Abstract
This paper presents a new estimation method for the parameters of a model generating times series. Given some conditions on the form of the power spectral density associated to the process, it is possible to indirectly recover parameter estimates from wavelet variances (WV) associated to the process. We propose an optimization criterion based on a standardized distance between the sample WV estimates and the model based WV, as is done e.g. with the generalized method of moments and therefore call the new estimator the generalized method of Wavelet Moments (GMWM) estimator. Moreover, it can be computed using simulations, so that it is very straightforward to implement in practice, since the only specification that is needed for a given model, is the data generating process. We derive its asymptotic properties for inference and perform a simulation study to compare the GMWM estimator to the MLE and another estimator with different models. We also use it to estimate the stochastic error’s parameters in accelerometer and gyroscopes composing inertial navigation systems by means of a sample of over 8000000 measurements, [...]
Generalized Method of Wavelet Moments

Stéphane Guerrier,
Research Center for Statistics,
University of Geneva, Geneva, Switzerland.
E-mail: Stephane.Guerrier@unige.ch,

Yannick Stebler and Jan Skaloud,
Geodetic Engineering Laboratory,
Swiss Federal Institute of Technology, Lausanne, Switzerland.
E-mail: jan.skaloud@epfl.ch, yannick.stebler@epfl.ch, and

Maria-Pia Victoria-Feser,
Research Center for Statistics,
University of Geneva, Geneva, Switzerland.
E-mail: Maria-Pia.VictoriaFeser@unige.ch

May 13, 2011
Abstract

This paper presents a new estimation method for the parameters of a model generating times series. Given some conditions on the form of the power spectral density associated to the process, it is possible to indirectly recover parameter estimates from wavelet variances (WV) associated to the process. We propose an optimization criterion based on a standardized distance between the sample WV estimates and the model based WV, as is done e.g. with the generalized method of moments and therefore call the new estimator the generalized method of Wavelet Moments (GMWM) estimator. Moreover, it can be computed using simulations, so that it is very straightforward to implement in practice, since the only specification that is needed for a given model, is the data generating process. We derive its asymptotic properties for inference and perform a simulation study to compare the GMWM estimator to the MLE and another estimator with different models. We also use it to estimate the stochastic error’s parameters in accelerometer and gyroscopes composing inertial navigation systems by means of a sample of over 800’000 measurements, for which no other estimation method can be used.

Keywords: State space models, generalized method of moments, time series, wavelet variance, indirect inference
1 Introduction

We propose a new estimator for the parameters of complex models which involve some type of dependent structure between observations. Such models encompass models for multivariate dependent variables or dynamical systems for example. In this paper, we concentrate on univariate series \{Y_t; t \in \mathbb{Z}\} that are stationary or non-stationary but with stationary backward differences of order \(d\). We denote by \(F_\theta\) the model associated to \{Y_t; t \in \mathbb{Z}\} and suppose that it generates the observed outcome \{y_t, t = 1, \ldots, T\}. An important useful tool for understanding underlying features of the process is the Power Spectral Density (PSD) computed in the frequency-domain. It is given by \(S_{F_\theta}(f) = \Delta t \sum_{\tau=-\infty}^{\infty} C_{F_\theta}(\tau) e^{-i2\pi f \tau \Delta t}\) for \(|f| \leq f_N = 1/(2\Delta t)\) (\(f_N\) is the Nyquist frequency) with \(C_{F_\theta}(\tau) = \text{cov}_{F_\theta}[Y_{t+\tau}, Y_t]\) is the Auto-Correlation Function (ACF) of \{Y_t\} such that \(\sum_{\tau=-\infty}^{\infty} C_{F_\theta}^2(\tau) < \infty\).

The estimation method we propose for the parameters \(\theta\) of the model \(F_\theta\) is based on a decomposition of the variance process \(\text{var}_{F_\theta}[y_t] = \int_{-1/f_N}^{1/f_N} S_{F_\theta}(f) \, df\) given by the Wavelet Variance (WV) which is thoroughly described in Percival and Walden (2000) (see also Section 2). Namely, for processes with PSD proportional to \(|f|^{-\alpha}\), \(f\) being the frequency, when \(-1 < \alpha < 3\) the WV uniquely defines the PSD (see Van Vliet and Handel 1982) so that sample information collected via WV estimates provides sufficient information for the estimation of the model’s parameters \(\theta\). We propose an optimization criterion based on a standardized distance between the sample WV estimates and the model based WV, as is done e.g. with the generalized method of moments (Hansen 1982) and therefore call the new estimator the generalized method of Wavelet Moments (GMWM) estimator. We show that it is asymptotically normal when \(\alpha < 1/2\). Moreover, it can be computed using simulations like with indirect inference (Smith 1993, Gourieroux et al. 1993, Gallant and Tauchen 1996), so that it is very straightforward to implement in practice, since the only specification that is
needed for a given model, is the data generating process.

