Chapter 4: Generalized Method of Wavelet Moments

Jan Skaloud* & Stéphane Guerrier†

*École Polytechnique Fédérale de Lausanne; †University of Geneva

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

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

EPFL - Winter 2020

Introduction

General Framework:

- It is based on the **Generalized Method of Wavelet Moments (GMWM)** of Guerrier et al. 2013, which is a new statistical approach to estimate the parameters of (complex) time series models.
- This methodology is **robust** (potentially applicable for FDI purposes) and is able to **automatically select a suitable model** (or rank models).
- Original application
  - The GMWM is able to estimate efficiently time series models which are commonly used to describe the errors of inertial sensors.
  - This calibration approach provides considerable improvements (in terms of navigation performance) compared to existing methods.
- The GMWM approach is completely general and therefore suitable for applications in other domains as oscillators, environmental modeling, economy, to name a few.

IMU Calibration - Motivation

Possible causes: due to sensors itself (analog to digital conversion, electronic imprecision possibly w.r.t to environmental conditions) and the sensor assembly (i.e., relative alignment between sensors)

Errors types

- Deterministic errors (usual mitigation $\rightarrow$ lab calibration)
  - Sensor mean biases, mean scale factors (w.r.t. temperature, g, ...)
  - Non-orthogonality between sensors
- Random components (usual mitigation $\rightarrow$ sensor fusion, filtering)
  - Time independent (switch on random bias)
  - Time uncorrelated (white noise)
  - Time correlated (stationary or non-stationary)

Correct stochastic sensor error modeling implies:

- Correct stochastic assumptions for inference (confidence of derived quantities)
- **Better trajectory estimation (in real-time or post-processing)**

Effect on position of error model - Motivation

Emulation setting:

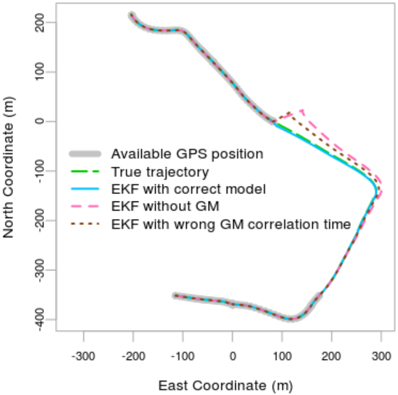
- Suppose the following model for inertial sensors (WN + GM):

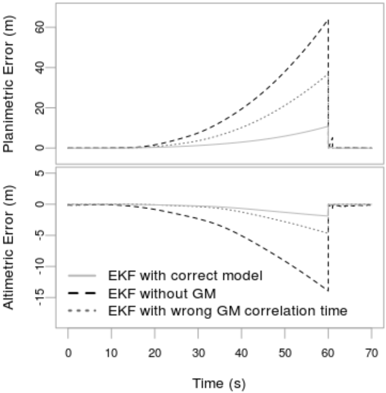
$$Y_t = \exp(-\beta\Delta t) Y_{t-1} + \varepsilon_t, \;\; \varepsilon_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_{GM}(1 - \exp(-2\beta\Delta t)))$$
$$X_t \overset{iid}{\sim} \mathcal{N}(0, \sigma^2), \;\; Z_t = X_t + Y_t.$$

- Consider the following three models:

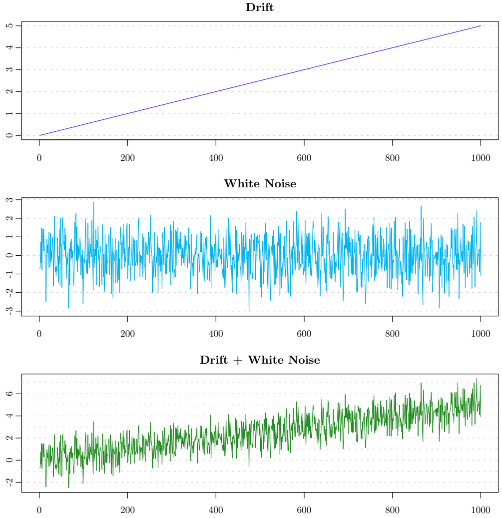
Sensor	**Scenario**	β	σ_{GM}	σ_{WN}
Acc.	Correct model	10^{-4}	50.0	70.0
	Wrong β	10^{-2}	50.0	70.0
	Without GM	–	–	70.0
Gyro.	Correct model	10^{-4}	10.0	30.0
	Wrong β	10^{-2}	10.0	30.0
	Without GM	–	–	30.0

Effect on position of error model - Motivation





An easy latent time series model - Motivation



Remarks:

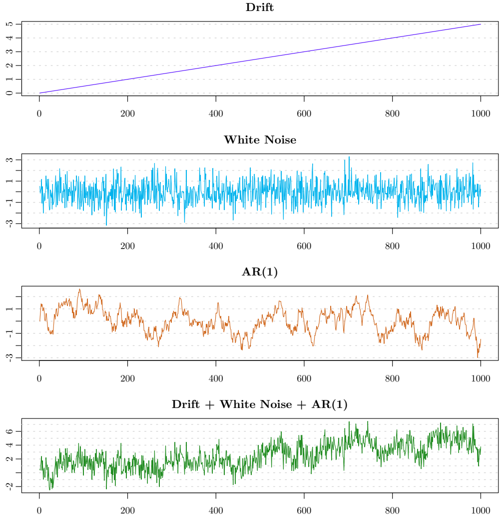
- Simple linear regression model:

$$y_t = \omega t + \varepsilon_t$$

$$\varepsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$$

- MLE is perfectly fine.
- **What if we add an AR1 process?**

Adding an autoregressive process - Motivation



Remarks:

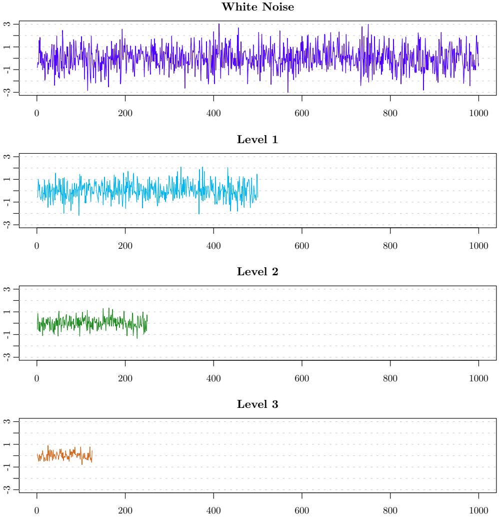
- Not a linear regression model but a **state space model**.
- Computing the likelihood is not an easy task (Kalman filter).
- **MLE (in fact EM-KF) fails**.

Estimation of Latent Time Series Models - Motivation

Existing methods:

- Transforming into a “non-latent” model (e.g. ARMA)
  - Does not work in general.
  - Tends to diverge when one latent time series is “close” to unit root.
  - Difficult to “inverse”.
- MLE of an associated state-space (possibly using EM algorithm)
  - Computationally intensive.
  - Systematically diverges with “complex” models.
  - A lot of work is needed for a new model (see Stebler et al. 2011).
- “Graphical” method
  - Limited to a few possible models.
  - Not consistent in general (see Guerrier, Molinari, and Stebler 2016).
  - “Inefficient” (see Guerrier, Molinari, and Stebler 2016).

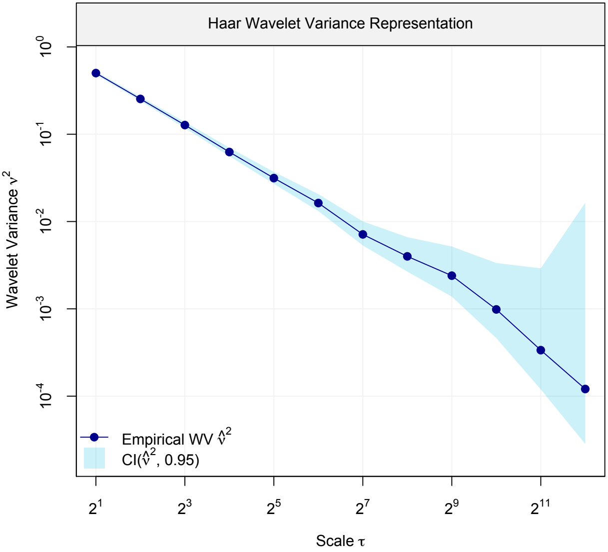
Looking differently at a time series using the Allan variance



Remarks:

- Proposed by Allan 1966.
- Originally intended to study the stability of atomic clocks.
- Allan Variance is closely related to the (Haar) Wavelet Variance.

The Wavelet (Allan) variance



Initial idea:

Match the WV:

- Exploit the relationship that exists between the model $F_{\boldsymbol{\theta}}$ and the WV $\boldsymbol{\nu}(\boldsymbol{\theta})$ (i.e. **mapping** $\boldsymbol{\theta} \mapsto \boldsymbol{\nu}(\boldsymbol{\theta})$).
- "Inverse" this mapping by minimizing some discrepancy between the empirical WV ($\hat{\boldsymbol{\nu}}$) and the theoretical WV $\boldsymbol{\nu}(\boldsymbol{\theta})$.
- This should provide an **approximation** of the point $\boldsymbol{\theta}(\hat{\boldsymbol{\nu}})$.

