Limits of the Allan Variance and Optimal Tuning of Wavelet Variance based Estimators

GUERRIER, Stéphane, et al.

Abstract

This article first demonstrates the inconsistency of the estimator based on the standard Allan Variance (AV) for composite stochastic processes. This result motivates the use of a recently developed estimator, called the Generalized Method of Wavelet Moments (GMWM) estimator. This estimator was previously shown to be consistent and asymptotically normally distributed under the settings of the present research. Moreover, and unlike Maximum Likelihood Estimators (MLE), this method is able to estimate parameters of complex composite stochastic processes. After establishing the link between the AV and GMWM estimators we prove that there always exists a GMWM estimator with smaller asymptotic variance than its AV based counterpart. As the GMWM may be biased in finite samples issued from simple models on which the AV or MLE can still be applied, we present several extensions to the GMWM which significantly enhance its performance. One of these extensions is a finite sample bias correction related to the principle of indirect inference which is very general and can be applied beyond the scope of the GMWM framework. Finally, the [...]
Limits of the Allan Variance and Optimal Tuning of Wavelet Variance based Estimators  
Stéphane Guerrier, Yannick Stebler, Jan Skaloud & Maria-Pia Victoria-Feser

Abstract—This article first demonstrates the inconsistency of the estimator based on the standard Allan Variance (AV) for composite stochastic processes. This result motivates the use of a recently developed estimator, called the Generalized Method of Wavelet Moments (GMWM) estimator. This estimator was previously shown to be consistent and asymptotically normally distributed under the settings of the present research. Moreover, and unlike Maximum Likelihood Estimators (MLE), this method is able to estimate parameters of complex composite stochastic processes. After establishing the link between the AV and GMWM estimators we prove that there always exists a GMWM estimator with smaller asymptotic variance than its AV based counterpart. As the GMWM may be biased in finite samples issued from simple models on which the AV or MLE can still be applied, we present several extensions to the GMWM which significantly enhance its performance. One of these extensions is a finite sample bias correction related to the principle of indirect inference which is very general and can be applied beyond the scope of the GMWM framework. Finally, the theoretical findings are supported by simulation studies that compares the finite sample performance of AV based estimators, MLE and the various different forms of GMWM estimators.

Index Terms—Composite Stochastic Processes, GMWM, Sensor Calibration, Signal Modeling.

I. INTRODUCTION
CHARACTERIZING signals with random behavior is a central topic in many engineering disciplines. In particular, consider a parametric model, say $F_{\theta}$, that is associated to an univariate Gaussian process $\{Y_t, t \in \mathbb{Z}\}$ that is stationary or non-stationary but with backward differences\(^1\) of order $d$, and let the time series $\{y_t, t = 1, \ldots, T\}$ be the corresponding observed outcome. We define $\theta \in \Theta \subseteq \mathbb{R}^p$ as the parameter vector of the model of interest where $\Theta$ is an open subset of $\mathbb{R}^p$. The methodology based on the Allan Variance (AV), noted $\phi_\tau^2$, with $\tau$ the averaging time, is a well-accepted approach for identifying noise structures with power-law behavior in a signal $\{y_t\}$. Although it was originally intended to characterize nonstationary phase and frequency noise in clocks and oscillators [1], it has been successfully applied for studying low-frequency spectral behavior of physical processes, as well as to problems dealing with different types of sensors. In particular, we find the modeling of errors corrupting inertial sensors [4, 12, 23, 24] or radio-astronomical instrumentation [17], and even physiological phenomena related to nerves activity [6, 7].

The AV for a process $\{Y_t\}$ with stationary second differences is defined by [11] as

$$\phi_\tau^2 = \operatorname{var} \left[ \frac{1}{2\tau^2} Y_{t+1} - 2Y_{t+1} + Y_{t+1} \right].$$

Linear regions in the $\log(\tau) - \log(\phi_\tau)$, where $\phi_\tau$ refers to the Allan Deviation (AD), are associated with regions of power-law behavior $f^\alpha$ in the signal’s Power Spectral Density (PSD) function $S_Y(f)$ where $f$ denotes the frequency. The PSD maps to $\phi_\tau^2$ by the following relationship:

$$\phi_\tau^2 = 4 \lim_{a \to -\infty} \int_0^a S_Y(f) \frac{\sin^4(\pi f \tau)}{(\pi f \tau)^2} df.$$  

(1)
Therefore, the AV is an alternative to the PSD which allows different power-law noises to be identified for building the model $F_{\theta}$ and estimate its parameters. The power-laws of PSD and AD, are related to each other by the following relationship:

$$S_Y(f) \propto f^\alpha \Leftrightarrow \phi_\tau \propto \tau^{\beta},$$

where $\beta = -(\alpha + 1)/2$ and $f = 1/\tau$. Tab. I summarizes important stochastic processes for $f^\alpha$ noise which have a linear representation in the $\log(\tau) - \log(\phi_\tau)$ plot.

When multiple stochastic processes are superposed in the process $\{Y_t\}$, their identification and the estimation of their parameters may become difficult using the method of the AV or PSD. With this respect, a new estimation method has recently been introduced in [12]. This method, called the Generalized Method of Wavelet Moments (GMWM), is specially well suited for the estimation of models $F_{\theta}$ composed of sums of different processes which are commonly employed in engineering to model $f^\alpha$ noise (we will refer to such processes as composite stochastic processes). The GMWM approach also offers an alternative to the likelihood-based estimation

\begin{table}[h]
\centering
\caption{Popular power-law processes with associated parameters. Note: the drift is removed by regression or filtering.}
\begin{tabular}{|l|l|l|l|}
\hline
Process & $f^\alpha$ & $\tau^\beta$ & Parameters \\
\hline
Quantization Noise & $f^2$ & $\tau^{-1}$ & $Q$ coefficient \\
White Noise & $f^0$ & $\tau^{-1/2}$ & $\sigma^2$ power \\
Bias Instability & $f^{-1}$ & $\tau^0$ & $B$ coefficient \\
Random Walk & $f^{-2}$ & $\tau^{0.5}$ & $\gamma^2$ power \\
Ramp (drift) & See Note & $\tau^1$ & $\omega$ ramp slope \\
\hline
\end{tabular}
\end{table}

\footnote{The first order backward difference of $Y_t$ is $Y_t^{(1)} = Y_t - Y_{t-1}$ and the backward difference of order $d$ is $Y_t^{(d)} = Y_t^{(d-1)} - Y_{t-1}^{(d-1)}$}

S. Guerrier & M.-P. Victoria-Feser are with the Research Center for Statistics, HEC-Genève, University of Geneva. Address: 40, Bd du Pont d’Arve, 1211 Geneva 4, Switzerland, Tel: +41 (0)22 379 88 02 Emails: Stephane.Guerrier@unige.ch, Maria-Pia.VictoriaFeser@unige.ch.
Y. Stebler & J. Skaloud are with the Geodetic Engineering Laboratory, École Polytechnique Fédérale de Lausanne (EPFL), Station 18, CH-1015 Lausanne, Switzerland. Emails: yannick.stebler@epfl.ch, jan.skaloud@epfl.ch
Manuscript received April 19, 2005; revised January 11, 2007.
This method is straightforward to implement and often the only feasible estimation method when the models \( F_\theta \) are complex. The GMWM exploits the unique relationship that exists between \( F_\theta \) and the vector of Wavelet Variance (WV), denoted \( \nu = [\nu_t]_{t \in \mathcal{D}} \) such that \( \mathcal{D} \) denotes the set of dyadic scales i.e. \( \mathcal{D} = \{ \tau_j | j \in \mathbb{N^*}, j \leq J \} \). Each element \( \nu_t \) can typically be estimated by the variance of Maximal Overlap (MO-) Discrete Wavelet Transform (DWT) coefficient series obtained by filtering \( \{y_t\} \) with some wavelet filter at level \( j \) \([10, 16]\). In the context of the GMWM, the filter usually taken is the Haar filter, implying that the resulting WV \( \nu_t \) is equal to half the AV, \( \nu^2_t \). The GMWM is based on the principle that the WV vector for the set of scales \( \tau \subseteq \mathcal{D} \) implied by \( F_\theta \), noted as \( \nu(\theta) \), is a mapping

\[
\theta \rightarrow \nu(\theta), \ \theta \in \Theta.
\]

In some sense, the GMWM approach inverses this mapping by trying to approximate the point \( \theta(\hat{\nu}) \) where \( \hat{\nu} = [\hat{\nu}_t]_{t \in \mathcal{T}} \) is the WV computed on \( \{y_t\} \). In other words, the GMWM aims to find the value of \( \theta \) implied by \( \hat{\nu} \). For that, the distance between \( \hat{\nu} \) and \( \nu(\theta) \) has to be minimized. The solution of this optimization problem corresponds to the point where \( \nu(\theta) \) is the closest possible approximation of \( \hat{\nu} \). The GMWM has already been successfully applied to the modeling of inertial sensor errors in \([20]\). This work has shown by means of simulation and emulations that the GMWM clearly outperforms AV estimators as well as Maximum Likelihood Estimators (MLE) (based on Expectation-Maximization algorithm \([19]\)) and other variance-based approaches (e.g. Hadamard or Total variance). In this context, the practical implications of the GMWM estimators are considerable because of the rapidly growing employment of small and inexpensive Micro-Electro-Mechanical System (MEMS) - Inertial Measurement Units (IMUs) whose error behavior are rather complex. Such sensors are used in a very wide range of applications such as 3D input devices, robotic, virtual reality, vehicle stability control, and so forth \([5]\). They are often used together with satellite navigation system (e.g. GPS) and the performance of the overall system very much depends on the relevance of the state-space modeling of the inertial errors. In the case of MEMS-IMUs there are often represented by composite stochastic processes for which traditional estimators (e.g. the AV or the MLE) are often not suitable.