For illustrative purposes, but without loss of generality, we consider here as an example Gaussian processes and a useful representation is given by a state-space model of the form

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \mathbf{w}_{t+1} + \mathbf{u}_{t+1}$$

(1)

with measurements

$$y_{t+1} = h \mathbf{x}_{t+1} + v_{t+1}$$

(2)

where $\mathbf{x}_t$ 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 $t+1$, $\mathbf{w}_t$ is a $q \times 1$ multivariate Gaussian white noise vector, i.e. $\mathbf{w}_t \sim \mathcal{N}(0, \mathbf{Q})$, $\mathbf{u}_t$ 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 $\mathbf{x}_t$ into the output $y_t$, and $v_t$ is a one-dimensional noise such that $v_t \sim \mathcal{N}(0, \sigma^2_{WN})$.

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 can be 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 $f(y; \mathbf{\theta}) = \partial F_{\mathbf{\theta}}(y) / \partial y$, with $\mathbf{\theta}$ the $p \times 1$ vector of parameters containing the unknown elements of $\Phi, \mathbf{Q}, h$ and possibly also $\mathbf{u}_{t+1}$ together with $\sigma^2_{WN}$. The maximization step can be very complex and finding the MLE is not always a simple task.

State space models, or indeed dynamic models, are used in many fields such as ecology, medical and environmental sciences, economics and finance, or engineering applications. For example, sensor calibration in which the behaviour of the errors
affecting signals has to be well understood and modelled are at the core of many engineering problems. 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 behaviour of an inertial sensor is given by the following state space model

$$x_{t+1} = e^{-\beta \Delta t} x_t + w_{t+1} + u_{t+1}$$
$$y_{t+1} = x_{t+1} + v_{t+1}$$

(3)

where $\Delta t$ is the time interval between two consecutive measurements, $u_t = \omega \Delta t$, $w_t \sim N(0, q)$ 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 of the sum of first order Gauss-Markov process with correlation time $\beta^{-1}$ and variance $\sigma^2_{GM}$, a drift with slope $\omega$ and a Gaussian white noise with variance $\sigma^2_{WN}$. Even with such a relatively simple model, the estimation task is non-trivial. Moreover, the task becomes even more challenging when the observed process size is large and the model is a mixture of three first-order Gauss-Markov random processes with different parameters. Such a model is found to be a suitable one for modelling the stochastic error in accelerometer and gyroscopes composing inertial navigation systems. In Section 5 we analyse a realization of size $833'685$ of such a process, and only our GMWM estimator can be used to estimate the model’s parameters.

This paper is organized as follows. In Section 2, we introduce the WV and 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 performance of our new estimator in the model defined by (3) to the MLE’s one as well as in a model used with financial data. In Section 5 we apply the new methodology to estimate parameters for the stochastic
error models in inertial sensors with a real dataset.

2 The Wavelet Variance

Basically, as pointed out by Percival and Guttorp (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 Guttorp (1994). The wavelet coefficients are built using wavelet filters \{h_{j,t}\}; j = 1, \ldots, J which for \(j = 1\) and for the MODWT satisfy

\[
\sum_{l=0}^{L_1-1} \tilde{h}_{1,l} = 0 \quad \sum_{l=0}^{L_1-1} \tilde{h}_{1,l}^2 \frac{1}{2} \quad \text{and} \quad \sum_{l=-\infty}^{\infty} \tilde{h}_{1,l} \tilde{h}_{1,l+2m} = 0
\]

where \(\tilde{h}_{1,l} = 0\) for \(l < 0\) and \(l \geq L_1\), \(L_1\) is the length of \(\tilde{h}_{1,t}\), and \(m\) is a nonzero integer. Let also \(\tilde{H}_1(f) = \sum_{l=0}^{L_1-1} \tilde{h}_{1,l} e^{-i2\pi fl}\) be the transfer function of \(\tilde{h}_{1,t}\). To obtain the \(j\)th level wavelet filters \(\{\tilde{h}_{j,t}\}\) of length \(L_j = (2^j - 1)(L_1 - 1) + 1\) one computes the inverse discrete Fourier Transform of

\[
\tilde{H}_j(f) = \tilde{H}_1(2^{j-1}f) \prod_{l=0}^{j-2} e^{i2\pi 2^l (L_1-1) L_1} \tilde{H}_1(\frac{1}{2} - 2^l f)
\]

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

\[
W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} y_{t-l}, \quad t \in \mathbb{Z}
\]
We define the WV at dyadic scales $\tau_j = 2^{j-1}$, as the variances of $\{\mathbf{W}_{j,t}\}$, i.e.

$$\nu^2(\tau_j) = \text{var}(\mathbf{W}_{j,t})$$

Notice that the WV are assumed not to depend on time. The condition for this property to hold is that the integration order $d$ for the series $\{Y_t\}$ to be stationary is such that $d \leq L_1/2$ and $\{\tilde{h}_{j,t}\}$ is based on a Daubechies (Daubechies 1992) wavelet filter; see Percival and Walden (2000), chapter 8. This is due to the fact that Daubechies wavelet filters of width $L_1$ contain an embedded backward difference filter of order $L_1/2$. In such a case, the series of wavelet coefficients $\{\mathbf{W}_{j,t}\}$ is stationary with PSD $S_{W_j}(f) = |\tilde{H}_j(f)|^2 S_{F_\theta}(f)$, $\cdot |$ denoting the modulus. This means that the variance of wavelet coefficients’ series is equal to the integral of its PSD (Serroukh et al. 2000)

$$\nu^2(\tau_j) = \int_{-1/2}^{1/2} S_{W_j}(f) = \int_{-1/2}^{1/2} |\tilde{H}_j(f)|^2 S_{F_\theta}(f) df$$

