2} \int_{(i+1)\Delta}^{(i+3)\Delta} \psi_{(i+1)\Delta}(u) \sigma_c(V_u) dW_u + (u_{i+2,k} - u_{i+1,k})/\Delta,$$

and $R(i+1)$ is a sum of **negligible residual terms** given by

$$R(i+1) = [b_c(V_{(i+1)\Delta}) - b_c(\hat{\tilde{V}}_i)] + \frac{1}{\Delta^2} \int_{(i+1)\Delta}^{(i+3)\Delta} \psi_{(i+1)\Delta}(s)(b_c(V_s) - b_c(V_{(i+1)\Delta}))ds.$$

The lag in (2) is to avoid some cumbersome correlations.

Spaces of approximation

b_c is estimated only on a compact subset A of the state space of (V_t). For simplicity

$$A = [0,1], \text{ and we set } \quad b_A = b_c 1_A. \tag{3}$$

Estimation strategy (model selection):

1) Take a family $S_m, m \in \mathcal{M}_n$ of finite dim. subspaces of $\mathbb{L}_2([0,1])$

2) Compute a collection of estimators $\hat{b}_m$ where for all m, $\hat{b}_m \in S_m$.

3) Data driven procedure chooses among the collection of estimators the final estimator $\hat{b}_{\hat{m}}$.

Here : **Trigonometric spaces**, $S_m, m \in \mathcal{M}_n$.

$S_m = \text{Span}(\varphi_1, \ldots, \varphi_{2m+1}) \subset \mathbb{L}_2([0,1])$ with

$\varphi_1(x) = 1_{[0,1]}(x)$,
$\varphi_j(x) = \sqrt{2}\cos(\mathbf{2\pi jx})1_{[0,1]}(x)$ for even j's
$\varphi_j(x) = \sqrt{2}\sin(\mathbf{2\pi jx})1_{[0,1]}(x)$ for odd j's, $j > 1$.

Dimension $D_m = 2m + 1 = \dim(S_m) \leq \mathcal{D}_n$ and
$\mathcal{M}_n = \{1, 3, \ldots, \mathcal{D}_n\}$.

Largest space in the collection has maximal dimension $\mathcal{D}_n$.

For all $x \in [0,1]$, $\sum_{j=1}^{2m+1} \varphi_j^2(x) = 2m + 1 = D_m$.

Thus, for any function $t \in S_m$, $\sup_{x \in [0,1]} |t(x)|^2 \leq D_m \int_0^1 t^2(x)dx$.

For each m, and for a function $t \in S_m$, we introduce the following contrast:

$$\gamma_{\mathbf{N}}(\mathbf{t}) = \frac{\mathbf{1}}{\mathbf{N}} \sum_{\mathbf{i=0}}^{\mathbf{N-1}} [\mathbf{H_{i+1}} - \mathbf{t}(\hat{\bar{\mathbf{V}}}_{\mathbf{i}})]^{\mathbf{2}}. \tag{4}$$

Then the mean squares estimators are defined as

$$\hat{\mathbf{b}}_{\mathbf{m}} = \arg\min_{\mathbf{t} \in \mathbf{S_m}} \gamma_{\mathbf{N}}(\mathbf{t}). \tag{5}$$

$$\hat{\bar{V}}_i = \frac{1}{k\delta} \sum_{j=0}^{k-1} \left(Z_{(ik+j+1)\delta} - Z_{(ik+j)\delta}\right)^2, \quad H_i = \frac{\hat{\bar{V}}_{i+1} - \hat{\bar{V}}_i}{\Delta}.$$

$$\text{Observations } Z_{\ell\delta} \ \text{ from } \begin{cases} dZ_t = d\log(S_t) = \sqrt{V_t}dW_t, \\ dV_t = b_c(V_t)dt + \sigma_c(V_t)dB_t \end{cases}$$

Well defined, the vector: $(\hat{b}_m(\hat{\hat{V}}_1), \ldots, b_m(\hat{\hat{V}}_N))$ and

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=0}^{N-1}(\hat{b}_m(\hat{\hat{V}}_i) - b_A(\hat{\hat{V}}_i))^2\right].$$

Thus, the error is measured via the risk $\mathbb{E}(\|\hat{b}_m - b_A\|_N^2)$ where

$$\|t\|_N^2 = \frac{1}{N}\sum_{i=0}^{N-1} t^2(\hat{\hat{V}}_i). \tag{6}$$

Assumptions.

