Chapter 3: Allan Variance

Jan Skaloud* & Stéphane Guerrier†

*École Polytechnique Fédérale de Lausanne; †University of Geneva

This document was prepared with the help of Dr. Davide Cucci, Dr. Roberto Molinari, Dr. Samuel Oros, Dr. Mucyo Karemera, Gaetan Bakalli, Cesare Miglioli, Lionel Voirol, Haotian Xu & Yuming Zhang

Material available online: https://gmwm.netlify.com

EPFL - Winter 2020

Estimation of composite stochastic processes

In general, stochastic calibration of sensors (or other observations) involves the estimation of the parameters of composite stochastic processes. These models are typically difficult to estimate because of their latent features. In the definition below we characterize the **class of composite stochastic processes** we shall consider here.

Definition 3.1 (Common class of processes for IMU calibration).

Let (X_t) be a sum of latent independent stochastic process such that:

- (X_t) is made of a sum which includes a subset or all processes in the set $\{$**QN**, **WN**, **AR1**, **RW**, **DR**$\}$ (i.e. see Definitions 2.3 to 2.7), where processes in the subset $\{$**QN**, **WN**, **RW**, **DR**$\}$ are included only once and process **AR1** can be included k times $(0 \leq k < \infty)$.
- Let $\mathcal{Q}$ denote an arbitrary compact subset of $\mathbb{R}^+$. Then, the innovation process for processes **WN**, **RW** and **AR1** have respective variances σ^2, γ^2 and ν^2 such that σ^2, γ^2 and $\nu^2 \in \mathcal{Q}$ and processes **QN** and **DR** have $Q^2 \in \mathcal{Q}$ and $|\omega| \in \mathcal{Q}$ respectively.

Estimation of composite stochastic processes

Estimation techniques for composite stochastic processes

There are **three main class of estimation techniques** that can be used for the estimation of the class of processes defined in the previous slide (i.e. Definition 3.1). The methods are the following:

- **Maximum likelihood based methods:** while these methods are in theory optimal there applicability is extremely limited due numerical reasons and tends to perform badly in practice. This method will be quickly reviewed in the following slides.
- **Allan variance-based methods:** this class of method is arguably the most popular approach for IMU calibration. However, they typically lead to inconsistent estimators and their finite sample performance is often much lower than GMWM-based technique(s). In this chapter, we will discuss in detail this method.
- **GMWM and related methods:** In our (*very biased*) opinion, these techniques are currently the best choice for the estimation of the parameters of the class of processes considered in Definition 3.1. As we will see, this method is in fact a formalized version Allan variance-based methods. The GMWM will be presented in the next chapter.

Maximum likelihood based methods

Maximum likelihood based approaches are generally inappropriate for the estimation of the class of processes considered Definition 3.1. In this course, we shall avoid a technical discussion on likelihood based method and refer the readers to Stebler et al. 2011 and Guerrier, Molinari, and Balamuta 2016 for more details. Instead, we will consider an example to illustrate the numerical issues of this technique.

Suppose we wish to estimate a composite process composed of a WN and an AR1 process, which we often simply denote as **WN() + AR1()**, i.e.

$$\begin{aligned} &Y_t = \phi_0 Y_{t-1} + Z_t, \quad Z_t \overset{iid}{\sim} \mathcal{N}\left(0, \nu_0^2\right), \\ &U_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_0^2\right), \quad X_t = Y_t + U_t. \end{aligned} \tag{3.1}$$

Then, we want to estimate the parameter $\boldsymbol{\theta}_0 = \left[\phi_0 \;\; \nu_0^2 \;\; \sigma_0^2\right]^T$.

Maximum likelihood based methods

Since the process is Gaussian we have

$$\mathbf{X} \sim \mathcal{N}\left(\mathbf{0}, \mathbf{\Sigma}(\boldsymbol{\theta}_0)\right),$$

where $\mathbf{X} := [X_1, ..., X_T]^T$ and $\mathbf{\Sigma}(\boldsymbol{\theta}_0) := \text{cov}(\mathbf{X})$. Since U_t and Y_t are independent, we have

$$\mathbf{\Sigma}(\boldsymbol{\theta}_0) = \text{cov}(\mathbf{X}) = \text{cov}(\mathbf{Y}) + \text{cov}(\mathbf{U}) = \sigma_0^2 \mathbf{I}_T + \frac{\nu_0^2}{1-\phi_0^2}\left[\phi_0^{|i-j|}\right]_{i,j=1,...,T}, \tag{3.2}$$

where $\mathbf{Y} := [Y_1, ..., Y_T]^T$, $\mathbf{U} := [U_1, ..., U_T]^T$ and where $\mathbf{I}_T$ denotes the identity matrix of dimension T. Note that the form of $\text{cov}(\mathbf{Y})$ is due to the autocovariance of an AR1 which has been discussed in Chapter 2.

Based on (3.2) we can now write the log-likelihood function of the modeled process considered here which, up to a constant, can be expressed as

$$\mathcal{L}\left(\boldsymbol{\theta}|\mathbf{X}\right) = -\log\left(\det\left(\mathbf{\Sigma}(\boldsymbol{\theta})\right)\right) - \mathbf{X}^T \mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}.$$

Maximum likelihood based methods

Therefore, we can following maximum likelihood estimator for $\boldsymbol{\theta}_0$:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmax}} \; \mathcal{L}\left(\boldsymbol{\theta} | \mathbf{X}\right) = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmin}} \; \log\left(\det\left(\boldsymbol{\Sigma}(\boldsymbol{\theta})\right)\right) + \mathbf{X}^T \boldsymbol{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{X}. \tag{3.3}$$

Unfortunately, the applicability of this estimator is essentially impossible when $T > 10^5$ since every evaluation of this the function $\mathcal{L}\left(\boldsymbol{\theta}|\mathbf{X}\right)$ requires to invert a $T \times T$ matrix, which entails a considerable (and often unrealistic) computational burden.

An alternative approach to compute maximum likelihood estimator for $\boldsymbol{\theta}_0$ is based on the EM-algorithm of Dempster, Laird, and Rubin 1977. If the process (Y_t) (or U_t) was observed we could easily estimate the parameters of (3.1) by considering separately the likelihood of both processes. Since (3.1) is a state-space model we could use the following approach:

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\text{argmax}} \; \mathcal{L}\left(\boldsymbol{\theta} | \mathbf{X}, \hat{\mathbf{Y}}(\boldsymbol{\theta})\right),$$

where $\hat{\mathbf{Y}}(\boldsymbol{\theta})$ denotes the estimation of $\mathbf{Y}$ (i.e. the states) based on a Kalman filter assuming $\boldsymbol{\theta}$ to be the correct parameter vector. Unfortunately, this approach suffers from similar computational limitations as (3.3).