Hence, there is therefore an implicit link between the WV and the parameters of the data generating model $F_\theta$. We exploit this connection when defining an estimator for $\theta$. For simple models, the link is explicit as shown in Section 3 for the random walk and in El-Sheimy, Hou, and Niu (2008) for other models. More general results on the uniqueness of the implicit link between the WV and the model parameters are given in Van Vliet and Handel (1982).

A consistent estimator for $\nu^2(\tau_j)$, for a finite (observed) process $\{y_t; t = 1, \ldots, T\}$, is given by the maximal-overlap estimator defined in Serroukh et al. (2000) (see also Percival 1995)

$$\hat{\nu}^2(\tau_j) = \frac{1}{M(T_j)} \sum_{t \in T_j} W_{j,t}^2$$

with $W_{j,t} = \sum_{l=0}^{L_j-1} \tilde{h}_{j,l} y_{t-l}, t \in T_j$ and where $T_j$ is the set of time indices for which the MODWT coefficients are free of end effects, and $M(T_j) = T - L_j + 1$ their
number. Moreover, Serroukh et al. (2000) show that under suitable conditions
\[ \sqrt{M(T_j)} (\nu^2(\tau_j) - \nu^2(\tau_j)) \] is asymptotically normal with mean 0 and variance \( S_{W_j}(0) = 2A_j \) with

\[ A_j = \int_{-1/2}^{1/2} S_{W_j}^2(f)df = \int_{-1/2}^{1/2} |\mathcal{F}_j(f)|^4 S_{\Phi}^2(f)df \] (6)

which can be estimated by means of

\[ \hat{A}_j = \frac{1}{2} \sum_{\tau=-M(T_j)}^{M(T_j)} \left[ \frac{1}{M(T_j)} \sum_{t \in T_j} W_{j,t} W_{j,t+|\tau|} \right] ^2 \] (7)

A particular choice for the wavelet filter is given by the Haar wavelet filter which first DWT filter \((j = 1)\) is

\[ \{h_{1,0} = 1/\sqrt{2}, \ h_{1,1} = -1/\sqrt{2} \} \] (8)

with length \( L_1 = 2 \). If the process is stationary with backward differences of order \( d > 1 \) one can use other wavelet filters such as Daubechies wavelet filters (Daubechies 1992). When WV is evaluated with Haar wavelet filters, it is actually equal to half the Allan variance (AV) (Allan 1966). In 1998, the IEEE standard IEEE Std 1293-1998 (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 is often used in engineering disciplines and physical sciences as a graphical approach for model building purposes. As for the AV, the WV is also used in 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.

The underlying stochastic processes \( F_\theta \) has a distinct signature in WV. The WV provides an octave-band estimate of the PSD \( S_{F_\theta} \) as

\[
\nu^2(\tau_j) \approx 2 \int_{1/2^{j+1}}^{1/2^j} S_{F_\theta}(f)df
\]

since the wavelet coefficients at scale \( \tau_j \) can be associated with frequencies in the passband \([1/2^{j+1}, 1/2^j]\). This comes from the fact that \( \tilde{h}_{j,l} \) can be seen as an approximate bandpass filter over this frequency range. This can be used to identify the presence of power-law processes in the signal. Indeed, power-law processes have a PSD which is proportional to \(|f|^{-\alpha}\), \( f \) being the frequency, hence leading, together with (9), to the following approximation

\[
\nu^2(\tau_j) \propto \tau_j^{-\alpha-1}
\]

This means that the slope of the line in a log-log plot of the WV versus scale allows deducing the exponent \( \alpha \). For example, a white noise has an exponent \( \alpha = 0 \) and is associated to linear region of the plot with a slope of \(-1/2\). To illustrate this, consider a series \( \{y_t; t = 1, \ldots, T\} \), generated from model (3) with \( \Delta t = 1, \beta = 0.05, \sigma_{GM}^2 = 16, \omega = 0.005, \sigma_{WN}^2 = 4, T = 6000 \), which realization is presented in the top panel of Figure 1. 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 of the individual processes within the model which WV are represented separately. Indeed, when the process is made up of the sum of
independent processes, i.e. $Y_t = \sum_k X_t^{(k)}$, we have that \(4\) can expanded to

$$\nu^2(\tau_j) = \int_{-1/2}^{1/2} |\tilde{H}_j(f)|^2 \left( \sum_k S_{X^{(k)}}(f) \right) df$$

$$= \sum_k \nu^2_k(\tau_j)$$

with $S_{X^{(k)}}$ the PSD of $X_t^{(k)}$ and $\nu^2_k(\tau_j)$ its WV at scale $\tau_j$. With such models, and in this case, the eye of an expert might be able to identify the different underlying processes. However, such “visual” separation of processes may not be always possible, not only when the process is not made up of the sum of independent processes, but also when it is.