Assume that the state space of (V_t) is a known open interval (r_0, r_1) of $\mathbb{R}^+$, $I = [r_0, r_1] \cap \mathbb{R}$ and

[A1] $0 \leq r_0 < r_1 \leq +\infty$, $\overset{\circ}{I} = (r_0, r_1)$, with $\boldsymbol{\sigma_c(v) > 0}$, for all $v \in \overset{\circ}{I}$.
$b_c \in C^1(I)$, b'_c bounded on I,
$\sigma_c^2 \in C^2(I)$, $(\sigma_c^2)'\sigma$ Lipschitz on I, $(\sigma_c^2)''$ bounded on I and
$\sigma_c^2(v) \leq \sigma_1^2, \forall v \in I$.

[A2] $\forall v_0, v \in \overset{\circ}{I}$, **scale density** $s(v) = \exp\left[-2\int_{v_0}^{v} b_c(u)/\sigma_c^2(u)du\right]$
satisfies $\int_{r_0} s(x)dx = +\infty = \int^{r_1} s(x)dx$; **speed density**
$m(v) = 1/(\sigma_c^2(v)s(v))$ satisfies $\int_{r_0}^{r_1} m(v)dv = M < +\infty$.

[A3] $\eta \sim \pi$, $\forall i, \mathbb{E}(\eta^i) < \infty$, where $\pi(v)dv = (m(v)/M)\mathbb{1}_{(r_0,r_1)}(v)dv$.

Under **[A1]-[A3]**, (V_t) **is strictly stationary with marginal distribution** π, ergodic and β**-mixing**, *i.e.* $\lim_{t\to+\infty} \beta_V(t) = 0$.

To prove our main result, we need the following **stronger mixing condition**:

[A4] The process (V_t) is **exponentially β-mixing,** *i.e.*, there exist constants $K > 0, \theta > 0$, such that, for all $t \geq 0$, $\beta_V(t) \leq Ke^{-\theta t}$.

[A4] satisfied in most standard examples.

Under [A1]-[A4], for fixed Δ, $(\bar{V}_i)_{i\geq 0}$ is a strictly stationary process. And we have:

Proposition 1 *Under* [A1]-[A4], *for fixed k and δ, $(\hat{\bar{V}}_i)_{i\geq 0}$ is strictly stationary and* $\beta_{\hat{\bar{\mathbf{V}}}}(\mathbf{i}) \leq \mathbf{c}\beta_{\mathbf{V}}(\mathbf{i\Delta})$ *for all $i \geq 1$.*

[A5] The process $(\hat{\tilde{V}}_i)_{i\geq 0}$ **admits a stationary density** π^* and there exist two positive constants π_0^* and π_1^* (independent of n, δ) such that $\forall m \in \mathcal{M}_n$, $\forall t \in S_m$,

$$\pi_0^* \|t\|^2 \leq \mathbb{E}(t^2(\hat{\tilde{V}}_0)) \leq \pi_1^* \|t\|^2. \tag{7}$$

The existence of the density π^* is easy to obtain.

The checking of (7) is more technical.

$$\|t\|_{\pi^*}^2 = \int t^2(x)\pi^*(x)dx, \quad \|t\|^2 = \int_0^1 t^2(x)dx \quad \text{and} \quad \|t\|_\infty = \sup_{x\in[0,1]} |t(x)|.$$

For a deterministic function $\mathbb{E}(\|\mathbf{t}\|_{\mathbf{N}}^{\mathbf{2}}) = \|\mathbf{t}\|_{\pi^*}^{\mathbf{2}} = \int \mathbf{t}^{\mathbf{2}}(\mathbf{x})\pi^*(\mathbf{x})\mathbf{dx}$.

