Wavelet-Variance-Based Estimation for Composite Stochastic Processes

Stéphane GUERRIER, Jan SKALOUD, Yannick STEBLER, and Maria-Pia VICTORIA-FESER

This article presents a new estimation method for the parameters of a time series model. We consider here composite Gaussian processes that are the sum of independent Gaussian processes which, in turn, explain an important aspect of the time series, as is the case in engineering and natural sciences. The proposed estimation method offers an alternative to classical estimation based on the likelihood, that is straightforward to implement and often the only feasible estimation method with complex models. The estimator furnishes results as the optimization of a criterion based on a standardized distance between the sample wavelet variances (WV) estimates and the model-based WV. Indeed, the WV provides a decomposition of the variance process through different scales, so that they contain the information about different features of the stochastic model. We derive the asymptotic properties of the proposed estimator for inference and perform a simulation study to compare our estimator to the MLE and the LSE with different models. We also set sufficient conditions on composite models for our estimator to be consistent, that are easy to verify. We use the new estimator to estimate the stochastic error’s parameters of the sum of three first order Gauss–Markov processes by means of a sample of over 800,000 issued from gyroscopes that compose inertial navigation systems.

SUPPLEMENTARY MATERIALS

Supplementary materials for this article are available online. Please go to www.tandfonline.com/r/JASA

KEY WORDS: Allan variance; Kalman filter; Signal processing; Time series.

1. INTRODUCTION

Let \( F_\theta, \theta \in \Theta \subseteq \mathbb{R}^p \) be the model associated to the univariate Gaussian time series \( \{Y_t; t \in \mathbb{Z}\} \) that is stationary or nonstationary but with stationary backward differences of order \( d \), and let \( \{y_t, t = 1, \ldots, T\} \) be the corresponding observed outcome. In this article, we propose a new estimation method for the parameter’s vector \( \theta \) that is based on the matching between the empirical and model-based wavelet variances (WV). We also set the conditions on \( F_\theta \) under which our estimator is consistent. In particular, we consider for the empirical WV the Maximal empirical and model-based wavelet variances (WV). We also set sufficient conditions on composite models for our estimator to be consistent, that are easy to verify. We use the new estimator to estimate the stochastic error’s parameters of the sum of three first order Gauss–Markov processes by means of a sample of over 800,000 issued from gyroscopes that compose inertial navigation systems. Supplementary materials for this article are available online.

with measurements

\[
Y_{t+1} = hx_{t+1} + u_{t+1},
\]

where \( x_t \) is a \( q \times 1 \) system state vector at time \( t \), \( \Phi \) is a \( q \times q \) coefficient or state-transition matrix from \( t \) to \( t+1 \), \( w_t \) is a \( q \times 1 \) multivariate WN vector, that is, \( w_t \sim \mathcal{N}(0, Q) \), \( u_t \) is a \( q \times 1 \) deterministic input vector, \( y_t \) is the one-dimensional output variable, \( h \) is a \( 1 \times q \) design vector which maps the state vector \( x_t \) into the output \( y_t \), and \( v_t \) is a one-dimensional noise such that \( v_t \sim \mathcal{N}(0, \sigma_{WN}^2) \).

The books by Harvey (1989) and Durbin and Koopman (2001) contain extensive accounts of state-space models and their applications. For linear and/or Gaussian state-space models, the Maximum Likelihood Estimator (MLE) is a natural choice for the estimation of the model’s parameters. The Kalman filter, MCMC, or other simulation smoother, together with the specification of a prior to start the latent process, are typically used for computing predictors of the state-variables and one-step-ahead predictors of the observations which are then used in an Expectation–Maximization (EM) algorithm of Dempster, Laird, and Rubing (1977) to compute the MLE (see also, e.g., Shumway and Stoffer 1982). The MLE is based on the density \( f_\theta(y) = \partial f_\theta(y)/\partial y \), with \( \theta \) the \( p \times 1 \) vector of parameters containing the unknown elements of \( \Phi, Q, h \) and possibly also \( u_{t+1} \) together with \( \sigma_{WN}^2 \). The maximization step can be very complex and finding the MLE is not always a simple task.

Moreover, even if new smoothers are regularly proposed (see, e.g., Durbin and Koopman 2002; McCausland, Miller, and Pelletier 2011) that improve the computational efficiency, the task becomes even more challenging when the observed process size is large and the model is, for example, a mixture of three first-order Gauss–Markov (GM) random processes (a reparameterization of AR(1) processes) with different parameters, a

Stéphane Guérrier is PhD student, Research Center for Statistics, HEC Genève, University of Geneva, Geneva, Switzerland (E-mail: Stephane.Guerrier@unige.ch). Jan Skaloud is Senior Scientist (E-mail: jan skaloud@epfl.ch) and Yannick Stebler is PhD student (E-mail: yannick.stbler@epfl.ch), Geodetic Engineering Laboratory, Swiss Federal Institute of Technology, Lausanne, Switzerland (E-mail: jan skaloud@epfl.ch; yannick.stbler@epfl.ch). Maria-Pia Victoria-Feser is Professor, Research Center for Statistics, HEC Genève, University of Geneva, Geneva, Switzerland (E-mail: Maria-Pia.VictoriaFeser@unige.ch). Stéphane Guérrier and Maria-Pia Victoria-Feser are partially supported by a Swiss National Fund Grant (no. 100018-131906). Yannick Stebler was partially supported by the European FP7-GALILEO-2007-GSA-1 grand CLOSE-SEARCH as well as by the project no. 103010.2 of the Swiss Commission for Technology and Innovation (CTI). The authors are grateful to two anonymous reviewers for their comment that helped to improve the content and the presentation of the article.

Published with license by American Statistical Association © Stéphane Guérrier, Jan Skaloud, Yannick Stebler, Maria-Pia Victoria-Feser

Journal of the American Statistical Association
September 2013, Vol. 108, No. 503, Theory and Methods
DOI: 10.1080/01621459.2013.799920

1021
model that is used in the case study Section 5. We also found in a simulation exercise in Section 4 that our estimator, unlike the classical MLE or the least squares estimator (LSE), is importantly less biased when the roots of an AR(1) process lie near the unit circle. However, a formal and more general treatment of the unit root problem is left for future research.

An important useful tool for understanding underlying features of a process is the Power Spectral Density (PSD), which for stationary processes, is given by

$$S_y(f) = \mathbb{E} \{ Y(t) \overline{Y}(t-\tau) \} = \sum_{l=-\infty}^{\infty} C_{\Delta l} \{ \tau \} e^{-2\pi i \Delta l f},$$

where $$\Delta l$$ is the Nyquist frequency. The PSD is related to the variance of a process after the WV can be interpreted as the variance of a process after the transfer function of the WV is the Power Spectral Density since $$(\tilde{H}_j)^2 = \mathbb{E} \{ \tilde{H}_j(f) \overline{\tilde{H}_j}(f) \}$$ for stationary processes. For wavelet filters of width $$\Delta l$$, the transfer function of the WV is given by

$$\tilde{H}_j(f) = \frac{1}{2} \sum_{l=-\infty}^{\infty} \tilde{h}_{j,l} e^{-2\pi i \Delta l f},$$

where $$\tilde{h}_{j,l}$$ is the $$l$$-th level wavelet filters defined in Percival and Walden (2000) (see also Section 2). The WV can be estimated from the sample $$\{ y_t ; t = 1, \ldots, T \}$$ using, for example, the maximal-overlap discrete wavelet transform estimator defined in Serroukh, Walden, and Percival (2000) (see also Percival 1995 and Equation (6) next), and for the matching of these empirical estimates to the WV implied by the model, we propose an optimization criterion based on a standardized distance between both WV estimates, as is done, for example, with the Generalized Method of Moments (GMM) (Hansen 1982). We call the new estimator the Generalized Method of Wavelet Moments (GMWM) estimator. We provide, among others, sufficient conditions on the model $$F_{\theta}$$ for the GMWM estimator to be consistent and asymptotically normal. Moreover, it can be computed using simulations as is done with indirect inference (Gourieroux, Monfort, and Renault 1993; Smith 1993; Gallant and Tauchen 1996), so that it is very straightforward to implement in practice.

While in this article we use wavelets (through their variances) for parametric estimation, wavelets have been used in many different statistical problems mainly for nonparametric estimation. To cite only a few, wavelets have been used in nonparametric density estimation (see, e.g., Donoho and Johnstone 1994; Kerykacharian and Picard 1992; Walter 1992; Donoho et al. 1996), inverse problems (see, e.g., Donoho 1995), nonparametric regression or curve fitting (see, e.g., Donoho and Johnstone 1994 and Donoho et al. 1995) also with correlated errors (see, e.g., Johnstone and Silverman 1997), or with locally stationary noise (see, e.g., Von Sachs and Macgibbon 2000).