Introduction: Wavelet Variance

Empirical WV:

- The WV (ν_j^2) is the **variance of wavelet coefficients** for the scale j.
- Wavelet coefficients ($W_{j,t}$) are weighted averages computed on the series Y_t.
- The weights are called wavelet filters h_j: e.g. the Haar wavelet filter.
- The wavelet filters give non-zero weights to observations at a given lag (window sizes of length L_j). Hence, there are as many WV as there are scales.
- The wavelet filters can be computed on consecutive windows, or on overlapping windows (to get $\widetilde{W}_{j,t}$ using $\tilde{h}_j$). Overlapping windows lead to more efficient estimators (such as the MODWT).

Definition: Wavelet Coefficients

Definition 4.1 (Wavelet Coefficients).

In a similar way to the AV, we can define the Wavelet Variance (WV) at dyadic scales (τ_j) for $j \in \{x \in \mathbb{N} : 1 \leq x < \log_2(T) - 1\}$. To do so, we first need to define the wavelet filters $h_{j,l}$ as "weights" having the following properties

$$\sum_{l=0}^{L_j-1} h_{j,l} = 0, \; \sum_{l=0}^{L_1-1} h_{1,l}^2 = \frac{1}{2} \; \text{ and } \; \sum_{l=-\infty}^{\infty} h_{1,l} h_{1,l+2m} = 0,$$

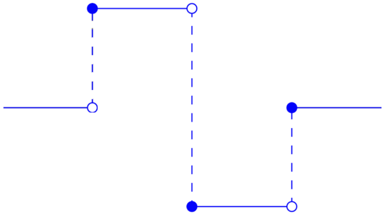
where $m \in \mathbb{N}^+$, $L_j = (2^j - 1)(L_1 - 1) + 1$ is the length of the filter at level j and L_1 is the length of the first level filter $h_{1,l}$.

Then, the wavelet coefficients $W_{j,t}$ are defined as

$$W_{j,t} = \sum_{l=0}^{L_j-1} h_{j,l} X_{t-l}.$$

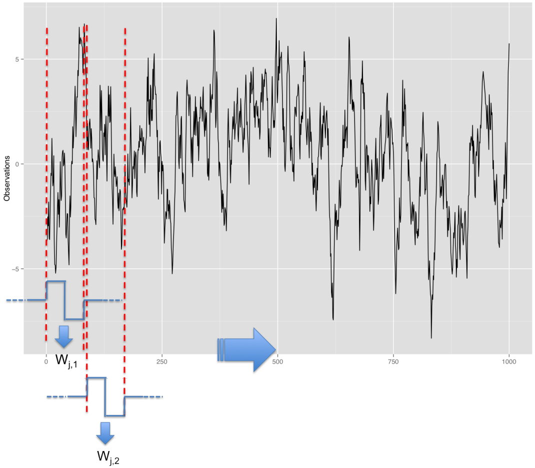
Wavelet Variance: Filters

The wavelet coefficients for scales j and $h_{j,l}$ are so called **wavelet filters** issued from a mother wavelet function (for example the Haar wavelet) and L_j is the length of the filter.



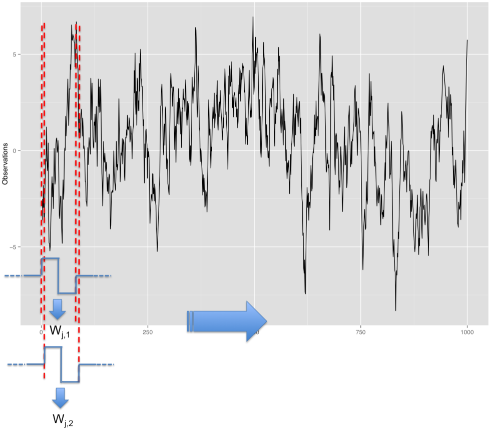
Example: Wavelet Coefficients

The wavelet filters can be applied in different ways such as the **Discrete Wavelet Transform (DWT)**:



Example: Wavelet Coefficients

The most useful (in the present context) is (arguably) the **Maximum Overlap Discrete Wavelet Transform (MODWT)**:



Definition: Wavelet Variance

Definition 4.2 (Wavelet Variance).

Once we have defined the wavelet coefficients $W_{j,t}$, we can now define the WV as being the variance of the wavelet coefficients at level j:

$$\nu_j^2 \equiv \text{Var}[W_{j,t}].$$

Lemma 4.3 (Relation between WV and AV).

The WV has an exact relationship to the AV when using the ***Haar wavelet filter*** *$h_{j,l}$, i.e.*

$$\nu_j^2 = 2\,\text{AVar}_j\,.$$

The proof of Lemma 4.3 is given in Percival and Guttorp 1994.

Theoretical WV

WV implied by $F_{\boldsymbol{\theta}}$:

Lemma 4.3 implies that the (Haar) WV is equivalent (up to a scaling factor) to the AV. Therefore, the theoretical WV can be computed using Lemmas 3.3 and 3.5. One possible definition of the theoretical WV (see also Zhang 2008 and Lemma 3.5) is the following:

$$\nu_j = f\left(\boldsymbol{\theta}\right) = \int_{-1/2}^{1/2} |\widetilde{H}_j(f)|^2 S_{F_{\boldsymbol{\theta}}}(f) df.$$

Example:

Consider an AR(1) and let $\tau_j = 2^j$, then the theoretical (Haar) WV of such process is given by

$$\nu_j = \frac{\left(\frac{\tau_j}{2} - 3\phi - \frac{\tau_j \phi^2}{2} + 4\phi^{\frac{\tau_j}{2}+1} - \phi^{\tau_j+1}\right)\sigma^2}{\frac{\tau_j^2}{2}\left(1-\phi\right)^2\left(1-\phi^2\right)}.$$

WV of latent time series models

A very useful property:

A direct consequence of Corollary 3.4 and Lemma 4.3 is that, similarly to the AV, the WV of latent time series models is simply the sum of individual latent WV. For example, suppose we have

$$Y_t = X_t^{(1)} + \ldots + X_t^{(k)},$$

then the PSD of Y_t is

$$S_{Y_t} = S_{X_t^{(1)}} + \ldots + S_{X_t^{(k)}},$$

so the WV of Y_t is given by

$$\nu_{Y_t,j} = \int_{-1/2}^{1/2} |\widetilde{H}_j(f)|^2 \left(\sum_{i=1}^{k} S_{X_t^{(i)}} \right) df = \sum_{i=1}^{k} \nu_{X_t^{(i)},j}.$$

Wavelet Variance Estimator

Definition 4.4 (Wavelet Variance Estimation).

An unbiased estimator for the WV issued from a certain wavelet transform is given by the **MODWT estimator** of Percival 1995 (and further in Serroukh, Walden, and Percival 2000):

$$\hat{\nu}_j^2 = \frac{1}{M_j(T)} \sum_{t=L_j}^{T} W_{j,t}^2,$$

where $M_j(T) = T - L_j + 1$.

Remark 1 (Data contamination).

Mondal and Percival 2012 underlines how the above estimator can become **severely biased under data contamination**. For this purpose they propose a robust estimator of WV.

M-Estimation of WV

In Guerrier et al. 2020 we introduced an alternative robust estimator of the WV. This estimator is advantageous from theoretical standpoint compared to the one proposed in Mondal and Percival 2012. In addition, this alternative estimator is more statistically efficient (see e.g. the simulation study presented in Guerrier et al. 2020). This estimator is based on Huber's Proposal 2 and thus, this M-estimator of WV ($\tilde{\nu}_j^2$) is the solution for ν_j^2 of

$$\frac{1}{M_j}\sum_{t=1}\underbrace{\omega^2\left(r_{j,t};\nu_j^2,c\right)\,r_{j,t}^2 - a_\omega(\nu_j^2,c)}_{\psi(W_{j,t},\nu_j^2,c)} = 0,$$

where

- $r_{j,t} \equiv \frac{W_{j,t}}{\nu_j}$
- $\omega(\cdot)$ are the weights given through a bounded function
- c is a tuning constant for robustness/efficiency trade-off
- $a_\omega(\cdot) \equiv \mathbb{E}\left[\omega^2\left(r_{j,t};\nu_j^2,c\right)\,r_{j,t}^2\right]$ is a correction term for Fisher consistency

Properties of the Wavelet Variance Estimator

Lemma 4.5 (Consistency).

Let (X_t) be such that:

- $(X_t - X_{t-1})$ *is a (strongly) stationary process,*
- $(X_t - X_{t-1})^2$ *has absolutely summable covariance structure,*
- $\mathbb{E}\left[(X_t - X_{t-1})^4\right] < \infty$ *for some* $\delta > 0$.

Defining $\hat{\boldsymbol{\nu}} \equiv [\hat{\nu}_j^2]_{j=1,\dots,J}$, *with* J *being a bounded quantity, then we have*

$$\hat{\boldsymbol{\nu}} \xmapsto{\mathcal{P}} \boldsymbol{\nu}\,.$$

The proof of Lemma 4.5 can be found in Appendix A ▸ Go to Appendix A.
The same result holds for the robust estimator $\tilde{\nu}_j^2$, see Proposition 3.1. of Guerrier et al. 2020.

Properties of the Wavelet Variance Estimator

Compare to consistency, asymptotic normality requires stronger conditions given in the following lemma. J is also bounded.

Lemma 4.6 (Asymptotic normality).

Let (X_t) be such that:

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

Then we have,

$$\sqrt{T}\left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\right) \overset{\mathcal{D}}{\longmapsto} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma}$ *is the asymptotic covariance matrix of* $\hat{\boldsymbol{\nu}}$ *with elements* $\sigma_{ij}^2 \equiv \sum_{h=-\infty}^{\infty} \operatorname{cov}\left(W_{i,0}W_{j,0}, W_{i,h}W_{j,h}\right)$.