The purpose of this paper is twofold. First, we formally confirm the results of \([20]\) by showing that the estimator based on the standard AV is in general inconsistent and that there always exist a GMWM estimator with smaller (asymptotic) variance. Second, we show that for simple models where the AV or the MLE can also be used, the GMWM estimation may have lower finite sample performance. With this respect, we introduce several improvements to the classical GMWM estimator such that its performance almost reaches the ones of the MLE. The rest of the paper is organised as follows. In Sec. II, we summarize the most important notations and conventions used throughout the text and define the main stochastic processes that are considered in this work for building \( F_\theta \). Sec. III, demonstrates the important consistency properties of the AV-based estimator. The outcoming results justify the use of the GMWM framework whose theoretical framework is described in Sec. IV. The relationship between the AV and the GMWM is demonstrated in Sec. V in which we also compare on a theoretical basis the AV with the GMWM-based estimator. Sec. VI handles the second important issue of this paper by proposing several extensions to the classical GMWM. A simulation study is then presented in Sec. VII that compares the performance of the classical with the modified GMWM estimator. Finally, Sec. VIII concludes.

II. IMPORTANT NOTATIONS, CONVENTIONS AND DEFINITIONS

A. Conventions and Notations

Conventions

- \( \{x_t\} \) refers to a sequence of values indexed by integer \( t \)
- \( x_t \) refers to the \( t \)th value of a sequence
- \( Y_t \) refers to a random variable indexed by integer \( t \)
- \( y_t \) refers to a realization of \( Y_t \) indexed by integer \( t \)

Important Notations

- \( F_\theta \) data generating model parameterized by \( \theta \)
- \( S_Y(\cdot) \) power spectral density function of \( \{Y_t\} \)
- \( \nu(\tau) \) wavelet variance at scale \( \tau \)
- \( \nu(\theta) \) wavelet variance implied by \( \theta \) assuming that \( F_\theta \) corresponds to the true data generating process
- \( \phi(\tau) \) Allan deviation of \( \{Y_t\} \) at scale \( \tau \)
- \( \phi(\theta) \) Allan deviation implied by \( \theta \) assuming that \( F_\theta \) corresponds to the true data generating process
- \( \phi^2(\tau) \) Allan variance of \( \{Y_t\} \) at scale \( \tau \)
- \( \phi^2(\theta) \) Allan variance of \( \{Y_t\} \) implied by \( \theta \)
- \( \tau_j \) dyadic wavelet/Allan variance scale (\( \tau_j = 2^{j-1} \))
- \( J \) maximum number of scales \( \tau_j \) for \( \tau \in \mathcal{D} \)
- \( J \) subset of \( \mathcal{D} \) such that \( \tau \subseteq \mathcal{D} \)
- \( \theta \) \((p \times 1)\) parameter vector such that \( \theta \in \Theta \subset \mathbb{R}^p \)
- \( |\cdot| \) denotes the cardinality
- \( \text{mod} \) denotes the modulus
- \( \text{dim} \) denotes the dimension of a vector

B. Stochastic Processes Definitions

In the sequel, we define the main stochastic processes considered in this work (see also Tab. I).

(P.1) **Gaussian White Noise (WN)** with parameter \( \sigma^2 \in \mathbb{R}^+ \). This process is defined as

\[
Y_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2)
\]

and has the following implied Haar WV at scale \( \tau \):

\[
\nu(\sigma^2) = \frac{\sigma^2}{2\tau}
\]

(P.2) **Quantization Noise (QN)** or rounding error (see e.g. \([15]\)) with parameter \( Q^2 \in \mathbb{R}^+ \). This process \( Y_t \) has a PSD of the form

\[
S_Y(f) = 4Q^2 \sin^2 \left( \frac{\pi f}{2\Delta t} \right) \Delta t, \ f < \frac{\Delta t}{2}
\]
and has the following implied Haar WV at scale $\tau$:
\[
\nu_\tau(Q) = \frac{3Q^2}{2\tau^2}
\]

(P.3) Drift with parameter $\omega \in \Omega$ where $\Omega$ is either $\mathbb{R}^+$ or $\mathbb{R}^-$. This process is defined as:
\[
Y_t = \omega t
\]
and has the following implied Haar WV at scale $\tau$:
\[
\nu_\tau(\omega) = \frac{\tau^2 \omega^2}{4}
\]

(P.4) Random walk (RW) with parameter $\gamma^2 \in \mathbb{R}^+$. This process is defined as:
\[
Y_t = \sum_{i=1}^{T} \gamma Z_t
\]
where $Z_t \overset{iid}{\sim} N(0,1)$ and this process has the following implied Haar WV at scale $\tau$:
\[
\nu_\tau(\sigma^2) = \left(\frac{2\tau^2 + 1}{12\tau}\right) \gamma^2
\]

(P.5) Auto-Regressive AR(1) process with parameter $\rho \in \Phi$ where $\Phi$ is either $(-\infty, -1)$ or $(-1, +\infty)$ and $\upsilon^2 \in \mathbb{R}^+$. This process is defined as:
\[
Y_t = \rho Y_{t-1} + \varepsilon_t \text{ where } \varepsilon_t \overset{iid}{\sim} N(0, \upsilon^2)
\]
and it has the following implied Haar WV at scale $\tau$:
\[
\nu_\tau(\rho, \upsilon^2) = \left(\frac{\tau^2 (1 - \rho^2) (1 - \rho^2)}{\tau^2 + 4} \right)
\]

III. ALLAN VARIANCE BASED ESTIMATION

As stated in Eq. (1), the AV of the process $\{Y_t, t \in \mathbb{Z}\}$ is linked to its PSD. Therefore, if $Y_t$ is such that $Y_t \sim F_\theta$, then $\phi_\tau^2 = f(\theta, \tau)$ and we can denote $\phi_\tau(\theta)$ as the AD implied by the model $F_\theta$. Analytical relationships for $f(\cdot)$ can be computed for the stochastic processes listed in Tab. I. Regressions on the linear parts of the $\log(\tau)$ - $\log(\phi_\tau(\theta))$ plot with slopes $\beta$ enable to estimate the parameter $\theta$ of the associated $\tau^\beta$ (or $\upsilon^\alpha$) power-law process (see e.g. [13]). For notational simplicity we shall omit the term $\tau$ in $f(\theta, \tau)$ and write instead $f(\theta)$ and we let $\tau \subseteq D$ denote the set scales considered.

In this section, we define precisely the principle of AV-based estimation for a single process as well as composite processes, respectively. Note that we consider only processes producing power-law noises that have a linear representation in the $\log(\tau)$ - $\log(\phi_\tau(\theta))$ plot (e.g. the processes listed in Tab. I) as the processes with more complex AD representation can not be rigorously estimated with this approach. In this section we will show under which conditions AV-based estimators are consistent. To derive these results we will make the following assumptions:

(A.1) $\hat{\phi}_\tau$ is a consistent estimator of $\phi_\tau$, $\tau \in \tau$.

(A.2) If $\{Y_t\}$ is a composite process then all sub-processes are independent.

(A.3) The AV of each processes composing $\{Y_t\}$ are such that $\hat{\phi}_\tau^2 > 0$, $\tau \in \tau$.

Remark A. Assumption (A.2) is very useful in practice as it enables to treat the AV (or WV) of each of the processes producing a composite process separately. This assumption is almost always implicitly made by practitioners even if most of the time it is not clearly stated. More formally, if $\{Y_t\}$ is composed of $M$ independent processes we may write $Y_t = \sum_{m=1}^{M} Y_{t}^{(m)}$ where $Y_{t}^{(m)}$ represents the $m^{th}$ sub-process. By the properties of the PSD we have that $S_{Y}^{(f)} = \sum_{m=1}^{M} S_{Y}^{(m)}(f)$ and using Eq. (1) (or Eq. (14) for the WV) we obtain that the AV (or the WV) of $\{Y_t\}$ is simply the sum of the AV (or the WV) of each of the sub-processes.

Remark B. Assumption (A.3) is in fact very intuitive and Processes (P.1) to (P.5) as well as ARIMA models and other common time series models satisfy this assumption except when their variance is equal to 0. However, this assumption will be necessary to show the inconsistency of the estimation methods based on the AV.

