Journal of the American Statistical Association

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/uasa20

Robust Two-Step Wavelet-Based Inference for Time Series Models

Stéphane Guerrier, Roberto Molinari, Maria-Pia Victoria-Feser & Haotian Xu

To cite this article: Stéphane Guerrier, Roberto Molinari, Maria-Pia Victoria-Feser & Haotian Xu (2022) Robust Two-Step Wavelet-Based Inference for Time Series Models, Journal of the American Statistical Association, 117:540, 1996-2013, DOI: 10.1080/01621459.2021.1895176

To link to this article: https://doi.org/10.1080/01621459.2021.1895176

JOURNAL OF THE AMERICAN STATISTICAL ASSOCIATION
2022, VOL. 117, NO. 540, 1996–2013: Theory and Methods
https://doi.org/10.1080/01621459.2021.1895176

Robust Two-Step Wavelet-Based Inference for Time Series Models

Stéphane Guerrier[a], Roberto Molinari[b], Maria-Pia Victoria-Feser[a], and Haotian Xu[a]

[a]University of Geneva, Geneva, Switzerland; [b]Auburn University, Auburn, AL

ABSTRACT

Latent time series models such as (the independent sum of) ARMA(p, q) models with additional stochastic processes are increasingly used for data analysis in biology, ecology, engineering, and economics. Inference on and/or prediction from these models can be highly challenging: (i) the data may contain outliers that can adversely affect the estimation procedure; (ii) the computational complexity can become prohibitive when the time series are extremely large; (iii) model selection adds another layer of (computational) complexity; and (iv) solutions that address (i), (ii), and (iii) simultaneously do not exist in practice. This paper aims at jointly addressing these challenges by proposing a general framework for robust two-step estimation based on a bounded influence M-estimator of the wavelet variance. We first develop the conditions for the joint asymptotic normality of the latter estimator thereby providing the necessary tools to perform (direct) inference for scale-based analysis of signals. Taking advantage of the model-independent weights of this first-step estimator, we then develop the asymptotic properties of two-step robust estimators using the framework of the generalized method of wavelet moments (GMWM). Simulation studies illustrate the good finite sample performance of the robust GMWM estimator and applied examples highlight the practical relevance of the proposed approach.

ARTICLE HISTORY
Received November 2019
Accepted February 2021

KEYWORDS
Generalized method of wavelet moments; Large-scale time series; Scale-based analysis of variance; Signal processing; State-space models; Wavelet variance

1. Introduction

As for many other fields of statistical research, time series analysis is also facing different challenges due to the increasing amounts of data being recorded over time within a wide variety of contexts going from biology and ecology to finance and engineering. Among others, these challenges include (i) the need for more complex (parametric) models (e.g., to deal with possible long-term non-stationary features, see below), (ii) the need for computationally efficient (or simply feasible) methods to analyze data and estimate such models as well as (iii) the need to deal with the increased probability of observing measurement errors in the form of contamination (outliers), model deviations, etc. In this context, there is currently a large variety of models and statistical methods available to analyze and draw conclusions from time series (see, e.g., Percival and Walden 2006; Durbin and Koopman 2012; Shumway and Stoffer 2013, for an overview). However, many of the existing methods may become unfeasible, for example, when estimating even slightly complex models on large time series data. Examples of such complex models are those that are characterized by (the independent sum of) ARMA and rounding error models, which we refer to as *latent* time series models, which therefore include a wide range of state-space models. Indeed, when dealing with moderately large sample sizes and a relatively large number of latent components, including models with long-term non-stationary features (e.g., intrinsically stationary models such as drifts and random walks), estimation and inference can become very challenging. In addition, the use of robust methods to perform inference when the data suffers from contamination is often a daunting task even for relatively simple (parametric) settings and moderate sample sizes, without considering more complex models in larger data settings (for motivating examples and robust inferential approaches, see, e.g., Maronna, Martin, and Yohai 2006; Maronna et al. 2019).