Under [A5], norms $\|.\|$ **and** $\|.\|_{\pi^*}$ **are equivalent** for functions in S_m

Proposition 2 *Assume that* $N\Delta \geq 1$ *and* $1/k \leq \Delta$. *Assume that* [A1]-[A5] *hold and consider a model* S_m *in the collection of models with* $\mathcal{D}_n \leq O(\sqrt{N\Delta}/\ln(N))$ *where* $\mathcal{D}_n$ *is the maximal dimension. Then the estimator* $\hat{b}_m$ *of* b *is such that*

$$\mathbb{E}(\|\hat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}_{\mathbf{A}}\|_{\mathbf{N}}^{\mathbf{2}}) \leq \mathbf{7}\|\mathbf{b}_{\mathbf{m}} - \mathbf{b}_{\mathbf{A}}\|_{\pi^*}^{\mathbf{2}} + \mathbf{K}\frac{\mathbb{E}(\sigma^{\mathbf{2}}(\mathbf{V_0}))\mathbf{D}_{\mathbf{m}}}{\mathbf{N}\boldsymbol{\Delta}} + \mathbf{K}'\boldsymbol{\Delta},$$

where $b_A = b\mathbb{1}_{[0,1]}$, b_m *is the orthogonal projection of* b *on* S_m *and* K *and* K' *are some positive constants.*

Note that the condition on $\mathcal{D}_n$ implies that $\sqrt{N\Delta}/\ln(N)$ must be large enough.

Rates.

If $b_A \in \mathcal{B}_{\alpha,2,\infty}([0,1])$, $\alpha \geq 1$, and $\|b_A\|_{\alpha,2,\infty} \leq L$.

and $\|b_m - b_A\|_{\pi^*}^2 \leq \pi_1^* \|b_m - b_A\|^2$

Choose $D_m = (N_n \Delta_n)^{1/(2\alpha+1)}$, we obtain

$$\mathbb{E}(\|\hat{\mathbf{b}}_{\mathbf{m}} - \mathbf{b}_{\mathbf{A}}\|_{\mathbf{n}}^{\mathbf{2}}) \leq \mathbf{C}(\alpha, \mathbf{L}, \pi_{\mathbf{1}}^*)(\mathbf{N}_{\mathbf{n}} \boldsymbol{\Delta}_{\mathbf{n}})^{-\mathbf{2}\alpha/(\mathbf{2}\alpha+\mathbf{1})} + \mathbf{K}' \boldsymbol{\Delta}_{\mathbf{n}}.$$

$$\begin{aligned} (N_n \Delta_n)^{-2\alpha/(2\alpha+1)} &= T_n^{-2\alpha/(2\alpha+1)} \\ &= \text{the } \textbf{optimal nonparametric rate} \text{ proved by} \\ &\quad \text{Hoffmann (1999) for direct observations of } V. \end{aligned}$$

Second term: study of cases in which it is negligible.

Model selection

$$\hat{m} = \arg \min_{m \in \mathcal{M}_n} \left[\gamma_n(\hat{b}_m) + \text{pen}(m) \right], \qquad (8)$$

with pen(m) a penalty to be properly chosen. We denote by $\tilde{b} = \hat{b}_{\hat{m}}$ the resulting estimator and we need to determine pen such that, ideally,

$$\mathbb{E}(\|\tilde{b} - b_A\|_N^2) \leq C \inf_{m \in \mathcal{M}_n} \left(\|b_A - b_m\|^2 + \frac{\mathbb{E}(\sigma^2(V_0)) D_m}{N\Delta} \right) + \text{negligible terms},$$

with C a constant which should not be too large.