A. AV-based Estimation for a Single Process

Suppose that the process $\{Y_t\}$ generated by $F_\theta$ with $\theta$ a unique parameter has linear representation in a $\log(\tau)$ - $\log(\phi_\tau(\theta))$ plot for the set of scales $\tau \subseteq G$ where
\[
G = \{\tau_k, \tau_{k+1}, ..., \tau_{k+h} | k, h \in \mathbb{N}^+, k + h \leq J\}
\]
denotes all possible sets which contain adjacent scales having a cardinality larger or equal than $p$ (i.e. 1 in this case). As a consequence of the linear relationship that exists between $\log(\tau)$ and $\log(\phi_\tau(\theta))$ we can write
\[
\log(\phi_\tau(\theta)) = f(\theta) + \beta \log(\tau), \quad \tau \in \tau
\]
where the function $f(\cdot)$ as well as the constant $\beta$ are known and depend through Eq. (1) on the assumed model $F_\theta$. However, when the estimated AD, noted $\hat{\phi}_\tau$, is computed from an observed finite sample $\{y_t, t = 1, \ldots, T\}$, we obtain
\[
\log(\hat{\phi}_\tau) = \log(\phi_\tau(\theta)) + \varepsilon_\tau, \quad \tau \in \tau
\]
where $\varepsilon_\tau$ are residuals. This linear relationship leads to some sort of AV-based “least-squares” estimator for $\theta$, noted $\hat{\theta}_{AV}$, which can be characterised as the solution of
\[
\hat{\theta}_{AV} = \arg\min_{\theta \in \Theta} \sum_{\tau \in \tau} \varepsilon_\tau^2.
\]
Moreover, since $f(\cdot)$ and $\beta$ are known for a given $F_\theta$ satisfying Eq. (2), $\hat{\theta}_{AV}$ can be defined explicitly as
\[
\hat{\theta}_{AV} = f^{-1}\left\{\frac{1}{|\tau|} \sum_{\tau \in \tau} \left[\log(\hat{\phi}_\tau) - \beta \log(\tau)\right]\right\},
\]
In the following example we present how Eq. (4) can be applied to processes (P.1), (P.2) or (P.3).
Example 1. If $Y_t$ follows (P.1), (P.2) or (P.3), then by using Eq. (4), we obtain the following estimators for the corresponding $\theta$: 
\[
\hat{\sigma}_AV^2 = \exp \left\{ \frac{2}{|\tau|} \sum_{\tau \in \tau} \log \left( \hat{\phi}_\tau \right) + \frac{1}{2} \log (\tau) \right\}, \\
\hat{Q}_AV = 1 + \exp \left\{ \frac{2}{|\tau|} \sum_{\tau \in \tau} \log \left( \hat{\phi}_\tau \right) + \log (\tau) \right\}, \\
\hat{\omega}_AV = \sqrt{2} \exp \left\{ \frac{1}{|\tau|} \sum_{\tau \in \tau} \log \left( \hat{\phi}_\tau \right) - \log (\tau) \right\}.
\]

The following Proposition 1 demonstrates that in these cases the AV-based estimator $\hat{\theta}_{AV}$ is consistent (the proof is given in App. A).

**Proposition 1.** For any $\tau \in G$ and if $Y_t$ follows either (P.1), (P.2) or (P.3), the AV-based estimator $\hat{\theta}_{AV}$, as defined in Eq. (4), is a consistent estimator of $\theta$ under Assumption (A.1).

When considering simple processes such as (P.1), (P.2) or (P.3), using Eq. (4) is certainly not the best possible option for estimating $\theta$ as this estimator is far less efficient than standard MLE. We will come back to this statement later in Sec. VII.

Regarding (P.4), the AD is often approximated by $\sqrt{2} \nu (\gamma^2) \approx \sqrt{3} \gamma \tau$ in the engineering literature (see e.g. [4]). This approximation then satisfies Eq. (2) and the AV estimator for $\gamma^2$ can hence be defined as
\[
\hat{\gamma}_AV^2 = \exp \left\{ \frac{2}{|\tau|} \sum_{\tau \in \tau} \log \left( \hat{\phi}_\tau \right) - \frac{1}{2} \log (\tau) \right\}.
\]

Although $\hat{\gamma}_AV^2$ may provide satisfactory results in practice and specially for large scales, the next Corollary demonstrates that it does not converge to $\gamma^2$ but to $\hat{\gamma}_AV^2 > \gamma^2$ (the proof is given in App. B).

**Corollary 1.** If $\sup_{\tau \in \tau} (\tau) < \infty$ and under Assumption (A.1), we have that $\hat{\gamma}_AV^2$ as defined in Eq. (5) is such that $\hat{\gamma}_AV^2, \gamma^2 \rightarrow \hat{\gamma}_AV^2 > \gamma^2$ and therefore that $\hat{\gamma}_AV^2$ is not a consistent estimator of $\gamma^2$.

From Corollary 1 one can deduce that a consistent estimator of $\gamma^2$ can be obtained by introducing the following correction:
\[
\hat{\gamma}_b = c_AV \hat{\gamma}_AV,
\]
with
\[
c_AV^{-1} = \exp \left\{ \frac{2}{|\tau|} \sum_{\tau \in \tau} b_AV (\tau) \right\}.
\]
However, we must again point out that for such a simple process using Eq. (5) or (6) is certainly not the best option to estimate the variance of a random walk.

**B. AV-based Estimation for Composite Stochastic Processes**

In this section, we consider the situation in which we use Eq. (4) to estimate the parameters of a model $F_\theta$ made of at least two processes among (P.1), (P.2), (P.3) and (P.4). Let $M$ denote the number of stochastic processes such that $2 \leq M \leq 4$. Let also $\tau_m \subset G$, $m = 1, ..., M$ denote the scales on which the $m^{th}$ element of $\theta$, say $\theta_m$, will be estimated (i.e. apply Eq. (4) to $\tau_m$). Thus, we have $\hat{\theta} = \left[ \hat{\theta}_m \right]_{m=1,..,M}$ where
\[
\hat{\theta}_m = f^{-1} \left\{ \frac{1}{|\tau_m|} \sum_{\tau \in \tau_m} \log \left( \hat{\phi}_\tau \right) - \beta \log (\tau) \right\}.
\]

As demonstrated in the following proposition, when $\{Y_t\}$ is a composite stochastic process (with $M > 1$), $\hat{\theta}$ is not consistent (the proof is given in App. C).

**Proposition 2.** Suppose that $\hat{\phi}_\tau$ satisfies Assumption (A.1) and that the process $\{Y_t\}$ is a composite stochastic process composed of $M > 1$ different sub-processes satisfying Assumption (A.2). Assume further that at least one of them is either (P.1), (P.2), (P.3) or (P.4) with parameter $\theta$ and the other process(es) satisfy Assumption (A.3). Then, the estimator $\hat{\theta}_{AV}$ as defined in Eq. (7) is not a consistent estimator of $\theta$, for any $\tau \subset G$.

As demonstrated in Proposition 2, AV-based estimators are generally inconsistent when applied to composite processes and their use is not justified for the estimation of single processes (as they are clearly outperformed by MLE). In this context, we believe that such methods are not optimal when used in practice.

The next example illustrates the case of a simple composite process $Y_t$ composed of a sum of (P.1) and (P.4) that we wish to estimate using the AV estimation method.

**Example 2.** Assume that we observe a process $\{Y_t, t \in \mathbb{Z}\}$ which is driven from the following model $F_\theta$
\[
Y_t = Y_{RW,t} + Y_{WN,t},
\]
where $Y_{WN,t}$ and $Y_{RW,t}$ are as defined in (P.1) and (P.4), respectively. Therefore, we have $\theta = [\sigma^2 \gamma^2]$ and the theoretical AV of this system is given by
\[
\hat{\phi}_\tau^2 = \frac{6\sigma^2 + (2\tau^2 + 1) \gamma^2}{6\tau}.
\]

The true log ($\hat{\phi}_\tau^2$) value is unknown in practice and only its estimated quantity $\log \left( \hat{\phi}_\tau^2 \right)$ is available. Suppose we are interested in estimating $\sigma^2$. In this case, the AV methodology would consist in applying a linear regression on the $j$ first scales in order to estimate $\sigma^2$, i.e.
\[
\log \left( \hat{\phi}_\tau^2 \right) = \log \left( \phi_\tau^2 \right) + \varepsilon_\tau, \quad \tau = 1, ..., j.
\]

This concept is illustrated in Fig. 1 showing the AV sequence of a simulated signal (presented in the top panel) $\{y_t, t = 1, \ldots, T\}$ with $T = 10^6$ issued from Eq. (8). In this case, linear regression would be performed on the $j = 5$ first scales (“full” black dots in the lower panel) for estimating $\sigma^2$. Since we are interested in estimating $\sigma^2$, $\log \left( \phi_\tau^2 \right)$ in Eq. (10) is replaced by the AV implied by such process, i.e.
\[
\log \left( \phi_\tau^2 \right) = \log \left( \frac{\sigma^2}{\tau} \right) = \log (\sigma^2) - \log (\tau).
\]
which demonstrates that in our example $\hat{\sigma}_{AV}^2$ is a consistent estimator of $\sigma^2$ if $\gamma^2 = 0$.

2) If $\gamma^2 > 0$, then $\frac{2(i+1)\gamma^2}{6i} > 0$ in Eq. (13) and we have

$$\lim_{T \to \infty} \hat{\sigma}_{AV}^2 = \sigma^2 + c, \quad \text{with} \quad c > 0.$$ 

This last inequality confirms that the conventional AV methodology does not provide a consistent estimator for $\sigma^2$ when $\gamma^2 > 0$. In other words, $\hat{\sigma}_{AV}^2$ is a consistent estimator of $\sigma^2$ if and only if $\gamma^2 = 0$. This illustrates the results of Propositions 1 and 2.