The proof of Lemma Appendix B can be found in Appendix B ▸ Go to Appendix B. The same result holds for the robust estimator $\tilde{\nu}_j^2$, see Lemma 3.1. of Guerrier et al. 2020.

Confidence Interval of the Wavelet Variance Estimator

Based on the asymptotic normality results (Lemma 4.6), we can construct the $(1-\alpha)$-confidence intervals for $\hat{\nu}_j$ as

$$\mathrm{CI}\left(\nu_j, \alpha\right) = \left[\hat{\nu}_j \pm z_{1-\frac{\alpha}{2}} \frac{\sigma_{jj}}{T}\right],$$

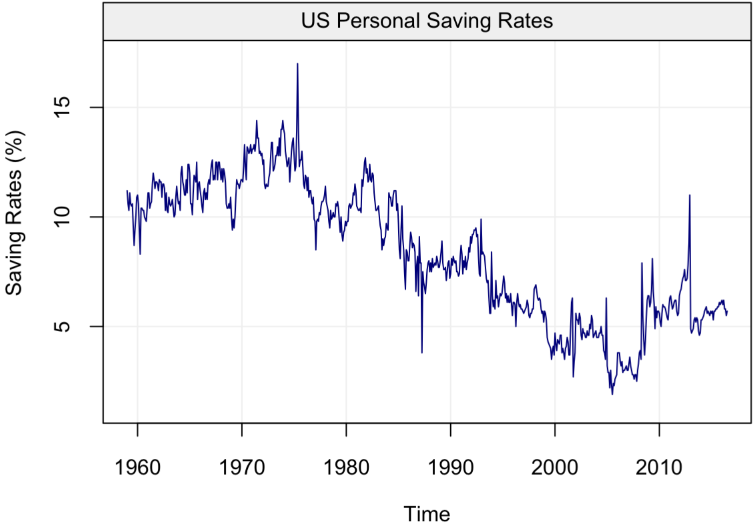
where $z_{1-\frac{\alpha}{2}} \equiv \mathbf{\Phi}^{-1}\left(1-\frac{\alpha}{2}\right)$ is the $(1-\frac{\alpha}{2})$ quantile of a standard normal distribution. However, the Long-Run Variance σ_{jj} is usually unknown. Many methods has been proposed to consistently estimate under mild conditions (see e.g. Newey and West 1986). However, for finite T, CI based on the asymptotic normality would be problematic, because the lower limit of CI can be negative. Alternatively, to avoid this problem, we can use following asymptotic result

$$\eta \frac{\hat{\nu}_j}{\nu_j} \xmapsto{\mathcal{D}} \chi^2_\eta,$$

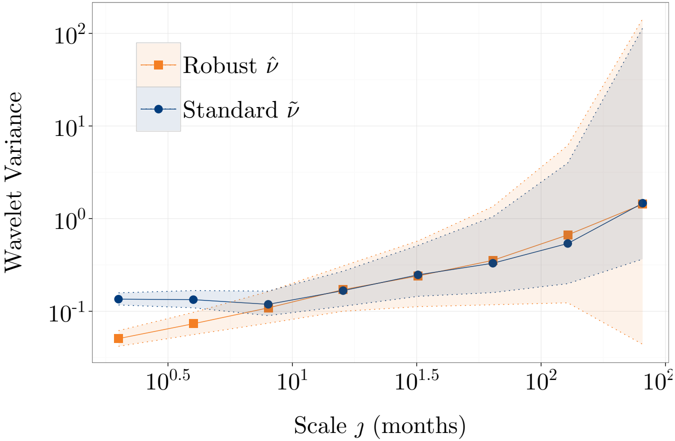
where η is a constant which can be estimated by $\max\left\{\frac{M_j}{2^j}, 1\right\}$. Moreover, this method can also avoid the estimation of Long-Run Variance σ_{jj}. Then the confidence interval is

$$\mathrm{CI}\left(\nu_j, \alpha\right) = \left[\frac{\eta\hat{\nu}_j}{F^{-1}_{\chi^2_\eta}(\alpha/2)}, \frac{\eta\hat{\nu}_j}{F^{-1}_{\chi^2_\eta}(1-\alpha/2)}\right].$$

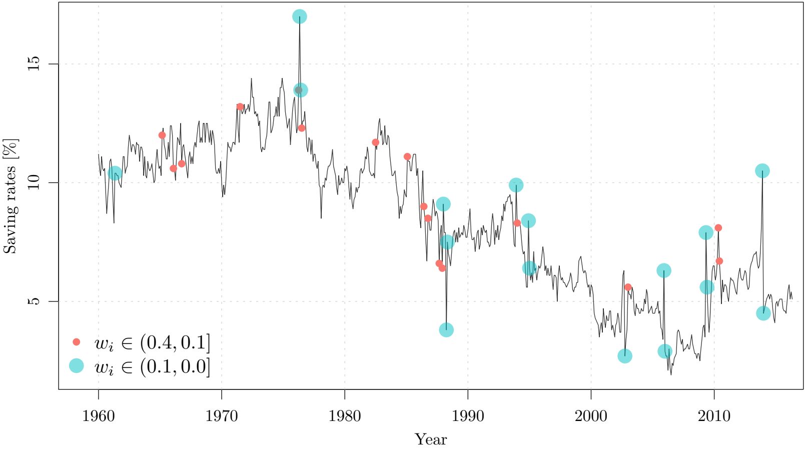
Application to US Saving Rates



Application to US Saving Rates



Application to US Saving Rates



Inverse Mapping

How to find the parameters implied by the observed WV

In a similar way to the method based on the AV, the idea would be to find the model parameter values implied by the observed WV, i.e.

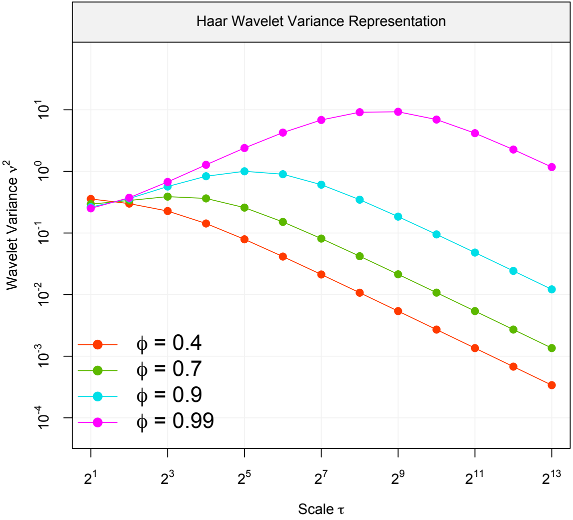
$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}(\hat{\boldsymbol{\nu}}),$$

where $\boldsymbol{\theta}(\boldsymbol{\nu})$ is the inverse of the function $\boldsymbol{\nu}(\boldsymbol{\theta})$ (the theoretical WV).

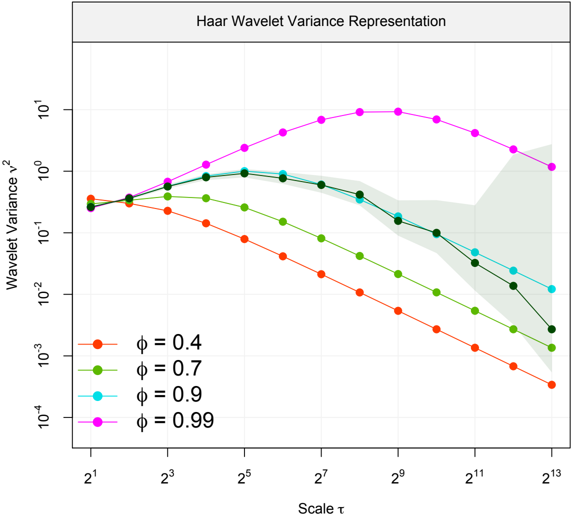
Problems with inverse mapping

The inverse $\boldsymbol{\theta}(\boldsymbol{\nu})$ can often be difficult to find when considering the parametric model that characterizes the theoretical WV $\boldsymbol{\nu}(\boldsymbol{\theta})$ which are often complicated when considering composed stochastic processes as those characterizing errors coming from inertial sensors.

Principle of the GMWM



Principle of the GMWM



The Generalized Method of Wavelet Moments

Definition 4.7 (Generalized Method of Wavelet Moments).

The GMWM estimator is obtained as follows

$$\hat{\boldsymbol{\theta}} = \underset{\boldsymbol{\theta} \in \boldsymbol{\Theta}}{\operatorname{argmin}} \, (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}))^T \boldsymbol{\Omega} (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})),$$

where

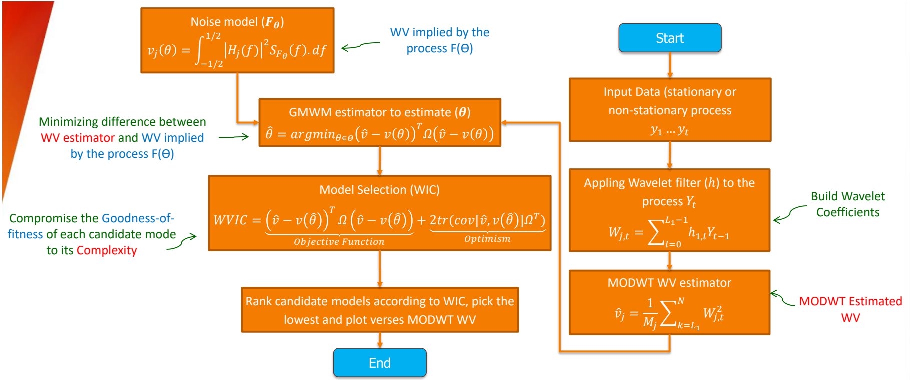
- $\boldsymbol{\Omega}$ is a positive-definite weighting matrix
- $\boldsymbol{\nu}(\boldsymbol{\theta})$ is the theoretical WV implied by a model with parameter vector $\boldsymbol{\theta}$.