We almost reach this aim for the estimation of b.

Theorem 1 *Assume that* [A1]-[A5] *hold,* $1/k \leq \Delta$, $\Delta \leq 1$ *and* $N\Delta \geq 1$. *Consider the collection of models with maximal dimension* $\mathcal{D}_n \leq O(\sqrt{N\Delta}/\ln(N))$. *Then the estimator* $\tilde{b}$ *of* b *where* $\hat{m}$ *is defined by (8) with*

$$\mathrm{pen}(\mathbf{m}) \geq \kappa\sigma_{\mathbf{1}}^{\mathbf{2}}\frac{\mathbf{D_m}}{\mathbf{N\Delta}}, \tag{9}$$

where κ *is a universal constant, is such that*

$$\begin{aligned}\mathbb{E}(\|\tilde{b} - b_A\|_N^2) \quad &\leq \quad C \inf_{m\in\mathcal{M}_n} \left(\|b_m - b_A\|_{\pi^*}^2 + \mathrm{pen}(m)\right) \\ &\qquad + K\left(\Delta + \frac{1}{N\Delta} + \frac{1}{\ln^2(N)k\Delta}\right).\end{aligned}$$

Proof relies on the following **Bernstein-type Inequality**:

Lemma 1 *Under the assumptions of Theorem 1, for any positive numbers ϵ and v, we have*

$$\mathbb{P}\left[\sum_{i=0}^{N-1} t(\hat{\bar{V}}_i) Z^{(1)}_{(i+1)\Delta} \geq N\epsilon, \|t\|_N^2 \leq v^2\right] \leq \exp\left(-\frac{N\Delta\epsilon^2}{2\sigma_1^2 v^2}\right).$$

W is a Brownian motion with respect to the augmented filtration $\mathcal{F}_s = \sigma((B_u, W_u), u \leq s, \eta)$.

Conclusion about technicalities associated with dependency:

1) Assumptions on the diffusion to ensure stationarity, mixing...

2) Martingale properties give the control of the centered empirical process: no loss due to mixing in the penalty.

3) Coupling and variance inequality for equivalence of empirical and theoretical norms and for residual terms.

Discrete time version

(with fixed sample step, set to 1) of the stochastic volatility model.

$$\begin{cases} Y_i = \exp(X_i/2)\eta_i, \\ X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1}, \end{cases} \tag{10}$$

(η_i) and (ξ_i) independent sequences of i.i.d. r.v.'s (noise processes).

Only $Y_1, \ldots, Y_n$ are observed,

while process of interest is $U_i = \exp(X_i/2)$, and in particular the functions $b(.)$ and $\sigma(.)$.

For $Y_i = \log(S_{i+1}/S_i)$:

$$\mathbf{Y_i} \sim_{\mathcal{L}} \mathbf{U_i}\eta_\mathbf{i},$$

$U_i = (\int_i^{i+1} V_s ds)^{1/2}$ and η_i i.i.d. $\mathcal{N}(0,1)$.

$\Rightarrow$ first equation of the continuous time model

$=$ first equation of (10) (exact discretization in distrib.) with specific Gaussian distribution for η.

$\Rightarrow$ Tools for estimating the common density of the U_i's common to both models.

But the second equations of both models: **same idea** of a time dynamics, but **do not coincide.**

Transformation into an Error-in-variables model.

$$\begin{cases} Z_i = X_i + \varepsilon_i \\ X_{i+1} = b(X_i) + \sigma(X_i)\xi_{i+1} \end{cases} \tag{11}$$

where

$$\begin{cases} \varepsilon_i = \ln(\eta_i^2) - \mathbb{E}(\ln(\eta_i^2)) \\ Z_i = \ln(Y_i^2) - \mathbb{E}(\ln(\eta_i^2)). \end{cases}$$

Here $\mathbb{E}(\ln(\eta_i^2))$ known + (η) and (ξ) are independent.