Introduction: Allan Variance

- The Allan Variance (AV) is a statistical technique originally developed in the mid-1960s to study the stability of precision oscillators (see e.g. Allan 1966).
- It can provide information on the types and magnitude of various superimposed noise terms (i.e. composite stochastic processes).
- This method has been adapted to characterize the properties of a variety of devices including inertial sensors (see El-Sheimy, Hou, and Niu 2008).
- The AV is a measure of variability developed for long term memory processes and can in fact be interpreted as a Haar wavelet coefficient variance (see Percival and Guttorp 1994). We will discuss this connection further on.

Definition: Allan Variance

Definition 3.2 (Allan Variance).

We consider the AV at dyadic scales (τ_j) starting from local averages of the process which can be denoted as

$$\bar{X}_t^{(j)} := \frac{1}{\tau_j} \sum_{i=1}^{\tau_j} X_{t-\tau_j+i}\,,$$

where $\tau_j := 2^j$, $j \in \{x \in \mathbb{N} \,:\, 1 \leq x < \log_2(T) - 1\}$ therefore determines the number of consecutive observations considered for the average. Then, the AV is defined as

$$\text{AVar}_j\,(X_t) := \frac{1}{2}\,\mathbb{E}\left[\left(\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)}\right)^2\right].$$

Remark 1 (Alternative scale definition).

The definition of the AV is actually valid for $\tau_j = \left\lfloor 2^j \right\rfloor$ with $j \in \{x \in \mathbb{R} \,:\, 1 \leq x < \log_2(T) - 1\}$. In some case, it could use to consider this alternative definition (see e.g. El-Sheimy, Hou, and Niu 2008) but we shall restrict ourself here to the case where $j \in \{x \in \mathbb{N} \,:\, 1 \leq x < \log_2(T) - 1\}$.

Properties of the Allan Variance

Remark 2 (Notation of the Allan Variance).

For notational simplicity, we may sometimes replace $\text{AVar}_j(X_t)$ by simply ϕ_j^2 when the dependence of the AV to the process (X_t) is evident.

As highlighted earlier, the AV is, among others, a widely and commonly used approach in engineering for sensor calibration as it is linked to the properties of the process (X_t) as shown in the following lemma (see e.g Percival and Walden 2006 for a proof).

Lemma 3.3 (AV connection to PSD).

For a stationary process (X_t) with PSD $S_X(f)$ we have

$$\phi_j^2 := \text{AVar}_j(X_t) = 4\int_0^\infty \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_X(f) df.$$

Therefore, this result establishes a direct connection between the AV and PSD. A natural question is therefore whether the mapping PSD $\mapsto$ AV is one-to-one. Greenhall 1998 (see Theorem 1) showed that this is actually not the case. This is illustrated in the following slides.

Spectral Ambiguity of the AV

Consider two (independent) stochastic processes (X_t) and (Y_t) with respective PSD $S_X(f)$ and $S_Y(f)$. Suppose that $S_X(f) \neq S_Y(f)$, then the two processes will have the same AV if

$$\Delta := \int_0^\infty \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} \Phi(f) df = 0,$$

where $\Phi(f) := S_X(f) - S_Y(f)$. To show that it is possible that $\Delta = 0$ when $\Phi(f) \neq 0$, we will use the following critical identity:

$$\sin^4(x) = \sin^2(x) - \frac{1}{4}\sin^2(2x). \tag{3.4}$$

First, we note that Δ may be expressed using (3.4) as follows:

$$\begin{aligned}
\Delta &= \int_0^\infty \frac{\sin^4(\tau_j \pi f)}{(\tau_j \pi f)^2} \Phi(f) df \\
&= \lim_{n \to -\infty} \int_{2^n}^\infty \frac{\sin^2(\tau_j \pi f) - \frac{1}{4}\sin^2(2\tau_j \pi f)}{(\tau_j \pi f)^2} \Phi(f) df.
\end{aligned}$$

Spectral Ambiguity of the AV

Second, by the change of variable $u = 2f$ in the second term we obtain

$$\Delta = \lim_{n \to -\infty} \left[\int_{2^n}^{\infty} \frac{\sin^2(\tau_j \pi f)}{(\tau_j \pi f)^2} \Phi(f) df - \frac{1}{2} \int_{2^{n+1}}^{\infty} \frac{\sin^2(\tau_j \pi u)}{(\tau_j \pi u)^2} \Phi(f) du \right].$$

Now suppose that $\Phi(f) = 2\Phi(2f)$. In this case, we have $\Phi(f) = 2\Phi(u)$ and therefore we obtain

$$\Delta = \lim_{n \to -\infty} \int_{2^n}^{2^{n+1}} \frac{\sin^2(\tau_j \pi f)}{(\tau_j \pi f)^2} \Phi(f) df = 0.$$

Remark 3.

This result demonstrates that the mapping from PSD to Allan variance is not necessarily one-to-one. Greenhall 1998 showed that in the continuous case (i.e. $\tau_j \in \mathbb{R}$) $\Delta = 0$ if and only if $\Phi(f) = 2\Phi(2f)$. However, the "only if" part of this results (while conjectured) is unknown in the discrete case and is currently being investigated.

Properties of the Allan Variance

One reason of explaining the widespread use of the Allan variance for sensor calibration is due to the following additivity property, which is particularly convenient to identify composite stochastic processes (see Definition 2.9).

Corollary 3.4 (Additivity of the AV).

Consider two (independent) stochastic processes (X_t) and (Y_t) with respective PSD $S_X(f)$ and $S_Y(f)$. Suppose that we observe the process $Z_t = X_t + Y_t$. Then, we have

$$\text{AVar}_j\left(Z_t\right) = \text{AVar}_j\left(X_t\right) + \text{AVar}_j\left(Y_t\right).$$

Proof: The proof of this result is direct from Lemma 3.3. Indeed, since $S_Z(f) = S_X(f) + S_Y(f)$, we have

$$\begin{aligned}
\text{AVar}_j\left(Z_t\right) &= 4\int_0^\infty \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_Z(f) df \\
&= 4\int_0^\infty \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_X(f) df + 4\int_0^\infty \frac{\sin^4(\pi f \tau_j)}{(\pi f \tau_j)^2} S_Y(f) df \\
&= \text{AVar}_j\left(X_t\right) + \text{AVar}_j\left(Y_t\right).
\end{aligned}$$

■

Properties of the Allan Variance

While Lemma 3.3 is an important results which is very convenient to determine the theoretical AV of a certain stochastic process. However, the applicability of this results is often limited since the integral defined in (3.3) can be intractable. An alternative to Lemma 3.3 has been proposed by Zhang 2008 and is far advantageous from a computational standpoint.