In response to the above challenges and limitations, this article proposes an alternative and general robust inference framework for (complex) parametric time series models. This framework is based on a two-step approach where the first (and most important) step consists in the proposal of a robust estimator of the Wavelet Variance (WV) (see, e.g., Percival 1995) with adequate asymptotic properties based on minimal conditions, while the second step integrates these results within the generalized method of wavelet moments (GMWM) framework (see Guerrier et al. 2013) for the purpose of inference on time series models.[1] With regard to the first step, the WV has been widely used within the natural and physical sciences for analysis of variance, model building and prediction (see, e.g., Percival and Walden 2006, for an overview), and more recently within other fields of research (see, e.g., Gallegati 2012; Xie and Krishnan 2013; Foufoula-Georgiou and Kumar 2014; Jia et al. 2015; Ziaja et al. 2016; Abry et al. 2018, to mention a few). In this context, the properties of the standard estimators of WV have been

[1]The method (with relative graphical tools) is implemented partly in the `simts` R package on CRAN and in a more complete manner in an open source R package, available at *github.com/SMAC-Group/gmwm*; see also Clausen et al. (2018); Bakalli et al. (2018); Radi et al. (2019)

CONTACT Maria-Pia Victoria-Feser *maria-pia.victoriafeser@unige.ch* Research Center for Statistics, GSEM, University of Geneva, Uni Mail, 40 Boulevard du Pont-d'Arve, 1211 Geneva, Switzerland.

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

developed (see Percival 1995; Serroukh, Walden, and Percival 2000) while their limitations in presence of contamination or outliers have been underlined in Mondal and Percival (2012a) where a robust estimator was also put forward for this purpose. However, as explained further on, the use of an alternative M-estimator is preferable in different settings, including the two-step framework considered in the present paper.

With respect to parametric inference for time series, robust methods (including two-step approaches) have been proposed in abundance over the past decades and a detailed overview of these can be found, for example, in Maronna, Martin, and Yohai (2006), Maronna et al. (2019, chap. 8) while Appendix A also provides a short literature review. More specifically, a general approach in this context is to maximize *asymptotic* efficiency of the resulting estimators with respect to a specific model (or class of models). For this purpose, a traditional approach consists in placing estimation (and inference) in the framework of bounded-influence estimators that are obtained by bounding the corresponding estimating equations (Hampel et al. 1986), such as the maximum likelihood estimator (MLE) equations or indeed those of other efficient (nonrobust) estimators. In this setting, a correction factor is often required to ensure consistency of the resulting robust estimator and needs to be defined for each model considered for a given time series. The computation of this model-dependent correction factor generally requires numerical approximations of possibly high-dimensional integrals or multiple one-dimensional integrals (for conditionally unbiased estimators, see Künsch 1984a), whose dimensions or number grow proportionally to the sample size. These consistency corrections can therefore not only be too burdensome to compute, relying on more computationally intensive simulation-based procedures to approximate these quantities, but also are rarely accounted for in the derivation of the statistical properties of the resulting estimator. On the other hand, robust two-step estimators (based for example on robust filtering methods or on robust autocovariance/autocorrelation estimators), while possibly paying a price in terms of asymptotic efficiency, deliver not only computational advantages but also allow to compare candidate models under a different perspective. Indeed, the first-step estimators provide a set of estimates that remain common to all candidate models and are resistant to outliers through weights (for weighted estimators) that are computed *only once* independently from the candidate models. Using these common estimates, aside from guaranteeing that the second step estimators inherit their robustness property, the consequence is that the model comparison in terms of fit (or prediction accuracy) is made solely on the structural features of the candidate models. This provides an alternative approach to model comparison based on robust estimators since, in the weighted estimating equation setting, the estimation weights are relative to each model under consideration that is, by assumption, the correct one. On the other hand, a two-step approach alleviates the latter hypothesis in the computation of the robust weights and hence provides an alternative route to robust model comparison. Moreover, this allows to develop graphical robust model selection tools which, in the context of this paper, rely only on a few quantities (even for large samples as seen further on and in Section 5) as well as to build robust model selection criteria in a relatively straightforward manner.