This article is organized as follows. In Section 2, we introduce the WV and explain how it is used mainly in the engineering literature. In Section 3, we present the estimator based on the WV and derive its statistical properties. A simulation study is then presented in Section 4 that compares the finite sample performance of the GMWM estimator to the MLE and the LSE. In Section 5, we apply the new methodology to estimate parameters for the stochastic error model in inertial sensors with a real dataset of size 833,685.

2. THE WAVELET VARIANCE

Basically, as pointed out by Percival and Guttrop (1994), the WV can be interpreted as the variance of a process after it has been subject to an approximate bandpass filter. WV can be built using wavelet coefficients issued from a modified Discrete Wavelet Transform (DWT) (Mallat 1999; Percival and Walden 2000) called the Maximal Overlap DWT (MODWT); see Greenhall (1991), Percival and Guttrop (1994). The wavelet coefficients are built using wavelet filters $$\{ \tilde{h}_{j,l} \}; j = 1, \ldots, J$$ which for $$j = 1$$ and for the MODWT satisfy

$$\sum_{l=0}^{L_j-1} \tilde{h}_{j,l} = 0; \quad \sum_{l=0}^{L_j-1} \tilde{h}_{j,l}^2 = \frac{1}{2}; \quad \text{and} \quad \sum_{l=0}^{\infty} \tilde{h}_{j,l} \tilde{h}_{j,l+2n} = 0,$$

where $$\tilde{h}_{j,l} = 0$$ for $$l < 0$$ and $$l \geq L_j$$ is the length of $$\tilde{h}_{j,l}$$, and $$m$$ is a nonzero integer. Let also $$\tilde{H}_j(f) = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} e^{-2\pi i \Delta l f}$$ be the transfer function of $$\tilde{h}_{j,l}$$. To obtain the $$j$$-th level wavelet filters $$\{ \tilde{h}_{j,l} \}$$ of length $$L_j = (2^j - 1)(L_1 - 1) + 1$$, one computes the inverse discrete Fourier Transform of

$$\tilde{H}_j(f) = \tilde{H}_j(2^{-j-1}f) \sum_{l=0}^{L_j-1} e^{2\pi i \Delta l f}(L_1 - 1) \tilde{H}_j \left( \frac{1}{2} - 2^{j-1}f \right).$$

The MODWT filter is actually a rescaled version of the DWT filter $$h_{j,l}$$, that is, $$\tilde{h}_{j,l} = h_{j,l}/2^{j/2}$$. Filtering an infinite sequence $$\{ Y_t ; t \in \mathbb{Z} \}$$ using the wavelet filters $$\{ \tilde{h}_{j,l} \}$$ yields the MODWT wavelet coefficients

$$\tilde{W}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} Y_{t-l} \quad \forall t \in \mathbb{Z}.$$
with $W_{j,t} = \sum_{l=0}^{L_j-1} f_{j,l} y_{j-l,t} \in (L_j; T)$ and $M_j = T - L_j + 1$, is a consistent estimator for $\nu^2(\tau_j)$; see Serroukh, Walden, and Percival (2000) who also show that under suitable conditions (see also Theorem 3 in the supplementary materials), \( \sqrt{M_j}(\nu^2(\tau_j) - \nu^2(\tau_j)) \) is asymptotically normal with mean 0 and variance

\[
S_{W_j}(0) = 2 \int_{-1/2}^{1/2} S^2_{W_j}(f) df = 2 \int_{-1/2}^{1/2} \left| \mathcal{H}_j(f) \right|^4 S^2_{W_{hj}}(f) df.
\]

Equation (7) can be estimated by means of

\[
\hat{S}_{W_j}(0) = \sum_{\tau=-M_j}^{M_j} \left[ \frac{1}{M_j} \sum_{j=L_j}^{T} W_{j,t+\tau} \right]^2
\]

and the asymptotic properties of Equation (8) are given in Percival and Walden (2000), p. 312.

A particular choice for the wavelet filter is the Haar wavelet filter whose first DWT filter $(j=1)$ is

\[
[h_{1,0} = 1/\sqrt{2}, \ h_{1,1} = -1/\sqrt{2}]
\]

with length $L_1 = 2$. If the process is stationary with backward differences of order $d > 1$, then other wavelet filters such as Daubechies wavelet filters can be used (Daubechies 1992). When the WV is evaluated with Haar wavelet filters, it is actually equal to half the Allan variance (AV) (Allan 1966). In 1998, the IEEE Standard 1293 (1998) introduced the AV technique as a noise identification method which can be used for determining the characteristics of underlying stochastic processes affecting signals. AV and WV are often used in engineering disciplines and physical sciences as a graphical approach for mining the characteristics of underlying stochastic processes as a noise identification method which can be used for deter-

\[
tation (Schieder and Kramer 2008, 2012). Let \( \phi(\cdot) = (\phi_1(\cdot), \phi_2(\cdot)) \) with $\text{dim}(\phi_1(\cdot)) = p$ (and $\text{dim}(\phi_2(\cdot)) = J - p$), and let also $g(\theta) = \mathbb{E}[\hat{\phi} - \phi(\theta)] = (g_1(\theta), g_2(\theta))$ (i.e., the same split as with $\phi(\theta)$). Consider the following conditions:

(C.1) If $\Omega$ is estimated by $\hat{\Omega}$, then $\hat{\Omega} \overset{P}{\rightarrow} \Omega$, that is, the weighting matrix converges to a population weighting matrix.

(C.2) $\hat{\theta}$ is a consistent estimator of the WV $\nu^2 = [\nu^2(\tau_j)]_{j=1,...,J}$.

(C.3) $\theta \in \Theta$ with $\Theta$ being an open subset of $\mathbb{R}^p$. Then, there exists a bijective, twice continuous differentiable function $k : \mathbb{R}^p \rightarrow \Theta$, such that $y = k^{-1}(\theta) \in \mathbb{R}^p$, $k(\gamma) = [k_1(\gamma_j)]_{j=1,...,p}$, $| (\partial k_y) k(\gamma) | = \prod_{j=1}^{p} (\partial k_\gamma) k_j(\gamma_j) > 0$ (i.e., the determinant of the Jacobian matrix is nonnegative) and if $p = 2$, $(\partial k_\gamma) k_j(\gamma_j) > 0$, $\forall j$ (i.e., the Jacobian matrix is positive-definite definite).

(C.4) $g_1(\theta)$ is twice continuously differentiable.

(C.5) $| (\partial k_\gamma) g_1(\theta) |$ is nonnegative.

(C.6) $| g_1(\theta) | \rightarrow \infty$ whenever $|\theta| \rightarrow \infty$.  

3. GMWM ESTIMATOR

We propose to estimate the model’s parameters using an estimator which combines on the one hand the WV and on the other hand the Generalized Least Squares (GLS) principle, using the relationship given in Equation (5). More precisely, we propose to find $\hat{\theta}$ such that the WV implied by the model, say $\hat{\phi}(\theta)$, matches the empirical WV, say $\hat{\phi}$, and solves the following GLS optimization problem

\[
\hat{\theta} = \arg\min_{\theta \in \Theta} (\hat{\phi} - \phi(\theta))^T \Omega (\hat{\phi} - \phi(\theta))
\]

in which $\Omega$, a positive-definite weighting matrix, is chosen in a suitable manner (see next). Equation (10) defines the GMWM estimator. $\phi(\cdot) = [\phi_j(\cdot)]_{j=1,...,J}$ is a binding function between $\theta$ and $\nu^2 = [\nu^2(\tau_j)]_{j=1,...,J}$ such that $\phi(\theta) = \nu^2$, and $\hat{\phi} = \hat{\nu}^2$ are two estimators.

In Theorem 3 given in the supplementary materials (see Section C), we use the results of Giraitis and Taqqu (1997) on limit theorems for bivariate Appell polynomials to set the conditions for $\hat{\phi}$ to be asymptotically multivariate normal. In particular, $S_{W_j}(f) \leq C|f|^{-\alpha}, f \in [-1/2, 1/2], \alpha < 1/2$. This result generalizes the results of Serroukh, Walden, and Percival (2000) to the multivariate case, while Serroukh and Walden (2000a, 2000b) generalize the results of Serroukh, Walden, and Percival (2000) to wavelet covariances.