Lemma 3.5 (AV connection to ACF).

For a stationary process (X_t) with variance σ_X^2 and ACF $\rho(h)$ we have

$$\text{AVar}_j\left(X_t\right) = \frac{\sigma_X^2}{\tau_j^2}\left(\tau_j\left[1-\rho(\tau_j)\right] + \sum_{i=1}^{\tau_j-1} i\left[2\rho(\tau_j-i)-\rho(i)-\rho(2\tau_j-i)\right]\right).$$

The proof of this result is instructive and is presented in Xu et al. 2017.

Properties of the Allan Variance

Remark 4.

Using Lemma 3.5, the exact form of the AV for different stationary processes, such as the general class of ARMA models, can easily be derived. Moreover, Zhang 2008 provided the theoretical AV for non-stationary processes such as the random walk and ARIMA models for which the AV, as mentioned earlier, represents a better measure of uncertainty compared to other methods.

Remark 5.

Lemma 3.5 was extended to non-stationary processes in Xu et al. 2017.

Example: In Appendix A we provide an example on the derivation of the theoretical AV of an MA(1) process (see ▸ Go to Appendix A)

Estimation of the Allan Variance

Several estimators of the AV have been introduced in the literature. The most commonly is (probably) the Maximum-Overlapping AV (MOAV) estimator proposed by Percival and Guttorp 1994, which is defined as follows:

Definition 3.6 (Maximum-Overlapping AV Estimator).

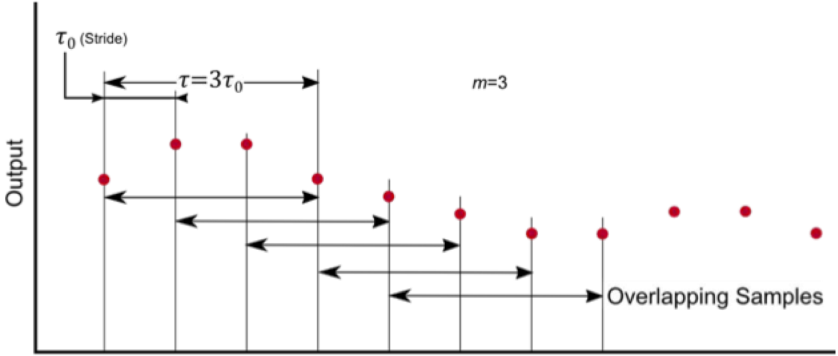
The MOAV is defined as:

$$\hat{\phi}_j^2 := \widehat{\text{AVar}}_j\left(X_t\right) = \frac{1}{2\left(T - 2\tau_j + 1\right)} \sum_{k=2\tau_j}^{T} \left(\bar{X}_k^{(j)} - \bar{X}_{k-\tau_j}^{(j)}\right)^2. \tag{3.5}$$

We will now study the properties of this estimator through the following lemmas.

Maximum-Overlapping AV Estimator

Concept of Allan Variance with overlapping samples



after Freescale Simiconductor, Application note #AN5087 Rev. 0,2/2015

Consistency of the Maximum-Overlapping AV Estimator

Lemma 3.7 (Consistency).

Let (X_t) be such that:

- $(X_t - X_{t-1})$ *is a (strongly) stationary process,*
- $(X_t - X_{t-1})^2$ *has absolutely summable covariance structure,*
- $\mathbb{E}\left[(X_t - X_{t-1})^4\right] < \infty$,

Then, we have

$$\widehat{\text{AVar}}_j\left(X_t\right) \xrightarrow{\mathcal{P}} \text{AVar}_j\left(X_t\right).$$

The proof of Lemma 3.7 is given in Appendix B Go to Appendix B

Remark 6 (Connection to Wavelet Variance).

This result is closely related by the results of Percival 1995 on the wavelet variance. We shall explore the connection between the AV and wavelet variance in the next section.

Asymptotic Normality of the MOAV Estimator

Compare to consistency, the asymptotic normality requires stronger conditions given in the following lemma.

Lemma 3.8 (Asymptotic normality).

Let (X_t) be such that:

- $(X_t - X_{t-1})$ *is a (strongly) stationary process.*
- $(X_t - X_{t-1})$ *is strong mixing process with mixing coefficient* $\alpha(n)$ *such that* $\sum_{n=1}^{\infty} \alpha(n)^{\delta/2+\delta} < \infty$ *for some* $\delta > 0$.
- $\mathbb{E}\left[(X_t - X_{t-1})^{4+\delta}\right] < \infty$ *for some* $\delta > 0$.

Then, under these conditions we have that

$$\sqrt{T}\left(\widehat{\text{AVar}}_j\left(X_t\right) - \text{AVar}_j\left(X_t\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_T^2/T),$$

where $\sigma_T^2 := \sum_{h=-\infty}^{\infty} \text{cov}\left(\left(\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)}\right)^2, \left(\bar{X}_{t+h}^{(j)} - \bar{X}_{t+h-\tau_j}^{(j)}\right)^2\right)$.

The proof of Lemma 3.8 is given in Appendix C Go to Appendix C

Confidence Interval of the MOAV Estimator

Based on the asymptotic normality results (Lemma 3.8), we can construct the $1-\alpha$ confidence intervals for $\widehat{\text{AVar}}_j\left(X_t\right)$ as

$$\text{CI}\left(\text{AVar}_j\left(X_t\right)\right) = \left[\widehat{\text{AVar}}_j\left(X_t\right) \pm z_{1-\frac{\alpha}{2}}\frac{\sigma_T}{T}\right],$$

where $z_{1-\frac{\alpha}{2}} := \mathbf{\Phi}^{-1}\left(1-\frac{\alpha}{2}\right)$ is the $\left(1-\frac{\alpha}{2}\right)$ quantile of a standard normal distribution.