GMWM as an Extremum Estimator

The GMWM is part of the **class of extremum estimators** (see Definition 2.21), i.e.

$$\hat{Q}_n(\boldsymbol{\theta}) \equiv -(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}))^T \boldsymbol{\Omega} (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})).$$

Flowchart GMWM



Thanks to Dr. Ahmed Radi!

A small example...

A simulated example:

Let $(Y_t)\ : t = 1, ..., 10^5$ be a simulated signal composed of a:

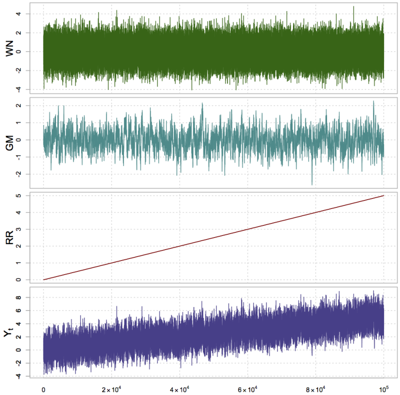
- First-order Gauss-Markov:
  $Y_t = \exp(-\beta\Delta t)Y_{t-1} + \varepsilon_t,\ \ \varepsilon_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma_G^2\left(1 - \exp(-2\beta\Delta t)\right)\right)$
- Gaussian White Noise: $X_t \overset{iid}{\sim} \mathcal{N}\left(0, \sigma^2\right)$
- Drift (rate ramp): $R_t = \omega t$

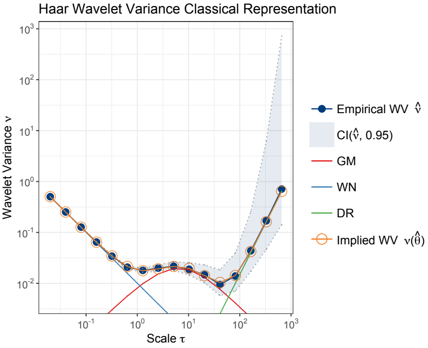
The observed process is therefore $Z_t = Y_t + X_t + R_t$ (which we write GM + WN + DR) and we have that $Z_t \sim F_{\boldsymbol{\theta}}$ where $\boldsymbol{\theta} = (\sigma, \sigma_G, \beta, \omega)$.

A small example...

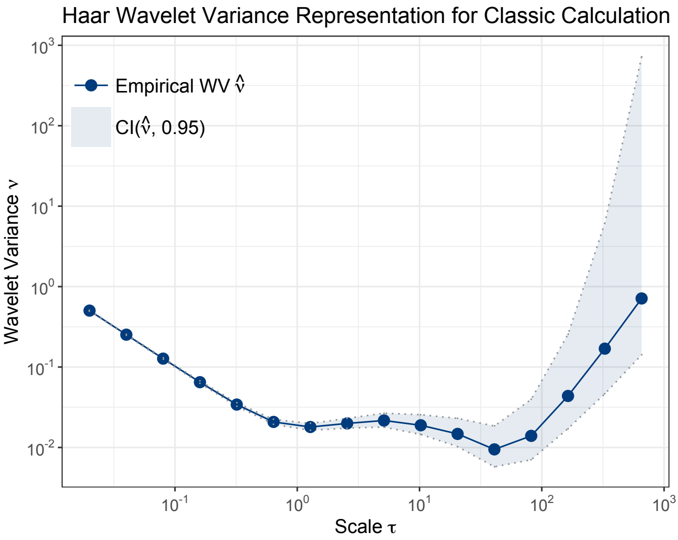
Simulation parameters:

$$\boldsymbol{\theta} = (\sigma^2, \beta, \sigma_G^2, \omega) = (1,\ 0.6,\ 0.1,\ 5 \cdot 10^{-5})$$

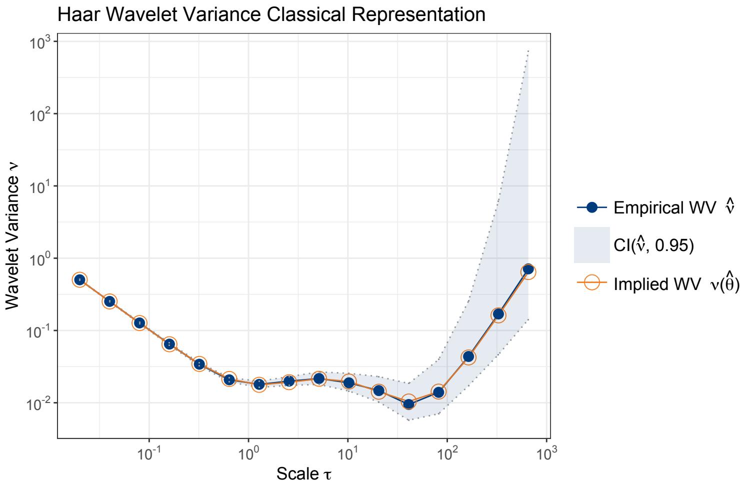




A small example...



GMWM estimation results



GMWM estimation results

Estimated parameters:

	θ_0	$\hat{\theta}$	$\mathrm{IC}\left(\theta_0, 0.95\right)$
σ^2	1.00	1.00	$(0.99; 1.01)$
β	0.60	0.58	$(0.55; 0.61)$
σ_G^2	10^{-1}	$1.07 \cdot 10^{-1}$	$\left(0.99 \cdot 10^{-1}; 1.12 \cdot 10^{-1}\right)$
ω	$5 \cdot 10^{-5}$	$4.87 \cdot 10^{-5}$	$\left(4.67 \cdot 10^{-5}; 5.07 \cdot 10^{-5}\right)$

Goodness of fit test:

$$\min_{\boldsymbol{\theta} \in \boldsymbol{\Theta}} \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)^T \boldsymbol{\Omega} \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right) \xrightarrow[T \to \infty]{\mathcal{D}} \chi^2_{m-p}$$

In this case, the p-value of the test is ≈ 1. Therefore we cannot reject that $F_{\boldsymbol{\theta}}$ is the right model.

Simulation

Setting:

The model "White Noise + Gauss Markov + Drift" is in fact commonly used for inertial sensors. We simulate similarly to Stebler et al. 2011 $B = 100$ signals of length $T = 6000$ from this model with $\boldsymbol{\theta}_0 = (\sigma, \sigma_G, \beta, \omega) = (2, 4, 0.05, 0.005)$.

Estimator errors - comparison GMWM vs. MLE (EM-KF)

	GMWM		MLE (EM-KF)	
	RMSE	R-RMSE	RMSE	R-RMSE
σ_G^2	0.96	0.06	74.57	4.66
β	$4.63 \cdot 10^{-3}$	0.09	0.04	0.85
σ^2	0.11	0.03	0.16	0.04
ω	$2.79 \cdot 10^{-4}$	0.06	0.12	23.58

Practical Exercise: Coding a Simple GMWM Estimator

Exercise:

1. Simulate an AR(1) with the parameters and sample size of your choice.
2. Write a function that computes $\hat{Q}_n(\theta)$ as defined in the previous slide with $\mathbf{\Omega} = \mathbf{I}$.
3. Find the GMWM estimator for the data you simulate in Step 1. Are the estimated parameters close to the correct ones?
4. Compare the results with the ones obtained with the `gmwm()` function.

Consistency of the GMWM

Lemma 4.8.

Suppose that:

(C1) *The function* $\boldsymbol{\nu}(\boldsymbol{\theta})$ *is such that* $\boldsymbol{\nu}(\boldsymbol{\theta}) = \boldsymbol{\nu}(\boldsymbol{\theta}_0)$ *if and only if* $\boldsymbol{\theta} = \boldsymbol{\theta}_0$

(C2) *The set* $\boldsymbol{\Theta}$ *is compact*

(C3) *The function* $\boldsymbol{\nu}(\boldsymbol{\theta})$ *is continuous in* $\boldsymbol{\theta}$

(C4) $\hat{\boldsymbol{\nu}} \overset{\mathcal{P}}{\to} \boldsymbol{\nu}$

(C5) $\boldsymbol{\Omega}$ *is positive-definite and "consistent" (if estimated).*

Then, we have:

$$\hat{\boldsymbol{\theta}} \xmapsto{\mathcal{P}} \boldsymbol{\theta}_0.$$

Proof (Sketch): Consistency of the GMWM

Remark 2.

Identifiability Condition (C1) is equivalent to Condition 2.24 and, as seen earlier, it can be verified by checking the assumptions in Theorem 2.26. In the latter all assumptions are generally satisfied considering the most commonly employed time series models while showing that $J(\boldsymbol{\theta})$ is nonnegative (or nonpositive) is not necessarily straightforward, especially when dealing with latent models.

