

Wavelet-Based Moment-Matching Techniques for Inertial Sensor Calibration

Stéphane Guerrier¹, Juan Jurado², Mehran Khaghani³, Gaetan Bakalli⁴, Mucyo Karemera⁵,
Roberto Molinari⁶, Samuel Orso⁷, John Raquet⁸, Christine Schubert⁹,
Jan Skaloud¹⁰, Haotian Xu¹¹, and Yuming Zhang¹²

Abstract—The task of inertial sensor calibration has required the development of various techniques to take into account the sources of measurement error coming from such devices. The calibration of the stochastic errors of these sensors has been the focus of increasing amount of research in which the method of reference has been the so-called “Allan variance (AV) slope method” which, in addition to not having appropriate statistical properties, requires a subjective input which makes it prone to mistakes. To overcome this, recent research has started proposing “automatic” approaches where the parameters of the probabilistic models underlying the error signals are estimated by matching functions of the AV or wavelet variance with their model-implied counterparts. However, given the increased use of such techniques, there has been no study or clear direction for practitioners on which approach is optimal for the purpose of sensor calibration. This article, for the first time, formally defines the class of estimators based on this technique and puts forward theoretical and applied results that, comparing with estimators in this class, suggest the use of the Generalized method of Wavelet moments (GMWM) as an optimal choice. In addition

to analytical proofs, experiment-driven Monte Carlo simulations demonstrated the superior performance of this estimator. Further analysis of the error signal from a gyroscope was also provided to further motivate performing such analyses, as real-world observed error signals may show significant deviation from manufacturer-provided error models.

Index Terms—Allan variance (AV), autonomous regression method for AV, generalized method of wavelet moments (GMWM), inertial measurement unit (IMU), slope method, stochastic error, Wavelet variance (WV).

I. INTRODUCTION

THE identification of a probabilistic time series model and the estimation of its relative parameters for the error signal issued from various sensors, such as inertial sensors, is a key challenge in many fields of engineering that has led to a great deal of research being produced. Aside from the size of the calibration data which can entail computational burdens for the mentioned estimation tasks, the stochastic errors of these signals are often complex since they can be characterized by composite (latent) stochastic processes where different underlying models contribute to the observed error signal. Although different approaches exist to perform estimation for the parameters of these processes, the currently adopted standard method for modeling the stochastic error of inertial sensors is the “Allan variance slope method” (AVSM) [1] that relies on the Allan variance (AV) which is widely accepted as being a quantity of reference for the calibration of the stochastic errors issued from (low-cost) inertial sensors. Indeed, the AVSM relies on the fact that certain stochastic processes contributing to the overall signal (such as white noises and random walks) are identifiable based on the slope of the plot. Based on this property, the AVSM requires practitioners to 1) make a log-log plot of the empirical AV of an error signal, 2) detect the regions of the plot which best represent an assumed model, and 3) estimate the parameters of the latter by estimating the slope of the AV within the selected region (based on which model parameters can be found). This approach is currently widely practiced in industry and academia making it a method of reference for inertial sensor calibration. Despite its popularity, however, the AVSM procedure is lengthy and prone to (human) errors as well as having been proven to be statistically inconsistent [2], thereby implying that, being among others subjective in nature, the resulting parameter estimates can be severely biased and do not improve as the length of the observed signal increases.

Manuscript received November 27, 2019; revised February 7, 2020; accepted March 12, 2020. Date of publication April 8, 2020; date of current version September 15, 2020. This work was supported in part by the SNSF Professorships under Grant 176843, in part by the Innosuisse-Boomerang under Grant 37308.1 IP-ENG, in part by the National Science Foundation (NSF) under Grant SES-1534433 and Grant SES-1853209, in part by the National Center for Advancing Translational Sciences/National Institute of Health (NCATS-NIH) under Grant UL1 TR002014, and in part by U.S. Air Force. The Associate Editor coordinating the review process was Vedran Bilas. (Stéphane Guerrier, Juan Jurado, and Mehran Khaghani are contributed equally to this work.) (Corresponding author: Stéphane Guerrier.)

Stéphane Guerrier, Gaetan Bakalli, Mucyo Karemera, Samuel Orso, Haotian Xu, and Yuming Zhang are with the Faculty of Science, Geneva School of Economics and Management, University of Geneva, 1205 Geneva, Switzerland (e-mail: stephane.guerrier@unige.ch; gaetan.bakalli@unige.ch; mucyo.karemera@unige.ch; samuel.orso@unige.ch; haotian.xu@unige.ch; yuming.zhang@unige.ch).

Juan Jurado is with the U.S. Air Force Test Pilot School, Edwards, CA 93523 USA (e-mail: juan.jurado.1@us.af.mil).

Mehran Khaghani is with Leica Geosystems, 9435 Heerbrugg, Switzerland (e-mail: mehran.khaghani@alummi.epfl.ch).

Roberto Molinari is with the Department of Statistics, Pennsylvania State University, State College, PA 16801 USA (e-mail: rum415@psu.edu).

John Raquet is with IS4S-Dayton, Integrated Solutions for Systems, Inc., Beavercreek, OH 45324 USA (e-mail: john.raquet@is4s.com).

Christine Schubert is with the Department of Mathematics and Statistics, Air Force Institute of Technology, Dayton, OH 45324 USA (e-mail: christine.schubert@afit.edu).

Jan Skaloud is with the Geodetic Engineering Laboratory, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: jan.skaloud@epfl.ch).

DISTRIBUTION STATEMENT A. Approved for public release; Distribution is unlimited 412TW-PA-19483.

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

Digital Object Identifier 10.1109/TIM.2020.2984820

0018-9456 © 2020 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

For the abovementioned reasons (i.e., statistical inconsistency and subjective nature of the AVSM), the literature has proposed different alternatives that either make this procedure autonomous or use the AV in a way to deliver consistent estimations (or both). Among these proposals, we can find those that make use of the linearity of logarithmic transforms of the AV to apply regression methods to estimate the parameters of the stochastic models assumed for the observed error signals. A recent example is given in [3] where a regularized regression approach is applied to the logarithm of the AV with base 10. Another approach is based on a linear transformation of the AV, more specifically the Haar wavelet variance (WV), where a generalized least square approach inverts the mapping between the model-implied WV and the empirical WV [4]. These moment-matching techniques belong to the class of the generalized method of moments estimators (introduced in [5]) where the considered ‘‘moments’’ are either the AV, the WV, or *functions* of one of these quantities. Given the presence of different moment-matching approaches based on the WV (AV) in order to automatize the sensor calibration process, it is important to understand how these approaches compare both from a theoretical as well as a practical point of view. Indeed, it would be appropriate to study these methods so that practitioners have some criteria that would allow them to choose the approach that best suits their requirements.

For the abovementioned reasons, this article intends to study the properties of this class of moment-matching approaches and put forward a proposed optimal moment-matching technique for inertial sensor calibration. Based on this goal, this article is organized as follows. In Section II, we summarize the notational convention used throughout this article. In Section III, we discuss the class of moment-matching estimators based on (functions of) the WV and formally compare them. Section IV compares calibration parameter estimation results using some existing moment-matching approaches based on the calibration of an accelerometer and a gyroscope from a STIM-300 inertial measurement unit (IMU). Finally, in Section VI, we summarize our findings and conclusions.

II. NOTATIONAL CONVENTION

Conventions

(x_t)	refers to a sequence of values indexed by integer t .
x_t	refers to the t th value of a sequence.
Y_t	refers to a random variable indexed by integer t .
y_t	refers to a realization of Y_t indexed by integer t .
T	refers to sample size of random variable Y_t .
\mathbb{R}_+	refers to the set of positive real numbers.
\mathbb{R}_-	refers to the set of negative real numbers.
$C^1(\mathcal{A}, \mathcal{B})$	refers to the set of functions from the set \mathcal{A} to the set \mathcal{B} whose first derivatives are continuous.
$\ \mathbf{A}\ _S$	denotes the matrix spectral norm.
$\ \mathbf{x}\ _A^2$	denotes the squared Mahalanobis distance, i.e., $\ \mathbf{x}\ _A^2 := \mathbf{x}^T \mathbf{A} \mathbf{x}$ where $\mathbf{x} \in \mathbb{R}^q$ and $\mathbf{A} \in \mathbb{R}^{q \times q}$.

$\|\mathbf{x}\|_2$ denotes the l_2 -norm of vector $\mathbf{x} \in \mathbb{R}^q$, i.e., $\|\mathbf{x}\|_2 := (\sum_{i=1}^q x_i^2)^{1/2}$.

$\|\mathbf{A}\|_S$ denotes the matrix spectral norm.