In Theorem 1 next, we state the conditions on $\hat{\Omega}$, $\hat{\phi}$, and $\phi(\theta)$ (hence on $F_{\theta}$) under which $\hat{\theta}$ is a consistent estimator of $\theta$. For that, we follow the methodology used for GMWM (see, e.g., Harris and Märtys 1999) and use the results of Komunjer (2012). Let $\phi(\theta) = (\phi_1(\theta), \phi_2(\theta))$ with $\text{dim}(\phi_1(\theta)) = p$ (and $\text{dim}(\phi_2(\theta)) = J - p$), and let also $g(\theta) = \mathbb{E}[\hat{\phi} - \phi(\theta)] = (g_1(\theta), g_2(\theta))$ (i.e., the same split as with $\phi(\theta)$). Consider the following conditions:
(C.7) For every $c \in \mathbb{R}^p$, the equation $g_1(\theta) = c$ has countably many (possibly zero) solutions in $\mathbb{R}^p$.

(C.8) For $p = 2$, $(\partial/\partial \theta) g_1(\theta)$ is positive definite, and for $p > 2$, the set of points $\theta_i$ for which $	ext{rank}[(\partial/\partial \theta) g_1(\theta)] < p - 1$ is bounded.

Theorem 1. For $\Omega$ satisfying (C.1), under (C.2), for $g(\theta)$ or $-g(\theta)$ satisfying (C.4)–(C.8), and if there exists a function $f$ satisfying (C.3), then $\hat{\theta}$ defined in Equation (10) is consistent.

The proof is given in the supplementary materials (see Section E). Condition (C.3) defines a reparameterization from $\theta \in \Theta$ to $\gamma \in \mathbb{R}^p$. Hence, in practice, for a given model $F_\theta$, one has first to check that a function $k = [k_j]_{j=1,\ldots,p}$ exists that satisfies (C.3). For variance parameters $0 < \sigma^2$, the $k_j$'s are the logarithm, and for other restricted parameters with restrictions of the form $a < \theta < b$, one can use the logit function on $(\theta_j - a)/(b - a)$, that is, $k_j(\gamma_j) = (b - a) \exp(\gamma_j)/(1 + \exp(\gamma_j)) + a$.

Condition (C.3) is convenient but actually too restrictive. The following corollary actually extends the conditions of Theorem 1.

Corollary 1. If the bijective, twice continuous differentiable function $k$ does not satisfy the conditions on the Jacobian matrix in (C.3), then $\hat{\theta}$ defined in (10) is consistent under conditions (C.1), (C.2), and (C.4)–(C.8) in which $(\partial/\partial \theta) g_1(\theta)$ is replaced by $(\partial/\partial \gamma) g_1(\theta)/(\partial/\partial \gamma) k(\gamma)$.

The proof follows directly from the proof of Theorem 1.

In Theorem 2 below, we provide the consistency conditions of the GMWM when $F_\theta$ is made of the sum of $L$ independent processes. Let their respective WV be $\phi_{j0}(\theta_{j0})$ and partition $\theta$ as $\theta = [\theta_{j0}]_{j=1,\ldots,L}$, we need the following set of conditions:

(C.9) The functions $\phi_{j0}(\theta_{j0})$ are linearly independent with all linear combinations of the other $\phi_{j0}(\theta_{j0})$, $l \neq l'$, $l, l' = 1, \ldots, L$, and for all $\tau_j$ that is, one cannot find $a_m, b_m \in \mathbb{R}, m = 1, \ldots, L$ such that $\phi_{j0}(\theta_{j0}) = \sum_{m=1}^{L} a_m \phi_{j0}(\theta_{j0}) + b_m$.

(C.10) For all $j$, $\phi_{j0}(\theta_{j0}) = \phi_{j0}(\theta_{j0})$ if and only if $\theta_{j0} = \theta_{j0}$ for all $l$.

(C.11) Each element of $\frac{\partial}{\partial \theta_{j0}} \phi_{j0}(\theta_{j0})$, where $\theta_{j0} = [\theta_{j0,m}], m = 1, \ldots, \dim(\theta_{j0})$ depends on $\tau_{j0}$ for at least $L - 1$ of the $l = 1, \ldots, L$.

Theorem 2. Let $F_\theta$ be the stochastic process associated to $\{Y_t, t \in \mathbb{Z}\}$ made of the sum of $L$ independent processes with respective WV $\phi_{j0}(\theta_{j0})$, $\hat{\theta}$ be the GMWM estimator of $\theta$, then for $\Omega$ satisfying (C.1), $\hat{\phi}$ satisfying (C.2), and $\phi_{j0}(\theta_{j0})$ satisfying conditions (C.9)–(C.11), $\hat{\theta}$ is a consistent estimator of $\theta$.

The proof is given in the supplementary materials (see Section E). Condition (C.10) can be verified for each subprocess WV using the conditions of Theorem 1. We do that in the supplementary materials (see Section F) for some of the processes presented in Table 3 (see Section A), and for the composite processes treated in this article, we show the consistency using the conditions in Theorem 2.

Obviously, the number of scales $J$ should be $J \geq p$ but at the same time, as discussed in Section 4, a too large $J$ can introduce variability in the estimator. The study of a possible optimal value for $J$ is out of the scope of the present article. However, we may notice that in Guerrier et al. (2013), the authors investigated through simulation studies different choices for the number of scales $J$ for mixtures of WN, QN, drift, RW, and AR(1) models, and found out that from a given number of scales, the variability of the GMWM does not really improve. This suggests that one could choose a number $J^*$ such that $J^* = \min(J, a \cdot p)$, with $a$ a bounded (positive) integer.

Then, under the conditions of Theorem 1 (or possibly Corollary 1) or Theorem 2 as well as Theorem 3, we have that $\sqrt{T}(\hat{\theta} - \theta)$ is asymptotically normal, with asymptotic covariance matrix $V_{\hat{\theta}}$ given by

$$V_{\hat{\theta}} = BV_{\hat{\phi}}B^T,$$

where

$$B = (D^T \Omega D)^{-1} D^T \Omega$$

and where $D = \partial \phi(\theta)/\partial \theta^T$, and $V_{\hat{\phi}}$ is the asymptotic covariance matrix of $\hat{\phi}$ with elements given in (A-3) (see Section C). When $\Omega = I$, then $V_{\hat{\theta}} = (D^T D)^{-1} D^T V_{\hat{\phi}} D (D^T D)^{-1}$. The most efficient estimator is obtained by choosing $\Omega = V_{\hat{\phi}}^{-1}$, leading then to $V_{\hat{\theta}} = (D^T V_{\hat{\phi}}^{-1} D)^{-1}$. In practice, the matrix $D$ is computed at $\hat{\theta}$.

The estimation of the WV covariance $\sigma_{ij}^2$ of $V_{\hat{\theta}}$ given in (A-3) is in general not straightforward. In the supplementary materials (see Section G), we show that under the assumption of a Gaussian process for $Y_l$, a suitable estimator is given by

$$\hat{\sigma}_{ij}^2 = \frac{1}{2} \sum_{\tau_{ij} = -M}^{M} \left[ \frac{1}{M_{ij}} \sum_{t = \max(L, \tau_{ij} - L, \tau_{ij} + L)}^{T} W_{i,t} W_{i,t+} \right]^2$$

$$+ \frac{1}{2} \sum_{\tau_{ij} = -M}^{M} \left[ \frac{1}{M_{ij}} \sum_{t = \max(L, \tau_{ij} - L, \tau_{ij} + L)}^{T} W_{i,t-} W_{i,t} \right]^2,$$

where $M_{ij} = \min(M_l, M_k)$ (see Equation (6)). Alternatively, when the process is not Gaussian or when the sample size is very large as is the case with the dataset analyzed in Section 4 so that the computation of Equation (13) is infeasible, one can use a parametric bootstrap to estimate $\text{cov}(\hat{\theta}, \hat{\theta})$. $Q$ samples of size $T$ are simulated from $F_\theta$ on which $Q$ WV $\hat{\theta}_q(t_k)$ and $\hat{\theta}_q(t_q), q = 1, \ldots, Q$, are computed and $\sigma_{ij}^2$ is estimated by their empirical covariance.

When $J > p$, that is, the number of WV is greater than the dimension of the parameter vector $\theta$, one can assess the goodness of fit of the model $F_\theta$ to the data by testing the hypotheses $H_0 : E[\phi(\theta)] = 0, H_1 : E[\phi(\theta)] \neq 0$ using the $\chi^2$-test statistic

$$T(\phi - \phi(\hat{\theta}))^T V_{\phi}^{-1} (\phi - \phi(\hat{\theta}))$$

which is asymptotically $\chi^2_{J - p}$ under $H_0$ (see Hansen 1982). It will be used to assess the fit of the postulated model in the case study in Section 5. The investigation of the finite sample properties of Equation (14) is left for future research.