Study of $J(\boldsymbol{\theta})$

The nonnegative or nonpositive property of $J(\boldsymbol{\theta})$ has been verified for a wide class of time series models (see Guerrier et al. 2013 and Guerrier et al. 2020)

Proof (Sketch): Consistency of the GMWM

While Condition (C2) is always assumed and, as discussed for Condition (C1), Condition (C3) is generally verified for all times series considered previously and is therefore nearly always verified.

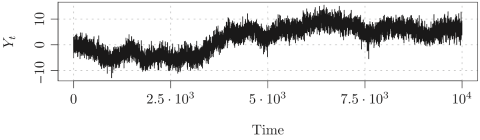
Condition (C4) is verified based on Lemma 4.5 and finally Condition (C5) is either assumed or can be verified using an appropriate estimator for $\boldsymbol{\Omega}$. Based on the latter two conditions, it is possible to verify that $\hat{Q}_n(\boldsymbol{\theta})$ converges uniformly in probability to $Q_0(\boldsymbol{\theta})$ (last condition of Theorem 2.24).

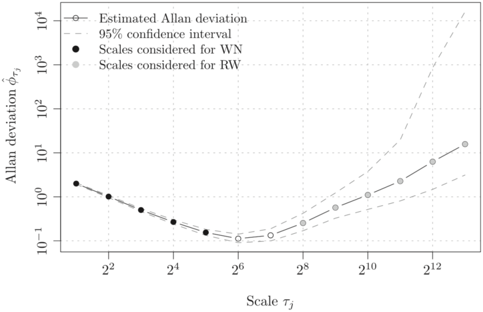
Based on these verifiable (verified) conditions, we have that

$$\hat{\boldsymbol{\theta}} \xmapsto{\mathcal{P}} \boldsymbol{\theta}_0,$$

thus concluding the proof. ■

Simulated Example: WN + RW





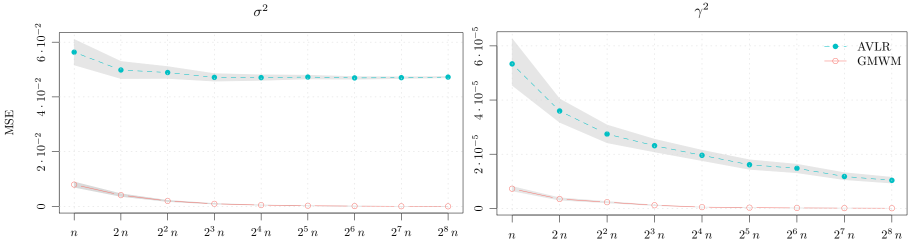
Simulation exercise

Consider the model "WN + RW" with parameters: $\sigma_0^2 = 4$ and $\gamma^2 = 0.01$. Simulation based on $B = 10^3$ Monte-carlo replications. (example coming from Guerrier, Molinari, and Stebler 2016)

Simulated Example: WN + RW

Consistency of AVLR and GMWM

If an estimator is consistent, then its Mean Squared Error (MSE) should tend to zero as the sample size (n) increases.



Practical Exercise: Simulation study

Exercise:

1. Show that the theoretical WV for the model WN() + RW() can be expressed as $\boldsymbol{\nu}(\boldsymbol{\theta}) = \mathbf{X}\boldsymbol{\theta}$.
2. In this case, show that the GMWM has a closed form solution.
3. Simulate a time series with the parameter $\sigma^2 = 4$, $\gamma^2 = 0.01$ and $T = 10^4$. Compare the AVLR, GMWM (based on the `gmwm()` function) and the GMWM (closed form). Which estimator is the best one?
4. Perform a simulation study to compare the three estimators. Are your results consistent with the ones of Guerrier, Molinari, and Stebler 2016 (presented in the previous slide)?

Conditions for Asymptotic Normality

Theorem 4.9.

In addition to Conditions (C1) to (C5), we assume:

(C6) $\boldsymbol{\theta}_0$ *is an interior point to* $\boldsymbol{\Theta}$

(C7) $\mathbf{H}(\boldsymbol{\theta}_0) \equiv \frac{\partial}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \boldsymbol{\nu}(\boldsymbol{\theta})\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$ *exists and is non-singular*

(C8) $\sqrt{T}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}) \xrightarrow[T \to \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$.

Then, we have

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow[T \to \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{V}),$$

where $\mathbf{V}$ *is the asymptotic covariance matrix of* $\hat{\boldsymbol{\theta}}$.

Proof (Sketch): Asymptotic Normality for the GMWM

The proof of asymptotic normality of the GMWM closely follows that of asymptotic normality for extremum estimators (see from Theorem 2.31 forwards).

Condition (C6) is another condition that inevitably needs to be assumed (as for Condition (C2)) while Condition (C7) is not necessary but guarantees the feasibility of Taylor expansion according to the approach used (we avoid technical details).

Knowing that Condition (C8) is verified (see Lemma 4.6), then we have that

$$\sqrt{T}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xmapsto[T\to\infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{V})\,.$$

■

Applications: the Environmental System Model

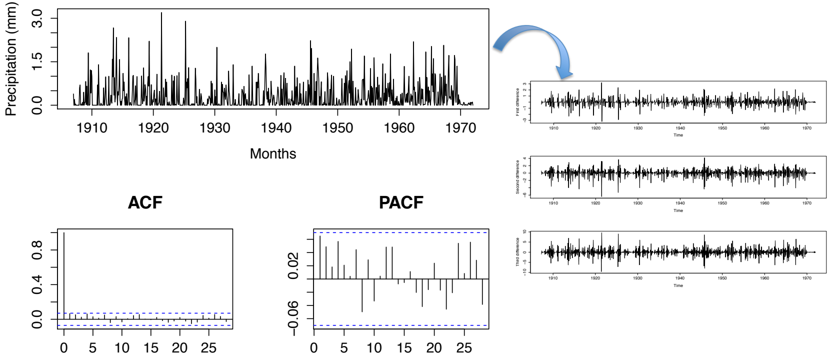


Figure: Monthly precipitation data from 1907 to 1972 taken form *Hipel and McLeod, Elsevier, 1994*

Applications: the Environmental System Model

For the precipitation phase, models which are usually considered are the white noise or AR(1). The ACF and PACF would suggest the first but let's test the AR(1)...

	ϕ	σ^2
MLE	$6.463 \cdot 10^{-2}$	$2.222 \cdot 10^{-1}$
CI	$[-5.702 \cdot 10^{-3}, 1.255 \cdot 10^{-1}]$	$[2.014 \cdot 10^{-1}, 2.413 \cdot 10^{-1}]$
GMWM	$5.384 \cdot 10^{-2}$	$2.205 \cdot 10^{-1}$
CI	$[-1.758 \cdot 10^{-2}, 1.255 \cdot 10^{-1}]$	$[1.984 \cdot 10^{-1}, 2.439 \cdot 10^{-1}]$
RGMWM	$3.892 \cdot 10^{-1}$	$1.016 \cdot 10^{-1}$
CI	$[3.008 \cdot 10^{-1}, 4.813 \cdot 10^{-1}]$	$[8.943 \cdot 10^{-2}, 1.133 \cdot 10^{-1}]$

Practical Exercise: Modelling US saving rates

Exercise:

1. Plot the US saving rates data-set (get the data with `data(savingrt, package = "simts")`)
2. Find a suitable model of the type "random walk + noise" using the `gmwm()`.
3. Compare the robust and non-robust estimators.

Model Selection using WVIC

In order to select a model among a set of candidate models for an observed time series, it is possible to estimate their out-of-sample prediction error and select the model that minimizes this error.

Definition 4.10 (Wavelet Variance Information Criterion (WVIC)).

The WVIC provides the out-of-sample prediction error between the WV implied by the selected model and the WV observed out-of-sample:

$$\text{WVIC} = \mathbb{E}\left[\mathbb{E}_0\left[\left(\hat{\boldsymbol{\nu}}^0 - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right)^T \boldsymbol{\Omega}\left(\hat{\boldsymbol{\nu}}^0 - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right)\right]\right].$$

Estimation of WVIC

An estimator of the WVIC is given by

$$\widehat{\text{WVIC}} = \left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right)^T \boldsymbol{\Omega}\left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right) + 2\,\text{tr}\left\{\widehat{\text{cov}}\left[\hat{\boldsymbol{\nu}}, \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right]\boldsymbol{\Omega}^T\right\}.$$

Model Selection using WVIC

Remark 3 (How to compute $\widehat{\text{WVIC}}$).

While the first term in $\widehat{\text{WVIC}}$ is given by the value of the GMWM objective function at the solution $\hat{\boldsymbol{\theta}}$, it is not straightforward to compute the second term (optimism):

$$\text{tr}\left\{\widehat{\text{cov}}\left[\hat{\boldsymbol{\nu}}, \boldsymbol{\nu}(\hat{\boldsymbol{\theta}})\right]\boldsymbol{\Omega}^T\right\}.$$

Computation of optimism term

There are various approaches to compute the optimism term:

- Compute the analytic form of the covariance term (see Guerrier et al., 2018).
- Compute the term via parametric bootstrap (see Guerrier, Molinari, and Skaloud 2015).
- Compute it using independent replicates (see Radi et al. 2017).

Model Selection using WV

Properties of the WVIC

Let us assume we have a set of candidate models $\mathcal{M}$ where $\hat{m}$ denotes the model selected using $\widehat{\text{WVIC}}$ and m^* denotes the best model within $\mathcal{M}$. Then

- The WVIC overfits meaning that asymptotically it selects $m^* \subset \hat{m}$ (i.e. it always contains the best model).
- The WVIC is loss efficient:

$$\frac{\hat{Q}_n(\hat{\boldsymbol{\theta}}_{\hat{m}})}{\hat{Q}_n(\hat{\boldsymbol{\theta}}_{m^*})} \xmapsto[T\to\infty]{\mathcal{P}} 1\,,$$

  i.e. the selected model performs as well as the best model asymptotically.