However, while robust two-step approaches have noticeable computational and model comparison advantages, these can nevertheless suffer from some theoretical and practical drawbacks. Indeed, when they are based on robust filtering, they may lead to biased estimators and lack asymptotic theory in order to perform adequate inference (see, e.g., Muler, Peña, and Yohai 2009). Alternatively, in the case where two-step estimators are based on robust moment estimators (e.g., autocovariance or autocorrelation), these are often computed in the framework of indirect inference (Smith 1993; Gourieroux, Monfort, and Renault 1993) or, similarly, in that of the (simulated) method of moments (McFadden 1989; Gallant and Tauchen 1996; Duffie and Singleton 1993). In these settings, two-step estimators can become computationally intensive in large samples since the number of auxiliary moments that are available in practice increase linearly with the sample size, thereby requiring a moment selection procedure which delivers additional uncertainty in the successive inference phase (see, e.g., Andrews 1999). However, the GMWM framework considerably reduces the latter drawback since its first-step moment is the WV which adequately summarizes the information in the spectral density (or autocovariance) function into $J < \log_2(T)$ moments (with T representing the sample size) thereby greatly reducing the number of auxiliary moments, consisting in the WV scales, even for large sample sizes while also allowing for inference on intrinsically stationary models. Being a natural extension of the GMWM framework, the Robust GMWM (RGMWM) relies on the proposed robust estimator of WV and makes use of its properties in order to derive the inferential properties of the RGMWM so as to perform adequate estimation and general inference procedures for a wide range of time series models under slight model deviations (such as the presence of outliers) in a computationally efficient manner.

Considering the above, the paper is organized as follows. Section 2 proposes the robust WV estimator that constitutes the first step of the RGMWM that is presented in Section 3. In both sections, the necessary asymptotic results are developed in order to perform robust estimation and inference for intrinsically stationary time series. Section 4 presents a simulation study where the robustness properties of the RGMWM estimator are compared with other estimators and Section 5 demonstrates the practical usefulness of the proposed method in an economic setting while another applied setting is presented in supplemental material E.

2. Robust WV

To introduce the WV, let us denote $(X_t \in \mathbb{R})$, $t = 1, \ldots, T$, as an *intrinsically* stationary time series that is either stationary or becomes so when a dth-order backward difference is applied to it. Moreover, let

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

denote the wavelet coefficients that result from a wavelet decomposition of the time series, where $(h_{j,t})$ is a known wavelet filter of length L_j at scale τ_j (commonly, dyadic scale $\tau_j = 2^j$), for $j = 1, \ldots, J$ and $J < \log_2(T)$. Since the filter sequence

$(h_{j,t})$ has specific properties, the wavelet coefficients can be seen as the result of a particular form of weighted moving-average taken over a number of observations proportional to L_j (which increases as j increases, see, for example, Percival and Walden 2006, for a general overview). Moreover, if $L_1 \geq$ d and (X_t) is a dth-order stationary time series, the resulting wavelet coefficients will be stationary. Based on the (stationary) wavelet coefficients, for $j = 1, \ldots, J$, the WV is defined as

$$\nu_j^2 := \text{var}(W_{j,t}) > 0 ,$$

that is, the variance of the wavelet coefficients, which can be expressed in vector form as $\boldsymbol{\nu} := [\nu_j^2]_{j=1,\ldots,J}$. Throughout this work, we require $\nu_j^2 > 0$, that is, the wavelet coefficients are nondegenerate, since we always normalize with ν_j^2. Moreover, we always consider $\nu_j = |\nu_j|$, that is, we only consider the positive root. While this definition holds independently from a possible parametric model underlying the time series (X_t), for the purpose of this work let us assume that the latter is generated from the parametric family of models $F_{\boldsymbol{\theta}_0}$, with unknown parameter vector denoted by $\boldsymbol{\theta}_0 \in \boldsymbol{\Theta} \subset \mathbb{R}^p$. In order to make the link between the WV and an assumed *stationary* parametric model $F_{\boldsymbol{\theta}}$ explicit, an alternative representation of this quantity is based on the Spectral Density Function (SDF) and is as follows:

$$\nu_j^2(\boldsymbol{\theta}) = \int_{-1/2}^{1/2} |H_j(f)|^2 S_{F_{\boldsymbol{\theta}}}(f) df ,$$

with $S_{F_{\boldsymbol{\theta}}}(f)$ denoting the theoretical SDF and $H_j(f)$ being the Fourier transform of the wavelet filters $(h_{j,t})$. The above equality is valid also when considering the theoretical spectral density of the wavelet coefficients themselves which, depending on the length of the wavelet filter, is defined even when the original time series (X_t) is intrinsically stationary (see, e.g., Percival and Walden 2006). Although the results of the present section hold without any parametric assumption for (X_t), in this paper we will generally assume that the true WV will depend on a parametric family of models $F_{\boldsymbol{\theta}}$. Therefore, we will use the notation $\boldsymbol{\nu}$ for the true WV (making its link with $F_{\boldsymbol{\theta}}$ implicit) while we will make use of the notation $\boldsymbol{\nu}(\boldsymbol{\theta})$ whenever it is appropriate to explicitly highlight the link with the underlying parametric model.

Based on the above definitions, each parametric time series model has a corresponding theoretical WV vector which summarizes the information contained in the SDF or AutoCovariance Function (ACF) into few quantities ($J < \log_2(T)$) for a wide range of intrinsically stationary time series models (e.g., SARIMA and many state-space models). Considering this, when compared to the SDF or ACF, the information loss of the WV can be offset by its ability to adequately represent a considerably wide range of intrinsically stationary time series. In this sense, aside from being discussed in Greenhall (1998) and Guerrier et al. (2013), supplementary material B includes additional results discussing its capability of adequately recovering the parameters for many time series models thereby supporting the usefulness of this quantity.

For the rest of this section, however, let us disregard any parametric assumption for (X_t) and simply assume that its corresponding wavelet coefficients at each level are stationary. With this in mind, several estimators have been proposed for the WV, the main one being the standard unbiased estimator of the WV proposed by Percival (1995) and defined as

$$\tilde{\nu}_j^2 := \frac{1}{M_j} \sum_{t=1}^{M_j} W_{j,t}^2 , \tag{1}$$

where M_j is the length of the wavelet coefficient process $(W_{j,t})$ at scale τ_j. The theoretical properties of this estimator were further studied in Serroukh, Walden, and Percival (2000) in which the conditions for its asymptotic properties are given. However, as highlighted by Mondal and Percival (2012a), the estimator of WV in (1) is not robust in the sense that it is considerably biased in the presence of outliers or different forms of data contamination. For this reason they put forward a robust estimator of the WV (developed for Gaussian time series specifically affected by scale-based contamination) by making use of a log-transformation of the (squared) wavelet coefficients to apply standard M-estimation theory for location parameters, with a chosen bounded estimating function. However, due to these transformations and the approximate corrections to reverse them, the asymptotic properties of the final estimator are also approximate. Therefore, considering the robustness issues of the standard estimator $\tilde{\nu}_j^2$ and the "limitations" of the latter robust estimator, the following paragraphs put forward a new M-estimator of WV that overcomes these issues and, as shown in Appendix C, performs generally better in finite sample settings.

2.1. M-Estimation of WV

Following the above discussion, this section generalizes the standard estimator of WV proposed in Percival (1995) to an M-estimator (Huber 1964) which can also be made robust by choosing a bounded score function, thereby delivering an appropriate framework for inference on this quantity. Let us therefore re-express (1) as an M-estimator which, using "argzero" to define the solution allowing a function to be equal to zero, is defined as follows:

$$\hat{\nu}_j^2 := \underset{\nu_j^2 \in \mathbf{N}}{\text{argzero}} \sum_{t=1}^{M_j} \psi(W_{j,t}, \nu_j^2), \tag{2}$$

where $\mathbf{N} \subset \mathbb{R}^+$ and $\psi(\cdot)$ is a score function (ψ-function) which can be unbounded or bounded with respect to $(W_{j,t})$. For bounded ψ-functions, popular choices include Huber's function and Tukey's Biweight function (see, e.g., Hampel et al. 1986). In order to determine the *infinitisemal* robustness of an M-estimator we must study its Influence Function (IF) and, based on standard properties of M-estimators in dependent data settings (see, e.g., Künsch 1984b), the following proposition states the sufficient conditions under which the WV estimator has a bounded IF and is therefore robust.