However, the so called "Long-Run Variance" σ_T^2 is usually unknown. Many methods have been proposed to consistently estimate it under mild conditions (see e.g. Newey and West 1986).

Remark 7.

Gaussian-based confidence intervals are often problematic with the AV as the lower limit of CI can very well be negative. Next, we will discuss an alternative method to construct the CI for such statistic.

Allan Variance log-log Representation

As illustrated in Lemmas 3.3 and 3.5 the AV depends on the properties of the stochastic process (X_t). We will see that "log-log" representation of the AV is often useful for the identify various processes that may compose (X_t).

For example, let's suppose that X_t is a white noise process. We showed in Appendix A that the theoretical AV of such process is given

$$\phi_j^2 := \text{AVar}_j(X_t) = \frac{\sigma^2}{\tau_j}.$$

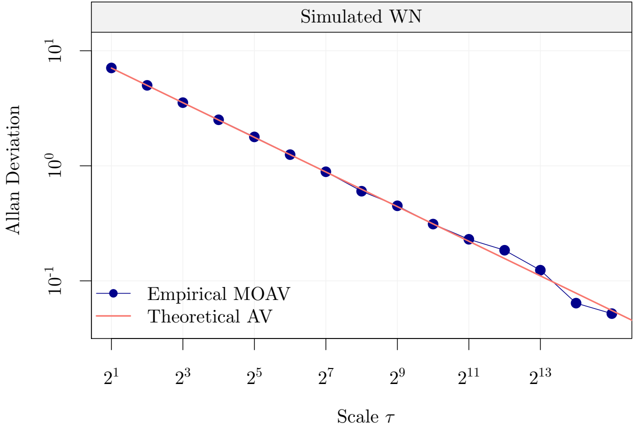
Therefore, we have that the Allan Deviation or AD (i.e. $\sqrt{\text{AVar}_j(X_t)}$ or ϕ_j) is such that

$$\log(\phi_j) = \log\left(\sqrt{\frac{\sigma^2}{\tau_j}}\right) = \log(\sigma) - \frac{1}{2}\log(\tau_j). \tag{3.6}$$

Thus, the log of the AD is linear in τ_j with a slope of $-1/2$ and with intercept $\log(\sigma)$. Let us start by considering a simple simulated example.

Allan Deviation of a WN process

Simulation based on a white noise process with $\sigma^2 = 1$ and $T = 10^5$.



Allan Variance log-log Representation

Suppose now that (X_t) is composite stochastic process composed of a WN (see 2.3) and RW (see 2.6). For simplicity, we assume that $X_t = Y_t + W_t$ where Y_t is a white noise process with variance σ^2 and W_t a random walk with variance γ^2. We already know that

$$\log\left(\text{AVar}_j(Y_t)\right) = \log\left(\sigma\right) - \frac{1}{2}\log(\tau_j).$$

and it can be shown (using for example Lemma 3.3) that

$$\text{AVar}_j(W_t) = \frac{1}{3}\gamma^2\tau_j,$$

and therefore we obtain

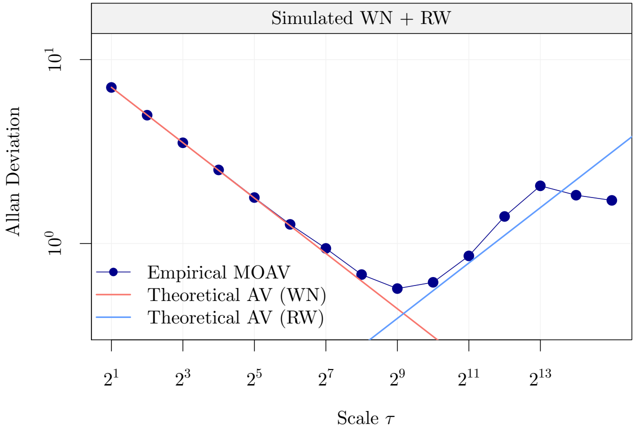
$$\log\left(\sqrt{\text{AVar}_j(W_t)}\right) = \log\left(\sqrt{\frac{1}{3}\gamma^2\tau_j}\right) = \log\left(\frac{1}{\sqrt{3}}\gamma\right) + \frac{1}{2}\log(\tau_j).$$

Thus, the log of the AD of (Z_t) is also linear in τ_j with a slope of $+1/2$. By Corollary 3.4 we also have that

$$\text{AVar}_j(X_t) = \text{AVar}_j(Y_t) + \text{AVar}_j(W_t) = \frac{\sigma^2}{\tau_j} + \frac{1}{3}\gamma^2\tau_j. \tag{3.7}$$

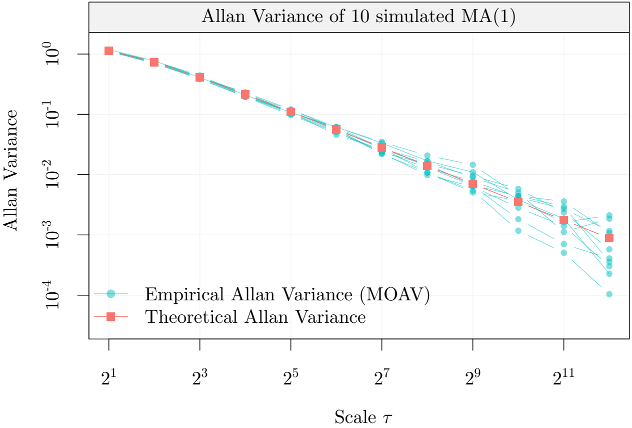
Allan Deviation of a WN + RW process

Simulation based on a white noise process with $\sigma^2 = 1$ and a random walk process with $\gamma^2 = 0.03^2$ and $T = 10^5$.



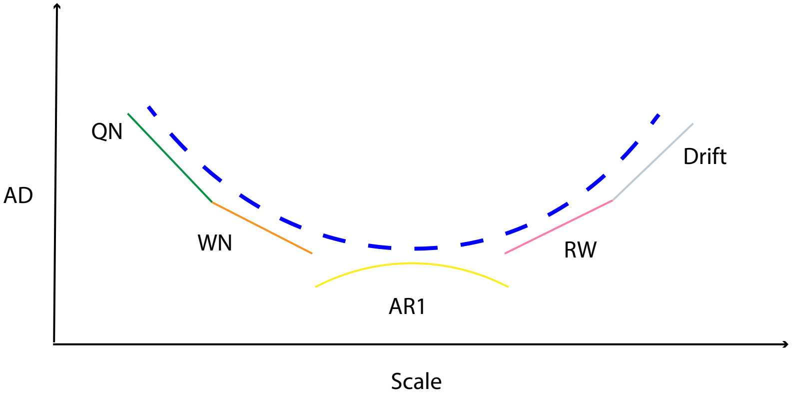
Simulated Example: MA(1)

Using the formula we derived in ▸ Example Theoretical AV MA(1). We simulated 10 MA(1) with $\theta = 0.9$ and $\sigma^2 = 1$. Their empirical AV (i.e. MOAV) are presented below together with the theoretical AV of this process in red.