The GMWM computational platform

Outline

1. **Visualizing the signal(s)**:
   - Is the signal disturbed (contaminated)?
   - Which models could best describe the visualized WV?
2. **Estimating the model(s)**
   - The `gmwm.imu()` function for parameter estimation
   - The options for estimation
3. **Inference**
   - Confidence intervals for the parameters
   - Goodness-of-fit of the model to the signal
4. **Model selection**
   - The efficient computation of the WVIC
   - Model selection from a set of user-specified models
   - Automatic model selection from a set of all sub-models of a single user-specified model

The GMWM package in action

Getting to know the package

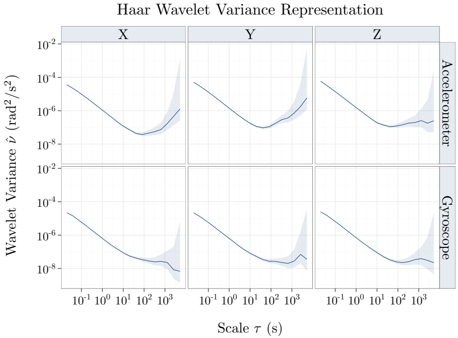
The package comes with multiple datasets

- To use the data within the `R` session:
  1. Load the `imudata` package (separate from the `gmwm` package)
  2. Make the dataset available in the session by typing `data(imu)`
- Load external data (e.g. binary data) with `read.imu()`

Visualizing the data

The function `wvar.imu()`

Once the inertial sensor error data has been entered, it is possible to plot the observed WV vs τ on a log-log scale.



R code:

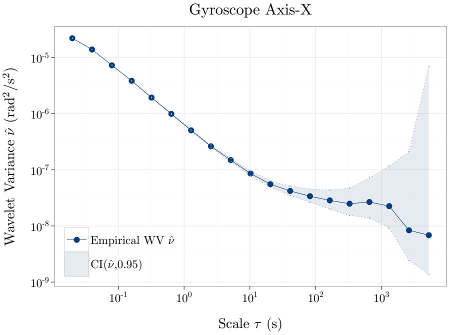
```
imu.obj = imu(imu6,
gyroscope = 1:3,
accelerometer = 4:6)
wv = wvar.imu(imu.obj)
plot(wv)
```

(Run time: 3.70 [sec])

Visualizing the data

The function `wvar()`

It is of course possible to plot the WV individually for an axis of accelerometer or gyroscope.



R code:

```
WV.gx = wvar(imu6[,1])
plot(WV.gx)
```

(Run time: 0.65 [sec])

GMWM estimation

The `gmwm()` function

The tailored function for IMU error modeling is `gmwm()`.

Main arguments of the `gmwm()` function

- `model`: The structure of the model is specified through this argument
- `data`: The signal to be modelled
- `compute.v`: The method for computing the weighting matrix $\boldsymbol{\Omega}$
  - `fast`: Estimated diagonal matrix
  - `diag`: Use a diagonal matrix with the asymptotic variance of $\hat{\boldsymbol{\nu}}$
  - `bootstrap`: Estimate $\boldsymbol{\Omega}$ through parametric bootstrap

GMWM estimation

The `model` argument

The GMWM package can estimate each specific model or any latent model made by a combination of all or a subset of the following models:

- `AR1()`: a first-order autoregressive process (reparametrization of Gauss-Markov process)
- `WN()`: white noise process
- `QN()`: quantization noise (rounding error)
- `RW()`: random walk process
- `DR()`: drift
- `AR()`: p-order autoregressive process
- `MA()`: q-order moving average process
- `ARMA()`: autoregressive-moving average processes

GMWM estimation

Latent model syntax

To specify a latent model we use the sign "+" between the available models while the `AR1()` model can be included more than once (say k times) and it can be specified as `k*AR1()`, for example:

- `ARMA()+WN()+RW()`
- `3*AR1()+DR()`

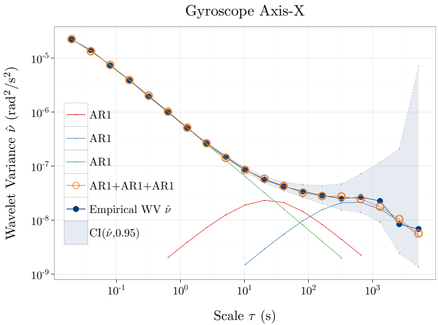
Convergence

If the `gmwm.imu()` function has problems of convergence, one can specify **starting values** for the parameters using the brackets in the syntax of each model (e.g. `WN(sigma2=0.5)`).

GMWM estimation

Visually assessing the fit

The function `plot()` applied to the object of a GMWM estimation allows to see how well the WV implied by the estimated model fits the observed WV.



R code:

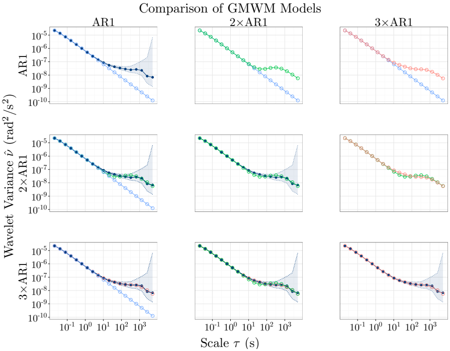
```
mod = gmwm.imu(3*AR1(),
imu6[,1]) plot(mod,
process.decomp = T)
```

(Run time: 0.56 [sec])

Model Comparison

Suppose we want to compare three models...

`AR1()`, `2*AR1()` and `3*AR1()`



R code:

```
Xt = imu[,1]
m1 = gmwm.imu(AR1(),Xt)
m2 = gmwm.imu(2*AR1(),Xt)
m3 = gmwm.imu(3*AR1(),Xt)
compare.models(m1, m2, m3)
```

(Run time: 5.20 [sec])

Inference

Output of estimation

Aside from the value of the estimated parameters, the output of the function `gmwm.imu()` provides other information for inference which include **confidence intervals** and **goodness of fit** test (GoF)

```
mod = gmwm.imu(2*AR1(), imu6[,1])
summary(mod, inference = T)
```

```
Model Information:
          Estimates         CI Low        CI High             SE
AR1    9.998700e-01 9.998336e-01 9.999064e-01 2.211547e-05
SIGMA2 5.223319e-11 4.343878e-11 6.102759e-11 5.346620e-12
AR1    1.265324e-01 1.265324e-01 1.265324e-01 1.765050e-08
SIGMA2 5.031185e-05 5.021566e-05 5.040805e-05 5.848209e-08

* The initial values of the parameters used in the minimization
of the GMWM objective function were  generated by the program
underneath seed: 1337.

Objective Function: 20.9297

Asymptotic Goodness of Fit:
Test Statistic: 903.8 on 15 degrees of freedom
The resulting p-value is: 0
```

WVIC model selection

Two options for model selection

- **Manual**: the function `rank.models()` allows the user to specify the set of models from which to select.
- **Automatic**: the function `auto.imu()` allows to specify one single model from which all sub-models are generated to select from.

R code:

```
rank.models(3*AR1()+WN(),
2*AR1()+QN(), imu6[,1], nested =
F, bootstrap = F, model.type="imu",
robust = F)
```

```
mod.res = auto.imu(imu6, model
= 3*AR1()+WN()+RW()+QN()+DR(),
bootstrap = F, robust = F)
```

WVIC model selection

Automatic IMU model selection

Suppose we want to apply the `auto.imu()` function to the X-axis gyroscope in the `imu` dataset and consider all model combinations within the `4*AR1() + WN() + RW()` model.

```
Xt = imu6[,1]
mod = 4*AR1()+WN()+RW()
mod.sel = auto.imu(Xt, model = mod)
summary(mod.sel)
```

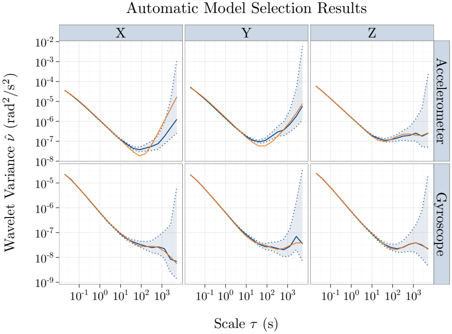
```
The model ranking for data column 1:
                             Obj Fun  Optimism   Criterion  GoF P-Value
1. AR1 AR1 AR1               17.2769    0.2974     17.5743            0
2. AR1 AR1 AR1 RW            17.2679    0.3139     17.5818            0
3. AR1 AR1 AR1 AR1           17.6022    0.4330     18.0351            0
4. AR1 AR1 RW                20.1044    0.4871     20.5915            0
5. AR1 AR1                   20.9864    0.2070     21.1933            0
...
19. RW                    28420.4972    0.0160  28420.5132            0
```

(Run time: 50.25 [sec] ≈ 14x faster than the bootstrap option 685.4 [sec])

WVIC Model Selection

Visualizing the Automatic IMU Model Selection

We can observe the results of the `auto.imu()` function by requesting a plot. The plot will contain the empirical WV in addition to the best implied WV that we find. In this particular case, we used the defaults of the `auto.imu()` (63 models $\times$ 6 columns).