C. Simulation Study

To further illustrate the inconsistency of AV-based estimators, we conducted a simple simulation study under the same settings as in Example 2. We simulated 500 signals $\{y_t, t = 1, \ldots, T\}$ under two models: the first is (P.1) with $\sigma^2 = 4$, while the second is a sum of (P.1) with $\sigma^2 = 4$ and (P.4) with $\gamma^2 = 0.01$. Three values for $T$ were used, namely $T = 10000$, $T = 100000$ and $T = 1000000$. According to Fig. 1, we computed the AD sequence, $\hat{\phi} = \frac{\tau_{AV}}{\tau}$, and performed the linear regression on the $j = 5$ first scales for each of the simulated signal $\{y_t\}$ in order to estimate $\hat{\sigma}^2$. The results of the estimation are presented in the boxplots of Fig. 2. The true value of the $\sigma^2$ is drawn as a thick line.

![Fig. 1. Simulated signal $\{y_t: t = 1, \ldots, T\}$ with $T = 10000$ issued from a sum of a white noise (with $\sigma^2 = 4$) and a random walk (with $\gamma_{AV}^2 = 0.01$) (see Eq. (8)) (top panel) and the estimated AV based on this simulated signal presented in a $\log(\tau)$ - $\log(\phi_{AV})$ form (lower panel). The points $\bullet$ represent the region on which the linear regression is performed in order to estimate the white noise power $\sigma^2$.](image1)

![Fig. 2. Result of the AV based estimation of $\sigma^2$ (true value depicted as a black horizontal line) when performed on 500 simulated signals composed of (P.1) with $\sigma^2 = 4$ (denoted as “WN” on the three first boxplots) and composed of the sum of (P.1) with $\sigma^2 = 4$ and (P.4) with $\gamma^2 = 0.01$ (denoted as “WN+RW” on the three next boxplots). Three signal lengths $T$ were used: 1000 (first and fourth boxplots), 10000 (second and fifth boxplots) and 100000 (third and sixth boxplots) samples.](image2)
the values of \( \hat{\sigma}_W^2 \), but this time based on signals issued from the second (WN+RW) model. There is a clear bias that does not vanish asymptotically which indicates the inconsistency of the estimator in this case. This empirically confirms the result of Proposition 2 and the developments of Example 2 claiming that the AV is an inconsistent estimator when several processes are present.

IV. THE STANDARD GMWM

A. The Wavelet Variance

The WV, \( \nu_{\tau_j} \), can be defined for \( j = 1, \ldots, J \) as

\[
\nu_{\tau_j} = \var(W_{j,t})
\]

where \( \{W_{j,t}, j = 1, \ldots, J, t \in \mathbb{Z}\} \) are series of wavelet coefficients issued from a MODWT of \( \{Y_t, t \in \mathbb{Z}\} \), i.e.

\[
W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{t-l}, t \in \mathbb{Z}.
\]

The MODWT wavelet filters \( \{\tilde{h}_{j,l}\} \) are actually a rescaled version of the DWT filter \( h_{j,l} \), i.e. \( h_{j,l} = \tilde{h}_{j,l}/2^{j/2} \). Note that the WV are assumed not to depend on time. The condition for this property to hold is that the differentiation order \( d \) for the series \( \{Y_t\} \) to be stationary is such that \( d \leq L_j/2 \) and \( \{\tilde{h}_{j,l}\} \) is based on a Daubechies wavelet filter (see [3] and [16], Chapter 8). In such a case, \( \{W_{j,t}\} \) is stationary with PSD \( S_{W_j}(f) = \text{mod} \left( \tilde{H}_j(f) \right)^2 S_{F_\theta}(f) \) where \( S_{F_\theta}(\cdot) \) is the PSD generated by \( F_\theta \) and \( \tilde{H}_j(\cdot) \) the transfer function of \( \tilde{h}_{j,l} \). This means that the variance of the \( \{W_{j,t}\} \) series is equal to the integral of its PSD [18], i.e.

\[
\nu_{\tau_j}(\theta) = \int_{-1/2}^{1/2} S_{W_j}(f) \, df = \int_{-1/2}^{1/2} \text{mod} \left( \tilde{H}_j(f) \right)^2 S_{F_\theta}(f) \, df
\]

Hence, as for the AV there is an implicit link between the WV and the parameters of \( F_\theta \). The GMWM exploits this connection when defining an estimator for \( \theta \), namely by matching a sample estimate of \( \nu_{\tau_j} \) together with the model-based expression of the WV given by Eq. (14). For WV based on Haar wavelet filters and for the processes considered in Sec. II, the integral in Eq. (14) can be solved analytically (see [12] for their expression which were derived using the general results of [25]).

For a finite (observed) process \( \{y_t, t = 1, \ldots, T\} \), the MODWT WV estimator given by

\[
\hat{\nu}_{\tau_j} = \frac{1}{M_j} \sum_{t=T}^{T} W_{j,t}^2
\]

with \( W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{t-l}, t \in (L_j; T) \) and \( M_j = T - L_j + 1 \), is a consistent estimator for \( \nu_{\tau_j} \) and is asymptotically normally distributed (see [18]). With this respect, [12] demonstrated that under some regularity conditions, the asymptotic distribution of \( \hat{\nu}_{\tau} \) associated to the set of scales \( \tau \) is given by

\[
\sqrt{T} \left( \hat{\nu}_{\tau} - \mathbb{E}[\nu_{\tau}] \right) \overset{d}{\rightarrow} \mathcal{N}(0, \Sigma_{\nu_{\tau}})
\]

which extends the univariate results of [18]. The elements of the WV covariance matrix

\[
\Sigma_{\nu_{\tau}} = \left[ \sigma_{ml}^{2} \right]_{m,l=1,\ldots,J}
\]

can be obtained through \( \sigma_{ml}^{2} = 2\pi S_{W_m W_l}(0) \) for \( m, l = 1, \ldots, J \) with

\[
S_{W_m W_l}(f) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \rho_{W_m W_l}(\tau) e^{-i f \tau}.
\]

and \( \rho_{W_m W_l}(\tau) = \text{cov} \left[ W_m, W_{l,t+\tau} \right] \) is the cross-covariance function. [12] show that under the assumption of a Gaussian process for \( \{Y_t\} \), a possible estimator for \( \sigma_{ml}^{2} \) is given by

\[
\hat{\sigma}_{ml}^{2} = \frac{1}{2} \sum_{\tau=-M(T_{ml})}^{M(T_{ml})} \left[ \frac{1}{M(T_{ml})} \sum_{t \in T_{ml}} W_{m,t} W_{l,t+\tau} \right]^2
\]

\[
+ \frac{1}{2} \sum_{\tau=-M(T_{ml})}^{M(T_{ml})} \left[ \frac{1}{M(T_{ml})} \sum_{t \in T_{ml}} W_{m,\tau} W_{l,t} \right]^2
\]

where \( T_{ml} \) is the smallest set of time indices containing both the indices in \( T_m \) and \( T_l \) (see Eq. (15)), and \( M(T_{ml}) \) their number. However, as noted in [12], the elements of \( \Sigma_{\nu_{\tau}} \) can be estimated in a more computationally efficient manner by parametric bootstrap.

B. The GMWM Estimator

Let \( S = \left\{ \mathcal{X} \subseteq \mathcal{P}(\mathcal{D}) \mid |\mathcal{X}| \geq p \right\} \) denote all possible subsets of \( \mathcal{D} \) such that their cardinality is larger or equal than \( p \). Then, for the scales \( \tau_j \) contained in the set \( \eta \subset S \), the GMWM aims to find the value of \( \theta \in \Theta \subseteq \mathbb{R}^p \) by minimising the distance between the vector of empirical WV, denoted as \( \hat{\nu}_{\eta} \), and the WV implied by \( F_\theta \), denoted as \( \nu_{\eta}(\theta) \). This optimisation problem is written as

\[
\hat{\theta}_{\eta} = \arg\min_{\theta \in \Theta} ||\hat{\nu}_{\eta} - \nu_{\eta}(\theta)||^2_{\Omega}
\]

where the norm \( ||x||^2_{\Omega} \) denotes the quadratic form \( x^T \Omega x \) and \( \Omega \) is a positive definite weighting matrix chosen in a suitable manner (see [12] for details). The solution of Eq. (17) corresponds to the point where \( \nu_{\eta}(\theta) \) is the closest possible approximation of \( \hat{\nu}_{\eta} \) with respect to the considered norm. It was shown in [12, Corollary 3] that, under some regularity conditions, the GMWM estimator \( \hat{\theta}_{\eta} \) is consistent for the parameters of a stochastic process composed of the sum of one or more but only one of each of the processes \( P_1 \) to \( P_4 \), and \( k \in \mathbb{N} : k < \infty \) processes \( P_5 \). In addition, under some conditions defined in [12], \( \hat{\theta}_{\eta} \) has the following asymptotic distribution:

\[
\sqrt{T} \left( \hat{\theta}_{\eta} - \theta_{\eta} \right) \overset{d}{\rightarrow} \mathcal{N}(0, \Sigma_{\theta_{\eta}})
\]

where \( \Sigma_{\theta_{\eta}} = \text{var}(\hat{\nu}_{\eta} B^T), B = (D^T \Omega)^{-1} D^T \Omega, D = \partial \nu_{\eta}(\theta)/\partial \theta^T \) and \( \hat{\nu}_{\eta} \) is the asymptotic covariance matrix of \( \hat{\nu}_{\eta} \) as expressed in Eq. (16).
In the next section, we show how the AV estimator is related to the GMWM estimator in the case of a single process composing \( \{Y_t\} \).