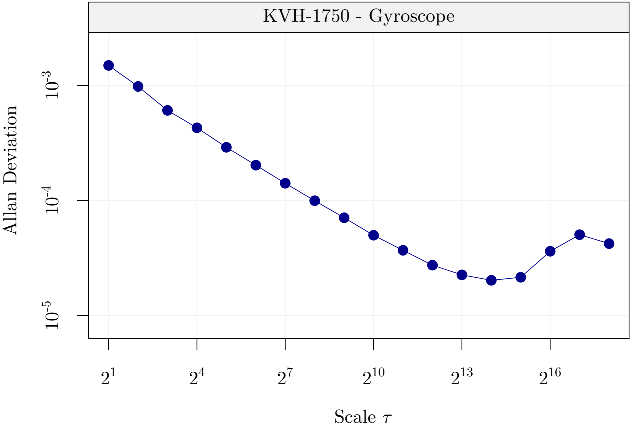
Parameter Estimation through the Allan Variance

The AV is a powerful technique to identify the processes considered in Definition 3.1. Indeed, the five processes we are considering are characterized in log-log representation as follows: *(i)* QN - linear with slope of -1; *(ii)* WN - linear with slope of -1/2; *(iii)* AR1 - curved shape with a slope $\in [-1/2, 1/2]$; *(iv)* RW - linear with slope of 1/2; *(v)* DR - linear with slope of 1.



Real Data: MOAV KVH-1750 (Gyro), 30 mins at 500Hz

The shape of the AD suggests that a WN + RW may be a reasonable approximation. But how could we estimate the parameters of this model?



AV-based Estimation

Main idea:

We have seen that there exist a mapping from $\boldsymbol{\theta}_0$ (model's parameters) to the AV (or AD), i.e.

$$\boldsymbol{\theta}_0 \mapsto \text{PSD}.$$

In general, the mapping from the PSD to the AV is not one-to-one, nevertheless, one may hope that in some case this mapping can be "inverted" (in some sense) so that the AV may be used to estimated $\boldsymbol{\theta}_0$. This is central of AV linear regression approach.

Let us assume that we want to estimate the parameter of a QN, WN, RW or DR process. Then the process (X_t) is such that it has a linear representation in a $\log(\phi_{\tau_j}) - \log(\tau_j)$ plot for the set of scales $\boldsymbol{\eta} \in \boldsymbol{\mathcal{G}}$ where

$$\boldsymbol{\mathcal{G}} = \left\{ \{\tau_k, \, ..., \, \tau_{k+h}\} \, | k, h \in \mathbb{N}^+, k + h \leq J \right\}$$

denotes all possible sets which contain adjacent scales having no. scales $|\boldsymbol{\eta}| > 0$.

AV-based Estimation

Then, there exists a linear relationship between $\log\left(\phi_{\tau_j}(\theta)\right)$ and $\log\left(\tau_j\right)$ we can write

$$\log\left(\phi_{\tau_j}\left(\theta\right)\right) = g\left(\theta\right) + \lambda \log\left(\tau_j\right), \;\; \forall \tau_j \in \boldsymbol{\eta}, \tag{3.8}$$

where the function $g(\cdot)$ as well as the constant λ are known and depend on the model. For a white noise, we have for example $g(\sigma) = \log(\sigma)$ and $\lambda = -1/2$, see (3.6). This linear relationship leads to the following "least-squares" estimator of θ, noted as $\hat{\theta}_{AV}$ and defined as

$$\hat{\theta}_{AV} := \underset{\theta \in \Theta}{\text{argmin}} \sum_{\tau_j \in \boldsymbol{\eta}} \left[\log\left(\hat{\phi}_{\tau_j}\right) - g\left(\theta\right) - \lambda \log\left(\tau_j\right)\right]^2, \tag{3.9}$$

which is an extremum estimator (see Definition 2.21) and admit the solution

$$\hat{\theta}_{AV} = g^{-1}\left\{\frac{1}{|\boldsymbol{\eta}|} \sum_{\tau_j \in \boldsymbol{\eta}} \left[\log\left(\hat{\phi}_{\tau_j}\right) - \lambda \log\left(\tau_j\right)\right]\right\}, \tag{3.10}$$

where $|\boldsymbol{\eta}|$ denotes the cardinality of the set $\boldsymbol{\eta}$. The derivation of (3.10) is instructive and is given in Appendix D Go to Appendix D

AV-based Estimation

Let us consider a simple example of the use of (3.10) with a WN process. We already derived that

$$\log\left(\mathrm{AVar}_j(Y_t)\right) = \log\left(\sigma\right) - \frac{1}{2}\log(\tau_j)$$

implying that $g(\sigma) = \log(\sigma)$ (and thus $g^{-1}(\sigma) = \exp(\sigma)$) and $\lambda = -1/2$. This leads to the following estimator of σ,

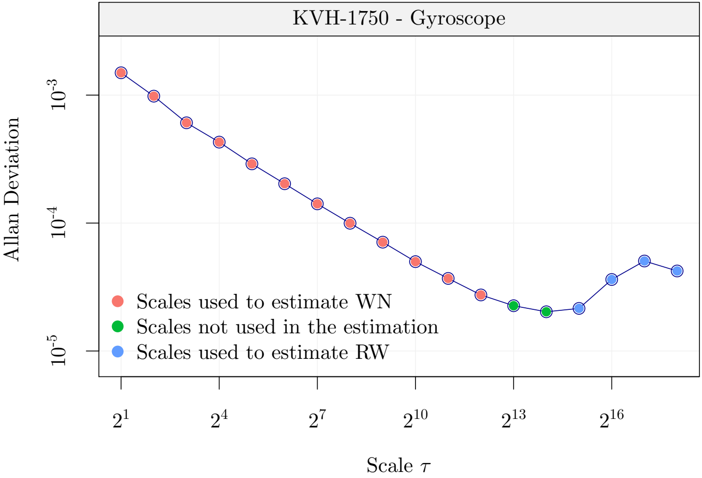
$$\hat{\sigma}_{AV} = \exp\left\{\frac{1}{|\boldsymbol{\eta}|}\sum_{\tau_j\in\boldsymbol{\eta}}\log\left(\hat{\phi}_{\tau_j}\right) + \frac{1}{2}\log\left(\tau_j\right)\right\}. \tag{3.11}$$