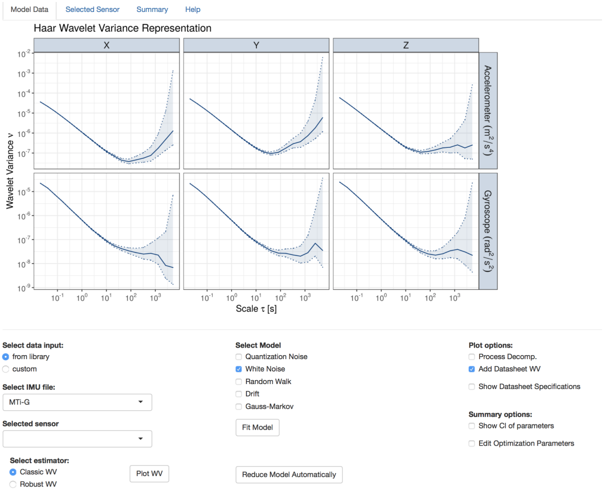


R code:

```
imu.obj = imu(imu6, gyroscope = 1:3,
accelerometer = 4:6, axis = c('X',
'Y', 'Z'))
auto.mod = auto.imu(imu.obj)
plot(auto.mod)
```

(Run time: 377.77 [sec])

Web-based Platform: gui4gmwm



Practical Exercise: Modelling Inertial Sensors

Exercise:

1. Find a suitable for error model to describe the signal of ln200 gyroscope. You can get the data as follows:

```r
# Load ln200 - gyro
data(ln200.gyro, package = "imudata")

# Let's pick gyro X
Xt = ln200.gyro[,1]
```

2. Same question for a KVH1750 accelerometer:

```r
# Load kvh data
data(kvh1750.acc, package = "imudata")

# Let's pick accel X
Xt = kvh1750.acc$`Acc. X`[[1]]$data
```

Can we do better than the GMWM?

The GMWM generally has the following advantages:

- It can easily estimate many state-space and latent time series models.
- It has convenient asymptotic properties.
- It is a feasible and computationally efficient approach compared to alternative methods (especially for large sample sizes).
- It can (easily) be extended to more complex modelling settings (e.g. statistically robust, multivariate, non-stationary, etc.).
- However, it is unclear if the method is (statistically) optimal...

Can we do better than the GMWM?


IEEE SENSORS JOURNAL, VOL. 16, NO. 1, JANUARY 1, 2016 163

Maximum Likelihood Identification of Inertial Sensor Noise Model Parameters

Janosch Nikolic, *Student Member, IEEE*, Paul Furgale, *Member, IEEE*, Amir Melzer, *Member, IEEE*, and Roland Siegwart, *Fellow, IEEE*

***Abstract*—Accurate visual-inertial localization and mapping systems require accurate calibration and good sensor error models. To this end, we present a simple offline method to automatically determine the parameters of inertial sensor noise models. The proposed methodology identifies noise processes across a large range of strength and time-scales, for example, weak gyroscope bias fluctuations buried in broadband noise. This is accomplished with a classical maximum likelihood estimator, based on the integrated process (i.e., the angle, velocity, or position), rather than on the angular rate or acceleration as is standard in the literature. This trivial modification allows us to capture noise processes according to their effect on the integrated process, irrespective of their contribution to rate or acceleration noise. The cause of the noise is not discussed in this article. The method is tested on different classes of sensors by automatically identifying the parameters of a standard inertial sensor noise model. The results are analyzed qualitatively by comparing the model's Allan variance to the Allan variance computed directly from sensor data. A simulation that resembles one of the devices under test facilitates a quantitative analysis of the proposed estimator. Comparison with a competing, state-of-the-art method shows the advantages of the algorithm.**

***Index Terms*—Sensor phenomena and characterization, maximum likelihood estimation, gyroscope and accelerometer noise model.**

I. Introduction

INERTIAL sensors are employed in countless applications, ranging from consumer electronics to autonomous vehicles and unmanned aerial systems. Many different types of sensors are used, from low-cost, multi-axis microelectromechanical systems (MEMS) devices with a footprint below 20mm^2, to ring laser gyroscopes (RLGs). In order to understand the characteristics of these sensors, both stochastic ("noise") and deterministic (sensitivity, axes misalignment, etc.) errors have to be considered.

When inertial measurements are combined with data from other sensors, such as precise landmark observations from a camera system for example, an accurate model for the different sources of error in the inertial sensor data is vital. Such a model facilitates the optimal design of an estimator, and allows us to verify the proper operation of all software and hardware components.

Fig. 1. Inertial sensor model with a deterministic and a random component. Here, the true angular rates ω are corrupted with deterministic errors, for example a scale factor that varies with temperature, as well as non-deterministic errors, such as additive broadband noise. This report presents a method to identify noise processes according to their contribution to the angular increments $\hat{\Omega}$.

A good sensor model often incorporates stochastic as well as deterministic components, as illustrated in Fig. 1. The stochastic part of the model includes errors such as broadband noise, or a slowly (randomly) varying bias, and describes them in a *probabilistic* sense. These errors can not be calibrated for, and limit the performance of the device fundamentally.

The instrumental noise figures are usually identified under static mechanical (i.e. non moving) and thermal conditions [1]. The stochastic model then provides an *upper bound* on the sensor performance, and crucial parameters for an estimator (i.e. noise densities).

This work focuses on how to identify such noise models automatically (for low-cost devices), since deriving a good noise model for an inertial sensor can be a difficult task for a non-expert. However, we also emphasise the importance of a good deterministic sensor model, since a stochastic model alone usually leads to over-confidence in the inertial data, especially when using uncalibrated sensors.

A. Existing Methods

Section II reviews the most commonly used stochastic models for inertial sensors. The Allan variance (AV) [2] and

Manuscript received August 2, 2015; revised September 1, 2015; accepted September 1, 2015. Date of publication September 3, 2015; date of current version December 10, 2015. This work was supported in part by the Armasuisse Science and Technology under Grant 8003501880 and in part by the Commission for Technology and Innovation under Grant 13394.1 PFFLENM. The associate editor coordinating the review of this paper and approving it for publication was Dr. Amitava Chatterjee.

The authors are with the Autonomous Systems Laboratory, Swiss Federal Institute of Technology, Zurich CH-8092, Switzerland (e-mail: janosch.nikolic@mavt.ethz.ch; paul.furgale@mavt.ethz.ch; amir.melzer@mavt.ethz.ch; rsiegwart@ethz.ch).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/JSEN.2015.2476668

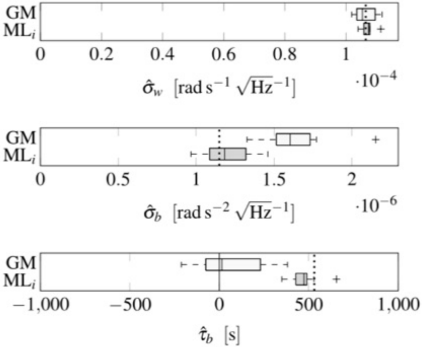


Fig. 11. Parameter estimates from two different methods: the generalized method of wavelet moments (GM) [7], [8], and integrated maximum likelihood (ML_i). The standard model parameters were estimated for 10 synthetic datasets, each 12 h long. Ground truth is indicated with a dotted line.

Can we do better than the GMWM?

5522 IEEE SENSORS JOURNAL, VOL. 16, NO. 14, JULY 15, 2016

Discussion on Maximum Likelihood-Based Methods for Inertial Sensor Calibration

Stéphane Guerrier, Roberto Molinari, and James Balamuta

Abstract—This letter highlights some issues which were overlooked in a recently published paper called maximum likelihood identification of inertial sensor noise model parameters. The latter paper does not consider existing alternative methods, which specifically tackle this issue in a possibly more direct manner and, although remaining a generally valid proposal, does not appear to improve on the earlier proposals. Finally, a simulation study rectifies the poor results of an estimator of reference in the same publication.

Index Terms—State-space, calibration, time series, generalized method of wavelet moments.

I. INTRODUCTION

THE identification of a probabilistic model and the estimation of its relative parameters for the error signal issued from inertial sensors is a key challenge in many fields of engineering that has led to a great deal of research being produced. Recently, [1] (hereinafter *paper*) put forward a new algorithm for the automatic identification of the stochastic error model parameters. Considering other existing methods, the authors claim to improve the estimation of the classical maximum likelihood estimator by considering "the maximum likelihood estimator based on the integrated process rather than on the angular rate or acceleration as is standard in the literature". This proposal is indeed an interesting one but tackles a problem which can often be solved in a more straightforward and stable manner through a direct maximum likelihood approach. In this letter, we first of all discuss some critical points of the *paper* and then rectify some results which compare the algorithm proposed in the *paper* with the Generalized Method of Wavelet Moments (GMWM) introduced by [2]. For these discussions, we only consider the case of uniformly sampled observations but these analyses can be eventually extended to other types of sampling scenarios.

II. DISCUSSION

A first observation which can be made on the *paper* is that the model put forward in it was already successfully estimated through the classical maximization of the likelihood function in [3], also applied within the framework of inertial sensors calibration. In the uniformly sampled setting, the simulation results in the *paper* do not seem to approach the results in [3] nor [4] which, moreover, are not cited nor considered for purposes of comparison. Indeed, the model considered in the *paper* is quite standard and can be defined as follows

$$X_t = \phi X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim \mathcal{N}(0, \sigma^2)$$
$$Y_t \sim \mathcal{N}(0, \gamma^2), \quad Z_t = X_t + Y_t \tag{1}$$

where (X_t) is a first-order Gaussian autoregressive process (simply a reparametrization of a Gauss-Markov process), (Y_t) is a Gaussian white noise process and (Z_t) the observed process. Although limited in scope, this model can be considered as useful in the setting of inertial sensor calibration and for this reason was among those considered in [3] where it was shown that it can be estimated through classical maximum likelihood approaches which, however, diverge in even slightly more complex scenarios.