V. LINK BETWEEN THE AV AND THE GMWM

If we consider the case of a signal \( \{Y_t, t \in \mathbb{Z}\} \) composed of a single process, the AV estimator \( \hat{\theta}_{\text{AV}} \) defined by Eq. (4) can be seen as a special case of the GMWM estimator \( \hat{\theta}_{\eta} \) developed in the previous section. This statement is formally shown in the next proposition (the proof is given in App. D).

**Proposition 3.** Let the process \( \{Y_t, t \in \mathbb{Z}\} \) follow either (P.1), (P.2) or (P.3) for any scale vector \( \tau \subset \mathcal{G} \). Let also \( \hat{\theta}_{\text{AV}} \) and \( \hat{\theta}_{\tau} \) denote the AV and GMWM estimator defined by Eq. (3) and (17), respectively. Then, we have that \( \hat{\theta}_{\text{AV}} = \hat{\theta}_{\tau} \) when the matrix

\[
\Omega^* = [\omega_{i,j}^*]_{i,j=1,...,|\tau|}
\]

is diagonal with diagonal elements

\[
\omega_{i,j}^* = \begin{cases} (\log(\phi_{t,i}) - \log(\phi_{t,j}))^2 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases} \tag{18}
\]

Proposition 3 also holds for (P.4) when using \( \phi_{t,\tau}(\theta) = \sqrt{c_{AV}}\gamma^{2T/3} \) in Eq. (18) and the consistent estimator \( \hat{\gamma}_{AV} = \frac{c_{AV}}{\text{var}(\phi_{t,\tau})} \).

Hence, when \( \{Y_t\} \) follows either (P.1), (P.2) or (P.3) (or (P.4) under the previously mentioned conditions), \( \hat{\theta}_{AV} \) (or \( \hat{\gamma}_{AV}^2 \)) are in fact GMWM estimators of \( \theta \). Therefore, they inherit their asymptotic properties, i.e. consistency and asymptotic normality. However, depending on the matrix \( \Omega \) on which the GMWM estimator is based, it can be more or less asymptotically efficient. Corollary 2 shows that there always exists a GMWM estimator that is asymptotically more efficient than the AV estimator given in Eq. (4). The proof is given in App. E.

**Corollary 2.** Consider the same setting as in Proposition 3. Then we have that there always exist at least one matrix \( \Omega^0 \) such that

\[
\lim_{T \to \infty} \frac{\text{var}(\hat{\theta}_{\tau})}{\text{var}(\hat{\theta}_{\text{AV}})} < 1, \quad \tau \subset \mathcal{G}
\]

where \( \hat{\theta}_{\tau} \) denotes the GMWM estimator of \( \theta \) based on the matrix \( \Omega^0 \).

Note again that Corollary 2 is also valid for process (P.4).

The next section presents some improved version of the GMWM estimator introduced in [12].

VI. MODIFIED FORMS OF THE GMWM

Simulation studies presented in [12, 20, 21] demonstrated that GMWM estimators can present some bias in finite samples. In such situations and for models \( F_{\theta} \) which are sufficiently simple to be well estimated by AV-based estimators or by MLE, the GMWM may have poorer performances. In this section, we present three methods that improve the GMWM performances. The first is related to the addition of standard moments of \( Y_t \) (e.g. expectation, variance) in the GMWM optimisation problem described by Eq. (17). We refer to such a version of the GMWM as the augmented GMWM. The second improvement concerns a procedure for selecting adequate scales over which the GMWM estimation can be computed. Finally, the third modification is a procedure which corrects the bias in the GMWM estimates resulting from finite samples. This last method is general and can therefore be applied beyond the scope of the GMWM framework.

A. Augmented GMWM

We define the augmented GMWM estimator based on the set of scales \( \eta \) as

\[
\hat{\theta}_{\eta} = \arg\min_{\theta \in \Theta} (\hat{\nu}_{\eta}^+ - \nu_{\eta}(\theta))^T \Omega^+ (\hat{\nu}_{\eta}^+ - \nu_{\eta}(\theta)) \tag{19}
\]

where \( \hat{\nu}_{\eta}^+ = [\hat{m}_{l, \eta}, \nu_{\eta}^+ = [\nu_l(\theta), \nu_l(\theta)] \) with \( \hat{m}_l \) and \( m_l \) denoting vectors containing \( l \) moments or functions of the moment of \( \{Y_t\} \). The former, \( \hat{m}_l \), contains the empirical moments while the latter, \( m_l \), the moments implied by \( F_\theta \).

As shown in Proposition 4, the augmented GMWM enjoys the same properties as the classical GMWM. The proof of this results is essentially identical to that of [12]. We shall therefore omit the proof.

**Proposition 4.** Under the condition of [12, Theorem 2] and if the matrix of the partial derivatives \( \partial \nu_{\eta}^+ / \partial \theta \) is of full rank, then for \( \eta \subset \mathcal{S} \), \( \hat{\theta}_{\eta} \) as defined in Eq. (19) provides a consistent estimator for \( \theta \).

In addition, if \( \sqrt{T} (\hat{\nu}_{\eta}^+ - \nu_{\eta}^+) \xrightarrow{\mathcal{L}} N(0, \Omega^\eta \nu_{\eta}^+ \nu_{\eta}^+ \Omega^\eta) \) then \( \hat{\theta}_{\eta}^+ \) is asymptotically normally distributed with covariance matrix \( \Omega^\eta = (\nu_{\eta}^+ \Omega^\eta \nu_{\eta}^+)^{-1} \nu_{\eta}^+ \Omega^\eta \nu_{\eta}^+ \Omega^\eta \) and \( \nu_{\eta}^+ = \partial \nu_{\eta}^+ / \partial \theta \).

B. Optimal Scales Selection

In its standard form, the computation of the GMWM estimator is based on all possible scales. However, this choice may not always be optimal as neighbouring scales are generally extremely correlated (an empirical illustration of this statement can e.g. be found in [20, Appendix B]). Selecting optimal scales for the GMWM estimation can for example be achieved by minimising the determinant of the asymptotic covariance matrix of the resulting estimator, denoted as \( \nu_{\eta}^+ \), i.e.

\[
\eta^* = \arg\min_{\eta \subset \mathcal{S}} \det (\nu_{\eta}^+) \tag{20}
\]

Clearly, the GMWM based on Eq. (20) is consistent and asymptotically normally distributed as a consequence of [12, Theorem 2]. Note that the same selection can also be done by taking into account the moments in the estimator (i.e. with the augmented GMWM).

C. Finite Sample Bias Correction

In this section we propose a method for bias correction in finite sample. This method uses the principle of indirect inference [8] which can be used to correct asymptotic bias. For
example, in the framework of robust statistics $M$-estimators often require the computation of correction terms that ensure the consistency of the estimators. When these correction terms are difficult to compute, indirect inference can be used to correct the asymptotic bias of an inconsistent $M$-estimator. This strategy was, for example, used in [14] who proposed a computationally efficient robust estimator for generalized linear latent variable models.

We will prove here that the principle of indirect inference can also be used to correct finite sample bias. Note that this idea is not totally new, as it was already partially investigated in [9]. This approach will be used to correct the finite sample bias of the GWMW, nevertheless the method is general and can therefore be employed outside of this context. To avoid unnecessary confusion, we use the notation $\hat{\theta}_0$ rather than $\theta$ to denote the true values of the parameter vector. Therefore, we define $\hat{\theta}$ as an estimator of $\theta_0 \in \Theta \subseteq \mathbb{R}^p$ based on a sample of size $T$ supposed to be generated from $F_{\theta_0}$. Assuming that $\mathbb{E} \{ \hat{\theta} \}$, where $\mathbb{E} \{ \cdot \}$ denotes the expectation under $F_{\theta_0}$, exists, we can always write
\[
\hat{\theta} = \theta_0 + Q b(\theta_0, T) + v(\theta_0, T) \tag{21}
\]
where $Q b(\theta_0, T) \equiv \mathbb{E} \{ \hat{\theta} - \theta_0 \}$ is a finite sample fixed bias that only depends on $\theta_0$ (and $T$), $Q$ is a $p \times p$ matrix of fixed coefficients that does not depend on $T$, and $v(\theta_0, T)$ is a random vector with expectation $0$ and variance $V_{\theta_0, T}$ and is such that Eq. (21) holds.

We define a bias corrected estimator $\hat{\theta}_c^H$ as:
\[
\hat{\theta}_c^H = \arg\min_{\theta \in \Theta} \| \theta - \hat{\theta}_c^H \|_F^2 \tag{22}
\]
where $\Phi_n$ is a symmetric positive definite matrix which converges almost surely to a symmetric positive definite matrix $\Phi$. Moreover, we have that
\[
\hat{\theta}_c^H F_{\theta_0} = \frac{1}{H} \sum_{h=1}^H \hat{\theta}_c^H F_{\theta_0} \tag{23}
\]
where $\hat{\theta}_c^H F_{\theta_0}$ denotes the value of $\hat{\theta}$ obtained on the $h$th simulated sample of size $T$ under $F_{\theta_0}$. Let also
\[
\lim_{H \to \infty} \hat{\theta}_c^H F_{\theta_0} = \hat{\theta}_c F_{\theta_0} \tag{24}
\]
Clearly, Eq. (22) is an indirect estimator in the sense of Gourieroux, Monfort and Trognon (see e.g. [8]). In order to determine the property of our bias correction method the following assumptions are required:

(A.1) The vector $\theta_0$ is such that $\theta_{0(j)} < \infty$, $j = 1, \ldots, p$.
(A.2) $\mathbb{E} \{ v(\theta, T) \} = 0$, $\forall T$, $\theta \in \Theta$.
(A.3) $\var \{ \hat{\theta} \} = V_{\theta, T} = T^{-1} V_{\theta}$.
(A.4) $\cov \{ v(\theta, T), b(\theta_c^H, T) \} = 0_{p \times p}$, $\forall T$.
(A.5) The bias function can be written as:
\[
b(\theta, T) = \left[ b_j \left( \theta_0(j), T \right) \right]_{j=1,\ldots,p}
\]
where
\[
b_j \left( \theta_0(j), T \right) = \sum_{i=0}^{q_j} a_{ij} \left( \frac{\theta_0(j)}{T} \right)^i < \infty
\]
(A.6) The matrix $Q$ and the vector $a_1 = \text{diag} \{ a_{1j} \}_{j=1,\ldots,p}$ are such that the inverse of the matrix $(I + \frac{1}{T} Q a_1)$ exists.