Figure 1: Simulated data process \((3)\) with $\beta = 0.05, \sigma^2_{GM} = 16, \omega = 0.005, \sigma^2_{WN} = 4, T = 6000$ (top panel) and graphical representation (log-log scale) of the Haar WV for the complete process (line “o”), the White Noise (“◊” line), first order Gauss-Markov (“□” line) and Drift (“△” line).
When a graphical representation is suitable, with Haar WV, the parameters of stochastic processes which PSD follows a power-law are sometimes estimated by means of linear regression on pre-identified linear regions in a log-log plot of scale $\tau_j$ versus WV $\nu^2(\tau_j)$. For example, this approach has been used for over 30 years as a standard routine measure of frequency stability in lasers (Fukuda et al. 2003) or atomic clocks (Allan 1987). More recently, the WV has also been used with optical sensors (Kebabian et al. 2005), various types of gas monitoring spectrometers (Bowling et al. 2003, Werle et al. 1993, Skrínški et al. 2009), sonic anemometer-thermometers (Loescher et al. 2005), inertial sensors (Guerrier 2009, El-Sheimy et al. 2008), radio-astronomical instrumentation (Schieder and Kramer 2001). The WV was also used in Percival and Guttorp (1994) to analyse vertical ocean shear measurements. In Fadel et al. (2004) it was employed to study the variability in heart-beat intervals. In Witcher (2004), discrete wavelet packet transforms are used to estimate one of the parameters of a seasonal long memory process for the analysis of atmospheric and economic time series. In Gebber et al. (2006), WV is exploited to study nerves activities. WV has also been applied in Earth orientation metrology in Gambis (2002) and with other types of (geo)physical data. However, the linear regression on identified linear regions of the WV plots provides reasonably estimated parameters only for a limited number of processes and is often biased (Stebler et al. 2011). In our simulated example presented in Figure 1, the graphical estimation of first-order Gauss-Markov process parameters mixed with other processes like Gaussian white noise is difficult. In the following Section, we propose instead a criterion based on a standardized distance between sample an model based WV that provides unbiased estimators of the model’s parameters.
3 GMWM estimator

We propose to estimate 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. Indeed, the WV provides a discrete octave-band approximation of the PSD (Percival and Walden 2000) which is linked to the model $F_\theta$ (see (4)). Moreover, Van Vliet and Handel (1982) have set the conditions on the PSD of the process so that the WV uniquely defines the PSD. The link is unique if $S_{F_\theta}(f) \approx C|f|^{-\alpha}$ with $-1 < \alpha < 3$. Greenhall (1998) studies the case $\alpha = 3$.

We propose to find $\hat{\theta}$ such that the WV implied by the model, say $\phi^*(\theta)$, matches the empirical WV, say $\hat{\phi}$, and solves the following GLS optimization problem

$$\hat{\theta} = \arg\min_{\theta \in \Theta} \left( \hat{\phi} - \phi^*(\theta) \right)^T \Omega \left( \hat{\phi} - \phi^*(\theta) \right)$$

in which $\Omega$, a positive definite matrix, is chosen in a suitable manner (see below). (10) defines GMWM estimator. More precisely, $\phi(\cdot) = [\phi_j(\cdot)]_{j=1,...,J}$ is a binding function between $\theta$ and $\nu^2 = [\nu^2(\tau_j)]_{j=1,...,J}$ such that $\phi(\theta) = \nu^2$, and $\hat{\phi} = \hat{\nu}$ and $\phi^*(\theta) = \phi(\hat{\theta})$ are two estimators. However, except in a few instances, the binding function is not known analytically and $\hat{\theta}$ cannot be constructed as $\phi^{-1}(\hat{\nu}^2)$. Such an exception happens for example in the simple random walk, i.e.

$$x_{t+1} = x_t + w_{t+1},$$

$$y_{t+1} = x_{t+1} \text{ with } w_t \sim N(0,\sigma_{RW}^2).$$

In this example $\theta = \sigma_{RW}^2$ and the PSD of $\{y_t\}$ is given by

$$S_{F_\theta}(f) = \left( \frac{\sigma_{RW}}{2\pi f} \right)^2$$

Then, using the Haar wavelet and (4), we have

$$\phi_j(\sigma_{RW}^2) = \nu^2(\tau_j) = \frac{\sigma_{RW}^2 \tau_j}{6}$$
Hence, in a log-log graph of $\nu(\tau_j)$ versus $\tau_j$, we have $\log(\nu(\tau_j)) = 0.5\log(\tau_j) + \log(\sigma_{RW}) - 0.5\log(6)$ and the random walk appears as a line with slope $1/2$. Moreover, at $\tau_j = 6$, we have that $\log(\nu(6)) = \log(\sigma_{RW})$. Other examples of analytical relationship between the WV and the model parameters can be found in e.g. El-Sheimy, Hou, and Niu (2008).