Proposition 2.1. Assuming that $(W_{j,t})$ is a strictly stationary process, the IF of the estimator of WV is bounded if $\psi(\cdot)$ is bounded.

The proof of this proposition can be found in supplementary material A.1 along with the definition of the IF.

Remark 2.1. The property of a bounded IF ensures "infinitesimal" robustness for *any* type of contamination. However, depending on the chosen bounded function, the resulting estimator may have a low breakdown point, which represents a different measure of robustness (see, e.g., Genton and Lucas 2003). Depending on the type of outliers, especially the ones that introduce a persistent effect (e.g., level-shift outliers), a high breakdown estimator would be preferred (see, e.g., Maronna, Martin, and Yohai 2006).

Given the intuitive result of Proposition 2.1, we therefore intend to deliver an M-estimator of the form in (2) whose ψ-function can be bounded and directly estimates the quantity of interest ν_j^2. The proposed approach is based on "Huber's Proposal 2" which was presented in Huber (1981) and was aimed at the estimation of the scale parameter of the residuals in the linear regression framework. Without loss of generality, we assume that $\mathbb{E}[W_{j,t}] = 0$ and consequently use this proposal by defining $r_{j,t} := W_{j,t}/\nu_j$ as the standardized wavelet coefficients, thereby defining the proposed estimator as

$$\hat{\nu}_j^2 := \underset{\nu_j^2 \in \mathbf{N}}{\text{argzero}} \left[\frac{1}{M_j} \sum_{t=1}^{M_j} \omega^2 \left(r_{j,t}; \nu_j^2, c \right) r_{j,t}^2 - a(\nu_j^2, c) \right], \quad (3)$$

where $\omega(\cdot)$ represents the weight function implied by the chosen ψ-function and $a(\nu_j^2, c)$ is a correction term to ensure Fisher consistency at the marginal distribution of the wavelet coefficients $(W_{j,t})$ (i.e. the true WV would be obtained if applying the estimator to the population from which $(W_{j,t})$ is issued). Indeed, this correction term is defined as

$$a(\nu_j^2, c) := \mathbb{E}[\omega^2 \left(r_{j,t}; \nu_j^2, c \right) r_{j,t}^2],$$

where the latter expectation is taken over the marginal distribution of $(W_{j,t})$ with variance ν_j^2. It can be noticed that, if the tuning constant $c \to \infty$, we have that $\omega(\cdot) \to 1$ and $a(\nu_j^2, c) \to 1$ such that $\hat{\nu}_j^2$ is the solution for ν_j^2 in the following estimating equation:

$$\frac{1}{M_j} \sum_{t=1}^{M_j} \frac{W_{j,t}^2}{\nu_j^2} - 1 = 0,$$

thereby delivering the estimator in (1). This property implies that the tuning constant c can be chosen to regulate the trade-off between robustness and efficiency of the resulting estimator. A discussion about the choice of the constant c can be found in supplemental material A.2 which highlights how this constant can be implicitly chosen by defining the level of statistical efficiency required with respect to the standard estimator. As mentioned, the term $a(\nu_j^2, c)$ depends on the marginal distribution of the stationary process which is assumed for the wavelet coefficients $(W_{j,t})$ and on the specific weight function $\omega(\cdot)$. The exact analytic form of this term therefore may be complicated to derive when considering distributions other than the Gaussian or other symmetric distributions.