Similarly for a RW, we obtain

$$\hat{\gamma}_{AV} = \sqrt{3}\exp\left\{\frac{1}{|\boldsymbol{\eta}|}\sum_{\tau_j\in\boldsymbol{\eta}}\log\left(\hat{\phi}_{\tau_j}\right) - \frac{1}{2}\log\left(\tau_j\right)\right\}. \tag{3.12}$$

Using these results let's estimate the parameters on the KVH-1750 gyro.

Reasonable Model: WN + RW



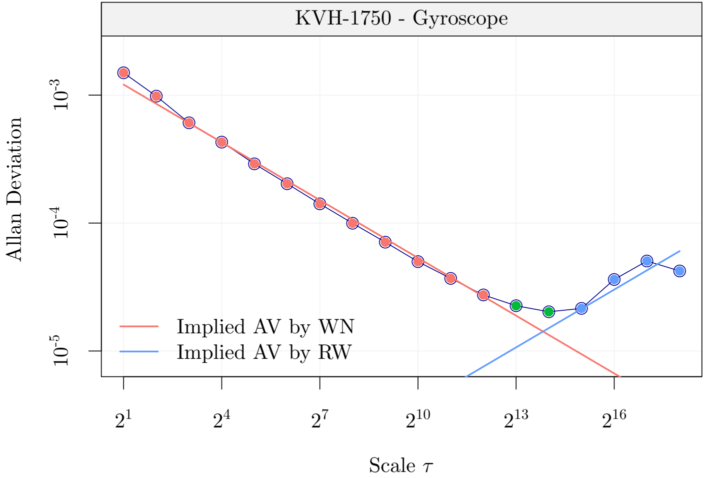
Example: MOAV KVH-1750 (Gyro) - Estimation

Using (3.11) and (3.12) on the selected scales, we obtain the following estimates:

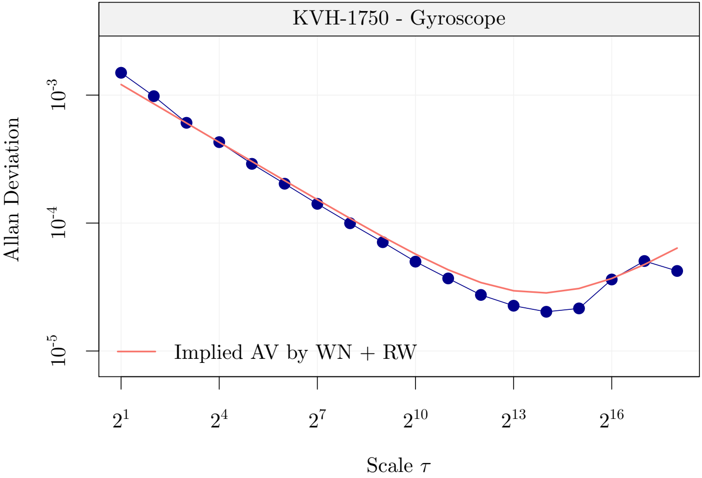
$$\hat{\sigma}_{AV} = \frac{1}{2} \exp\left\{ \frac{1}{12} \sum_{j=1}^{12} \log\left(\hat{\phi}_j\right) + \frac{1}{2} \log\left(2^j\right) \right\} \approx 1.710 \cdot 10^{-3}$$

$$\hat{\gamma}_{AV} = \sqrt{3} \exp\left\{ \frac{1}{4} \sum_{j=15}^{18} \log\left(\hat{\phi}_j\right) - \frac{1}{2} \log\left(2^j\right) \right\} \approx 2.043 \cdot 10^{-7}.$$

Example: MOAV KVH-1750 (Gyro) - Estimation



Example: MOAV KVH-1750 (Gyro) - Estimation



Properties of AV-based Estimation

Before studying the properties of the AV-based estimation technique, we will assess whether the MOAV estimator (see Definition 3.6 is consistent (see Lemma 3.7) and asymptotically normal (see Lemma 3.8) under the setting of given Definition 3.1. Lemma 1 of Guerrier, Molinari, and Stebler 2016 answered that question by showing that all the conditions of Lemmas 3.7 and 3.8 are satisfied this setting.

Since $\hat{\theta}_{AV}$ admits the analytically solution given in (3.10) the consistency of the AV-based estimator can be directly studied using the consistency of the AV and the continuous mapping theorem (see Theorem 2.23). Indeed, using the latter we have

$$\hat{\theta}_{AV} = g^{-1}\left\{\frac{1}{|\boldsymbol{\eta}|}\sum_{\tau_j \in \boldsymbol{\eta}}\left[\log\left(\hat{\phi}_{\tau_j}\right) - \lambda \log\left(\tau_j\right)\right]\right\}$$

$$\xmapsto{\mathcal{P}} g^{-1}\left\{\frac{1}{|\boldsymbol{\eta}|}\sum_{\tau_j \in \boldsymbol{\eta}}\left[\log\left(\phi_{\tau_j}\right) - \lambda \log\left(\tau_j\right)\right]\right\} = \theta^*\,,$$

and therefore if we can show that $\theta^* = \theta_0$ then the estimator is consistent.

Properties of AV-based Estimation

Corollary 3.9 (Consistency of AV-based Estimation).

The AV-based estimation technique is consistent if the process (X_t) is composed of either a QN, WN or DR process.

Proof: Simply show that $\theta^* = \theta_0$.

Corollary 3.10 (Inconsistency of AV-based Estimation).

The AV-based estimation technique is not consistent if the process (X_t) includes at least one RW process or contains more than one latent process.

Proof: It can be show at any model $\theta^* > \theta_0$, implying the consistency of the method.

A formal proof of these results are given in Guerrier, Molinari, and Stebler 2016.

Properties of AV-based Estimation

Remark 8 (Why is the AV-based estimation inconsistent?).