Important Notations

Θ	the parameter space.
θ	$(p \times 1)$ <i>generic</i> parameter vector such that $\theta \in \Theta \subset \mathbb{R}^p$.
θ_0	$(p \times 1)$ <i>true</i> parameter vector such that $\theta_0 \in \Theta \subset \mathbb{R}^p$.
F_θ	data generating model parameterized by θ .
\mathcal{J}	$\mathcal{J} := \{x \in \mathbb{N} \mid p \leq x < \log_2(T)\}$.
J	an element in the set \mathcal{J} , i.e., an integer denoting the the number of scales such that it is at least the same as the number of parameters but smaller than $\log_2(T)$.
\mathbf{v}	$(J \times 1)$ Wavelet variance or Allan variance vector.
$\mathbf{v}(\theta)$	$(J \times 1)$ Wavelet variance or Allan variance vector implied by θ assuming that F_θ corresponds to the true data generating process.
$\mathbf{f}(\cdot)$	a known vector-valued function such that $\mathbf{f} : \mathbb{R}_+^J \mapsto \mathcal{G} \subset \mathbb{R}^J$.
Ω	a positive definite matrix in $\mathbb{R}^{J \times J}$.
$\hat{\Omega}$	an estimate of the matrix Ω .
$ \cdot $	denotes the absolute value.
$\mathbf{A} \boxtimes \mathbf{B}$	we have that $\mathbf{A} \boxtimes \mathbf{B} := \mathbf{A} \mathbf{B} \mathbf{A}^T$ where $\mathbf{A} \in \mathbb{R}^{k \times d}$ and $\mathbf{B} \in \mathbb{R}^{d \times d}$.
T_j	number of wavelet coefficients at scale $j \in \mathbb{N} \setminus \{0\}$, $T_j := T - 2^j + 1$.

III. GENERALIZED METHOD OF WAVELET FUNCTIONAL MOMENTS

In order to formalize the framework of reference for this article, we first consider the time series $(X_t)_{t=1, \dots, T}$ which is supposedly generated by a composite stochastic process F_θ delivered by the sum of independent subprocesses. We let F_{θ_0} denote the true data-generating process, which is assumed known up to the value of θ_0 . The vector θ_0 is therefore the *true* parameter value which corresponds to a possible value in $\Theta \subset \mathbb{R}^p$. We let θ denote a *generic* parameter vector, which should therefore not be confused with the true parameter θ_0 . In order to discuss the estimation of θ_0 , let us consider the AV or WV which can be computed on the time series (X_t) for different (dyadic) scales of decomposition J . For the purpose of this work, we will, however, consider $J \in \mathcal{J}$ scales such that there are *at least* the same number of scales as of parameters. With this in mind, we introduce a class of estimators of θ_0 that we define as follows:

$$\hat{\theta} := \operatorname{argmin}_{\theta \in \Theta} \|\mathbf{f}(\hat{\mathbf{v}}) - \mathbf{f}(\mathbf{v}(\theta))\|_{\Omega}^2 \quad (1)$$

where $\hat{\mathbf{v}} \in \mathbb{R}_+^J$ and $\mathbf{v}(\theta) \in \mathbb{R}_+^J$ denote, respectively, a suitable estimator of the AV or WV computed on (X_t) and the model-based counterpart (i.e., the AV or WV implied by the assumed model F_θ). The vector-valued function $\mathbf{f}(\cdot)$ is such that $\mathbf{f} : \mathbb{R}_+^J \mapsto \mathcal{G} \subset \mathbb{R}^J$ and is assumed known. Moreover, $\Omega \in \mathbb{R}^{J \times J}$ is a positive definite matrix which, if estimated, shall be denoted as $\hat{\Omega}$ (instead of Ω) in order to emphasize the

stochastic nature of the matrix. Since the AV is a special case of the (Haar) WV (see [6]–[8] for details), we choose to call the class of estimators in (1) as “generalized method of wavelet functional moments” estimators (GMWFM). The latter is quite general and includes, among others, the generalized method of wavelet moments (GMWM) proposed in [4] or the autonomous regression method for Allan variance (ARMAV) of [3]. Indeed, the GMWM corresponds to the choice $\mathbf{f}(\mathbf{x}) = \mathbf{x}$, while the ARMAV is based on $\mathbf{f}_i(\mathbf{x}_i) = \log_{10}(\mathbf{x}_i)$, $i = 1, \dots, J$, where $\mathbf{f}_i(\cdot)$ and \mathbf{x}_i denote the i th element of $\mathbf{f}(\cdot)$ and \mathbf{x} , respectively. In addition, the GMWM and the ARMAV are based on different but relatively similar choices of the matrix $\mathbf{\Omega}$.

In this article, we investigate the requirements on the function $\mathbf{f}(\mathbf{x})$ to ensure that the estimator $\hat{\boldsymbol{\theta}}$ is consistent and asymptotically normally distributed. Moreover, we discuss whether an optimal choice for $\mathbf{f}(\mathbf{x})$ exists. For this purpose, we need to define a set of assumptions that will be used in order to investigate these properties. Therefore, let us study the first assumption regarding injectivity of the function $\mathbf{g}(\boldsymbol{\theta}) := \mathbf{f}(\mathbf{v}(\boldsymbol{\theta}))$ which can be found in the following.

Assumption A (Injectivity): The functions $\mathbf{f}(\cdot)$ and $\mathbf{v}(\cdot)$ are such that $\mathbf{f}(\cdot)$ is injective in \mathbb{R}_+^J and $\mathbf{v}(\cdot)$ is injective in Θ .

If this assumption holds, then a direct consequence is that $\mathbf{g}(\boldsymbol{\theta})$ is injective in Θ . More precisely, the first part of Assumption A is rather mild since the function $\mathbf{f}(\cdot)$ can be chosen in such a way as to respect this condition. However, the second part of the assumption can be challenging to prove. For example, [9] considered the injectivity of the function $\mathbf{v}(\cdot)$ and provide a series of results allowing to verify this property for various classes of latent time series models. The latter demonstrates that the second part of Assumption A would hold for the class of models considered in [4], with a few exceptions. For example, if the time series contains a drift with parameter ω , it is necessary to assume that the sign of ω is known (since $\mathbf{v}(\boldsymbol{\theta})$ only depends on ω^2). A general strategy to prove whether Assumption A holds for a specific model can be found in [10] (which is also used in [9]) while we prove the second requirement of Assumption A (i.e., $\mathbf{v}(\cdot)$ is injective in Θ) in Lemma 1 for the general model considered in [3]. The latter model is a composite model made by the sum of a 1) quantization noise with parameter $Q^2 \in \mathbb{R}_+$, 2) white noise with parameter $\sigma^2 \in \mathbb{R}_+$, 3) bias instability with parameter $B \in \mathbb{R}_+$, 4) random walk with parameter $\gamma^2 \in \mathbb{R}_+$, and 5) drift with parameter $\omega \in \mathbb{R}_+$.

Lemma 1: Let

$$\boldsymbol{\theta} := [Q^2 \ \sigma^2 \ B \ \gamma^2 \ \omega] \in \Theta \subset \mathbb{R}_+^5$$

and let c be a positive constant. Then, the function

$$\mathbf{v}_j(\boldsymbol{\theta}) := c \left(\frac{3Q^2}{2^{2j}} + \frac{\sigma^2}{2^j} + \frac{2 \log(2)}{\pi} B^2 + \frac{\gamma^2 2^{2j}}{3} + \omega^2 2^{2j-1} \right)$$

is injective in Θ .

Remark A: The positive constant c is simply related to the choice of the AV or (Haar) WV: in the case of the former, we have that $c = 1$ while for the Haar WV, we have $c = (1/2)$.

Proof: First, we notice that it is sufficient to show that

$$\mathbf{v}^*(\boldsymbol{\theta}) = \mathbf{v}^*(\boldsymbol{\theta}^*)$$

if and only if $\boldsymbol{\theta} = \boldsymbol{\theta}^*$, where $\mathbf{v}^*(\boldsymbol{\theta})$ denotes the first five elements of the vector $\mathbf{v}(\boldsymbol{\theta})$. Moreover, the function $\mathbf{v}^*(\boldsymbol{\theta})$ can be reparametrized as a function of $\boldsymbol{\beta}$ defined as

$$\boldsymbol{\beta} := [Q^2 \ \sigma^2 \ B^2 \ \gamma^2 \ \omega^2]$$

where the only difference with $\boldsymbol{\theta}$ is that the elements B and ω are squared. Since the latter elements are positive (the sign of ω is known and is assumed positive for this proof without loss of generality), the square function is also injective and this implies that if the WV is injective for their squares, by composition of injective functions it is also injective for the original values. Therefore, it is sufficient to show that

$$\mathbf{v}^*(\boldsymbol{\beta}) = \mathbf{v}^*(\boldsymbol{\beta}^*)$$

if and only if $\boldsymbol{\beta} = \boldsymbol{\beta}^*$. We start by computing the Jacobian matrix $\mathbf{J}(\boldsymbol{\beta})$ which is defined as

$$\mathbf{J}(\boldsymbol{\beta}) := \frac{\partial}{\partial \boldsymbol{\beta}^T} \mathbf{v}^*(\boldsymbol{\beta}) = c \begin{bmatrix} \frac{3}{2} & \frac{1}{2} & \frac{2 \log(2)}{\pi} & \frac{2}{3} & 2 \\ \frac{3}{2} & \frac{1}{2} & \frac{2 \log(2)}{\pi} & \frac{4}{3} & 8 \\ \frac{16}{3} & \frac{4}{3} & \frac{2 \log(2)}{\pi} & \frac{8}{3} & 32 \\ \frac{64}{3} & \frac{8}{3} & \frac{2 \log(2)}{\pi} & \frac{16}{3} & 128 \\ \frac{256}{3} & \frac{16}{3} & \frac{2 \log(2)}{\pi} & \frac{32}{3} & 512 \\ \frac{3}{1024} & \frac{1}{32} & \frac{2 \log(2)}{\pi} & \frac{3}{3} & \end{bmatrix}.$$