As a possible extension of the GMWM when analytical expressions for $\phi(\theta)$ in Equation (10) are too complicated to compute, one can resort to simulations to compute $\phi(\theta)$ and...
hence place the GMWM in the framework of indirect inference (Gourieroux, Monfort, and Renault 1993; Smith 1993; Gallant and Tauchen 1996). Basically, given a sample of observations \( \{y_t, t = 1, \ldots, T\} \) and a hypothetical model \( F_\theta \), \( \hat{\phi}_j \) is defined as the WV \( \hat{\nu}^j(\tau_j) \) estimated from the sample using Equation (6). Let also \( \hat{\phi}^*_j(\theta) = \frac{1}{\pi} \sum_{r=1}^R \hat{\phi}_{j,r}(\theta) \) with \( \hat{\phi}_{j,r}(\theta) \) being the WV estimates \( \hat{\nu}^j(\tau_j) \) computed on simulated series \( \{y^*_r(\theta), t = 1, \ldots, T\}, r = 1, \ldots, R \). Then, \( \hat{\phi} = [\hat{\phi}_j(\theta)]_{j=1,\ldots,J} \) and \( \hat{\phi}^*(\theta) = [\hat{\phi}^*_j(\theta)]_{j=1,\ldots,J} \) are used in Equation (10) to obtain an estimate \( \hat{\theta} \) of \( \theta \), whose properties are described in, for example, Gourieroux, Monfort, and Renault (1993). In particular, for \( R \) sufficiently large, \( V_\theta \approx B V_\theta^* B^T \). In that case, \( B \) can be computed numerically.

A step-by-step algorithm for computing the GMWM estimator (10) given a dataset of size \( T \) and a choice for \( F_\theta \), is given by the following steps:

- **Step 1.** Compute the empirical WV \( \hat{\phi} \) for a suitable number of scales \( J \geq \dim(\theta) \) using Equation (6).
- **Step 2.** Construct the matrix \( \Omega \). We recommend using for \( \Omega \) the inverse of a diagonal matrix whose elements are \( \hat{\phi}^*_j(\theta) \) given in Equation (8) for the considered scales. This choice ensures that condition (C.1) is satisfied.
- **Step 3.** Compute the model-based WV \( \phi(\theta) \).
- **Step 4.** Minimize the quadratic form defined in Equation (10) in \( \theta \) to obtain the GMWM estimator \( \hat{\theta} \).
- **Step 5.** Compute confidence intervals for \( \hat{\theta} \) by parametric bootstrap or using Equation (11) with \( \theta \) replaced by \( \hat{\theta} \).

Hence, the method is rather easy to implement in practice. Assuming that we have a time series, say \( y_t \), of length \( T \) and a model \( F_\theta \) made of the sum of processes satisfying Corollary 3, the estimation of \( \theta \) starts with the computation (only once) of the empirical WV using Equation (6) and their estimated variances using Equation (8) (that are used in \( \Omega \)). The model-based WV is then analytically built using the sum of the model-based WV (in Table 3) of the independent subprocesses in \( F_\theta \). Given all these ingredients, the quadratic form defined in Equation (10) is then optimized. This optimization problem is very simple. Confidence intervals for \( \hat{\theta} \) are, however, generally computed by parametric bootstrap. This task is quite easy to implement, but of course implies simulating data and repeating several times Steps 1–4. This can be taken in practice about a minute with massive datasets such as the one considered in the case study section.

If a simulation-based approach (indirect inference) is preferred, then Step 3 is not necessary and Step 4 is replaced by the following steps:

- **Step 4a.** Choose a sufficiently large value for \( R \) (\( R = 100 \) is usually a sensible choice)
- **Step 4b.** Minimize the quadratic form defined in Equation (10) in which \( \phi(\theta) \) is replaced by \( \phi^*(\theta) \), each element being computed using Equation (6) in which the observations are replaced by a simulated sample from \( F_\theta \) (for a given current value of \( \theta \) of size \( RT \)).

Step 5 can be then used for inference, for sufficiently large \( R \) (see, e.g., Genton and Ronchetti 2003).

### 4. SIMULATION STUDY

This section is dedicated to the finite sample performance evaluation of the GMWM estimator compared to the MLE and the LSE. For the GMWM estimator, one has to make a choice about \( \Omega \) in Equation (10). The choice is not straightforward in practice, as it is the case for the moment-based estimator GMM or Indirect-Inference-based estimators. Asymptotically, it is well known that the optimal choice for \( \Omega \) is any matrix proportional to \( V_\theta^{-1} \). This does not mean that this choice is better in finite samples, especially because the asymptotic covariance matrix needs to be estimated. A simple choice is the identity matrix, and in our case an intermediate solution is the inverse of the diagonal matrix made of the variance estimates of the MODWT using Equation (8). Indeed, in our simulations, we found that choosing \( \Omega = V_\theta^{-1} \) may lead in some cases to numerical instability, probably due to the estimation of the covariances using Equation (13). Anyway, asymptotically these three choices for \( \Omega \) lead to a consistent GMWM (provided the other conditions are also satisfied). When optimizing Equation (10), we use a quasi-Newton optimization method. For the MLE, we use the EM algorithm together with the Kalman smoother (EM-KF) as proposed by Shumway and Stoffer (1982) (see also Holmes 2010).

An important engineering application is sensor calibration in which the behavior of the errors affecting signals has to be well understood and modeled. In particular, gyroscopes and accelerometers which are part of inertial navigation systems used for space, aeronautical, ground, and underwater applications are subject to random errors that largely affect the quality of the positioning and navigation solution over time. A widely accepted model for the error behavior of an inertial sensor is given by the following state-space model

\[
x_{t+1} = e^{-\beta \Delta t} x_t + u_{t+1} + u_{t+1}
y_{t+1} = x_{t+1} + v_{t+1},
\]

where \( \Delta t \) is the time interval between two consecutive measurements, \( u_t = \omega \Delta t, u_t \sim N(0, q) \) with \( q = \sigma^2_{GM} (1 - e^{-2\beta \Delta t}) \), and \( v_t \sim N(0, \sigma^2_{WN}) \). The observed series is hence made up of the sum of a first-order GM process with correlation time \( \beta^{-1} \) and variance \( \sigma^2_{GM} \), a drift with slope \( \omega \), and a WN with variance \( \sigma^2_{WN} \). Even with such a relatively simple model, the estimation task is nontrivial.

For our simulation study, we consider simulated processes \( \{y_t, t = 1, \ldots, T\} \) from model (15) as well as an AR(1) with autocorrelation parameter near the boundary value of 1. From model (15), we actually generate three types of samples which correspond to three different (sub)models. In Model 1, we set \( \sigma^2_{WN} = \omega = 0 \), hence we have only a first-order GM, in Model 2, we set \( \omega = 0 \), hence we remove the drift only from the complete model (15) (Model 3). We generate 100 processes of size \( T = 6000 \) with \( \Delta t = 1 \) and \( \theta = (\sigma^2_{WN}, \sigma^2_{GM}, \beta, \omega) = (4, 16, 0.05, 0.005) \) at the complete model. For the submodels, the parameters are constrained accordingly and they are not estimated.

For both the GMWM (simulation based or not) and the EM-KF, the initial values for the optimizations were set to \( \theta_{\text{start}} = (1.0, 1.0, 1.0, 0.0) \), which is relatively far away from the true simulation values. We found that the choice for the starting
values is not a serious issue for the computation of the GMWM. The root mean-squared errors (RMSE) as well as the relative RMSE (relative to the true parameter value) are presented in Table 1 for Models 1–3, and for the GMWM and the EM-KF. We also tried a simulation-based version of the GMWM with $R = 100$ and found out that the RMSE are of the same order as for the GMWM with analytical WV (results not presented here). The sample WV were computed for $J = 12(< \log(6000)/\log(2) \approx 12.55)$ scales for all models. The results show that for the smaller models 1 and 2, the RMSE is smaller for the EM-KF than for the GMWM estimator, while the RMSE of the EM-KF explodes for the complete model 3. This last feature is actually well known with models with a drift component. For example, Hinrichsen and Holmes (2009) consider a multivariate state-space model composed of a drift with unknown rate and a random walk process to model the growth of an ecological population with observations that are subject to measurement error. They actually use a two-step estimation procedure by which the drift is first estimated by linear regression and then removed from the state equation leading to a simpler model that is well estimated by means of the EM-KF (see also Stebler et al. 2011).