A second observation which can be made considering these results is that although the algorithm described in the *paper* proposes an innovative and effective approach for the estimation of this simple model, it does not appear to improve over existing methods (e.g. [3]) and can often be solved in a more straightforward manner. Indeed, in the *paper* (X_t) and (Y_t) are assumed to be independent and Z_t is multivariate Gaussian. In this setting, let $\mathbf{Z} = (Z_1, \dots, Z_T)^T$, $\mathbf{Y} = (Y_1, \dots, Y_T)^T$, $\mathbf{X} = (X_1, \dots, X_T)^T$ and let us define $\mathbf{\Sigma}(\boldsymbol{\theta}) = \text{cov}(\mathbf{Z})$ where $\boldsymbol{\theta} = (\phi, \sigma^2, \gamma^2)^T$. We therefore have

$$\mathbf{\Sigma}(\boldsymbol{\theta}) = \gamma^2 \mathbf{I}_T + \frac{\sigma^2}{1-\phi^2}\left[\phi^{|i-j|}\right]_{i,j=1,\dots,T}. \tag{2}$$

Considering this form, the log-likelihood function considered here, up to a constant, can be expressed as

$$l(\boldsymbol{\theta}) = -\frac{1}{2}\log(|\mathbf{\Sigma}(\boldsymbol{\theta})|) - \frac{1}{2}\mathbf{Z}^T \mathbf{\Sigma}(\boldsymbol{\theta})^{-1}\mathbf{Z}. \tag{3}$$

The maximization of (3) is straightforward in this case and does not require the use of a Kalman filter. Of course, with very large sample sizes such as those considered in the *paper*, this approach may not be as suitable as the inversion of $\mathbf{\Sigma}(\boldsymbol{\theta})$ entails a considerable computational burden. In this sense, the method proposed in the *paper* has a computational advantage over the maximization of (3) since it is based on a log sampling scheme. However, this implies a loss of efficiency which has yet to be quantified and thereby does not clarify the advantages of this proposal, for example, over a simple moment-based estimate which would deliver a faster and more reliable approach. An example of such moment-based estimator would be to consider the autocovariances at lags

Manuscript received February 5, 2016; accepted May 4, 2016. Date of publication May 9, 2016; date of current version June 16, 2016. The associate editor coordinating the review of this paper and approving it for publication was Prof. Julian C. C. Chan.

S. Guerrier and J. Balamuta are with the Department of Statistics, University of Illinois at Urbana-Champaign, Champaign, IL 61820 USA (e-mail: stephane@illinois.edu; balamut2@illinois.edu).
R. Molinari is with the Research Center for Statistics, Geneva School of Economics and Management, University of Geneva, Geneva 1211, Switzerland (e-mail: roberto.molinari@unige.ch).

Digital Object Identifier 10.1109/JSEN.2016.2565389

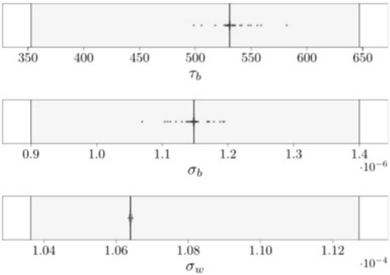


Fig. 1. Results for the simulation study described in Sec. III. Boxplots represent the results for the GMWM while the gray area represents the range of estimated values of the maximum likelihood approach presented in the *paper*.

Can we do better than the GMWM?

In short: Yes!

According to the type of model to be estimated, other estimators (if feasible) can estimate the parameters in a more statistically efficient manner. **However, the estimator will have to be modified (unlike the GMWM) depending on the model.**

Simulated Example:

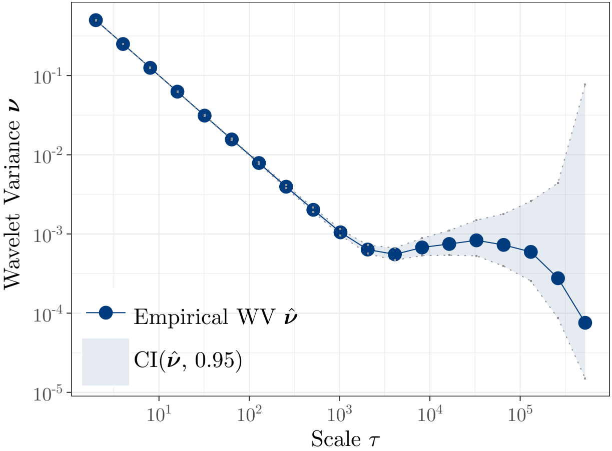
Let us consider a model made by a white noise (with parameter $\sigma^2 = 1$) and a first-order autoregressive model (with parameters $\phi = 0.9999$ and $\upsilon^2 = 10^{-6}$).

We simulate 100 time series of length $T = 10^6$ and estimate the parameters of the model using

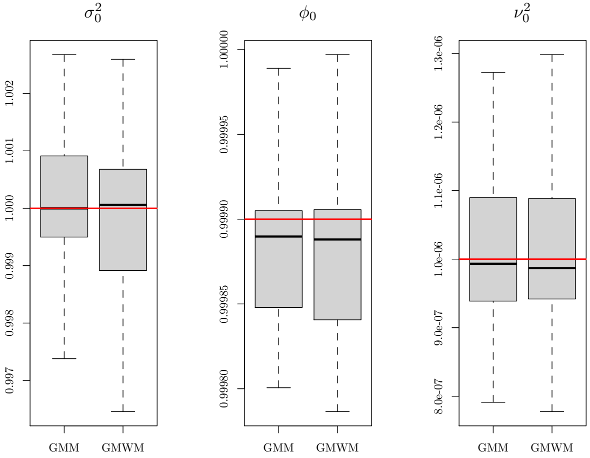
- GMWM
- GMM based on the first three lags of the autocovariance function ($\gamma(h)$ with $h = 0, \ldots, 3$)

Can we do better than the GMWM?

Below is the WV log-log plot of one of the simulated time series (it is similar to that of an IMU stochastic error signal)



Can we do better than the GMWM?



Although the GMM is slightly better for this model, the GMWM does not lose much efficiency and remains generally good and computationally efficient for many (complex) stationary and non-stationary models.

References I

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

Guerrier, S., R. Molinari, and J. Skaloud (2015). “Automatic Identification and Calibration of Stochastic Parameters in Inertial Sensors”. In: *Navigation* 62.4.

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

Guerrier, S. et al. (2013). “Wavelet-Variance-Based Estimation for Composite Stochastic Processes”. In: *Journal of the American Statistical Association* 108.503.

Guerrier, Stéphane et al. (2020). “Robust Two-Step Wavelet-Based Inference for Time Series Models”. In: *arXiv preprint arXiv:2001.04214*.

Mondal, Debashis and Donald B Percival (2012). “M-estimation of wavelet variance”. In: *Annals of the Institute of Statistical Mathematics* 64.1, pp. 27–53.

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

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

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

References II

Radi, A et al. (2017). "An Automatic Calibration Approach for the Stochastic Parameters of Inertial Sensors". In: *Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2017)*. IEEE, pp. 3053 –3060.

Serroukh, A., A. T. Walden, and D. B. Percival (2000). "Statistical Properties and Uses of the Wavelet Variance Estimator for the Scale Analysis of Time Series". In: *Journal of the American Statistical Association* 95.449, pp. 184–196.

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

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

Appendix A: Proof of Lemma 4.5

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

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

Therefore, we obtain

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

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

$$\hat{\nu}_j^2 = \frac{1}{M_j(T)} \sum_{t=L_j}^{T} W_{j,t}^2 \xmapsto{\mathcal{P}} \nu_j^2.$$

Proof of Lemma 4.5

Since

$$\hat{\nu}_j^2 \xmapsto{\mathcal{P}} \nu_j^2\,,$$

and given that J is a bounded quantity, we have that

$$\begin{aligned}\|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\|_2 &= \sum_{j=1}^{J} (\hat{\nu}_j^2 - \nu_j^2)^2 \\ &< J \max_j \left((\hat{\nu}_j^2 - \nu_j^2)^2\right).\end{aligned}$$

Based on the element-wise consistency, we have that

$$J \max_j \left((\hat{\nu}_j^2 - \nu_j^2)^2\right) \xmapsto{\mathcal{P}} 0,$$

and hence

$$\|\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\|_2 \xmapsto{\mathcal{P}} 0\,.$$

which concludes the proof. ■

▸ Return to Lemma 4.5

Proof of Lemma 4.6

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

$$\sqrt{T}\boldsymbol{a}^T\left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\right) \overset{\mathcal{D}}{\longmapsto} \mathcal{N}(0, \boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}),$$

then by the Cramér-Wold Device we have

$$\sqrt{T}\boldsymbol{s}^T\boldsymbol{\Sigma}^{-1/2}\left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\right) \overset{\mathcal{D}}{\longmapsto} \mathcal{N}(0, 1),$$

where $\boldsymbol{\Sigma}$ is the asymptotic covariance matrix of $\hat{\boldsymbol{\nu}}$, which concludes the proof. ■

▸ Return to Lemma 4.6