Log of $Y_i^2 \Rightarrow$ **sign of Y_i** can not be recovered.

Observations: $(Z_i)_{1 \leq i \leq n}$.

Quotient strategy for estimation:

$$\ell = bf, \qquad \hat{b} = \frac{\hat{\ell}}{\hat{f}}.$$

Density estimation for f + estimation of ℓ

in a **convolution model** − an **error in variable** model

Why do mixing problems vanish from important terms (for the rates).

What is the benchmark?

Projection estimator for density of X_1 when the process **is observed**,

$$\hat{f}_m = \sum_j \hat{a}_j \varphi_j, \quad \hat{a}_j = \frac{1}{n}\sum_{i=1}^{n} \varphi_j(X_i)$$

where φ_j is e.g. still the trigonometric basis.

$$\mathbb{E}(\hat{f}_m) = f_m = \sum_j a_j \varphi_j, \quad a_j = \langle f, \varphi_j \rangle.$$

Then

$$\mathbb{E}(\|\hat{f}_m - f_A\|^2) = \|f - f_m\|^2 + \mathbb{E}(\|f_m - \hat{f}_m\|^2)$$

and

$$\mathbb{E}(\|f_m - \hat{f}_m\|^2) = \mathbb{E}(\sum_j (\hat{a}_j - a_j)^2) = \sum_{\mathbf{j}} \mathrm{Var}\left(\frac{\mathbf{1}}{\mathbf{n}}\sum_{\mathbf{i=1}}^{\mathbf{n}} \varphi_{\mathbf{j}}(\mathbf{X_i})\right)$$

β-mixing variance inequality:

$$\mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i)\right) \leq \frac{4}{n}\int \varphi_j^2(x)b(x)d\mathbb{P}(x)$$

with

$$\sum_j \varphi_j^2 = 2m+1, \qquad \int b(x)d\mathbb{P}(x) \leq \sum_k \beta_k,$$

and

$$\sum_j \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n}\varphi_j(X_i)\right) \leq \sum_{\mathbf{k}} \beta_{\mathbf{k}} \frac{D_m}{n}.$$

This explains why $\mathrm{pen}(m) = \kappa \sum_{\mathbf{k}} \beta_{\mathbf{k}} \frac{D_m}{n}$

Lot of works on the subject (Lerasle (2009), Gannaz and Wintenberger (2010)).

Now for $Z_i = X_i + \varepsilon_i$, $f_Z = f \star f_\varepsilon$ (convolution).

$f_Z^* = f^* f_\varepsilon^*$ where $g^*(u) = \int e^{ixu} g(x) dx$

$f^* = f_Z^* / f_\varepsilon^* \Rightarrow \quad \hat{\mathbf{f}}^*(\mathbf{u}) = \dfrac{\frac{\mathbf{1}}{\mathbf{n}} \sum_{\mathbf{k=1}}^{\mathbf{n}} \mathbf{e}^{\mathbf{iuZ_k}}}{\mathbf{f}_\varepsilon^*(\mathbf{u})}$

$$\hat{f}_m(x) = \frac{1}{2\pi} \int_{-\pi m}^{\pi m} e^{-iux} \hat{f}^*(u) du$$

Fourier inversion with cutoff, for integrability.

Bias measured w.r.t. $f_m(x) = \dfrac{1}{2\pi} \displaystyle\int_{-\pi m}^{\pi m} e^{-iux} f^*(u) du$

Mean squared error:

$$\mathbb{E}(\|\hat{\mathbf{f}}_{\mathbf{m}} - \mathbf{f}\|^{\mathbf{2}}) = \|\mathbf{f} - \mathbf{f}_{\mathbf{m}}\|^{\mathbf{2}} + \mathbb{E}(\|\mathbf{f}_{\mathbf{m}} - \hat{\mathbf{f}}_{\mathbf{m}}\|^{\mathbf{2}}).$$