When the EM-KF behaves well (models 1 and 2), it has a better performance in terms of RMSE than the GMWM estimator. However, one can further improve the efficiency of the latter by decreasing the number of scales $J$ at which the WV are estimated. Indeed, in this example, $J = 12$ scales are used to estimate two or three parameters, and if more scales are added (supposing a larger $T$), this only introduces more variability in the GMWM estimator. An optimal choice (in terms of GMWM estimator’s efficiency) of the scales and their number is beyond the scope of the present article. The number of scales is obviously a function of the number of parameters $\rho$, but their choice (among the possible ones) depends on the model that is considered. For example, with model (15) without the drift element (i.e., Model 2), according to the WV graph in Figure 4, if the last three to four scales are ignored, then the WV are still able to capture information about the other model components. Actually, removing the last four scales improves the efficiency of the GMWM estimator in the simulation study for Model 2 (results not shown here).

An important feature of the GMWM estimator is that it can be used for models for which the MLE (or other classical estimators) is seriously biased. For example, maximum likelihood estimation of ARMA processes, and/or least squares estimation (based on the Yule–Walker equations) of AR processes, when the roots lie near the unit circle, provide estimators that can be seriously biased (see, e.g., Andrews 1993). With the GMWM estimator with Haar coefficients for the WV, we do not have this boundary problem, since, for example, with an AR(1) process, with autocorrelation parameter $\rho$, when the latter tends to the value of 1, we get a random walk which is a nonstationary process. However, the wavelet coefficients $W_{ij}$, actually are backward differences of order at least 1, so that the problem of nonstationarity disappears. Note also that taking first differences of a stationary AR(1) process is known to lead to the problem of overdifferencing, since the resulting series is not invertible (it becomes an ARMA(1,1)) which has more parameters than the original one. Consequently, its parameters will be difficult to estimate, and it will tend to degrade the quality of forecasts, so differencing is not always a suitable method to overcome the estimation problem.

As an illustration, we simulated 1000 AR(1) processes of size $T = 5000$ with $\rho$ ranging from 0.8 to 1 (and with residual variance $\sigma^2 = 4$) and estimated the LSE based on the Yule–Walker equations together with the GMWM ($J = 12$). Figure 1 presents the relative RMSE of the LSE and GMWM estimator for both parameters. The GMWM estimator of $\rho$ is less precise than the LSE, except when $\rho$ is close to 1. For $\sigma^2$, the RMSE of both estimators are quite similar for $\rho$ smaller than 0.9 but for higher values the bias and variance of the LSE explodes showing that there is a boundary effect that affects the LSE. The MLE provides RMSE that are sensibly smaller than the LSE (not presented here) but for values of $\rho$ larger than 0.95 the MLE is difficult if not impossible to compute.

<table>
<thead>
<tr>
<th>Model</th>
<th>GMWM RMSE</th>
<th>GMWM R-RMSE</th>
<th>EM-KF RMSE</th>
<th>EM-KF R-RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>$\sigma^2_{GM}$</td>
<td>2.00</td>
<td>0.13</td>
<td>1.26</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>8.14 · 10⁻³</td>
<td>0.16</td>
<td>4.09 · 10⁻³</td>
</tr>
<tr>
<td>2</td>
<td>$\sigma^2_{GM}$</td>
<td>1.92</td>
<td>0.12</td>
<td>1.55</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>1.05 · 10⁻²</td>
<td>0.21</td>
<td>4.94 · 10⁻³</td>
</tr>
<tr>
<td>3</td>
<td>$\sigma^2_{GM}$</td>
<td>0.96</td>
<td>0.06</td>
<td>74.57</td>
</tr>
<tr>
<td></td>
<td>$\beta$</td>
<td>4.63 · 10⁻³</td>
<td>0.09</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>$\sigma^2_{WN}$</td>
<td>0.11</td>
<td>0.03</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>$\omega$</td>
<td>2.79 · 10⁻⁴</td>
<td>0.06</td>
<td>0.12</td>
</tr>
</tbody>
</table>

Figure 1. Relative RMSE based on 1000 simulations from AR(1) processes with $\rho$ ranging from 0.8 to 1 and with residual variance $\sigma^2 = 4$, of the LSE relative to the GMWM estimator of $\rho$ (•) and of $\sigma^2$ (○).
5. CASE STUDY: INERTIAL SENSORS

An inertial navigation system exploits observation from usually an orthogonal triad of gyroscopes and accelerometers measuring angular rates and specific forces, respectively. Once initialized, the navigation is a dead-reckoning process (i.e., the solution at one epoch is computed from current observations as well as from the solution of the previous epoch) in which gyroscope signals (after suppressing Earth rotation) are integrated once to yield orientation, and accelerometers (after accounting for gravity and apparent forces) are integrated twice to get the velocity, and finally the position in three-dimensional space. The advantage of inertial navigation is in the autonomy (i.e., no external infrastructure is needed) and the large sensor bandwidth (unlike satellite positioning system), while its weakness lies in the time-dependent error behavior (due to the integration process). Indeed, the sensor signals are corrupted by random errors, making the resulting positioning and attitude (or orientation) error increase rapidly with time. To bound the error growth, inertial sensors are combined with other sensors (e.g., GPS, odometer, and altimeter) through optimal (e.g., Kalman) filtering. Such combined navigation systems are not only the base for manned and unmanned vehicle (e.g., spacecrafts, aircrafts, cars, and robots) trajectory and orientation determination, but also for pedestrian and indoor localization devices. The gyroscope and accelerometer error behavior is modeled through state augmentation in the system where the augmented states account for sensor errors usually modeled as stochastic processes. This requires that the parameters of the stochastic processes in the state-space model are carefully set up prior to filtering. The quality of this setup (or estimation) strongly influences the quality of the predicted trajectory. Traditionally, the Allan variance with ad hoc graphical estimation (see, e.g., IEEE Standard 1293 1998), or the KF-Self-Tuning approach (see Niu et al. 2007; Waegli 2009; Stebler et al. 2012), or even the MLE on state-space models (with simple models) (Stebler et al. 2011) are used to both determine the error’s model structure and to estimate their parameters. The GMWM estimator offers an alternative approach that is especially suitable when the stochastic error process is rather complex.

In this section, we consider the angular rate signal issued from a real Micro-Electro-Mechanical System (MEMS) gyroscope where the spectral structure of errors is often complex. Such sensors have a great potential in navigation due to ergonomic (smaller and lighter equipment) and economical considerations (El-Sheimy and Niu 2007). Therefore, MEMS sensors are increasingly used in a very wide range of applications such as 3D input devices, robotic, virtual reality, vehicle stability control and so forth, and research focuses on modeling and compensating for their large and variable measurement errors. To validate the use of the GMWM for modeling the MEMS’ errors, we first analyze the signal produced in static conditions at a frequency of 100 Hz (833, 685 measurements). We provide GMWM estimates on a rather complex model that cannot be estimated by an alternative standard method. To show the impact of the importance of model structure and estimation precision of its parameters in positioning, we then present an emulation study in which the trajectory of a helicopter performing airborne laser scanning is used and for which a nearly perfect (i.e., reference) trajectory can be computed. The emulated measurements along the real trajectory are corrupted with an error signal observed in static condition, and the three methods (AV, the KF-Self-Tuning Approach, and GMWM) are used to model the error signal. The predicted trajectories based on the corrupted measurements and estimated error model, together with GPS measurements, are then compared to determine the positioning error.

After removing the mean in the signal made of 833, 685 measurements produced in static conditions at a frequency of 100 Hz, the time-varying part of the sensor errors is directly available and is presented in the time domain in Figure 3 (upper panel), together with the Haar WV of this process with 95% confidence intervals (lower panel). The WV computed on the original signal give an indication of the underlying stochastic processes that are summed up to build up a composite process. A possible model is a mixture of three different first-order GM processes \( x_{t+1} = e^{-\beta}x_t + w_t, \) where \( w_t \sim \mathcal{N}(0, q) \) with \( q = \sigma^2_{GM}(1 - e^{-2\beta t}) \) and \( \Delta t = 0.01 \) s, the sampling interval. In such a case, the parameters to be estimated are \( \theta = (\sigma^2_1, \beta_1, \sigma^2_2, \beta_2, \sigma^2_3, \beta_3) \) and the GMWM is identifiable since a first order-Markov random process is a one-to-one reparameterization of an AR(1) process whose sum allows identifiable GMWM (see Corollaries 2 and 3). The GMWM estimates of the parameter set \( \theta \) and its corresponding 95% confidence intervals using the WV covariances estimated by means of a parametric bootstrap (400 samples) are given in Table 2. The suitability of the estimated model for the data at hand can be judged graphically by a matching of the empirical WV and the parametric WV using the estimates in Table 2, as it is done in Figure 3 (lower panel). It can be seen that the estimated process matches the observed one very well across WV scales, yielding an indication of a suitable fit. Moreover, one can also compute the goodness of fit test statistic in Equation (14) for competing models and test to what extent the chosen one yields a suitable fit. For that, we also considered a single GM process and the sum of two GM processes. The test statistics and corresponding p-values are 38436.9 (on 17 df) and \( p \)-value < 10^{-5} for the single GM process, 84787.01 (on 15 df) and \( p \)-value < 10^{-5} for a sum of two GM processes, and 16,81637 (on 13 df) and \( p \)-value = 0.208 for a sum of three GM processes, respectively, clearly indicating that the last model is suitable. The corresponding estimated WV are drawn in Figure 3 (lower panel).