Since $\mathbf{J}(\boldsymbol{\beta})$ does not depend on $\boldsymbol{\beta}$ we let $\mathbf{J} := \mathbf{J}(\boldsymbol{\beta})$, we, based on the mean-value theorem, write

$$\begin{aligned} \mathbf{v}^*(\boldsymbol{\beta}) - \mathbf{v}^*(\boldsymbol{\beta}^*) &= \mathbf{v}^*(\boldsymbol{\beta}) - [\mathbf{v}^*(\boldsymbol{\beta}) + \mathbf{J} \cdot (\boldsymbol{\beta}^* - \boldsymbol{\beta})] \\ &= \mathbf{J} \cdot (\boldsymbol{\beta} - \boldsymbol{\beta}^*). \end{aligned}$$

Since we have that

$$\det(\mathbf{J}) = c^5 \frac{84357 \log(2)}{1024\pi} > 0$$

the only solution of the equation

$$\mathbf{J} \cdot (\boldsymbol{\beta} - \boldsymbol{\beta}^*) = \mathbf{0}$$

is $\boldsymbol{\beta}^* = \boldsymbol{\beta}$, which concludes the proof. \blacksquare

Having discussed Assumption A which appears to be reasonable to assume in general (given the different cases in which it is verified), we now consider the other set of assumptions that are needed to prove the consistency of the estimator $\hat{\boldsymbol{\theta}}$.

Assumption B (Compactness): The set Θ is compact.

Assumption C (Consistency): For all $j \in \{1, \dots, J\}$, we have

$$|\hat{\mathbf{v}}_j - \mathbf{v}_j(\boldsymbol{\theta}_0)| = o_p(1).$$

Moreover, if $\mathbf{\Omega}$ is estimated by $\hat{\mathbf{\Omega}}$ then we have

$$\|\hat{\mathbf{\Omega}} - \mathbf{\Omega}\|_S = o_p(1)$$

where, $X_n = o_p(1)$ means that for all $\varepsilon > 0$ we have that

$$\lim_{n \rightarrow \infty} \Pr(|X_n| > \varepsilon) = 0.$$

Assumption D (Continuity): The function $\mathbf{g}(\boldsymbol{\theta}) := \mathbf{f}(\mathbf{v}(\boldsymbol{\theta}))$ is continuous in Θ .

Assumption B is a common regularity condition which is typically assumed for most estimation problems or is replaced by other types of constraints. Its main purpose is to ensure that certain quantities that we will consider in the proofs will be bounded in order to ensure convergence. Assumption C is rather mild, and lower level conditions equivalent to this assumption can, for example, be found in [11] for the WV (as well as in [12] under weaker conditions) or by combining these results with the work of [7] who showed the equivalence between the AV and WV. Finally, Assumption D requires the function $\mathbf{f}(\mathbf{v}(\boldsymbol{\theta}))$ to be continuous in Θ which is the case when both $\mathbf{f}(\cdot)$ and $\mathbf{v}(\cdot)$ are continuous within their respective composition domains. Since the function $\mathbf{v}(\boldsymbol{\theta})$ is continuous in Θ for nearly all models of interest (such as those considered in [9] or the model discussed in [3]), it is sufficient for $\mathbf{f}(\cdot)$ to be continuous in \mathbb{R}_+^J to satisfy this assumption. Based on these assumptions, we can state the following consistency result.

Theorem 1: Under Assumptions A to D, we have that

$$\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = o_p(1).$$

Proof: Let

$$Q(\boldsymbol{\theta}) := \|\mathbf{g}(\boldsymbol{\theta}_0) - \mathbf{g}(\boldsymbol{\theta})\|_{\Omega}^2$$

where $\mathbf{g}(\boldsymbol{\theta})$ is defined in Assumption D. Then, we have

$$Q(\boldsymbol{\theta}) \leq \|\Omega\|_S \|\mathbf{g}(\boldsymbol{\theta}_0) - \mathbf{g}(\boldsymbol{\theta})\|_2^2.$$

Therefore, Assumption A implies that $Q(\boldsymbol{\theta})$ has a unique minimum in $\boldsymbol{\theta} = \boldsymbol{\theta}_0$.

Next, Assumption D directly implies the continuity of the function $Q(\boldsymbol{\theta})$ in Θ . Moreover, from the continuous mapping theorem together with Assumptions C and D, we have $|\hat{\mathbf{g}}_j - \mathbf{g}_j(\boldsymbol{\theta}_0)| = o_p(1)$ for all $j \in \{1, \dots, J\}$. Then, following the same strategy as in [12] (Proposition 3.1.), we obtain:

$$\sup_{\boldsymbol{\theta} \in \Theta} |\widehat{Q}(\boldsymbol{\theta}) - Q(\boldsymbol{\theta})| = o_p(1)$$

where

$$\widehat{Q}(\boldsymbol{\theta}) := \|\mathbf{f}(\hat{\mathbf{v}}) - \mathbf{f}(\mathbf{v}(\boldsymbol{\theta}))\|_{\Omega}^2.$$

Therefore, Theorem 2.1 of [13] can be applied to obtain the consistency of $\hat{\boldsymbol{\theta}}$, therefore concluding the proof. ■

Theorem 1 implies that any GMWFM estimator is consistent under the same conditions needed to ensure the consistency of the GMWM estimator provided that the function $\mathbf{f}(\cdot)$ is both injective (see Assumption A) and continuous (see Assumption D). Therefore, the requirements on the function $\mathbf{f}(\cdot)$ are rather mild but we shall see that this function has a more relevant impact on the asymptotic distribution of the estimator. Before introducing this result, as for the result of consistency, we first state and discuss relevant assumptions.

Assumption E (Interior and Convex): The vector $\boldsymbol{\theta}_0$ is such that $\boldsymbol{\theta}_0 \in \text{Int}(\Theta)$ and Θ is convex.

Assumption F (Function Differentiability): The function $\mathbf{f}(\cdot)$ is such that $\mathbf{f} \in \mathcal{C}^1(\mathbb{R}_+^J, \mathcal{G})$ allowing us to define

$$\mathbf{F}(\boldsymbol{\theta}_0) := \left. \frac{\partial}{\partial \mathbf{x}^T} \mathbf{f}(\mathbf{x}) \right|_{\mathbf{x}=\mathbf{v}(\boldsymbol{\theta}_0)}.$$

Moreover, defining the matrices

$$\Omega^*[\boldsymbol{\theta}_0, \mathbf{F}] := \mathbf{F}(\boldsymbol{\theta}_0)^T \boxtimes \Omega$$

and

$$\mathbf{A}(\boldsymbol{\theta}_0) := \left. \frac{\partial}{\partial \boldsymbol{\theta}^T} \mathbf{v}(\boldsymbol{\theta}) \right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}$$

then, the matrix

$$\mathbf{H}[\boldsymbol{\theta}_0, \mathbf{v}, \Omega, \mathbf{F}] := \mathbf{A}(\boldsymbol{\theta}_0)^T \boxtimes \Omega^*[\boldsymbol{\theta}_0, \mathbf{F}]$$

exists and is nonsingular.

Assumption G: The estimator $\hat{\mathbf{v}}$ has the following asymptotic distribution:

$$\sqrt{T_J}(\hat{\mathbf{v}} - \mathbf{v}(\boldsymbol{\theta}_0)) \xrightarrow{T \rightarrow \infty} \mathcal{N}(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta}_0))$$

where $\mathbf{V}(\boldsymbol{\theta}_0) := \lim_{T \rightarrow \infty} \text{cov}(\sqrt{T_J} \hat{\mathbf{v}})$ is a positive-definite symmetric matrix.

The topological requirements of Assumption E are quite mild although stronger than necessary. Indeed, the fact that $\boldsymbol{\theta}_0$ is required to be an interior point of the convex space Θ is convenient (but not strictly necessary) to ensure that expansions (such as Taylor expansions) can be made between $\boldsymbol{\theta}_0$ and an arbitrary point in Θ . Assumption F contains different requirements but what it basically requires is that the function $\mathbf{f}(\cdot)$ is differentiable in such a way that it can be used to make Taylor expansions for the purposes of demonstrating the asymptotic normality of the estimator $\hat{\boldsymbol{\theta}}$. Based on these expansions, we obtain expressions that deliver the matrix $\mathbf{H}[\boldsymbol{\theta}_0, \mathbf{v}, \Omega, \mathbf{F}]$ which needs to be positive-definite in order for the estimator to have an asymptotic variance (and hence define an asymptotic distribution). Finally, Assumption G is required for any estimator which makes use of moments (such as the AV or WV) to deliver asymptotic normality of the estimator itself. This assumption is verified under few additional conditions compared to those required for Assumption C, as highlighted again in [11], [14], and, under weaker conditions, in [12]. Using these assumptions, we obtain the following result.