Since in most cases the binding function is not known or difficult to derive analytically, we actually propose to use simulations to compute $\hat{\phi}^*(\theta)$ and hence place our estimator in the framework of indirect inference (Smith 1993, Gourieroux et al. 1993, Gallant and Tauchen 1996). Basically, given a sample of observations $\{y_t, t = 1, \ldots, T\}$ and an hypothetical model $F_\theta$ (which can be derived a priori from a WV graph), define $\hat{\phi}_j$ as the WV $\hat{\nu}^2(\tau_j)$ estimated from the sample using equation (5), and $\hat{\phi}^*_j(\theta)$ as the WV estimate $\hat{\nu}_j^2(\tau_j)$ computed on a simulated series $\{y^*_t(\theta), t = 1, \ldots, R \cdot T\}, R \geq 1$ from $F_\theta$. Alternatively, one can compute $R$ WV estimates $\hat{\nu}_j^2(\tau_j)$ on simulated series $\{y^*_t(\theta), t = 1, \ldots, T\}$ and obtain $\hat{\nu}_j^2(\tau_j) = \frac{1}{R} \sum_{r=1}^{R} \hat{\nu}_j^2(\tau_j)$. Then $\hat{\phi} = [\hat{\phi}_j]_{j=1,\ldots,J}$ and $\hat{\phi}^*(\theta) = [\hat{\phi}^*_j(\theta)]_{j=1,\ldots,J}$ are used in equation (10) to obtain an estimate $\hat{\theta}$ of $\theta$. The consistency and asymptotic normality of the indirect estimator $\hat{\theta}$ is inherited from the consistency and asymptotic normality of the auxiliary estimator $\hat{\phi}$. In the following theorem, we state the conditions on the wavelet $W_{j,t}$ coefficients for $\hat{\phi}$ to be asymptotically multivariate normal using the results of Giraitis and Taqqu (1997) on limit theorems for bivariate Appell polynomials. The proof is given in the Appendix. This result generalizes the results of Serroukh, Walden, and Percival (2000) to the multivariate case.

**Theorem 1** Let the wavelet coefficients be linear stationary sequences $W_{j,t} = \sum_{s \in \mathbb{Z}} a_{j}(t-s)\xi_{s}$ in iid $\xi_{s}$ with mean zero, unit variance, $E[||\xi_{s}^4||] < \infty$, and with $\sum a_{j}^2(t) < \infty, j =$
Let the PSD of $W_{j,t}$ be such that there exist a $C$ and

$$S_{W_j}(f) \leq C|f|^{-\alpha_j}, f \in [-\pi, \pi]$$  \hspace{1cm} (11)

If

$$\alpha_j < 1/2, \forall j = 1, \ldots, J$$  \hspace{1cm} (12)

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

$$\sigma^2_{kl} = 2\pi S_{kl}(0)$$

and where $S_{kl}(f) = \frac{1}{2\pi} \sum_{\tau=\infty}^{\infty} \gamma_{kl}(\tau) e^{-i f \tau}$ are the cross spectral densities with cross-covariances $\gamma_{kl}(\tau) = \text{cov}(W_{k,t}, W_{l,t+t})$.

Notes. (A) The number of WV $J$ is fixed. (B) Conditions (11) and (12) ensure weak dependence for the process $W_{j,t}$. (C) If the process $\{y_t\}$ is non-stationary but with stationary backward differences of order $d$, the processes $W_{j,t}$ are stationary if $L_1 \geq 2d$. (D) Because of the relationship (4) in which $|H_j(f)|^2$ is model independent, condition (11) is satisfied when it is valid for $\{y_t\}$. (E) Adding $\alpha > -1$ to condition (11) insures that $\phi$ is bijective according to the results of Van Vliet and Handel (1982). (F) The conditions do not imply Gaussianity of the process. (G) Condition (11) is satisfied in particular when $S_{W_j}(f) = C|f|^{-\alpha_j}G_j(1/|f|)$ where $G_j(\cdot)$ are slowly varying functions at infinity (see Giraitis and Taqqu 1997). (H) For assumptions in the time domain, see Giraitis and Taqqu (1998).

The estimation of $\sigma^2_{kl}$ is in general not straightforward. In the Appendix, we show that under the assumption of a Gaussian process for $\{y_t\}$, a suitable estimator is
given by

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

(13)

where \( T_{kl} \) is the largest set of time indices containing both the indices in \( T_k \) and \( T_l \) (see (5)), and \( M(T_{kl}) \) their number. Alternatively, when the process is not Gaussian or when the sample size is very large as it is the case with the dataset analysed in Section 4 so that the computation of (13) is infeasible, one can use a parametric bootstrap to estimate \( \text{cov}(\hat{\nu}^2(\tau_k), \hat{\nu}^2(\tau_l)) \). \( Q \) samples of size \( T \) are simulated from \( F_\theta \) on which \( Q \) WV \( \hat{\nu}^2_q(\tau_k) \) and \( \hat{\nu}^2_q(\tau_l) \) are computed and \( \sigma_{kl}^2 \) is estimated by their empirical covariance.