It should be noted that the sum of three first-order GM random processes can be reparameterized as an ARMA(3,2) process (see, e.g., Granger and Morris 1976; Terasvirta 1977), so that one could, in principle, estimate the latter instead of the

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Estimates</th>
<th>IC(0.05)</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \beta_1 )</td>
<td>2.1720 · 10^2</td>
<td>(2.1465 · 10^2; 2.1906 · 10^2)</td>
</tr>
<tr>
<td>( \sigma_1 )</td>
<td>7.4521 · 10^{-3}</td>
<td>(7.4477 · 10^{-3}; 7.4688 · 10^{-3})</td>
</tr>
<tr>
<td>( \beta_2 )</td>
<td>6.0693 · 10^{-1}</td>
<td>(6.7890 · 10^{-1}; 7.8655 · 10^{-1})</td>
</tr>
<tr>
<td>( \sigma_2 )</td>
<td>2.9691 · 10^{-4}</td>
<td>(2.9494 · 10^{-4}; 2.9870 · 10^{-4})</td>
</tr>
<tr>
<td>( \beta_3 )</td>
<td>3.5563 · 10^{-3}</td>
<td>(3.5333 · 10^{-3}; 3.9086 · 10^{-3})</td>
</tr>
<tr>
<td>( \sigma_3 )</td>
<td>5.5127 · 10^{-4}</td>
<td>(5.4038 · 10^{-4}; 5.6223 · 10^{-4})</td>
</tr>
</tbody>
</table>
former. However, one should then be able to invert the estimated ARMA since very often, and in particular in the example at hand, a sum of first-order GM random processes is the one that we believe explains the real underlying process, which for gyroscopes reflect the short- and long-term correlated noise due to quantization and mechanical thermal noise phenomena present in MEMS sensors (see Kittel 1958; Drexler 1992). To recover the sum of GM processes’ parameters from an estimated ARMA process, and also their standard errors, several conditions need to be satisfied. First, the roots of the processes must lie outside the unit circle. Second, the Jacobian matrix of the transformation between the two parameterizations must be invertible to apply the delta method (Rao 1973; Benichou and Gail 1989). With the inertial sensor data, the estimated processes have roots that are near the unit circle, and moreover, the Jacobian matrix of the transformation evaluated at the estimated parameters is not invertible. In that case at least, estimating an ARMA process and converting the estimated model and inference to the sum of first-order GM processes is infeasible.

To illustrate the impact of the importance of model structure and estimation precision in device positioning, we also performed an emulation study that was presented in Stebler et al. (2012). Using extremely reliable equipments, the trajectory of a helicopter performing airborne laser scanning can be precisely determined. From this trajectory, we computed the theoretical perfect inertial measurements (i.e., accelerations and angular rotations) that should be observed along this trajectory. These perfect measurements were then corrupted with a real static error signal acquired such as the one analyzed above. Then, these “pseudo” measurements were used together with the real GPS measurements to compute a trajectory. At one point, we introduced relatively short artificial gaps in GPS observations of 1 min duration (note that, in practice, such gaps are very common) and compared the deviations of the trajectories obtained with different inertial error models estimated from the data. The two first estimated models are based on the AV and on the KF-Self-Tuning approach. A detailed discussion on these methods can be found for example in Stebler et al. (2012). The GMWM was also used as an alternative estimator and model building approach. The trajectories are depicted in Figure 2 (adapted from Stebler et al. 2012) along with the “true” trajectory. It can be seen that the GMWM-based model (black-dashed line) limits significantly the error growth during the GPS-signal outage compared with the other two benchmark methods which diverged from the “true” trajectory by several thousand of meters! The poor performances of the standard methods explains the recent explosion of the research conducted to determine the stochastic modeling of MEMS-type inertial sensors.

6. CONCLUDING REMARKS

In this article, we present a new estimator for the parameters of composite stochastic processes which is shown to be consistent for the class made of the sum of independent white-noise, drift, quantization noise, random walk, and $k < \infty$ AR(1). As demonstrated in Stebler et al. (2012), the GMWM has many advantages over existing alternative methods for applications in engineering or natural sciences. The enlargement of the class of models for which the GMWM is consistent involves verifying the conditions provided in Theorem 2 and more generally in Theorem 1 (with Corollary 1) and is left for future research. Also, one could use in principle another consistent estimator for
the Haar WV, such as the one proposed in Nason, von Sachs, and Kroisandt (2000).

As an anonymous referee pointed out, one could wonder why to use wavelets instead of making inference directly on the spectral density for which a similar estimation procedure could be used. We believe that our approach is more suitable for the following reasons. First, inference on the PSD would make the optimization of a least-squares-type measure (between the empirical and model-based PSD) more difficult to solve when the PSD has large variability over very narrow frequency bands. As shown in Percival and Walden (2000), the wavelet coefficients at scale \( \tau_j \) are associated with frequencies in the interval \([1/2^{j+1}, 1/2^j]\) and relation (5) can be approximated by

\[
v_\tau^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S(f) df.
\]

(16)

This means that the WV summarizes the information in the PSD using just one value per octave frequency band. Thus, it is particularly useful when the PSD is relatively featureless within each octave band. In the case of the widely used pure power law (particularly useful when the PSD is relatively featureless within each octave frequency band), using just one value per octave frequency band. Thus, it is particularly useful when the PSD is relatively featureless within each octave band.

The proofs and some additional results are given in the supplementary materials. This document is organized as follow. In Section A we provide the analytical version of the Haar WV for the processes considered in this article. In Section B we explain the graphical use of the WV for model building purposes. Theorem 3 which demonstrates the multivariate normality of the WV is given in Section C. The proofs of Theorems 1 and 2 are given in Sections D and E. Corollaries 2 and 3 which provides some additional results regarding the consistency of composite stochastic processes are given in Section F. Finally, in Section G we explain the construction of the covariance estimator of the WV presented given (13).

NOTES

A C++ program for computing the GMWM for the models treated in this article is available upon request from the authors.

REFERENCES


Daubechies, I. (1992), Ten Lectures on Wavelets, Philadelphia: SIAM. [1022,1023]


Drexler, K. E. (1992), Nanosystems—Molecular Machinery, Manufacturing, and Computation, New York: Wiley. [1028]


Supplementary Materials

Wavelet variance based estimation for composite stochastic processes

Stéphane Guerrier Jan Skaloud, Yannick Stebler & Maria-Pia Victoria-Feser

A  Parametric Wavelet Variances for some specific processes

The results presented in Table 3 are based on the developments provided in Zhang (2008) for the AV. Indeed, in the latter, the $n$ parameter corresponds to the scales $\tau_j$ and the analytical version of the AV must be divided by 2 to obtain the analytical version of the Haar WV. To obtain the analytical WV of sums of independent processes, one just has to sum the WV of each process. For more general models such as ARFIMA($p,q,d$) see Zhang (2008).

B  Graphical use of Wavelet Variances for model building purposes

If one assumes that the underlying processes that are summed up to build the observed signal are independent, then the resulting WV is the sum of the WV of each independent process. Elementary processes, such as AR(1), WN etc. have a known form and recognizable shape in the WV log-log scale plot (one actually takes the square root of the WV). For example, a WN process is characterized by a straight line with slope $-1/2$, a RW by a nearly straight line with slope $1/2$ and AR(1) by a quadratic-type shape. With some training, one can try to guess which processes may compose the observed signal (on the log-log scale graphic); see also e.g. El-Sheimy et al. (2008). Finding a reasonable model might take several attempts; one can either compare graphically the model fitted versus the empirical WV (log-log scale), or use the $\chi^2$-test (14).