(A.7) The expectations $\mathbb{E} \left[ (\theta_j)^k \right] = m_k < \infty$ exist for $k = 1, \ldots, \max_{j=1,\ldots,p} (q_j)$.
(A.8) The coefficients $a_{ij}$ are such that they satisfy for $j, k = 1, \ldots, p$:
\[
\sum_{i=3}^{q_j} a_{ij} \frac{\theta_{(j)i}}{T^{i-3}} \left( \hat{\theta}_{(j)i} \right)^{i-3} - \mathbb{E} \left[ \left( \hat{\theta}_{(j)i} \right)^{i-3} \right] = d_j < \infty
\]
\[
\cov \left[ \sum_{i=1}^{q_j} a_{ij} \frac{\theta_{(j)i}}{T^{i-1}} \left( \hat{\theta}_{(j)i} \right)^{i-1}, \sum_{i=1}^{q_k} a_{ik} \frac{\theta_{(k)i}}{T^{i-1}} \left( \hat{\theta}_{(k)i} \right)^{i-1} \right] = c_{jk} < \infty
\]
where $\hat{\theta}_{(j)i}$ denotes the $j$th element of $\hat{\theta}_c$.

(A.9) The finite sample binding function $\pi(\theta) = \theta + b(\theta, T)$ is uniformly continuous and one-to-one in $\theta$.

(A.10) The support of the distribution of $\hat{\theta}$ is included in $\pi(\Theta)$.

Remark C. Assumption (A.3) is used to determine the convergence rate of $\hat{\theta}_c^H$ and might be replaced by the condition that the convergence rate of $\hat{\theta}$ is $T^{-1/2}$.

Under this setting we have the following result whose proof is given in App. F.

Proposition 5. Under Assumptions, (A.1) to (A.10), the corrected estimator $\hat{\theta}_c^H$ as defined in Eq. (22) is such that
\[
\mathbb{E} \left[ \lim_{H \to \infty} \hat{\theta}_c^H \right] = \mathbb{E} \{ \hat{\theta} \} = \theta_0 + O(T^{-3/2}) ,
\]
\[
\var \left[ \lim_{H \to \infty} \hat{\theta}_c^H \right] = \var \{ \hat{\theta} \} = \var \{ \theta \} + O(T^2) .
\]

VII. SIMULATIONS

The simulation study we present in this section aims to illustrate the theoretical findings presented in the former sections of this article. First, we compare the finite sample performances of the various GWMW estimators with the AV-based estimator (when applicable) and the MLE for estimating simple processes, namely that of a white noise and an AR(1) process. Second, we compare the performances of the different GWMW estimators still in a simple scenario yet for which the AV-based estimator is not applicable and the MLE presents a severe bias that makes its use in practice impossible (see e.g. [12]). The second setting corresponds to a composite process composed of the sum of an AR(1), a white noise and a drift.

A. White noise process

In this simulation we consider a white noise process (i.e. (P.1)) of length $T = 10^3$ with parameter $\sigma^2 = 4$ and compare the performances of the following estimators of $\sigma^2$:

- The classical GWMW estimator $\hat{\sigma}_n^2$ as defined in Eq. (17) based on all possible scales (i.e. $\eta = D$) and with $\Omega = V_{\text{diag}}$ where the elements of $V_{\text{diag}}$ are defined as
\[
(V_{\text{diag}})_{i,j} = \begin{cases}
\tilde{\text{var}}(\hat{\nu}_T) & \text{if } i = j \\
0 & \text{if } i \neq j
\end{cases}
\]

- The AV-based estimator $\hat{\sigma}_n^2$ defined in Example (1) ;

- The classical GMWM estimator $\hat{\theta}_P$ as defined in Eq. (17) based on all possible scales (i.e. $\eta = D$) and with $\Omega = V_{\text{diag}}$ where the elements of $V_{\text{diag}}$ are defined as
\[
(V_{\text{diag}})_{i,j} = \begin{cases}
\tilde{\text{var}}(\hat{\nu}_T) & \text{if } i = j \\
0 & \text{if } i \neq j
\end{cases}
\]
• The classical GMWM estimator $\hat{\theta}_{\eta}$ as defined in Eq. (17) based on “optimal” scales, i.e. $\eta = \eta^o$, as defined in Eq. (20), and with $\Omega = V^{-1}_{\text{diag}}$ (for the corresponding scales);
• The augmented GMWM estimator $\hat{\theta}_D^+$ as defined in Eq. (19) with $\nu_\eta^+ = [\hat{m}_2 \hspace{0.5cm} \hat{\nu}_\eta]$ based on all possible scales and with $\Omega = V^{-1}_{\text{diag}}$ (augmented with the covariance of the second moment and obtained by parametric bootstrap);
• The augmented GMWM estimator $\hat{\theta}_D^+$ as defined in Eq. (19) with $\nu_\eta^+ = [\hat{m}_2 \hspace{0.5cm} \hat{\nu}_\eta]$ based on all possible scales and with $\Omega = \text{var}(\hat{\nu}_D)^{-1}$ (obtained by parametric bootstrap);
• The bias correct GMWM estimator $\hat{\theta}_D$ as defined in Eq. (22) in which $\theta = \hat{\theta}_D$, $\Phi_n = 1$ and $H = 100$;
• The MLE $\hat{\sigma}^2_{\text{ML}}$.

Fig. 3 shows the empirical distribution of the different estimators obtained with 500 bootstrap replications and Tab. II reports the bias, the variances, the MSEs and the relatives efficiencies (to the MLE) of the various estimators presented above.

As expected, it can be observed that the MLE performs best in this setting. The approach based on the AV is by far the worst estimator with an MSE that is about 9 times larger than the MSE of the MLE. The standard GMWM estimator appears to suffer from a slight bias but the augmented GMWM as well as the bias correction method presented in Sec. VI-C seem to greatly reduce this bias. However, the selection of “optimal” scales for the GMWM estimator does not improve the estimator’s performance but rather seem an important bias.

<table>
<thead>
<tr>
<th>bias</th>
<th>var</th>
<th>MSE</th>
<th>eff.</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\hat{\sigma}^2_{\text{AV}}$</td>
<td>$-1.32 \cdot 10^{-1}$</td>
<td>$2.82 \cdot 10^{-1}$</td>
<td>$2.99 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>$\hat{\theta}_{\eta^o}$</td>
<td>$-1.19 \cdot 10^{-1}$</td>
<td>$6.12 \cdot 10^{-2}$</td>
<td>$7.54 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>$\hat{\theta}_D$</td>
<td>$-2.29 \cdot 10^{-1}$</td>
<td>$1.13 \cdot 10^{-1}$</td>
<td>$1.65 \cdot 10^{-1}$</td>
</tr>
<tr>
<td>$\hat{\theta}_D^+$</td>
<td>$8.72 \cdot 10^{-3}$</td>
<td>$3.50 \cdot 10^{-2}$</td>
<td>$3.5 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>$\hat{\theta}_D^+$</td>
<td>$7.99 \cdot 10^{-3}$</td>
<td>$3.56 \cdot 10^{-2}$</td>
<td>$3.57 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>$\hat{\theta}_D^+$</td>
<td>$7.37 \cdot 10^{-3}$</td>
<td>$6.62 \cdot 10^{-2}$</td>
<td>$6.63 \cdot 10^{-2}$</td>
</tr>
<tr>
<td>$\hat{\sigma}^2_{\text{ML}}$</td>
<td>$1.07 \cdot 10^{-2}$</td>
<td>$3.26 \cdot 10^{-2}$</td>
<td>$3.27 \cdot 10^{-2}$</td>
</tr>
</tbody>
</table>

The two versions of the augmented GMWM presented in this simulation have an efficiency that is comparable to the MSE of the MLE.