Theorem 2: Under Assumptions A to G, the estimator $\hat{\boldsymbol{\theta}}$ has the following asymptotic distribution:

$$\sqrt{T_J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow{T \rightarrow \infty} \mathcal{N}(\mathbf{0}, \Sigma[\boldsymbol{\theta}_0, \mathbf{v}, \Omega, \mathbf{F}])$$

where

$$\Sigma[\boldsymbol{\theta}_0, \mathbf{v}, \Omega, \mathbf{F}] := \mathbf{B}[\boldsymbol{\theta}_0, \mathbf{v}, \Omega, \mathbf{F}] \boxtimes \mathbf{V}(\boldsymbol{\theta}_0)$$

and

$$\mathbf{B}[\boldsymbol{\theta}_0, \mathbf{v}, \Omega, \mathbf{F}] := \mathbf{H}[\boldsymbol{\theta}_0, \mathbf{v}, \Omega, \mathbf{F}]^{-1} \mathbf{A}(\boldsymbol{\theta}_0)^T \Omega^*[\boldsymbol{\theta}_0, \mathbf{F}].$$

Proof: Let $\Theta(T) := \{\mathbf{x} \in \mathbb{R}^p \mid \|\mathbf{x} - \boldsymbol{\theta}_0\|_2 \leq d(T)\}$, defining a nonstochastic set, where $d(T) = o(1)$. Moreover, we also define $\Theta^*(T) := \Theta \cap \Theta(T)$. Since $\hat{\boldsymbol{\theta}}$ is consistent by Theorem 1 (based on Assumptions A to D), there exists a function $d(T)$ such that

$$\begin{aligned} \hat{\boldsymbol{\theta}} &:= \underset{\boldsymbol{\theta} \in \Theta}{\text{argmin}} \|\mathbf{f}(\hat{\mathbf{v}}) - \mathbf{f}(\mathbf{v}(\boldsymbol{\theta}))\|_{\Omega}^2 \\ &= \underset{\boldsymbol{\theta} \in \Theta^*(T)}{\text{argmin}} \|\mathbf{f}(\hat{\mathbf{v}}) - \mathbf{f}(\mathbf{v}(\boldsymbol{\theta}))\|_{\Omega}^2 + o_p(1). \end{aligned} \quad (2)$$

Within the set $\Theta^*(T)$, which shrinks toward θ_0 as the sample size T increases, we can expand $\mathbf{f}(\hat{\mathbf{v}})$ and $\mathbf{f}(\mathbf{v}(\theta))$ around $\mathbf{v}(\theta_0)$ using a Taylor expansion to obtain

$$\begin{aligned}\mathbf{f}(\hat{\mathbf{v}}) &= \mathbf{f}(\mathbf{v}(\theta_0)) + \mathbf{F}(\theta_0)(\hat{\mathbf{v}} - \mathbf{v}(\theta_0)) + o_p(1) \\ \mathbf{f}(\mathbf{v}(\theta)) &= \mathbf{f}(\mathbf{v}(\theta_0)) + \mathbf{F}(\theta_0)(\mathbf{v}(\theta) - \mathbf{v}(\theta_0)) + o(1).\end{aligned}$$

Therefore, by combining this result with (2), we obtain

$$\begin{aligned}\hat{\theta} &= \operatorname{argmin}_{\theta \in \Theta^*(T)} \|\mathbf{F}(\theta_0)(\hat{\mathbf{v}} - \mathbf{v}(\theta))\|_{\Omega}^2 + o_p(1) \\ &= \operatorname{argmin}_{\theta \in \Theta^*(T)} \|\hat{\mathbf{v}} - \mathbf{v}(\theta)\|_{\Omega^*}^2 + o_p(1)\end{aligned}\quad (3)$$

where, similar to the definition of Assumption F,

$$\Omega^* := \Omega^*[\theta_0, \mathbf{F}] = \mathbf{F}(\theta_0)^T \boxtimes \Omega.$$

Next, we consider the following approximation of $\hat{\theta}$:

$$\tilde{\theta} := \operatorname{argmin}_{\theta \in \Theta} \|\hat{\mathbf{v}} - \mathbf{v}(\theta)\|_{\Omega^*}^2.$$

Proposition 2.2 of [12] implies, under the current assumption framework, that

$$\sqrt{T_J}(\tilde{\theta} - \theta_0) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma[\theta_0, \mathbf{v}, \Omega, \mathbf{F}]).$$

Since $\hat{\theta} = \tilde{\theta} + o_p(1)$, a direct application of Slutsky's theorem allows to conclude that the abovementioned results remains true for $\hat{\theta}$ and we obtain

$$\sqrt{T_J}(\hat{\theta} - \theta_0) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \Sigma[\theta_0, \mathbf{v}, \Omega, \mathbf{F}])$$

which concludes the proof. \blacksquare

An implication of this result (made evident in particular from (3) in the proof) is the fact that, no matter which choice is made for the function $\mathbf{f}(\cdot)$ and the matrix Ω (provided that they satisfy the previously mentioned assumptions), we can define a matrix Ω^* (that depends upon $\mathbf{f}(\cdot)$ and Ω) such that we can express the estimator as

$$\hat{\theta} := \operatorname{argmin}_{\theta \in \Theta} \|\hat{\mathbf{v}} - \mathbf{v}(\theta)\|_{\Omega^*}^2. \quad (4)$$

Therefore, as long as the matrix $\Omega^* := \Omega^*[\theta_0, \mathbf{F}]$ is positive definite, the estimator $\hat{\theta}$ is asymptotically normally distributed under the abovementioned assumptions, and the only aspect that is affected by the change of Ω^* is the efficiency of the resulting estimator. Consequently, the choice of a specific function $\mathbf{f}(\cdot)$ (which respects the required properties) only contributes to modifying the weighting matrix Ω^* , therefore delivering approximately the same results for any such function $\mathbf{f}(\cdot)$. The weighting matrix Ω^* is therefore crucial to the efficiency of the estimator $\hat{\theta}$. As shown in the following corollary, the optimal choice (in terms of asymptotic efficiency) of Ω^* is the inverse of $\mathbf{F}(\theta_0) \boxtimes \mathbf{V}(\theta_0)$. Although the true $\mathbf{V}(\theta_0)$ is unknown in practice, it can be consistently estimated by the estimator proposed in [15] or with the approach discussed in [12]. Moreover, the corollary of Theorem 2 shows how asymptotically optimal estimators can be constructed for the GMWFM.

Corollary 1: Under Assumptions A to G (i.e., the same conditions of Theorem 2), the estimator $\hat{\theta}$ based on the

function $\mathbf{f}(\cdot)$ and the matrix $\Omega^\circ := [\mathbf{F}(\theta_0) \boxtimes \mathbf{V}(\theta_0)]^{-1}$ is asymptotically efficient in the class of GMWFM estimators.

Proof: Under our assumptions, it is easy to verify that the asymptotic covariance matrix of $\hat{\theta}$ is given by

$$\Sigma[\theta_0, \mathbf{v}, \Omega^\circ, \mathbf{F}] = [\mathbf{A}(\theta_0)^T \boxtimes \mathbf{V}(\theta_0)^{-1}]^{-1}.$$

We proceed by demonstrating that the difference between the asymptotic covariance matrix in Theorem 2 and the above-mentioned covariance matrix leads to a positive semidefinite matrix. Following Section 5.2 in [13], it is easy to show that:

$$\begin{aligned}\Sigma[\theta_0, \mathbf{v}, \Omega, \mathbf{F}] - \Sigma[\theta_0, \mathbf{v}, \Omega^\circ, \mathbf{F}] \\ = \mathbf{H}[\theta_0, \mathbf{v}, \Omega, \mathbf{F}]^{-1} \boxtimes \mathbb{E}[\mathbf{W}\mathbf{W}^T]\end{aligned}$$

where

$$\begin{aligned}\mathbf{W} &:= \mathbf{A}(\theta_0)^T \Omega^*[\theta_0, \mathbf{F}]\mathbf{Z} \\ &\quad - \mathbf{H}[\theta_0, \mathbf{v}, \Omega, \mathbf{F}]\Sigma[\theta_0, \mathbf{v}, \Omega^\circ, \mathbf{F}]\mathbf{A}(\theta_0)^T \mathbf{F}(\theta_0)^T \Omega^\circ \mathbf{Z}\end{aligned}$$

and \mathbf{Z} is a random vector such that

$$\mathbb{E}[\mathbf{Z}\mathbf{Z}^T] = \mathbf{F}(\theta_0)\mathbf{V}(\theta_0)\mathbf{F}(\theta_0)^T = (\Omega^\circ)^{-1}.$$

The result follows since $\mathbb{E}[\mathbf{W}\mathbf{W}^T]$ is positive semidefinite, which concludes the proof. \blacksquare