The main reason for the inconsistency of this approach is that the effect of multiple superimposed stochastic processes cannot (perfectly) separated in the log-log representation. This suggests the following alternative definition of $\hat{\theta}_{AV}$,

$$\hat{\theta}^*_{AV} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \sum_{j=1}^{J} \left(\hat{\phi}_j - \phi_j(\boldsymbol{\theta}) \right)^2,$$

where $\phi_j(\boldsymbol{\theta})$ denotes the theoretical AV under the assumption that $\boldsymbol{\theta}$ is the correct parameter vector. We will see that this approach is a special case of the GMWM estimator and it leads to consistent and asymptotically normally distributed estimators.

Example: MOAV KVH-1750 (Gyro) - Estimation with $\hat{\theta}_{AV}^{*}$

Let us revisit the KVH-1750 (Gyro) dataset using this alternative AV-estimation approach. We show in (3.7) that

$$\text{AVar}_j(X_t) = \text{AVar}_j(Y_t) + \text{AVar}_j(W_t) = \frac{\sigma^2}{\tau_j} + \frac{1}{3}\gamma^2\tau_j.$$

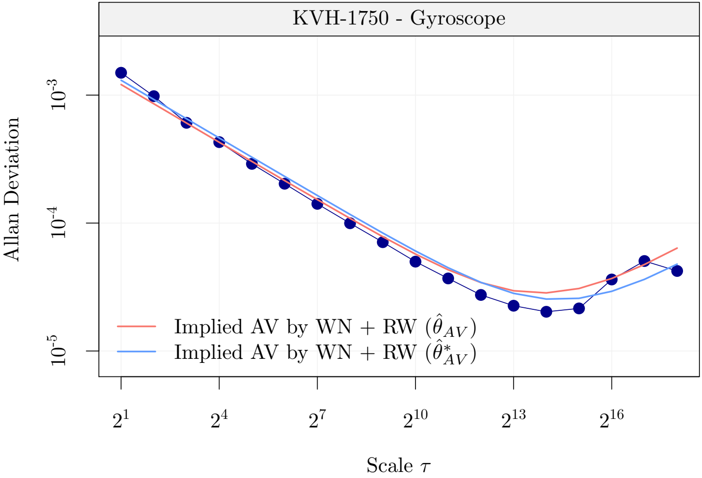
and therefore we can define the following estimator:

$$\hat{\boldsymbol{\theta}}_{AV}^{*} = [\hat{\sigma}_{AV}^{*} \quad \hat{\gamma}_{AV}^{*}]^T = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argmin}} \sum_{j=1}^{J} \left(\phi_j^2 - \frac{\sigma^2}{\tau_j} - \frac{1}{3}\gamma^2\tau_j\right)^2.$$

As noticed in Guerrier et al. 2019 this estimator has a simple solution. corresponding to the point estimate:

$$\hat{\sigma}_{AV}^{*} \approx 1.847 \cdot 10^{-3}$$
$$\hat{\gamma}_{AV}^{*} \approx 1.493 \cdot 10^{-7}.$$

Example: MOAV KVH-1750 (Gyro) - Estimation with $\hat{\theta}^*_{AV}$



References I

Allan, D. W. (1966). "Statistics of Atomic Frequency Standards". In: *Proceedings of the IEEE* 54.2, pp. 221–230.

Dempster, A. P., N. M. Laird, and D. B. Rubin (1977). "Maximum Likelihood from Incomplete Data via the EM Algorithm". In: *Journal of the royal statistical society. Series B (methodological)*, pp. 1–38.

El-Sheimy, N., H. Hou, and X. Niu (2008). "Analysis and Modeling of Inertial Sensors using Allan Variance". In: *IEEE Transactions on Instrumentation and Measurement* 57.1, pp. 140–149.

Greenhall, C. A. (1998). "Spectral Ambiguity of Allan Variance". In: *IEEE Transactions on Instrumentation and Measurement* 47.3, pp. 623–627.

Guerrier, S., R. Molinari, and J. Balamuta (2016). "Discussion on Maximum Likelihood-Based Methods for Inertial Sensor Calibration". In: *IEEE Sensors Journal* 16.14, pp. 5522–5523.

Guerrier, S., R. Molinari, and Y. Stebler (2016). "Theoretical Limitations of Allan Variance-based Regression for Time Series Model Estimation". In: *IEEE Signal Processing Letters* 23.5, pp. 597–601.

Guerrier, Stéphane et al. (2019). *Wavelet-Based Moment-Matching Techniques for Inertial Sensor Calibration*. arXiv: `1911.07049 [stat.ME]`.

Newey, Whitney K and Kenneth D West (1986). *A simple, positive semi-definite, heteroskedasticity and autocorrelationconsistent covariance matrix*.

Percival, D. B. and P. Guttorp (1994). "Long-memory processes, the Allan variance and wavelets". In: *Wavelet Analysis and its Applications*. Vol. 4. Elsevier, pp. 325–344.

References II

Percival, D. B. and A. T. Walden (2006). *Wavelet methods for time series analysis*. Cambridge university press.

Percival, D. P. (1995). "On Estimation of the Wavelet Variance". In: *Biometrika* 82.3, pp. 619–631.

Stebler, Y. et al. (2011). "Constrained Expectation-Maximization Algorithm for Stochastic Inertial Error Modeling: Study of Feasibility". In: *Measurement Science and Technology* 22.8, p. 085204.

Xu, H. et al. (2017). "A Study of the Allan Variance for Constant-Mean Nonstationary Processes". In: *IEEE signal processing letters* 24.8, pp. 1257–1260.

Zhang, N. F. (2008). "Allan Variance of Time Series Models for Measurement Data". In: *Metrologia* 45.5, p. 549.