Therefore, under the conditions of Theorem 1 (and note (E)), we have that \( \hat{\theta} \) is consistent and asymptotically normal, with asymptotic covariance matrix \( V_\hat{\theta} \) given by (see e.g. Genton and Ronchetti 2003)

\[
V_\hat{\theta} = BV_\hat{\phi}B^T + \frac{1}{R}BV_\hat{\phi^*}B^T
\]

(14)

where

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

(15)

and where the \( J \times p \) matrix \( D \) is the Jacobian matrix of the transformation \( \phi(\cdot) \), i.e. \( D = \partial \phi(\theta)/\partial \theta^T \), and where \( V_\phi \) and \( V_\phi^* \) are the asymptotic covariance matrices of respectively \( \phi \) and \( \phi^* \). When no simulations are needed to compute \( \hat{\theta} \), that is when an analytical expression exists for the WV as an explicit function of the elements of \( \theta \) and \( \Omega = I \), then \( V_\theta = (D^T D)^{-1} D^T V_\phi D (D^T D)^{-1} \). In the random walk example, choosing \( \tau_j = 6 \), we get \( D = 1 \) and \( V_{\hat{\sigma}^2_{RW}} = V_{\hat{\nu}^2(6)} \). When simulations are needed,
assuming $R$ to be sufficiently large, we have that $V_{\hat{\theta}} \approx BV_{\phi}B^T$ and the most efficient estimator is obtained by choosing $\Omega = V_{\phi}^{-1}$, leading then to $V_{\hat{\theta}} = \left(D^TV_{\phi}^{-1}D\right)^{-1}$. In practice, the matrix $D$ can be numerically approximated and computed at $\hat{\theta}$.

Finally, practical issues such as the starting point for the estimation procedure, the number of WV to be considered, the choice of the wavelet filter as well as other considerations are discussed in Section 4.

### 4 Simulation study

This section is dedicated to the performance evaluation of the GMWM estimator compared to the MLE or other estimators. For the GMWM estimator, we consider for $\Omega$ in (10) the diagonal matrix with diagonal elements given by the inverse of the sample variance estimates of the MODWT using (7). Indeed, in our simulations, we found that choosing $\Omega = V_{\phi}^{-1}$ may lead in some cases to numerical instability, probably due to the estimation of the covariances using (13). In all cases, we chose $R = 100$ for the data generation in the GMWM. 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). We consider simulated processes $\{y_t, t = 1, \ldots, T\}$ from model (3) as well as from an AR(1)-LARCH(1) model given by (see also Francq and Zakoïan 2010)

\begin{align*}
x_{t+1} &= (1 - \beta x_t)u_{t+1} \\
y_{t+1} &= \psi y_t + x_{t+1}
\end{align*}

where the $u_t$ are i.i.d $(0, \sigma_u^2)$. The linear ARCH (LARCH) model in (16) was first suggested (for infinite orders) by Robinson (1991) and further analysed by Giraitis et al. (2000) and Giraitis et al. (2004). Model (3) is typically used for stochastic
sensor error modeling while model (16)-(17) finds applications with financial data.

From model (3), we actually generate three types of samples which correspond to three different (sub)models. In Model 1, we set $\sigma_{WN}^2 = \omega = 0$, hence have an first order Gauss-Markov with White noise, in Model 2, we set $\omega = 0$, hence remove the drift only from the complete model (Model 3) (3). We generate 100 processes of size $T = 6000$ with $\Delta t = 1$ and $\theta = (\sigma_{WN}^2, \sigma_{GM}^2, \beta, \omega) = (4, 16, 0.05, 0.005)$ at the complete model. For the submodels, the parameters are constrained accordingly and not estimated.

For both the GMWM and the EM-KF, the initial values for the optimization 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, except that starting far away can make the computational time quite long. The root mean squared errors (RMSE) as well as the relative RMSE (relative to the true parameter value) are presented in Table 1. The WV where computed for $J = 12(< \log(6000)/\log(2) = 12.55)$ 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 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 (model 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 later by decreasing the number of scales $J$.
at which the WV are estimated. Indeed, in this example $J = 12$ scales are used to estimate 2 or 3 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 paper. It does not only depend on the ratio of the number of parameters to estimate and $J$ but also on the type of model that is considered. For example, with model (3) without the drift element (i.e. Model 2), according the the WV graph in Figure 1, if the last 3 to 4 scales are ignored, then the WV are still able to capture information about the other model’s components. Actually, removing the last 4 scales, improves the efficiency of the GMWM estimator in the simulation study for model 2 (results not shown here). On the other hand, with the model used for the analysis of the real example in Section 5 (see Figure 2), the last scales are very important in capturing the different model’s features, while the first ones could be dropped.

Table 1: RMSE and relative RMSE (R-RMSE) of the GMWM and EM-KF estimators for 100 simulated processes of size $T = 6000$ from model (3) (Model 3) with $\Delta t = 1$ and $\theta = (\sigma^2_{WN}, \sigma^2_{GM}, \beta, \omega) = (4, 16, 0.05, 0.005)$ and from submodels with $\sigma^2_{WN} = \omega = 0$ (Model 1) and with $\omega = 0$ (Model 2).