B. First order autoregressive process

In the second simulation we consider an AR(1) process (i.e. (P.5)) of length $T = 10^3$ with parameter $\sigma^2 = 4$. Then we compare the performances of the following estimators:

• The classical GMWM estimator $\hat{\theta}_D = [\hat{\phi}_D \hspace{0.5cm} \hat{\nu}_D^2]$ as defined in Eq. (17) based on all possible scales (i.e. $\eta = \mathcal{D}$) and with $\Omega = V^{-1}_{\text{diag}}$ (as defined in Sec. VII-A);
• The classical GMWM estimator $\hat{\theta}_{\eta^o} = [\hat{\phi}_{\eta^o} \hspace{0.5cm} \hat{\nu}_{\eta^o}^2]$ as defined in Eq. (17) based on “optimal” scales, i.e. $\eta = \eta^o$, as defined in Eq. (20), and with $\Omega = V^{-1}_{\text{diag}}$ (for the corresponding scales);
• The augmented GMWM estimator $\hat{\theta}_D^+$ as defined in Eq. (19) with $\nu_{\eta}^+ = [\hat{m}_2 \hspace{0.5cm} \hat{\nu}_{\eta}]$ based on all possible scales and with $\Omega = V^{-1}_{\text{diag}}$ (augmented with the covariance of the second moment and obtained by parametric bootstrap);
• The bias correction GMWM estimator $\hat{\theta}_D$ as defined in Eq. (22) in which $\theta = \hat{\theta}_D$, $\Phi_n = I_2$ and $H = 100$;
• The MLE $\hat{\theta}_{\text{ML}} = [\hat{\phi}_{\text{ML}} \hspace{0.5cm} \hat{\nu}_{\text{ML}}^2]$.  

Note that the estimator based on the AV was omitted from this comparison because an AR(1) process (with $\rho \neq (0, 1)$) does not have a linear representation in the $\log(\tau)$ - $\log(\phi_r)$ plot.

Fig. 4 shows the empirical distribution of the different estimators obtained with 500 bootstrap replications and Tab. III reports the bias, the variances, the MSEs and the relatives efficiencies (to the MLE) of the various estimators presented above. As in the previous simulation, the MLE performs best. It can be also observed that the standard GMWM estimator suffers from an important bias. The augmented GMWM $\hat{\theta}_D^+$ as well as the bias corrected estimator $\hat{\theta}_D$ are both able to largely reduce this bias. Again, the selection of “optimal” scales for the GMWM estimator does not improve the estimator’s performances. This may indicate that all scales contain some valuable information and that all of them should be kept in the estimation.
TABLE III

EMPIRICAL BIAS, VARIANCES, MSES AND RELATIVE EFFICIENCIES (TO MLE) BASED ON 500 BOOTSTRAP REPlications FOR AN AR(1) PROCESS (i.e. (P.5)) OF LENGTH T = 10^3 WITH PARAMETERS ρ = 0.95 and ν^2 = 4 FOR THE ESTIMATORS DEFINED IN SECTION VII-B.

\[
\begin{array}{cccc}
\text{ρ} & \text{bias} & \text{var} & \text{MSE} & \text{eff.} \\
\hline
\theta_D^0 & -9.73 \cdot 10^{-3} & 2.10 \cdot 10^{-4} & 3.05 \cdot 10^{-4} & 3.08 \\
\theta_D^\eta & -9.73 \cdot 10^{-3} & 2.10 \cdot 10^{-4} & 3.05 \cdot 10^{-4} & 3.08 \\
\theta_D^+ & -3.17 \cdot 10^{-3} & 1.12 \cdot 10^{-4} & 1.22 \cdot 10^{-4} & 1.23 \\
\theta_{ML} & -2.93 \cdot 10^{-4} & 1.65 \cdot 10^{-4} & 1.66 \cdot 10^{-4} & 1.67 \\
\sum & -2.70 \cdot 10^{-3} & 9.18 \cdot 10^{-5} & 9.91 \cdot 10^{-5} & 1.00 \\
\end{array}
\]

| \text{ν^2} & \text{bias} & \text{var} & \text{MSE} & \text{eff.} |
|-----------------|-----------------|-----------------|-----------------|-----------------|
| \theta_D^0 & 7.37 \cdot 10^{-3} & 4.68 \cdot 10^{-2} & 4.68 \cdot 10^{-2} & 1.44 |
| \theta_D^\eta & 7.37 \cdot 10^{-3} & 4.68 \cdot 10^{-2} & 4.68 \cdot 10^{-2} & 1.44 |
| \theta_D^+ & 1.25 \cdot 10^{-2} & 4.88 \cdot 10^{-2} & 4.90 \cdot 10^{-2} & 1.51 |
| \theta_{ML} & 9.64 \cdot 10^{-3} & 4.61 \cdot 10^{-2} & 4.62 \cdot 10^{-2} & 1.42 |
| \sum & 3.26 \cdot 10^{-3} & 3.25 \cdot 10^{-2} & 3.25 \cdot 10^{-2} & 1.00 |

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{fig4.png}
\caption{Empirical distribution of the different estimators (defined in section VII-B) obtained with 500 bootstrap replications for an AR(1) process (i.e. (P.5)) of length T = 10^3 with parameter ρ = 0.95 and ν^2 = 4.}
\end{figure}

C. A composite stochastic process

In this last simulation we consider realizations of a composite process composed of the sum of an AR(1), a white noise and a drift, of length T = 10^3 with parameters ρ = 0.95, ν^2 = 4, σ^2 = 16 and ω = 0.04. We compare the performances of the following estimators:

- The classical GMWM estimator \( \hat{\theta}_D^0 = \left[ \hat{\rho}_D, \hat{\nu}_D, \hat{\sigma}_D^2, \hat{\omega}_D \right] \) as defined in Eq. (17) based on all possible scales (i.e. \( \eta = D \)) and with \( \Omega = V_{\text{diag}}^{-1} \) (as defined in section VII-A);

- The classical GMWM estimator \( \hat{\theta}_D^\eta = \left[ \hat{\rho}_\eta, \hat{\nu}_\eta, \hat{\sigma}_\eta^2, \hat{\omega}_\eta \right] \) as defined in Eq. (17) based on “optimal” scales, i.e. \( \eta = \eta^* \) as defined in Eq. (20), and with \( \Omega = V_{\text{diag}}^{-1} \) (for the corresponding scales);

- The augmented GMWM estimator \( \hat{\theta}_D^+ = \left[ \hat{\rho}_D^+, \hat{\nu}_D^+, \hat{\sigma}_D^2, \hat{\omega}_D^+ \right] \) as defined in Eq. (19) with \( \hat{\nu}_\eta = [m_1, m_2] \hat{\nu}_\eta \) (note that here \( m_2 \) refers to variance of the process) based on all possible scales and with \( \Omega = V_{\text{diag}}^{-1} \) (augmented with the covariance of the first two moments and obtained by parametric bootstrap);

- The bias correct GMWM estimator \( \hat{\theta}_D = \left[ \hat{\rho}_D, \hat{\nu}_D, \hat{\sigma}_D^2, \hat{\omega}_D \right] \) as defined in Eq. (22) in which \( \theta = \theta_D, \Phi_n = I_4 \) and \( H = 100 \).

In order to compute the augmented GMWM estimator we need define the considered moments implied by \( F_\theta \). Although this is very easy for the processes considered previously this is slightly more complicated for the process at hand. Therefore, we provide here the expectation and variance of the considered composite stochastic process. Let \( \{X_t\}, \{U_t\} \) and \( \{Z_t\} \) denote respectively an AR(1), a white noise and a drift. Let also \( \{Y_t\} \) denote the process at hand so that we may write \( Y_t = X_t + U_t + Z_t, t = 1, ..., T \). The expectation and variance of the drift can be computed as follow

\[
\begin{align*}
\mathbb{E}[Z_t] &= \frac{1}{T} \sum_{t=1}^{T} \omega t = \frac{\omega (T+1)}{2} \\
\mathbb{E}[Z_t^2] &= \frac{1}{T} \sum_{t=1}^{T} (\omega t)^2 = \omega^2 \left( \frac{T^2}{3} + \frac{T}{2} + \frac{1}{6} \right) \\
\text{var}(Z_t) &= \mathbb{E}[Z_t^2] - \mathbb{E}^2[Z_t] = \frac{\omega^2 (T^2 - 1)}{12} 
\end{align*}
\]

Using the above results and since the processes are independent we have that

\[
\begin{align*}
m_1(\theta) &= \mathbb{E}[Y_t] = \mathbb{E}[X_t] + \mathbb{E}[U_t] + \mathbb{E}[Z_t] = \frac{\omega (T+1)}{2} \\
m_2(\theta) &= \text{var}(Y_t) = \text{var}(X_t) + \text{var}(U_t) + \text{var}(Z_t) \\
&= \frac{\omega^2 (T^2 - 1)}{12}
\end{align*}
\]

Fig. 5 shows the empirical distribution of the different estimators obtained with 500 bootstrap replications and Tab. IV reports the bias, the variances and the MSEs of the various versions of the GMWM estimators presented above. It can be observed that the augmented GMWM estimator performs best and able to estimate the parameter \( \omega \) far more precisely than the other estimators. As the standard estimator does not seem to be biased, the bias correction method does not improve the estimation. Like in the two previous simulations, the estimator based on “optimal” scales does not perform better than the classical GMWM employing all scales. Again, it shall be stressed that both the AV-based estimator as well as the MLE are not applicable and would fail to provide meaningful results in this case.

VIII. CONCLUSIONS

In this paper, we have studied the recently developed estimator, called the GMWM estimator [12] that is particularly suited for the estimation of parameters of complex composite stochastic processes. It has already been successfully applied to the modeling of inertial sensor errors in [20], and shown
by means of simulation and emulations to clearly outperform classical approaches such as AV-based estimators. In this context, the practical implications of the GMWM estimators are considerable because of the rapidly growing employment of small and inexpensive MEMS-IMUs whose error behavior are rather complex. However, what was not established until now, is that the estimator based on the standard AV for composite stochastic processes is not consistent, which explains its poor performance in emulation studies. In the case of simple processes, the AV estimator is actually shown to be a particular GMWM estimator and that there always exists a GMWM estimator with smaller asymptotic variance than its AV-based counterpart. We also provide extended versions of the GMWM estimator that enhance its finite sample performance (bias and variance reduction). The most important one, is a finite sample bias correction related to indirect inference which is also very general and can be applied beyond the scope of the GMWM framework.