Remark 2.2. In the case where the wavelet coefficients are assumed to come from a Gaussian distribution, the correction term $a(\nu_j^2, c)$ can be expressed as $a_\psi(c)$ since it only depends on the value of the tuning constant c, and can be found explicitly using the results of Dhrymes (2005). This is the case when the time series is itself Gaussian (see, e.g., Percival 2016, for a discussion) or can be assumed as an approximation given the averaging nature of the wavelet filter. On the other hand, if the marginal distribution of the wavelet coefficients is non-Gaussian, the term $a(\nu_j^2, c)$ could eventually be numerically approximated once (based on the standardized simulated values from the assumed distribution with known or estimated parameters) and used for all scales τ_j and would need to be accounted for in the subsequent inference phase.

Having defined the M-estimator of WV in (3), let us now list a set of conditions which allows us to derive the asymptotic properties of this estimator. Firstly, letting $\nu_{j,0}^2$ represent the true WV at scale τ_j, we define the following conditions for the ψ-function:

(C1) The Bouligand derivative of $\psi(\cdot)$ is continuous almost everywhere.

(C2) $\mathbb{E}[\psi(W_{j,t}, \nu_j^2)] = 0$ if and only if $\nu_j^2 = \nu_{j,0}^2$.

The first condition is a technical requirement that allows us to perform an expansion in order to represent $\hat{\nu}_j^2$ in an explicit form (in addition to the Tukey Biweight function, for example, the Huber ψ-function also respects this condition, under a stationary Gaussian assumption, as shown in Lemma B.1 in Appendix B where the definition of Bouligand derivative is also given). On the other hand, Condition (C2) requires the true WV to be identifiable through the chosen ψ-function. This condition is verified when choosing the Huber and Tukey Biweight ψ-functions and assuming the wavelet coefficient process $(W_{j,t})$ is stationary and Gaussian as shown in supplementary material A.3.

Given these technical conditions on the properties of the ψ-function, we can now study the process-related conditions for which we define the filtration $\mathcal{F}_t := (\dots, \epsilon_{t-1}, \epsilon_t)$ where ϵ_t are iid random variables. With this definition we can now deliver the first process-related condition.

(C3) $(W_{j,t})$ is a strictly stationary process and can be represented as

$$W_{j,t} = g_j(\mathcal{F}_t),$$

where $g_j(\cdot)$ is an $\mathbb{R}$-valued measurable function such that $W_{j,t}$ is well defined.

Although not necessarily expressed in this form, Condition (C3) is commonly assumed when studying asymptotics within a time series setting and is respected by a very general class of time series models (see Wu 2005). In our case, this condition is necessary in order to apply the functional dependence measure defined in Wu (2011).

At this stage, we further denote $W_{j,t}^\star = g_j(\mathcal{F}_t^\star)$ as being the coupled version of $W_{j,t}$, where $\mathcal{F}_t^\star := (\dots, \epsilon_{-1}, \epsilon_0^\star, \epsilon_1, \dots, \epsilon_t)$, with ϵ_0 and $\epsilon_0^\star$ also being iid random variables. Hence, the two processes $(W_{j,t})$ and $(W_{j,t}^\star)$ depend on filtrations that only differ by one element, that is, ϵ_0 and $\epsilon_0^\star$, therefore implying that $W_{j,t}^\star = W_{j,t}$ for $t < 0$. Based on this definition, we can define the functional dependence measure given in Wu (2011)

$$\delta^j_{t,q} := \| W_{j,t} - W^\star_{j,t} \|_q,$$

where $\|Z\|_q := (\mathbb{E}[|Z|^q])^{1/q}$ for $q > 0$. This dependence measure can be interpreted as the expected impact of the innovation $\epsilon^\star_0$ on the moments of $W^\star_{j,t}$ with respect to its "original" path given by $W_{j,t}$. Using this definition, we provide the final process-related condition.

(C4) The process $(W_{j,t})$ is such that $\sum_{t=0}^{\infty} \delta^j_{t,4} < \infty$ for $j = 1, \dots, J$.