Corollary 1 shows that, under suitable conditions, any estimator belonging to the class of GMWFM estimators can be asymptotically optimal provided that it is based on the function $\mathbf{f}(\cdot)$ and the matrix $\Omega^\circ := [\mathbf{F}(\theta_0)\mathbf{V}(\theta_0)\mathbf{F}(\theta_0)^T]^{-1}$. This implies that there exists an infinite number of possible efficient estimators (based on different functions $\mathbf{f}(\cdot)$ and matrix Ω°) leading to the same optimal asymptotic covariance matrix $[\mathbf{A}(\theta_0)^T \boxtimes \mathbf{V}(\theta_0)^{-1}]^{-1}$. In the case where $\mathbf{f}(\mathbf{x}) = \mathbf{x}$, the matrix Ω° has the simplest expression given by $\Omega^\circ = \mathbf{V}(\theta_0)^{-1}$ since $\mathbf{F}(\theta_0) = \mathbf{I}$, therefore also suggesting that its (consistent) estimation is more straightforward in practice. The choice of the function $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ presents several other advantages compared to possible alternative choices. For example, this function allows the estimator to be solved analytically for various commonly used models. This is of particular importance for inertial sensor calibration as most models considered in this field allow for such a closed-form solution. Indeed, suppose that there exists a matrix \mathbf{X} that does not depend on θ such that $\mathbf{v}(\theta)$ can be expressed as $\mathbf{v}(\theta) = \mathbf{X}\mathbf{h}(\theta)$ for all $\theta \in \Theta$, where $\mathbf{h}(\cdot)$ is an injective vector-valued function such that $\mathbf{h} : \Theta \mapsto \mathcal{H} \subset \mathbb{R}^p$. This is, for example, the case for the model considered in Lemma 1. Indeed, denoting $a := 2 \log(2)/\pi$, the function $\mathbf{v}(\theta)$ can be expressed as follows:

$$\mathbf{v}(\theta) = c \underbrace{\begin{bmatrix} \frac{3}{2^2} & \frac{1}{2^1} & a & \frac{2^1}{3} & 2^1 \\ \frac{3}{2^4} & \frac{1}{2^2} & a & \frac{2^2}{3} & 2^3 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \frac{3}{2^{2J}} & \frac{1}{2^J} & a & \frac{2^J}{3} & 2^{2J-1} \end{bmatrix}}_{\mathbf{X}} \underbrace{\begin{bmatrix} Q^2 \\ \sigma^2 \\ B^2 \\ \gamma^2 \\ \omega^2 \end{bmatrix}}_{\mathbf{h}(\theta)}. \quad (5)$$

Hence, taking the parameter vector defined in Lemma 1, the vector-valued function $\mathbf{h}(\boldsymbol{\theta})$ is the identity for all elements except for B and ω for which it is the square function. Since all parameters (are assumed to) belong to \mathbb{R}_+ , we have that the function $\mathbf{h}(\boldsymbol{\theta})$ is injective which, in general, allows us to write

$$\hat{\boldsymbol{\theta}} := \operatorname{argmin}_{\boldsymbol{\theta} \in \Theta} \|\hat{\boldsymbol{v}} - \mathbf{v}(\boldsymbol{\theta})\|_{\Omega}^2 = \mathbf{h}^{-1}(\hat{\boldsymbol{v}})$$

where

$$\hat{\boldsymbol{v}} := \operatorname{argmin}_{\boldsymbol{v} \in \mathcal{H}} \|\hat{\boldsymbol{v}} - \mathbf{X}\boldsymbol{v}\|_{\Omega}^2$$

and where we let $\boldsymbol{v} := \mathbf{h}(\boldsymbol{\theta})$. Moreover, since the function $\|\hat{\boldsymbol{v}} - \mathbf{X}\boldsymbol{v}\|_{\Omega}^2$ is differentiable in \boldsymbol{v} , we have

$$\hat{\boldsymbol{v}} := \operatorname{argmin}_{\boldsymbol{v} \in \mathcal{H}} \|\hat{\boldsymbol{v}} - \mathbf{X}\boldsymbol{v}\|_{\Omega}^2 = \operatorname{argzero}_{\boldsymbol{v} \in \mathcal{H}} \mathbf{X}^T \Omega (\hat{\boldsymbol{v}} - \mathbf{X}\boldsymbol{v}).$$

Therefore, we have that

$$\mathbf{X}^T \Omega \mathbf{X} \hat{\boldsymbol{v}} = \mathbf{X}^T \Omega \hat{\boldsymbol{v}}$$

which corresponds to the standard (weighted) least-squares equations. Since Ω is invertible, \mathbf{X} is full (column) rank, and $\mathbf{X}^T \Omega \mathbf{X}$ is nonsingular, thus we obtain

$$\hat{\boldsymbol{v}} = (\mathbf{X}^T \Omega \mathbf{X})^{-1} \mathbf{X}^T \Omega \hat{\boldsymbol{v}} \quad (6)$$

which provides a simple plug-in estimator for $\boldsymbol{\theta}_0$ defined as

$$\hat{\boldsymbol{\theta}} = \mathbf{h}^{-1}[(\mathbf{X}^T \Omega \mathbf{X})^{-1} \mathbf{X}^T \Omega \hat{\boldsymbol{v}}]. \quad (7)$$

The abovementioned closed-form solution is therefore a first advantage of choosing $\mathbf{f}(\mathbf{x}) = \mathbf{x}$. Moreover, there are a few practical advantages stemming from this setting: the first of which is the fact that, given a closed form solution for this class of models, no optimization is required to compute the estimates, therefore delivering computationally fast solutions. In addition, even if the model of interest contains a subset of this class of models, this closed-form solution can be used as an approximate method to quickly obtain “good” starting values that can increase the computational efficiency of the optimization procedure required to solve (1). Finally, the abovementioned form allows to obtain the exact form of the asymptotic variance of $\hat{\boldsymbol{v}}$ up to the value of $\mathbf{V}(\boldsymbol{\theta}_0)$ (i.e., the asymptotic variance of $\hat{\boldsymbol{v}}$) which, using the delta method, would allow us to obtain the exact variance of $\hat{\boldsymbol{\theta}}$ for this class of models. Indeed, Lemma 2 shows that asymptotic properties of $\hat{\boldsymbol{\theta}}$ discussed in Theorems 1 and 2 are in this case satisfied under nearly no assumptions.

Lemma 2: Consider the estimator $\hat{\boldsymbol{\theta}}$ defined in (1) with the choice $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ and the model studied in Lemma 1. Then, the estimator $\hat{\boldsymbol{\theta}}$ is such that $\|\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = o_p(1)$. In addition, if $\boldsymbol{\theta}_0 \in \operatorname{Int}(\Theta)$, the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ is given by

$$\sqrt{T_J}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, [\mathbf{G}(\boldsymbol{\theta}_0)\mathbf{M}] \boxtimes \mathbf{V}(\boldsymbol{\theta}_0))$$

where

$$\begin{aligned} \mathbf{V}(\boldsymbol{\theta}_0) &:= \lim_{T \rightarrow \infty} \operatorname{cov}(\sqrt{T_J} \hat{\boldsymbol{v}}) \\ \mathbf{M} &:= (\mathbf{X}^T \Omega \mathbf{X})^{-1} \mathbf{X}^T \Omega \\ \mathbf{G}(\boldsymbol{\theta}_0) &:= \left(\frac{\partial}{\partial \boldsymbol{\theta}^T} \mathbf{h}(\boldsymbol{\theta}) \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} \right)^{-1}. \end{aligned}$$

Proof: Using Theorem 2.1. of [16], we have

$$\sqrt{T_J}(\hat{\boldsymbol{v}} - \mathbf{v}(\boldsymbol{\theta}_0)) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{V}(\boldsymbol{\theta}_0)). \quad (8)$$

Therefore, using (5), we have

$$\hat{\boldsymbol{\theta}} = (\mathbf{X}^T \Omega \mathbf{X})^{-1} \mathbf{X}^T \Omega \hat{\boldsymbol{v}} \xrightarrow{\mathcal{P}} \mathbf{h}(\boldsymbol{\theta}_0)$$

where $\xrightarrow{\mathcal{P}}$ is used to denote “convergence in probability.”

By the continuous mapping theorem, we obtain

$$\hat{\boldsymbol{\theta}} = \mathbf{h}^{-1}(\hat{\boldsymbol{v}}) \xrightarrow{\mathcal{P}} \mathbf{h}^{-1}[\mathbf{h}(\boldsymbol{\theta}_0)] = \boldsymbol{\theta}_0$$

which verifies the first statement of Lemma 2. Using (6) and (8), we directly obtain

$$\sqrt{T_J}(\hat{\boldsymbol{v}} - \mathbf{h}(\boldsymbol{\theta}_0)) \xrightarrow[T \rightarrow \infty]{\mathcal{D}} \mathcal{N}(\mathbf{0}, \mathbf{M} \boxtimes \mathbf{V}(\boldsymbol{\theta}_0)).$$

In addition, we have the following through a Taylor expansion:

$$\begin{aligned} &\sqrt{T_J}(\hat{\boldsymbol{v}} - \mathbf{h}(\boldsymbol{\theta}_0)) \\ &= \sqrt{T_J}(\mathbf{h}(\hat{\boldsymbol{\theta}}) - \mathbf{h}(\boldsymbol{\theta}_0)) \\ &= \sqrt{T_J}(\mathbf{h}(\boldsymbol{\theta}_0) + \mathbf{G}(\boldsymbol{\theta}_0)^{-1}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) - \mathbf{h}(\boldsymbol{\theta}_0)) + o_p(1) \end{aligned}$$

which allows, using Slutsky’s theorem, to verify the second statement of Lemma 2 and concludes the proof. ■

Remark B: In Lemma 2, Ω is supposedly a known quantity but it could straightforwardly be replaced by a consistent estimator (see Assumption C). The requirement that $\boldsymbol{\theta}_0$ lies inside the parameter space allows to make a first-order approximation of $\mathbf{h}(\boldsymbol{\theta})$. In the case there is no interest in estimating the bias instability parameter B and the drift parameter ω , this requirement can be relaxed as $\mathbf{h}(\boldsymbol{\theta})$ is the identity function [see (5)].