To illustrate this, consider a series $\{y_t; t = 1, \ldots, T\}$, generated from model (15) with $\Delta t = 1$, $\beta = 0.05, \sigma^2_{GM} = 16, \omega = 0.005, \sigma^2_{WN} = 4, T = 6000,$
Table 3: Analytical PSD and WV implied by WN, RW, Drift, QN, AR(1), ARMA(1,1) and ARIMA(0,1,1) models. We use $\tau$ for $\tau_j$.  

<table>
<thead>
<tr>
<th>Model</th>
<th>$\theta$</th>
<th>$\nu^2(\tau)$ (Haar)</th>
</tr>
</thead>
<tbody>
<tr>
<td>White Noise (WN)</td>
<td>$\sigma^2$</td>
<td>$\frac{\sigma^2}{2\pi}$</td>
</tr>
<tr>
<td>Random Walk (RW)</td>
<td>$\omega$</td>
<td>$\frac{\omega^2}{2\pi}$</td>
</tr>
<tr>
<td>Drift</td>
<td>$\sigma^2$</td>
<td>$\frac{\sigma^2}{2\pi\sin^2(\frac{\pi f}{2})}$</td>
</tr>
<tr>
<td>Quantization Noise (QN)</td>
<td>$\rho, \sigma^2$</td>
<td>$\frac{\rho^2}{2\pi\sin^2(\frac{\pi f}{2})}$</td>
</tr>
<tr>
<td>AR(1)</td>
<td>$\rho, \psi, \sigma^2$</td>
<td>$\frac{\rho^2}{2\pi\sin^2(\frac{\pi f}{2})}$</td>
</tr>
<tr>
<td>ARMA(1,1)</td>
<td>$\rho, \psi, \sigma^2$</td>
<td>$\frac{\rho^2}{2\pi\sin^2(\frac{\pi f}{2})}$</td>
</tr>
<tr>
<td>ARIMA(0,1,1)</td>
<td>$\rho, \psi, \sigma^2$</td>
<td>$\frac{\rho^2}{2\pi\sin^2(\frac{\pi f}{2})}$</td>
</tr>
</tbody>
</table>

$0 < \rho < 1, |\psi| < 1, Q > 0$  

$0 < \sigma^2 < 0, |\sigma| < 1, Q > 0$
whose realization is presented in the top panel of Figure 4. From the output \{y_t\}, WV at scales of size $2^j$, $j = 1, 2, \ldots, J \leq \log(T)/\log(2)$ are computed and then plotted in the same figure. They actually correspond to the sum (on the log-log scale graphic) of the individual processes within the model. The WV of the individual processes are represented separately on the same Figure.

![Simulated process and Haar WV representations](image)

**Figure 4**: Simulated data process (15) with $\Delta t = 1, \beta = 0.05, \sigma^2_{GM} = 16, \omega = 0.005, \sigma^2_{WN} = 4, T = 6000$ (top panel) and graphical representation (log-log scale) of the Haar WV for the complete process (line “o”), the White Noise (“o” line), first order GM (“□” line) and Drift (“△” line).

The WV is used in many applied fields for model building purposes, for example to study variation of soil properties (Lark and Webster, 2001), canopy structure (Bradshaw and Spies, 1992), industrial production index (Gallegati and Gallegati, 2007), surface temperature and albedo (Lindsay
et al., 1996) or sea level pressure changes (Torrence and Compo, 1998). The definition of WV was extended to wavelet packets in Gabbanini et al. (2004) and was employed for determining crack widths on the Brunelleschi dome of the Santa Maria del Fiore cathedral in Florence.

C Multivariate normality of WV

**Theorem 3.** Let the stationary or non-stationary but with backward differences of order d time series \(\{Y_t; t \in \mathbb{Z}\}\) be such that the corresponding wavelet coefficients \(W_{j,t}, j = 1, \ldots, J, J < \infty\) are stationary with

\[
W_{j,t} = P_s^2 Z_{a,j}(t-s) \sim \text{iid } \mathcal{N}(0, 1),
\]

where \(E[|\xi_s^4|] < \infty\), and with \(\sum a_j^2(t) < \infty, j = 1, \ldots, J\). Let the PSD of \(W_{j,t}\) be such that there exists a \(C\) and

\[
S_{W_j}(f) \leq C|f|^{-\alpha_j}, f \in [-1/2, 1/2]
\]

If

\[
\alpha_j < 1/2, \forall j = 1, \ldots, J
\]

then \(\sqrt{T}(\hat{\phi} - E[\hat{\phi}]) \sim \mathcal{N}(0, \mathbf{V}_\phi)\) as \(T \to \infty\), where \(\mathbf{V}_\phi = [\sigma_{kl}^2]_{k,l=1,\ldots,J}\) with

\[
\sigma_{kl}^2 = \sum_{\tau=-\infty}^{\infty} \text{cov}(W_{k,t}^2, W_{l,t+\tau}^2)
\]

**Proof.** Theorem 3 is in fact a special case of Theorem 6.2 of Giraitis and Taqqu (1997), which applies to quadratic forms given by bivariate Appell polynomials

\[
Q^j_T = \sum_{t,s=1}^{T} b_j(t-s)P_{m_j,n_j}(X_{j,t}, X_{j,s})
\]

in which \(X_{j,t} = W_{j,t}, m_j = n_j = 1, b_j(\tau) = \delta(\tau) = 1/M_j\) if \(\tau = 0\) and 0 otherwise, so that \(Q^j_T = \hat{\nu}_{j}^2(\tau_j), \forall j = 1, \ldots, J\). The first condition in the A2 of Theorem 6.2 of Giraitis and Taqqu (1997) is on \(X_{j,t}\), i.e. \(W_{j,t}\), namely (A-1). The second is on the PSD of \(b_j(\tau)\), and since the form of \(b_j(\tau)\) implies that \(\beta_j = 0, \forall j\), then the second condition is also satisfied together with the last condition (i.e. (2.13) in Giraitis and Taqqu, 1997) which reduces to \(\alpha_j = \alpha < 1/2\), i.e. (A-2).
Notes. (A) The number of WV $J$ is fixed. (B) Conditions (A-1) and (A-2) ensure weak dependence for the process $W_{j,t}$. (C) The conditions do not imply Gaussianity of the process. (D) For assumptions in the time domain, see Giraitis and Taqqu (1998), Theorem 5.1.

D Proof of Theorem 1

Proof. We write $g_1(\theta) = g_1(k(\gamma)) = \tilde{g}_1(\gamma)$, so that $(\partial/\partial \gamma)\tilde{g}_1(\gamma) = (\partial/\partial \theta)g_1(\theta) (\partial/\partial \gamma)k(\gamma)$. Then conditions (C.3) to (C.8) ensure the conditions on $\tilde{g}_1(\gamma)$, $\gamma \in \mathbb{R}^p$ for $\gamma$ to be globally identified, i.e. $\tilde{g}_1(\gamma) = 0$ if and only if $\gamma = \gamma_0$ according to Theorem 1 in Komunjer (2012) for the case $p \neq 2$. Moreover, if $p = 2$, the Gale-Nikaido-Fisher-Rothenberg conditions, see for example Fisher (1965) or Rothenberg (1971), are satisfied and global identification also holds. Since $k$ is bijective, then global identification for $\gamma$ implies $g_1(\theta) = 0$ if and only if $\theta = \theta_0$ (with $\theta_0 = k(\gamma_0)$). Moreover, if $g(\theta) = 0$ then $\theta = \theta_0$ (because $g_1(\theta) = 0$) and if $\theta = \theta_0$ then $g(\theta) = g(\theta_0) = 0$ because of the consistency of $\hat{\theta}$. Finally, since the VW $\hat{\phi}_j$ have mean $\nu^2(\tau_j)$ and finite variance (7) the condition 1.2 in Harris and Mátyás (1999) is satisfied by the weak law of large numbers, and together with condition (C.1), the assumptions of Theorem 1.1 in Harris and Mátyás (1999) are satisfied so that $\hat{\theta}$ is consistent.

E Proof of Theorem 2

Proof. The proof is straightforward. First recall that the WV of the sum of independent processes is equal to the sum of the WV of each sub process. Condition (C.10) insures identifiability of each sub process and condition (C.9) insures that a modification of the WV of one sub process cannot be compensated, in the sum, by a modification in one or more other sub processes. In order for the sum of all sub processes WV to be uniquely determined (in $\theta$), one has to ensure that there is no more than one sub process for which one of the parameters governs vertical translation. This is ensured by condition (C.11).