Squared bias $\|f - f_m\|^2 = \int_{|u|\geq \pi m} |f^*(u)|^2 du$

Variance term

$$\begin{aligned}
\mathbb{E}(\|f_m - \hat{f}_m\|^2) &= \operatorname{Var}\left(\frac{1}{n}\sum_{k=1}^{n}\int_{-\pi m}^{\pi m}\frac{e^{iuZ_k}}{f_\varepsilon^*(u)}du\right) \\
&= \frac{1}{n^2}\sum_{k,\ell=1}^{n}\int_{-\pi m}^{\pi m}\int_{-\pi m}^{\pi m}\frac{\operatorname{cov}(\mathbf{e^{iuZk}}, \mathbf{e^{ivZ_\ell}})}{f_\varepsilon^*(u)f_\varepsilon^*(-v)}du
\end{aligned}$$

For $k \neq \ell$

$$\begin{aligned}
\operatorname{cov}(e^{iuZk}, e^{ivZ_\ell}) &= \mathbb{E}(e^{i(uX_k - vX_\ell) + i(u\varepsilon_k - v\varepsilon_\ell)}) - \mathbb{E}(e^{iu(X_k+\varepsilon_k)})\mathbb{E}(e^{-iv(X_\ell+\varepsilon_\ell)}) \\
&= \operatorname{cov}(\mathbf{e^{iuXk}}, \mathbf{e^{ivX_\ell}})\mathbf{f}_\varepsilon^*(\mathbf{u})\mathbf{f}_\varepsilon^*(\mathbf{-v})
\end{aligned}$$

This yields

$$\mathbb{E}(\|f_m - \hat{f}_m\|^2) \leq \underbrace{\frac{1}{n}\int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*(u)|^2}}_{\textbf{usual deconvolution variance bound}} + \underbrace{\mathrm{Var}\left(\frac{1}{n}\sum_{k=1}^{n}\int_{-\pi m}^{\pi m} e^{iuX_k}du\right)}_{\textbf{standard variance of a mixing process}}$$

If $|f_\varepsilon^*(u)| \sim C(1+|u|)^{-\gamma}$, **main variance term** $= O\left(\frac{m^{2\gamma+1}}{n}\right)$.

Second variance term $= O\left(\frac{m}{n}\right)$ with mixing or independence$\Rightarrow$
Negligible.

(see Comte, Dedecker, Taupin (2008)).

More generally

$$|f_\varepsilon^*(u)| \sim c(1+|u|)^{-\gamma}\exp(-\mu|u|^\delta)$$

Examples: Gaussian case $|f_\varepsilon^*(u)| = \exp(-u^2/2)$, $\gamma = 0$, $\delta = 2$.

Case $\log(\mathcal{N}(0,1)^2)$: $|f_\varepsilon^*(u)| \sim \sqrt{2/e}\exp(-\pi|u|/2)$, $\gamma = 0$, $\delta = 1$.

$\Rightarrow$ **Nonstandard variance orders**,

$\Rightarrow$ **Nonstandard rates** of convergence for well-chosen m.

Model (Cutoff) Selection.

$$\text{pen}(m) = \frac{\kappa}{n} m^\omega \int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*(u)|^2} \text{ where } \omega = \begin{cases} 0 & \text{if} \quad 0 \le \delta < 1/3 \\ \inf(\frac{3\delta-1}{2}, \delta) & \text{if} \quad \delta > 1/3 \end{cases}$$

$$\hat{m} = \arg\min_{m} \left\{ -\|\hat{f}_m\|^2 + \mathrm{pen}(m) \right\}.$$

We get for β-mixing with coefficients of X such that $\beta_k \leq ck^{-(1+\theta)}$ with $\theta > 3$, we get

$$\mathbb{E}(\|\hat{f}_{\hat{m}} - f\|^2) \leq C \inf_{1 \leq m \leq m_n} \left(\|f - f_m\|^2 + \mathrm{pen}(m) \right) + \frac{C}{n}$$

where m_n must be cautiously bounded.