<table>
<thead>
<tr>
<th></th>
<th>GMWM</th>
<th>EM-KF</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>RMSE</td>
<td>R-RMSE</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{GM}$</td>
<td>1.55</td>
<td>0.10</td>
</tr>
<tr>
<td>$\beta$</td>
<td>5.33 · 10^{-3}</td>
<td>0.11</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{GM}$</td>
<td>1.84</td>
<td>0.12</td>
</tr>
<tr>
<td>$\beta$</td>
<td>1.01 · 10^{-2}</td>
<td>0.20</td>
</tr>
<tr>
<td>$\sigma^2_{WN}$</td>
<td>0.12</td>
<td>0.03</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma^2_{GM}$</td>
<td>1.37</td>
<td>0.09</td>
</tr>
<tr>
<td>$\beta$</td>
<td>5.06 · 10^{-3}</td>
<td>0.10</td>
</tr>
<tr>
<td>$\sigma^2_{WN}$</td>
<td>0.11</td>
<td>0.03</td>
</tr>
<tr>
<td>$\omega$</td>
<td>3.63 · 10^{-4}</td>
<td>0.07</td>
</tr>
</tbody>
</table>

For general AR(p)-LARCH(q) models, Francq and Zakoïan (2010) provide the conditions under which the MLE is an inconsistent model’s parameters estimator.
Table 2: Bias and RMSE of the GMWM for 100 simulated processes of size $T = 1000$ from model (16)-(17) with $\theta = (\psi, \beta, \sigma_u) = (0.9, -0.5, 1)$. The bias and RMSE of the MLE and the WLSE of Francq and Zakoïan (2010) for the same model are reported from their Table 1 computed using 500 simulated processes.

<table>
<thead>
<tr>
<th></th>
<th>QMLE</th>
<th></th>
<th></th>
<th>WLSE</th>
<th></th>
<th></th>
<th>GMWM</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
<td>Bias</td>
<td>RMSE</td>
</tr>
<tr>
<td>$\psi = 0.9$</td>
<td>-0.327</td>
<td>9.265</td>
<td>0.002</td>
<td>0.022</td>
<td>-0.008</td>
<td>0.022</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta = -0.5$</td>
<td>2.009</td>
<td>43.999</td>
<td>-0.028</td>
<td>0.058</td>
<td>-0.045</td>
<td>0.373</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\sigma_u = 1$</td>
<td>2.686</td>
<td>2.756</td>
<td>-0.019</td>
<td>0.076</td>
<td>-0.032</td>
<td>0.325</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

In particular, for model (16)-(17) the MLE is inconsistent if $u_t$ is Gaussian, so that Francq and Zakoïan (2010) propose an alternative weighted least-squares estimator (WLSE) and other variations of it. We compare the performance of the GMWM to their estimator (and to the MLE) in this case and in the same simulation setting. 100 series of length $T = 1000$ are simulated with $u_t$ Gaussian and with $\theta = (\psi, \beta, \sigma_u) = (0.9, -0.5, 1)$. Table 2 provides the RMSE of the MLE, the WLSE and the GMWM. While the GMWM is less efficient than the WLSE, it provides consistent estimates of the model’s parameters (the biases are relatively small). The MLE has as expected a bad behaviour in terms of RMSE. For this model, the WLSE is therefore more appropriate, but the GMWM estimator is more generally and rather more straightforwardly applicable to other models for which alternative estimators to the MLE are not available.

5 Case study: Inertial sensors

An inertial navigation system exploits observation from an orthogonal triad composed of gyroscopes and accelerometers measuring angular rates and accelerations, respectively. Once initialized, the navigation is a dead-reckoning process (i.e. the solution at one epoch is computed from the solution of the previous epoch) in which gyroscope signals are integrated once to yield orientation, and accelerometers are in-
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), 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 error increase rapidly with time. To bound the error growth, inertial sensors are combined with other sensors (e.g. GPS, odometer, altimeter) through optimal (e.g. Kalman) filtering. The gyroscope and accelerometer error behaviour is modeled through state augmentation in the system. This requires that the parameters of the stochastic processes characterizing sensor errors in the state-space model is carefully setup prior to filtering. The GMWM estimator can typically be used for such purposes in a sensor calibration procedure where the reference signals are known or available.