Appendix A: Example on the Theoretical AV of an MA(1)

In this example we will derive the theoretical AV of an MA(1) process, which is defined as

$$X_t = \varepsilon_t + \theta \varepsilon_{t-1}, \quad \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),$$

where $|\theta| < 1$ and $\sigma^2 < \infty$. The variance of this process is given by

$$\sigma_X^2 = \text{var}\left(X_t\right) = \text{var}\left(\varepsilon_t\right) + \theta^2\, \text{var}\left(\varepsilon_{t-1}\right) = \left(1 + \theta^2\right)\sigma^2.$$

Similarly, the autocovariance is given by

$$\gamma(h) = \text{cov}\left(X_t, X_{t-h}\right) = \begin{cases} \sigma_X^2 & \text{if } h = 0 \\ \theta\sigma^2 & \text{if } |h| = 1 \\ 0 & \text{if } |h| > 1. \end{cases}$$

Appendix A: Example on the Theoretical AV of an MA(1)

From the autocovariance we obtain

$$\rho(h) = \text{corr}\left(X_t, X_{t-h}\right) = \begin{cases} 1 & \text{if } h = 0 \\ \frac{\theta}{1+\theta^2} & \text{if } |h| = 1 \\ 0 & \text{if } |h| > 1. \end{cases}$$

We can now apply the formula given in Lemma 3.5, which leads to

$$\begin{aligned} \text{AVar}_j\left(X_t\right) &= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + \sum_{i=1}^{\tau_j - 1} i\left[2\rho(\tau_j - i) - \rho(i) - \rho(2\tau_j - i)\right]\right) \\ &= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + 2\sum_{i=1}^{\tau_j - 1} i\rho(\tau_j - i) - \sum_{i=1}^{\tau_j - 1} i\rho(i) - \sum_{i=1}^{\tau_j - 1} i\rho(2\tau_j - i)\right) \\ &= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + 2(\tau_j - 1)\rho(1) - \rho(1)\right) \\ &= \frac{\left(1+\theta^2\right)\sigma^2}{\tau_j^2}\left(\tau_j + (2\tau_j - 3)\frac{\theta}{1+\theta^2}\right). \end{aligned}$$

Return to the Properties of the AV

Appendix B: Proof of Lemma 3.7

The proof of the result is direct from Theorem 2.28. Let $Z_t = X_t - X_{t-1}$, then since Z_t is stationary with mean zero then so is $(X_t - X_{t-h})$ for all $h \in \mathbb{Z}$. This directly imply that $\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)}$ is also stationary (since it based on a linear combination of stationary processes) and so is $Y_t := (\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)})^2$ (since it is based on a time invariant transformation of a stationary process). Moreover, there exist constant c_h such that

$$\sum_{h=-\infty}^{\infty} \gamma_Y(h) = \sum_{h=-\infty}^{\infty} c_{|h|}\gamma_{Z^2}(h).$$

Therefore, we obtain

$$\sum_{h=-\infty}^{\infty} |\gamma_Y(h)| = \sum_{h=-\infty}^{\infty} |c_{|h|}\gamma_{Z^2}(h)| \leq \sup_{k=1,\dots,\infty} c_k \sum_{h=-\infty}^{\infty} |\gamma_{Z^2}(h)| < \infty,$$

since both terms are bounded. Using the same approach we have that $\mathbb{E}\left[Y_t^2\right]$ is bounded since $\mathbb{E}\left[Z_t^4\right]$ bounded. Thus, we can apply Theorem 2.28 on the process Y_t, i.e.

$$\widehat{\text{AVar}}_j\left(X_t\right) = \frac{1}{2}\bar{Z}_t \xmapsto{\mathcal{P}} \frac{1}{2}\mathbb{E}[Z_t] = \text{AVar}_j\left(X_t\right),$$

which concludes the proof. ■

▸ Return to Lemma 3.7

Appendix C: Proof of Lemma 3.8

The proof of the result is direct from Theorem 2.32. Let $Z_t = X_t - X_{t-1}$, then since Z_t is stationary with mean zero then so is $(X_t - X_{t-h})$ for all $h \in \mathbb{Z}$. This directly imply that $\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)}$ is also stationary (since it based on a linear combination of stationary processes) and so is $Y_t := (\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)})^2$ (since it is based on a time invariant transformation of a stationary process). And since (Y_t) is a borel-measurable function of Z_t, we have (Y_t) is also a strong mixing process with mixing coefficient $\alpha^*(n) \leq \alpha(n)$, hence $\sum_{n=1}^{\infty} \alpha^*(n)^{\delta/2+\delta} < \infty$ for some $\delta > 0$. Moreover, since $(\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)})$ is a linear function of Z_t, and $\mathbb{E}\left[Z_t^{4+\delta}\right] < \infty$, by triangle inequality, we have $\mathbb{E}\left[(\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)})^{4+\delta}\right] < \infty$. Thus, we can apply Theorem 2.32 on the process Y_t, i.e.

$$\sqrt{T}\left(\widehat{\text{AVar}}_j\left(X_t\right) - \text{AVar}_j\left(X_t\right)\right) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_T^2/T),$$

where $\sigma_T^2 := \sum_{h=-\infty}^{\infty} \text{cov}\left(\left(\bar{X}_t^{(j)} - \bar{X}_{t-\tau_j}^{(j)}\right)^2, \left(\bar{X}_{t+h}^{(j)} - \bar{X}_{t+h-\tau_j}^{(j)}\right)^2\right)$, which concludes the proof. ■

Return to Lemma 3.8

Appendix D: Derivation of Equation (3.10)

The derivation of the AV estimator closed form solution is pretty straightforward. From (3.9), we just first take the first derivative with respect to $\boldsymbol{\theta}$, which yields:

$$
\begin{aligned}
&\frac{\partial}{\partial \boldsymbol{\theta}} \sum_{\tau_j \in \boldsymbol{\eta}} \left[\log\left(\hat{\phi}_{\tau_j}\right) - g\left(\theta\right) - \lambda \log\left(\tau_j\right)\right]^2 = 2 \sum_{\tau_j \in \boldsymbol{\eta}} \left[\log\left(\hat{\phi}_{\tau_j}\right) - g\left(\theta\right) - \lambda \log\left(\tau_j\right)\right] g'\left(\theta\right) = 0 \\
&\Leftrightarrow \frac{1}{|\boldsymbol{\eta}|} g'\left(\theta\right) g\left(\theta\right) = g'\left(\theta\right) \sum_{\tau_j \in \boldsymbol{\eta}} \left[\log\left(\hat{\phi}_{\tau_j}\right) - \lambda \log\left(\tau_j\right)\right] \\
&\Leftrightarrow \hat{\boldsymbol{\theta}}_{AV} = g^{-1} \left\{ \frac{1}{|\boldsymbol{\eta}|} \sum_{\tau_j \in \boldsymbol{\eta}} \left[\log\left(\hat{\phi}_{\tau_j}\right) - \lambda \log\left(\tau_j\right)\right] \right\}.
\end{aligned}
$$

▸ Return to Equation (3.10)