Aside from the computational advantage of using the function $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ in (1) for the class of models for which $\mathbf{v}(\boldsymbol{\theta})$ can be expressed as $\mathbf{X}\mathbf{h}(\boldsymbol{\theta})$ (or as a “good” starting value for other models), there is another potential advantage of using the identity function for the purposes of estimation which relates to their bias. Indeed, the standard estimators of AV or WV are unbiased (see, e.g., [11]), meaning that $\mathbb{E}[\hat{\boldsymbol{v}}] = \mathbf{v}(\boldsymbol{\theta}_0)$ and implying that $\mathbb{E}[\hat{\boldsymbol{\theta}}] = \mathbf{h}(\boldsymbol{\theta}_0)$ based on (6). A first implication of these properties is that, aside from consistency, it is possible to show that the estimates of most of the parameters in models whose theoretical WV can be expressed as $\mathbf{X}\mathbf{h}(\boldsymbol{\theta})$ are unbiased (i.e., when $\mathbf{h}_i(\cdot)$ is the identity or a linear function). However, if the theoretical WV cannot be expressed in the latter form, formal proofs to determine the finite sample behavior of the estimators defined in (1) may be hard to derive. Nevertheless, an intuitive argument would support the employment of the function $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ since it directly makes use of unbiased estimators of the WV to match their theoretical counterpart (which is a desirable property in order to achieve unbiasedness with respect to the parameter of interest $\boldsymbol{\theta}_0$).

To better highlight the concepts behind the abovementioned reasoning, we compare the following two asymptotically equivalent estimators:

- 1) $\hat{\boldsymbol{\theta}}$ based on the choice of function $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ and a (nonrandom) weight matrix Ω_1
- 2) $\hat{\boldsymbol{\theta}}$ based on another choice of function, such that $\mathbb{E}[\mathbf{f}(\hat{\boldsymbol{v}})] \neq \mathbf{f}(\mathbb{E}[\hat{\boldsymbol{v}}])$ and a (nonrandom) weight matrix Ω_2 .

We then consider the expected value of the objective function of the first estimator $\hat{\theta}$

$$\mathbb{E}[\|\hat{\mathbf{v}} - \mathbf{v}(\theta)\|_{\Omega_1}^2] = \|\mathbb{E}[\hat{\mathbf{v}} - \mathbf{v}(\theta)]\|_{\Omega_1}^2 + \text{tr}(\Omega_1 \text{var}(\hat{\mathbf{v}})).$$

Since the second term of the abovementioned equation does not depend on θ , we let $b_1 := \text{tr}(\Omega_1 \text{var}(\hat{\mathbf{v}}))$ and we can write

$$\mathbb{E}[\|\hat{\mathbf{v}} - \mathbf{v}(\theta)\|_{\Omega_1}^2] = \|\mathbf{v}(\theta_0) - \mathbf{v}(\theta)\|_{\Omega_1}^2 + b_1$$

since $\mathbb{E}[\hat{\mathbf{v}}] = \mathbf{v}(\theta_0)$. Therefore, this function is unbiased in the sense that it is minimized at the true value θ_0 .

Following the argument in [17], it is therefore expected that the bias of this estimator will be large if bias in the objective function of the corresponding estimator is large. Next, we consider the objective function of the second estimator and, recalling that $\mathbf{g}(\theta) := \mathbf{f}(\mathbf{v}(\theta))$ as defined in Assumption D, we define $\Delta(\theta_0) := \mathbb{E}[\mathbf{f}(\hat{\mathbf{v}})] - \mathbf{g}(\theta_0)$ and $b_2 := \text{tr}(\Omega_2 \text{var}(\mathbf{f}(\hat{\mathbf{v}})))$. Using these definitions, we obtain

$$\begin{aligned} \mathbb{E}[\|\mathbf{f}(\hat{\mathbf{v}}) - \mathbf{g}(\theta)\|_{\Omega_2}^2] &= \|\mathbb{E}[\mathbf{f}(\hat{\mathbf{v}}) - \mathbf{g}(\theta)]\|_{\Omega_2}^2 + b_2 \\ &= \|\mathbf{g}(\theta_0) + \Delta(\theta_0) - \mathbf{g}(\theta)\|_{\Omega_2}^2 + b_2. \end{aligned}$$

Moreover, by applying the mean value theorem, it is possible to assess the order of $\Delta(\theta_0)$

$$\begin{aligned} \Delta(\theta_0) &= \mathbb{E}[\mathbf{f}(\hat{\mathbf{v}})] - \mathbf{g}(\theta_0) \\ &= \mathbb{E}[\mathbf{g}(\theta_0) + \mathbf{F}(\mathbf{v}(\theta^*))(\hat{\mathbf{v}} - \mathbf{v}(\theta_0))] - \mathbf{g}(\theta_0) \\ &= T_J^{-1/2} \mathbb{E}[\sqrt{T_J} \mathbf{F}(\mathbf{v}(\theta^*))(\hat{\mathbf{v}} - \mathbf{v}(\theta_0))] = \mathcal{O}(T_J^{-1/2}) \end{aligned}$$

where $\theta^* \in \Theta$, $\mathbf{v}(\theta^*)$ is on the line connecting $\mathbf{v}(\theta_0)$ and $\hat{\mathbf{v}}$, and the term $\sqrt{T_J} \mathbf{F}(\mathbf{v}(\theta^*)) \mathbf{F}(\mathbf{v}(\theta^*)) (\hat{\mathbf{v}} - \mathbf{v}(\theta_0))$ is $\mathcal{O}_p(1)$ by the continuous mapping theorem and Assumption C. Therefore, we have

$$\begin{aligned} \underset{\theta \in \Theta}{\text{argmin}} \mathbb{E}[\|\mathbf{f}(\hat{\mathbf{v}}) - \mathbf{g}(\theta)\|_{\Omega_2}^2] \\ = \underset{\theta \in \Theta}{\text{argmin}} \|\mathbf{g}(\theta_0) + \Delta(\theta_0) - \mathbf{g}(\theta)\|_{\Omega_2}^2 \end{aligned}$$

implying that bias of the objective function is of order $\mathcal{O}(T_J^{-1/2})$ and, consequently, the bias of $\hat{\theta}$ is also of order $\mathcal{O}(T_J^{-1/2})$. As a result, we expect estimators based on the choice $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ to have relatively small biases compared to other choices of $\mathbf{f}(\mathbf{x})$. An important example is when $\mathbf{f}(\mathbf{x})$ is a convex/concave function. More specifically, using Jensen's inequality, we have the following:

$$\begin{cases} \mathbb{E}[\mathbf{f}(\hat{\mathbf{v}})] > \mathbf{f}(\mathbb{E}[\hat{\mathbf{v}}]), & \text{if } \mathbf{f} \text{ is strictly convex} \\ \mathbb{E}[\mathbf{f}(\hat{\mathbf{v}})] < \mathbf{f}(\mathbb{E}[\hat{\mathbf{v}}]), & \text{if } \mathbf{f} \text{ is strictly concave.} \end{cases} \quad (9)$$

Having delivered different theoretical results and arguments comparing the use of different functions $\mathbf{f}(\cdot)$ for the purpose of parameter estimation as defined in (1), Section IV performs some simulation studies where, using different models for the stochastic error of the sensors, we compare the performance of different GMWFM estimators.

TABLE I
TRUE PARAMETER VALUES FOR MONTE CARLO SIMULATIONS

θ	Gyro		Accel	
	Value	Units	Value	Units
QN (Q)	-	-	1.79×10^{-6}	m/s
WN (σ)	1.57×10^{-1}	deg/ $\sqrt{\text{hr}}$	4.70×10^{-2}	m/s/ $\sqrt{\text{hr}}$
RW (γ)	1.34	deg/hr/ $\sqrt{\text{hr}}$	4.35×10^1	m/s/hr/ $\sqrt{\text{hr}}$
DR (ω)	-	-	4.14×10^1	m/s/hr/hr

IV. SIMULATION RESULTS

In this section, we compare the estimation performance (in terms of bias and variance) across two GMWFM estimators, namely, the GMWM and ARMAV, as well as the standard AVSM as a reference. To carry out this comparison, we make use of the parameter estimates (based on the GMWM) of the stochastic processes identified from the real calibration data coming from the X -axis accelerometer and X -axis gyroscope of a STIM-300 IMU [18]. Hence, the parameter estimates on this real calibration data were considered as being the true parameter values for simulation purposes and their values (along with the respective models) are presented in Table I. For each sensor (with respective stochastic models), two Monte Carlo simulation settings were considered based on two different sample sizes (long and short) and with sampling frequency fixed at 250 Hz. In the ‘‘long’’ signal setting, the sample size was set to $T = 6.3 \times 10^6$, corresponding to 7 h of calibration data, while in the ‘‘short’’ signal setting, the sample size was set to $T = 2.5 \times 10^6$, corresponding to 2.5 h of calibration data. Each estimation was repeated 3000 times, and the empirical distributions of the three estimators are presented in Fig. 1.