As an example, we applied the GMWM estimator to the angular rate signal issued from a low-cost gyroscope where the spectral structure of errors is often quite complex. The signal was recorded in static conditions at a frequency of 100 Hz and provided 833'685 measurements. After having removed the mean, the time-varying part of the sensor errors is directly available and is presented in the time domain on the top of Figure 2, together with the Haar WV of this process with 95% confidence intervals. The WV computed on the original signal give an indication of the underlying stochastic processes that are summed up to build it up. A possible model is a mixture of three first-order Gauss-Markov random processes. In such a case, the parameters to be estimated are \( \theta = \{ \sigma_1^2, \beta_1, \sigma_2^2, \beta_2, \sigma_3^2, \beta_3 \} \) and the resulting error model can be written in a synthetic discrete form as

\[
y_{t+1} = \sum_{g=1}^{3} \varepsilon_t^{(g)}
\]
Figure 2: Gyroscope observed error process (top panel) and graphical comparison (log-log scale) between the Haar WV (line “o”) computed from the observed signal and the synthetic signal using the estimated parameters (line “∇”). The synthetic signal is the mean of the WV computed on 100 simulated processes using the estimated parameters. The WV of the estimated three first-order Gauss-Markov processes are represented by respectively the “□”, “◊” and “△” lines.

where

\[ x_{t+1}^{(g)} = e^{-\beta_g \Delta t} x_t^{(g)} + w_t^{(g)}, \quad w_t^{(g)} \sim \mathcal{N}(0, q_g) \]

and \( w_t^{(g)} \) follows process-noise variance

\[ q_g = \sigma_{GM,g}^2 (1 - e^{-2\beta_g \Delta t}) \]

and \( \Delta t = 0.01 \).

The GMWM estimates of the parameter set \( \theta \) and its corresponding 95% confi-
Table 3: Estimated parameters with associated 95% confidence intervals for the mixture of three first-order Gauss-Markov random processes with the Gyroscope signal data.

<table>
<thead>
<tr>
<th></th>
<th>Estimates</th>
<th>IC(·, 0.95)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\sigma_1$</td>
<td>1.302 · 10^{-2}</td>
<td>(1.302 · 10^{-2}; 1.302 · 10^{-2})</td>
</tr>
<tr>
<td>$\beta_1$</td>
<td>2.502 · 10^2</td>
<td>(2.502 · 10^2; 2.502 · 10^2)</td>
</tr>
<tr>
<td>$\sigma_2$</td>
<td>2.608 · 10^{-4}</td>
<td>(2.606 · 10^{-4}; 2.610 · 10^{-4})</td>
</tr>
<tr>
<td>$\beta_2$</td>
<td>8.980 · 10^{-3}</td>
<td>(8.965 · 10^{-3}; 8.995 · 10^{-3})</td>
</tr>
<tr>
<td>$\sigma_3$</td>
<td>9.487 · 10^{-3}</td>
<td>(9.442 · 10^{-3}; 9.531 · 10^{-3})</td>
</tr>
<tr>
<td>$\beta_3$</td>
<td>4.011 · 10^{-6}</td>
<td>(3.922 · 10^{-6}; 4.099 · 10^{-6})</td>
</tr>
</tbody>
</table>

...
of this study. Theoretically, the GMWM estimator together with the estimating criterion (10) could serve as a basis for a formal model fit statistic, but this idea is left for further research.
References


Holmes, E. (2010, December). Derivation of the em algorithm for constrained and
unconstrained multivariate autoregressive state-space (marss) models. Technical report, Northwest Fisheries Science Center, NOAA Fisheries 2725 Montlake Blvd E., Seattle, WA 98112.


Strus, J., M. Kirkpatrick, and J. Sinko (2008). Gps/imu development of a high
accuracy pointing system for maneuvering platforms.


Appendix

Proof of Theorem 1

Theorem 1 is in fact a special case of Theorem 6.1 of Giraitis and Taqqu (1997) under the conditions of their Theorem 2.4, which applies to quadratic forms given by bivariate Appell polynomials

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

in which \( X_{j,t} = W_{j,t}, m_j = n_j = 1, b_j(\tau) = \delta(\tau) = 1/M(T_j) \) if \( \tau = 0 \) and 0 otherwise, so that \( Q_T^j = \hat{\nu}_j^2(\tau_j), \forall j = 1, \ldots, J. \) The conditions of Theorem 2.4 of Giraitis and Taqqu (1997) are on the PSD of \( X_{j,t} \) (i.e. \( W_{j,t} \)), namely (11), on the PSD of \( b_j(\tau) \), namely \( S_{b_j(\tau)}(f) \leq C|f|^{-\beta_j}, \) and on a function of \( \alpha_j \) (in (11)) and \( \beta_j \), namely \( 2\alpha_j + 2\beta_j < 1 \) (see (2.13) in Giraitis and Taqqu 1997, with \( m = n = 1 \)). Since the form of \( b_j(\tau) \) implies that \( \beta_j = 0, \forall j, \) the conditions are satisfied and (12) in Theorem 1 holds.
Covariance estimation of WV estimators

We first note that $M(T_j)/T \to 1$ as $T \to \infty$ so that we use $T$ instead of $M(T_j)$ below. We have that

$$\sigma^2_{kl} = \text{cov}[\hat{\nu}^2(\tau_k), \hat{\nu}^2(\tau_l)]$$

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

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

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

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

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

$$\sigma^2_{kl} = 2 \int_{-1/2}^{1/2} S_{W_k,W_l}(f)^2 df = 2 \int_{-1/2}^{1/2} S_{W_k,W_l}(f)^2 df$$  \hspace{1cm} (18)$$

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 (18), we follow Whitcher, Guttorp, and Percival (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 that to obtain symmetric covariances, it is important to estimate them this way.