**APPENDIX A**

*Proof of Proposition 1:* Consider the vector $\hat{\phi} = \left[\hat{\phi}_\tau\right]_{\tau \in \tau}$. Since $\hat{\phi}_\tau$ is consistent by assumption (A.1) $\tau \in \tau$, i.e. $\hat{\phi}_\tau \xrightarrow{P} \phi$, we also have that $g\left(\hat{\phi}_\tau\right) \xrightarrow{P} g\left(\phi\right)$ if $g\left(\cdot\right)$ is a continuous function by the continuous mapping theorem (see e.g. [2, Theorem 1.14]). For $\hat{\theta}_{AV}$ defined by (4), we may write $\hat{\theta}_{AV} = k\left(\hat{\phi}\right)$. Then for (P.1), (P.2) or (P.3), $k\left(\cdot\right)$ is continuous (see Example 1), such that $\hat{\theta}_{AV} = k\left(\hat{\phi}\right) \xrightarrow{P} k\left(\phi\right) = \theta$ which concludes the proof.

**APPENDIX B**

*Proof of Corollary 1:* First, we apply the same approach as used in the proof of Proposition 1. Since $\hat{\gamma}^2_{AV}$ as defined in (5) can be written as $\hat{\gamma}^2_{AV} = k\left(\hat{\phi}\right)$ with $k\left(\cdot\right)$ a continuous function, and since $\hat{\phi} \xrightarrow{P} \phi = [\phi_\tau]_{\tau \in \tau}$, with
\begin{align*}
\phi_\tau &= \sqrt{\gamma^2 \left( \frac{\tau}{3} + \frac{1}{6\tau^2} \right)}, \text{ then by using (5) we have} \\
\lim_{T \to \infty} \hat{\gamma}_{AV}^2 &= 3 \exp \left\{ \frac{2}{\tau} \sum_{\tau \in \tau} \left[ \log (\phi_\tau) - \frac{1}{2} \log (\tau) \right] \right\} \\
&= 3 \exp \left\{ \frac{2}{\tau} \sum_{\tau \in \tau} \left[ \log \left( \sqrt{\gamma^2 \left( \frac{2\tau^2 + 1}{6\tau^2} \right)} \right) \right] \right\} \\
&= \hat{\gamma}_2^2
\end{align*}

Now, one can always write
\begin{align*}
\gamma^2 &= 3 \exp \left\{ \frac{2}{\tau} \sum_{\tau \in \tau} \left[ \log \left( \frac{\gamma}{\sqrt{3}} \right) \right] \right\} \\
&= 3 \exp \left\{ \frac{2}{\tau} \sum_{\tau \in \tau} \left[ \log \left( \sqrt{\gamma^2 \left( \frac{2\tau^2 + 1}{6\tau^2} \right)} \right) - b_{AV}(\tau) \right] \right\} \\
\text{and} \\
b_{AV}(\tau) &= \log \left( \sqrt{\gamma^2 \left( \frac{2\tau^2 + 1}{6\tau^2} \right)} \right) - \log \left( \frac{\gamma}{\sqrt{3}} \right) \\
&= \log \left( \frac{1 + \frac{1}{2\tau^2}}{\gamma} \right) > 0
\end{align*}

since \( \sup_{\tau \in \tau} (\tau) < \infty \). Hence \( \gamma^2 < \hat{\gamma}_2^2 \) which concludes the proof.

\section*{APPENDIX C}

\textbf{Proof of Proposition 2}: The proof is straightforward since any combination of a process chosen among (P.1), (P.2), (P.3) and (P.4) with any other different process(es) satisfying assumption (A.3) does not have an exact linear representation of the form defined in (2). So we may apply the same rational as in Corollary 1 to show the inconsistency of the estimator for \( \hat{\theta} \). Therefore \( \hat{\theta} \) can not be a consistent estimator of \( \theta \).

\section*{APPENDIX D}

\textbf{Proof of Proposition 3}: The proof is straightforward. Indeed, for processes (P.1), (P.2) or (P.3), we have that
\begin{align*}
\hat{\theta}_{AV} &= \arg \min_{\theta \in \Theta} \sum_{\tau \in \tau} \varepsilon_\tau^2 \\
&= \arg \min_{\theta \in \Theta} \sum_{\tau \in \tau} \left[ \log (\phi_\tau) - \log (\phi_\tau (\tau)) \right]^2 \\
&= \arg \min_{\theta \in \Theta} (\hat{\nu}_\tau - \nu_\tau (\theta))^T \Omega^* (\hat{\nu}_\tau - \nu_\tau (\theta))
\end{align*}
where the elements of \( \Omega^* \) are given by (18). Therefore, we have that \( \hat{\theta}_{AV} = \hat{\theta} \) as they are the solutions of the same problem.

\section*{APPENDIX E}

\textbf{Proof of Corollary 2}: Since \( \hat{\theta}_{AV} = \hat{\theta}^*_\tau \) with \( \hat{\theta}^*_\tau \) being the GMWM estimator based on \( \Omega^* \), and since the GMWM estimator with minimal asymptotic variance is \( \hat{\theta}^*_\tau \) where \( \Omega^\circ = \nabla_{\nu^\circ}^{-1} = \Omega^* \), we verify the result of Corollary 2.
We now consider the various terms in Eq. (F-3). We start by considering the $j^{th}$ element of $E\left[u\right]$ and we let
\[ u_j = \sum_{i=3}^{T_0} \frac{a_{ij}}{T^2} \left( \hat{\theta}_{(i,j)} - \hat{\theta}_{(i,j)}^c \right) \]
so we obtain
\[
E\left[u_j\right] = \mathbf{E} \left[ \frac{a_{ij}}{T^2} \left( \theta_{(j,i)} - \hat{\theta}_{(j,i)}^c \right) \right] + \mathbf{E} \left[ u_j \right] \\
= \frac{a_{ij}}{T^2} \left( \theta_{(j,i)} - \mathbf{E} \left[ \hat{\theta}_{(j,i)}^c \right] \right) + O \left( T^3 \right) \\
= \frac{a_{ij}}{T^2} \left( \theta_{(j,i)} - \mathbf{E} \left[ \hat{\theta}_{(j,i)}^c \right] \right) \var\left[ \hat{\theta}_{(j,i)} \right] + O \left( T^3 \right) \\
= \frac{a_{ij}}{n^2} \left( \theta_{(j,i)} - \mathbf{E} \left[ \hat{\theta}_{(j,i)}^c \right] \right) \var\left[ \hat{\theta}_{(j,i)} \right] + O \left( T^3 \right) \\
= \frac{a_{ij}}{T^2} \left( \theta_{(j,i)} - \mathbf{E} \left[ \hat{\theta}_{(j,i)}^c \right] \right) \left( \theta_{(j,i)} + \mathbf{E} \left[ \hat{\theta}_{(j,i)} \right] \right) + O \left( T^3 \right) \\
= \frac{a_{ij}}{T^2} \left( \theta_{(j,i)} - \mathbf{E} \left[ \hat{\theta}_{(j,i)}^c \right] \right) \left( \theta_{(j,i)} + \mathbf{E} \left[ \hat{\theta}_{(j,i)} \right] \right) + O \left( T^3 \right)
\]
Using Eq. (F-3) we may write
\[
E\left[ \hat{\theta}_j \right] - \theta_0 = \mathbf{B} E\left[u\right]
\]
with $\mathbf{B} = \left( \mathbf{I} + \frac{1}{T^2} \mathbf{Q} \right)^{-1} \mathbf{Q}$, note that Assumption (A.6)* guarantees that $\mathbf{B}$ exists. Now, we can also write
\[
\left( \theta_{(j,i)} - \mathbf{E} \left[ \hat{\theta}_{(j,i)}^c \right] \right) = - \left( \mathbf{B} E\left[u\right] \right)_{(j)}
\]
Since that matrix $\mathbf{Q}$ is fixed and does not depend on $T$ the terms $E\left[ \hat{\theta}_j \right]$ and $E\left[u\right]$ converge at the same rate and using (F-4) we have that (at least) $E\left[u\right] = O \left( T^2 \right)$. Moreover, since the term $\left( \theta_{(j,i)} + \mathbf{E} \left[ \hat{\theta}_{(j,i)} \right] \right)$ is bounded by Assumptions (A.1)*, (A.5)* and (A.7)* we have that
\[
\frac{a_{ij}}{T^2} \left( \theta_{(j,i)} - \mathbf{E} \left[ \hat{\theta}_{(j,i)}^c \right] \right) \left( \theta_{(j,i)} + \mathbf{E} \left[ \hat{\theta}_{(j,i)} \right] \right) = O \left( T^4 \right)
\]
So, by substituting Eq. (F-5) in Eq. (F-4) we obtain that
\[
E\left[u_j\right] = O \left( T^3 \right)
\]
and, we obtain that
\[
E\left[ \hat{\theta}_j \right] - \theta_0 = O \left( T^3 \right)
\]
Finally, by Assumptions (A.9)* and (A.10)* the solution of Eq. (22) is unique which concludes the proof. 

ACKNOWLEDGMENT

The authors would like to thank Roberto Molinari for his helpful comments and suggestions.

REFERENCES