As shown in Fig. 1, it would appear that the GMWM approach delivers the best overall performances in terms of bias and dispersion, whereas the ARMAV and AVSM approaches have alternating performances according to the parameter of interest. When comparing the different lengths of the signals (i.e., long and short settings), we can observe that the bias and variance of all GMWFM estimators appear to marginally improve as confirmed in Fig. 2 that summarizes the overall estimation performance of each approach in terms of root mean square error (RMSE). From Figs. 1 and 2, the GMWM displays the best performance overall (i.e., lower RMSE) across all parameters regardless of sample length. In contrast, AVSM tends to exhibit larger RMSE, especially when estimating the parameter of the quantization noise (Q). In the accelerometer simulations, the ARMAV remains close to the performance of the GMWM while, in the gyroscope simulations, the ARMAV shows similar performance to the AVSM when estimating the white noise parameter (σ), while all three approaches perform similarly when estimating random walk (γ).

The simulation results in this section, therefore, seem to confirm the conclusions made in the conclusions made in Section III based on the developed theoretical results. Indeed, it would appear that the choice of the function $\mathbf{f}(\mathbf{x}) = \mathbf{x}$ would

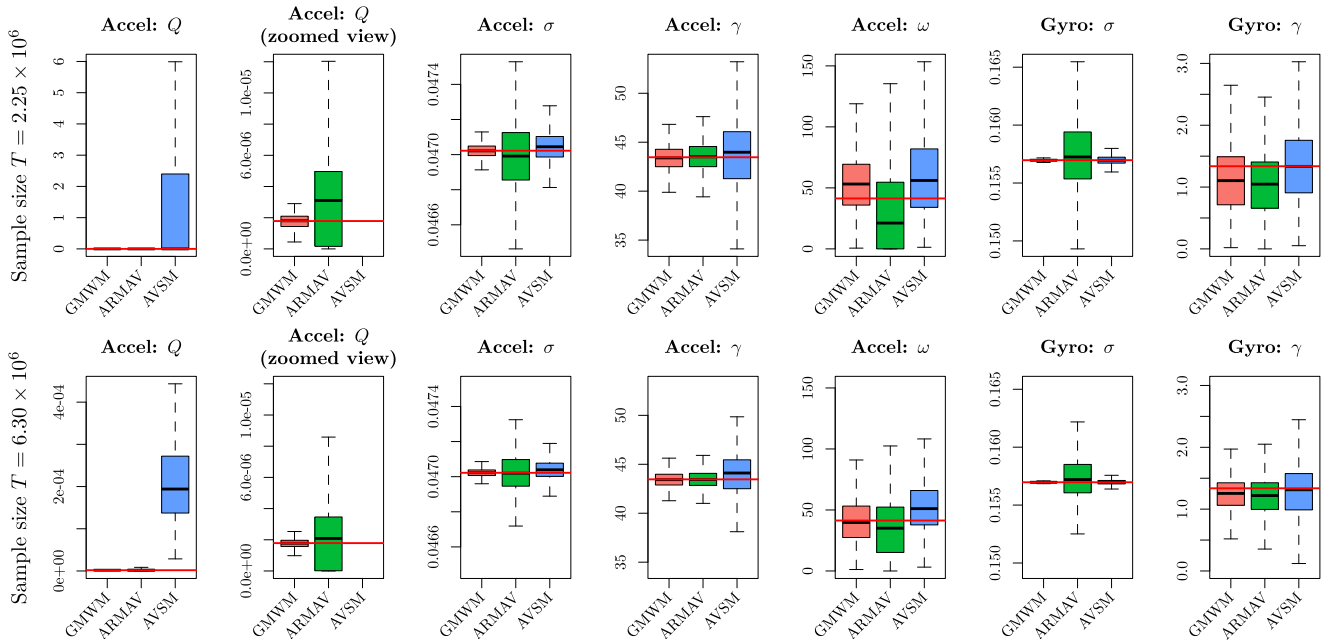


Fig. 1. Empirical distribution of the estimations of the GMWM, ARMAV, and AVSM approaches for the parameters of the stochastic error of an accelerometer and a gyroscope for signal lengths of $T = 2.5 \times 10^6$ (top row) and $T = 6.3 \times 10^6$ (bottom row), respectively.

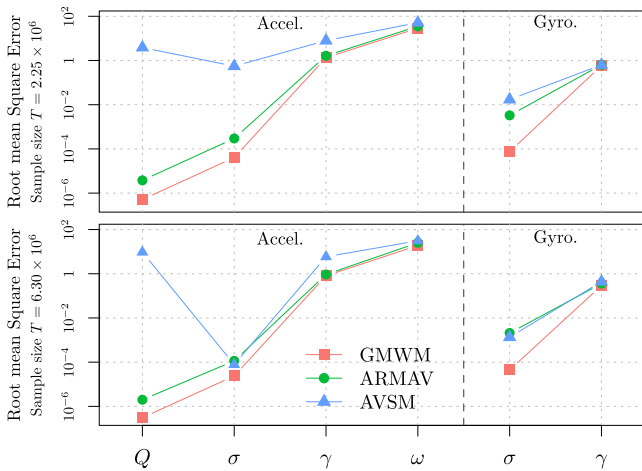


Fig. 2. Comparison of estimated root mean squared error between GMWM, ARMAV, and AVSM approaches for the parameters of the stochastic error of an accelerometer and a gyroscope for signal lengths of $T = 2.5 \times 10^6$ (top row) and $T = 6.3 \times 10^6$ (bottom row), respectively.

be the optimal choice when considering a GMWFM estimator for (automatic) sensor calibration.

V. REAL-WORLD APPLICATION

Following the establishment of GMWM as the optimal moment-matching technique for inertial sensor calibration, we demonstrate in this section the necessity and benefits of performing such calibration, even when rather extensive calibration information is provided by the sensor manufacturer. To this end, we analyze 3.5 h of static data collected on a Navchip IMU [19] using GMWM and compare the resulting model with that suggested by the manufacturer. Another example with more detailed discussions can be found in [20].

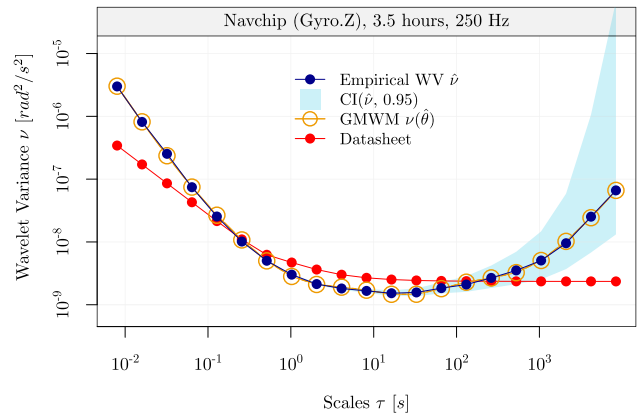


Fig. 3. Comparison of the fitness of models estimated by GMWM and provided in the datasheet for the gyroscope inside Navchip IMU.

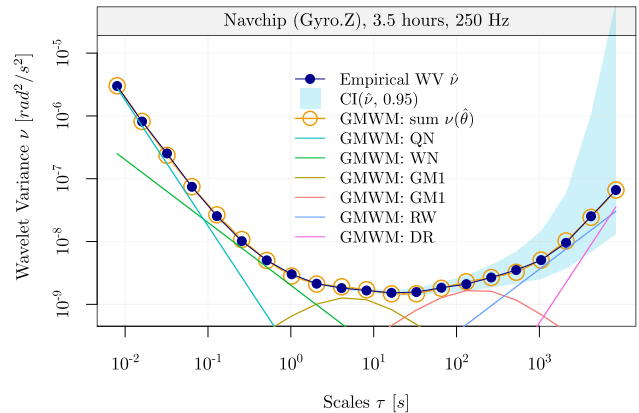


Fig. 4. Decomposition of the model suggested by GMWM in terms of WV for the gyroscope inside Navchip IMU.

Fig. 3 shows the empirical WVs computed from the data, resulted from the error model provided by the manufacturer in the datasheet, and the one from the error model estimated

by GMWM. While the model estimated by GMWM closely matches the collected data in terms of WV, the model suggested in the datasheet is far from them and too optimistic at most scales. The model estimated by the GMWM is composed of quantization noise, white noise, two first-order Gauss–Markov processes, random walk, and drift. Fig. 4 shows the decomposition of the implied WV by each of these models, revealing the important contribution of all of them.

VI. CONCLUSION

This article discussed the properties and performance of a general class of estimators denoted as GMWFM. Being based on moment-matching techniques through the use of different functions of the WV, these estimators put forward different approaches to perform (automatic) sensor stochastic calibration. Given the variety of proposed functions that build this class of estimators, this article analyzed and proved the properties of such estimators, thereby suggesting that the optimal estimator in this class is the one based on the identity function which corresponds to the GMWM. These conclusions are supported by the simulation study which consequently suggests that the GMWM should be the preferred estimator among the GMWFM estimators for stochastic calibration of inertial sensors.