In term of the mixing study, much thinner results can be proved, not detailed here.

Now: Conditions are required for X to be β-mixing, in an autoregressive and heteroskedastic model. See e.g. Doukhan (1994).

For $\ell = bf$, same principle:

$$\hat{\ell}_m = \frac{1}{2\pi n} \sum_{k=1}^{n} Z_{k+1} \int_{-\pi m}^{\pi m} \frac{e^{-iuZ_k}}{f_\varepsilon^*(u)} du$$

New variance term $= \mathbb{E}(\mathbf{Z_1^2}) \dfrac{\displaystyle\int_{-\pi m}^{\pi m} \frac{du}{|f_\varepsilon^*(u)|^2}}{n}$

Same orders as previously but unbounded $\Rightarrow$ additional technical difficulties.
Moreover Z_{k+1} and Z_k $\Rightarrow$ two different indices, to split into odd/even terms.

$$Z_{k+1} = X_{k+1} + \varepsilon_{k+1} = b(X_k) + \sigma(X_k)\xi_{k+1} + \varepsilon_{k+1},$$

while $Z_k = X_k + \varepsilon_k$. Many results are obtained in two steps by conditioning by X.

Risk bound for one estimator holds for $\theta > 1$.

$$\text{pen}_\ell(m) = \mathbb{E}(Z_1^2)\text{pen}(m)$$

and under much stronger mixing conditions $\theta > 14$, + moment conditions

$$\hat{m}_\ell = \arg\min_m \left(-\|\hat{\ell}_m\|^2 + \text{pen}_\ell(m)\right)$$

$$\mathbb{E}(\|\hat{\ell}_{\hat{m}_\ell} - \ell\|^2) \le C \inf_{1\le m\le m_n} \left(\|\ell - \ell_m\|^2 + \text{pen}_\ell(m)\right) + \frac{C}{n}$$

References.

Part 1.

• "Nonparametric estimation for a stochastic volatility model", with V. Genon-Catalot and Y. Rozenholc. *Finance and Stochastics* **14**, n^{o}1, 49-80, 2010.

• "Nonparametric adaptive estimation for integrated diffusions," with V. Genon-Catalot and Y. Rozenholc. *Stochastic Processes and Their Applications* **119**, n^{o}3, 811-834, 2009.

• "Penalized nonparametric mean square estimation of the coefficients of diffusion processes," with V. Genon-Catalot and Y. Rozenholc, *Bernoulli* **13**, n^{o}2, 514-543, 2007.

Part 2.

- "Adaptive estimation of the dynamics of a discrete time stochastic volatility model", with C. Lacour and Y. Rozenholc. *Journal of Econometrics* **154**, n^{o}1, 59-73, 2010.

- "Adaptive density deconvolution for dependent inputs with measurement errors", with J. Dedecker and M.-L. Taupin. *Mathematical Methods of Statistics* **17**, n^{o}2, 87-112, 2008.

- "Nonparametric estimation of the regression function in an errors-in-variables model", with M.-L. Taupin, *Statistica Sinica* **17**, n^{o}3, 1065-1090, 2007.

• Berbee, H. (1979). Random walks with Stationary Increments and Renewal Theory. Mathematical Centre Tracts 112. Amsterdam: Mathematisch Centrum.

• Dedecker, J. and Prieur, C. (2005). New dependence coefficients. Examples and applications to statistics. *Probab. Theory Relat. Fields* **132** (2), 203236.

• Doukhan, P. (1994). *Mixing: Properties and examples.* Lecture Notes in Statistics (Springer). New York: Springer-Verlag.

• Pardoux, E. and Veretennikov, A. Yu. (2001). On the Poisson equation and diffusion approximation. I. *Ann. Probab.* **29**, 3, 1061-1085.

• Viennet, G. (1997). Inequalities for absolutely regular sequences: application to density estimation. *Probab. Theory Relat. Fields* **107** (4), 467492.