Hence, $\theta$ is identifiable so that (C.3) to (C.8) of Theorem 1 are satisfied together with conditions (C.1) and (C.2), so that $\hat{\theta}$ defined in (10) is a consistent estimator of $\theta$. 

5
Consistency for some processes

We consider here WV built using Haar filters.

**Corollary 2.** Under conditions (C.1) and (C.2), the parameters of the following processes

1. **White noise (WN) with parameter** $\sigma^2 \in \mathbb{R}^+$

2. **Quantization noise (QN) (or rounding error, see e.g. Papoulis and Pillai, 1991) with parameter** $Q^2 \in \mathbb{R}^+$

3. **Drift with parameter** $\omega \in \Omega$ where $\Omega$ is either $\mathbb{R}^+$ or $\mathbb{R}^-$

4. **Random walk (RW) with parameter** $\gamma^2 \in \mathbb{R}^+$

5. **AR(1) process with parameter** $\rho \in \Phi$ where $\Phi$ is either $(-\infty, -1)$ or $(-1, +\infty)$ and $\nu^2 \in \mathbb{R}^+$

**can be consistently estimated by the GMWM estimator $\hat{\theta}$.**

**Proof.** First, we consider a WN process with parameter $\sigma^2 \in \mathbb{R}^+$. An appropriate transformation for $\sigma^2$ to be spanned in $\mathbb{R}$ is the logarithm function which satisfies condition (C.4). Moreover, if we let the function $g_1(\theta)$ be defined as the WV on the first scale (i.e. $\tau = 1$) we have $g_1(\theta) = E[\phi_1] - \sigma^2/2$. Clearly, this function is twice continuously differentiable (and therefore satisfying condition (C.4)) and $||g_1(\theta)||$ tends to infinity whenever $\sigma^2$ goes to infinity (satisfying condition (C.6)). As $\frac{\partial}{\partial \sigma^2} g_1(\theta) = -\frac{1}{2}$ condition (C.7) is also verified. Condition (C.5) is verified since $-|g_1(\theta)|$ is non-negative and as $p = 1$ condition (C.8) is verified.

Second, we consider a QN with parameter $Q^2 \in \mathbb{R}^+$. Using the logarithm for the parameter $Q^2$ spans it in $\mathbb{R}$ and satisfies (C.4). Moreover, if we let the function $g_1(\theta)$ be defined as the WV on the first scale (i.e. $\tau = 1$) we have $g_1(\theta) = E[\phi_1] - \frac{3}{2} Q^2$. This function satisfies conditions (C.4) to (C.7) and also (C.8) as $p = 1$.

Third, we consider a drift with parameter $\omega$ with known sign, then such process may be reparametrized as $\omega^* = |\omega| \in \mathbb{R}^+$. As for the previous processes the logarithm transformation can be used which verifies condition (C.4). Let the function $g_1(\theta)$ be defined as the WV on the first scale (i.e. $\tau = 1$) we have $g_1(\theta) = E[\phi_1] - \frac{1}{4} (\omega^*)^2$. This function verifies conditions (C.4) to (C.8) as $\omega^* \in \mathbb{R}^+$ and $p = 1$. 

Then, we consider a RW with parameter $\gamma^2 \in \mathbb{R}^+$. As for the WN process one can use the logarithm to span $\gamma^2$ in $\mathbb{R}$ (which verifies (C.4)). Let the function $g_1 (\theta)$ be defined as the WV on the first scale (i.e. $\tau = 1$) we have $g_1 (\theta) = \mathbb{E} [\phi_1] - \frac{1}{4} \gamma^2$. Clearly, this function verifies conditions (C.4) to (C.7) and as $p = 1$ condition (C.8) is also verified.

Finally, we consider an AR(1) process with parameter $\rho \in \Psi$ where $\Psi$ is either $(-\infty, -1)$ or $(-1, +\infty)$ and $\nu^2 \in \mathbb{R}^+$. The logarithm function may be used to span $\nu^2$ in $\mathbb{R}$ and the function $k(x) = \log (x + 1)$ which spans $\rho$ in $\mathbb{R}$ if $\Psi$ is $(-1, +\infty)$ (if $\Psi$ is $(-\infty, -1)$ we may use the function $k(x) = \log (1 - x)$). These functions satisfy condition (C.4). Let $g_1 (\theta)$ be defined as the model based WV on the scales 2 and 3 (i.e $\tau_2 = 2$ and $\tau_3 = 4$), then

$$g_1 (\theta) = \left( \begin{array}{c} \mathbb{E} [\phi_2] - \frac{(1 - 3\rho^2 + 3\rho^2 - \rho^4)\nu^2}{4(1 - \rho^2)(1 - \rho^2)} \\ \mathbb{E} [\phi_3] - \frac{(2 - 3\rho - 2\rho^2 + 4\rho^3 - \rho^5)\nu^2}{2(1 - \rho^2)(1 - \rho^2)} \end{array} \right)^T$$

This function verifies conditions (C.4), (C.6) and (C.7). Condition (C.5) is also satisfied as

$$\left| -(\partial / \partial \theta)g_1 (\theta) \right| = \frac{\nu^2 (2\rho + 3)}{64 (\rho + 1)^2}$$

is non-negative with $\nu^2 \in \mathbb{R}^+$ and $\rho \in \Psi$ in addition $-(\partial / \partial \theta)g_1 (\theta)$ is positive definite (verifying (C.8)).

Therefore, under condition (C.1) and (C.2) the processes (P.1) to (P.5) can be consistently estimated by the GMWM estimator. \(\square\)

**Corollary 3.** Assuming independent processes, 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.3) and $k < \infty$ processes (P.5) can be consistently estimated by the GMWM estimator under condition (C.1), (C.2) and if $\psi_j \neq \psi_l$, $\psi_j \neq 1$ and $\psi_j \neq 0 \forall j, l = 1, \ldots, k, j \neq l$ where $\psi_j$ denotes the parameter $\psi$ of the $j^{th}$ process (P.5).

**Proof.** Under the restrictions imposed on the parameters $\psi_j$ of the AR(1) processes condition (C.9) is satisfied as none of the WV of the different processes are linearly dependent. Moreover, as the processes (P.1) to (P.5) defined in Corollary 2 have parameters that can be consistently estimated by the GMWM estimator condition (C.10) is automatically satisfied. Finally it can easily be verified that differentiating the WV of the process at hand
with respect to any element of $\theta$ provides a function that depends on $\tau$ thus verifying condition (C.11). Then the GMWM estimator $\hat{\theta}$ is consistent. □

G Covariance estimation of WV estimators

We first note that $M_{jk}/T \to 1$ as $T \to \infty$ so that we use $T$ instead of $M_{jk}$ below. We have that

$$
\sigma_{kl}^2 = \text{cov} \left[ \hat{\nu}^2(\tau_k), \hat{\nu}^2(\tau_l) \right]
$$

$$
\approx \text{cov} \left[ \frac{1}{T} \sum_{t=1}^{T} W_{k,t}^2, \frac{1}{T} \sum_{s=1}^{T} W_{l,s}^2 \right]
$$

$$
= \frac{1}{T^2} \sum_{t=1}^{T} \sum_{s=1}^{T} \text{cov} [W_{k,t}^2, W_{l,s}^2]
$$

$$
= \frac{1}{T} \sum_{\tau=-T}^{T} \left( 1 - \frac{|\tau|}{T} \right) \text{cov} [W_{k,t}^2, W_{l,t+\tau}^2]
$$

$$
= \frac{2}{T} \sum_{\tau=-T}^{T} \left( 1 - \frac{|\tau|}{T} \right) \text{cov} [W_{k,t}, W_{l,t+\tau}]^2
$$

the last equality being obtained using Isserlis theorem (Isserlis, 1918) for Gaussian processes $W_{k,t}$. Moreover, as $T \to \infty$ we have

$$
\sigma_{kl}^2 = 2 \int_{-1/2}^{1/2} S_{W_k,W_l}(f)^2 df = 2 \int_{-1/2}^{1/2} S_{W_l,W_k}(f)^2 df \quad (A-4)
$$

Note that we can also write $2 \int_{-1/2}^{1/2} S_{W_k,W_l}(f)^2 df = \int_{-1/2}^{1/2} S_{W_k,W_l}(f)^2 df + \int_{-1/2}^{1/2} S_{W_l,W_k}(f)^2 df$. To estimate the quantity in (A-4), we follow Whitcher et al. (2000) to obtain (13). Note that with finite $T$, the left and right elements of the sum in the right-handside of (13) are not equal, so to obtain symmetric covariances, it is important to estimate them this way.
References