REFERENCES

- [1] *IEEE Standard Specification Format Guide and Test Procedure for Single-Axis Interferometric Fiber Optic Gyros*, Standard 952-1997, Feb. 1998, pp. 1–84.
- [2] S. Guerrier, R. Molinari, and Y. Stebler, “Theoretical limitations of allan variance-based regression for time series model estimation,” *IEEE Signal Process. Lett.*, vol. 23, no. 5, pp. 597–601, May 2016.
- [3] J. Jurado, C. M. Schubert Kabban, and J. Raquet, “A regression-based methodology to improve estimation of inertial sensor errors using allan variance data,” *Navigation*, vol. 66, no. 1, pp. 251–263, Jan. 2019.
- [4] S. Guerrier, J. Skaloud, Y. Stebler, and M.-P. Victoria-Feser, “Wavelet-variance-based estimation for composite stochastic processes,” *J. Amer. Stat. Assoc.*, vol. 108, no. 503, pp. 1021–1030, Sep. 2013.
- [5] L. P. Hansen, “Large sample properties of generalized method of moments estimators,” *Econometrica*, vol. 50, no. 4, pp. 1029–1054, Jul. 1982.
- [6] P. Flandrin, “Wavelet analysis and synthesis of fractional brownian motion,” *IEEE Trans. Inf. Theory*, vol. 38, no. 2, pp. 910–917, Mar. 1992.
- [7] D. B. Percival and P. Guttorp, “Long-memory processes, the allan variance and wavelets,” in *Wavelet Anal. Appl.*, vol. 4, pp. 325–344, Apr. 1994.
- [8] D. B. Percival, “A wavelet perspective on the allan variance,” *IEEE Trans. Ultrason., Ferroelectr., Freq. Control*, vol. 63, no. 4, pp. 538–554, Apr. 2016.
- [9] S. Guerrier and R. Molinari, “On the identifiability of latent models for dependent data,” 2016, *arXiv:1607.05884*. [Online]. Available: <http://arxiv.org/abs/1607.05884>
- [10] I. Komunjer, “Global identification in nonlinear models with moment restrictions,” *Econ. Theory*, vol. 28, no. 4, pp. 719–729, Aug. 2012.
- [11] D. P. Percival, “On estimation of the wavelet variance,” *Biometrika*, vol. 82, no. 3, pp. 619–631, 1995.
- [12] S. Guerrier and R. Molinari, “Fast and robust parametric estimation for time series and spatial models,” 2016, *arXiv:1607.05861*. [Online]. Available: <http://arxiv.org/abs/1607.05861>
- [13] W. K. Newey and D. McFadden, “Large sample estimation and hypothesis testing,” in *Handbook Econometrics*, vol. 4, 1994, pp. 2111–2245.
- [14] A. Serroukh, A. T. Walden, and D. B. Percival, “Statistical properties and uses of the wavelet variance estimator for the scale analysis of time series,” *J. Amer. Stat. Assoc.*, vol. 95, no. 449, pp. 184–196, Mar. 2000.
- [15] D. W. K. Andrews, “Heteroskedasticity and autocorrelation consistent covariance matrix estimation,” *Econometrica*, vol. 59, no. 3, pp. 817–858, May 1991.
- [16] S. Guerrier, R. Molinari, M.-P. Victoria-Feser, and H. Xu, “Robust two-step wavelet-based inference for time series models,” 2020, *arXiv:2001.04214*. [Online]. Available: <https://arxiv.org/abs/2001.04214>
- [17] C. Han and P. C. B. Phillips, “GMM with many moment conditions,” *Econometrica*, vol. 74, no. 1, pp. 147–192, Jan. 2006.
- [18] Sensoror AS. (Apr. 2013). *STIM 300 Inertial Measurement Unit Datasheet*. [Online]. Available: <https://www.sensoror.com/products/inertial-measurement-units/stim300/>
- [19] Intersense. *Navchip Inertial Measurement Unit Datasheet*. Accessed: Feb. 10, 2015. [Online]. Available: <http://www.intersense.com/pages/16/246>
- [20] S. Guerrier, R. Molinari, and J. Skaloud, “Automatic identification and calibration of stochastic parameters in inertial sensors,” *Navigation*, vol. 62, no. 4, pp. 265–272, Dec. 2015.



Stéphane Guerrier received the M.Sc. degree in navigation techniques and geomatics engineering from the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland, in 2008, for his work on the integration of redundant MEMS-IMUs with GPS, and the Ph.D. degree in statistics from the University of Geneva, Geneva, Switzerland, in 2013.

He was a Visiting and Assistant Professor with the University of California, Santa Barbara (UCSB), University of Illinois at Urbana-Champaign (UIUC), and Penn State. He is currently an Assistant Professor of statistics and data science with GSEM, University of Geneva. His research focuses on time series, computational statistics, and model selection.



Juan Jurado received the B.S.E.E. degree from Texas A&M University, College Station, TX, USA, in 2006, the M.S.F.T.E. degree from the U.S. Air Force Test Pilot School, Edwards, CA, USA, in 2013, and the M.S.E.E. and Ph.D. degrees from the Air Force Institute of Technology, Dayton, OH, USA, in 2012 and 2019, respectively.

He is currently the Director of Education with the U.S. Air Force Test Pilot School. Previously, he served as the Director of engineering for the 413th Flight Test Squadron and oversaw various C-130, V-22, and H-1 flight test programs. His research interests include aircraft performance modeling, online sensor calibration, image processing, visual-inertial navigation, and statistical sensor management for multisensor navigation problems.



Mehran Khaghani is currently a Research and Development Engineer with Leica Geosystems, Heerbrugg, Switzerland, working on sensor fusion algorithms for Global Navigation Satellite System (GNSS) and Inertial Navigation System (INS)-based systems. He is a Former Post-Doctoral Researcher with the University of Geneva, Geneva, Switzerland, and EPFL, Lausanne, Switzerland, with his research mainly focused on stochastic error modeling for inertial sensors and vehicle dynamic model (VDM)-based navigation for unmanned aerial vehicles (UAV).



Gaetan Bakalli received the M.Sc. degree in statistics from the University of Geneva, Geneva, Switzerland, in 2016, where he is currently pursuing the Ph.D. degree in statistics.

He worked in various positions in Finance for HSBC and Cargill with the University of Geneva. His research interests are time series, model selection, as well as applied statistics in finance and engineering.



Mucyo Karemera received the Ph.D. degree in mathematics from the University of Geneva, Geneva, Switzerland, in 2016.

He was a Post-Doctoral Researcher with the Department of Statistics, Pennsylvania State University, State College, USA. He is currently a Post-Doctoral Researcher with the Research Center for Statistics, University of Geneva. His research interests include computational statistics, time series, algebra, and topology.



Roberto Molinari received the master's degree in international affairs from the LUISS Guido Carli University, Rome, Italy, in 2007, which brought to experiences in the United Nations Economic Commission for Europe (UNECE), Organization for Economic Cooperation and Development (OECD), and Ernst & Young, and the Ph.D. degree in statistics from the University of California at Santa Barbara, Santa Barbara, CA, USA, in 2016.

He was a Visiting Professor in statistics with the University of California. He was also a Consultant and then a Research Assistant Professor with Penn State University, State College, PA, USA. His research interests are in robust time series model estimation, data privacy, spatial statistics, and model selection as well as applied statistics in the field of economics, finance, and medicine.



Samuel Orso received the Ph.D. degree in statistics from the University of Geneva, Geneva, Switzerland, in 2019.

He is currently a Post-Doctoral Researcher with the Geneva School for Economics and Management, University of Geneva. His research interests include computational statistics, robust estimation, and finite-sample inference.



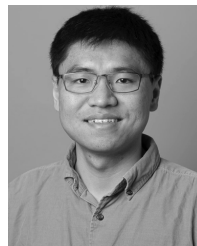
John Raquet is currently the Director of the Dayton Office of Integrated Solutions for Systems (IS4S). Formerly, he was the Director of the Autonomy and Navigation Technology (ANT) Center, Air Force Institute of Technology, Dayton, OH, USA, where he was also a Professor of electrical engineering. He has been developing navigation system technology for more than 30 years.



Christine Schubert is currently a Professor of statistics with the Air Force Institute of Technology (AFIT), Dayton, OH, USA. She has over 20 years of experience working as a Statistician, developing statistical methodologies and conducting analysis across a wide range of government applications. She joined the Faculty at AFIT in 2010, where she researches and consults in the areas of detection and classification, AI, data analytics, navigation, network analysis, and structural health monitoring.



Jan Skaloud is currently the Senior Scientist and a Lecturer with the Swiss Federal Institute of Technology Lausanne (EPFL), Lausanne, Switzerland, and an Adjunct Professor with the Department of Geomatics Engineering, University of Calgary, Calgary, AB, Canada. He directs research in the field of satellite and inertial navigation, sensor calibration, and modern mapping techniques.



Haotian Xu received the M.Sc. degree in statistics from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2015. He is currently pursuing the Ph.D. degree with the Research Center for Statistics, University of Geneva, Geneva, Switzerland.

His research interests are in robust statistics, time series, and high-dimensional statistics.



Yuming Zhang received the B.Sc. degree in mathematics and statistics from the University of Illinois at Urbana-Champaign, Champaign, IL, USA, in 2017. She is currently pursuing the Ph.D. degree in statistics with the University of Geneva, Geneva, Switzerland.

Her research interests include computational statistics, high-dimensional statistics, time series, as well as applied statistics in engineering, medicine, and Earth sciences.