This condition can be interpreted as a requirement for the difference in fourth moments between $(W_{j,t})$ and its coupled version $(W^\star_{j,t})$ to be summable with respect to t, implying that the innovation $\epsilon^\star_0$ has a limited impact in time on how much $W^\star_{j,t}$ deviates from $W_{j,t}$ and, hence, the process $(W_{j,t})$ is a "stable" process (see Wu 2011). An example of how Conditions (C3) and (C4) are verified using the Haar wavelet filter in the case where (X_t) follows a causal ARMA process is given in Supplementary Material A.4. It should finally be noted that Condition (C4), which places constraints up to the fourth moments of the process, could eventually be relaxed by using results such as those presented in Fonseca, Mondal, and Zhang (2019) but this investigation is left for future work.

Remark 2.3. In the case where the chosen filter for the wavelet decomposition belongs to the Daubechies family, Conditions (C3) and (C4) can be placed directly on the dth-order difference of the process (X_t), that we will denote as Δ_t, instead of the process $(W_{j,t})$. The advantage of this consists in the fact that, if the length of the wavelet filter at the first level is such that $L_1 \geq d$ (where d is the required differencing such that (Δ_t) is stationary), then we only need (Δ_t) to respect these conditions to ensure that all levels of decomposition respect them too. Indeed, in the case of a Daubechies wavelet filter with N vanishing moments, all levels of wavelet coefficients simply correspond to a deterministic linear combination of (Δ_t) and we would therefore have $g_j(\cdot) = \gamma_j g(\cdot)$ and $\delta^j_{t,q} = \lambda_j \delta_{t,q}$ where $g(\cdot)$ and $\delta_{t,q}$ would be uniquely related to Δ_t and constants γ_j and λ_j only depend on the chosen wavelet filter. For example, the process $(W_{j,t})$ in Conditions (C3) and (C4) would be replaced by the process $(\Delta_t = X_t - X_{t-1})$ in the case of the Haar wavelet filter and hence we would have $\Delta_t = g(\mathcal{F}_t)$ and $\delta_{t,q} = \|\Delta_t - \Delta^\star_t\|_q$. If other families of wavelet filters were to be considered, then the verification of these conditions may be less straightforward but can generally be assumed to hold if the original process directly respects them.

We can now determine the asymptotic properties of the proposed M-estimator of WV in (3). For this reason, let us further define $\boldsymbol{W}_t := [W_{j,t}]_{j=1,\dots,J}$ as the vector of wavelet coefficients at time t and $\hat{\boldsymbol{\nu}} := [\hat{\nu}^2_j]_{j=1,\dots,J}$ as the vector of estimated WV using the proposed M-estimator in (3). Moreover, we define the projection operator (see Wu 2011) as

$$\mathcal{P}_t \cdot := \mathbb{E}[\cdot|\mathcal{F}_t] - \mathbb{E}[\cdot|\mathcal{F}_{t-1}], \quad t \in \mathbb{Z}.$$

This operator therefore represents a measure of how much the conditional expectation of a process can change once the immediately previous information is removed. As for the previously defined functional dependence measure, intuitively the projection operator should not be too sensitive if the underlying process is stable (in the sense of Wu 2011). It must be underlined that this operator can be applied to vectors: considering for example $\boldsymbol{X}_t \in \mathbb{R}^d$ where each dimension follows Condition (C3) and with $q \geq 2$, we have $\|\mathcal{P}_0 \boldsymbol{X}_t\|_q \leq \|\boldsymbol{X}_t - \boldsymbol{X}^*_t\|_q$.

Using the above definitions, we finally define the quantities $\boldsymbol{D}_0 := \sum_{t=0}^{\infty} \mathcal{P}_0 \psi(\boldsymbol{W}_t, \boldsymbol{\nu})$ and $\boldsymbol{M} := \mathbb{E}[-\partial/\partial\boldsymbol{\nu}\, \psi(\boldsymbol{W}_t, \boldsymbol{\nu})]$ to deliver the following theorem on the asymptotic distribution of the proposed estimator $\hat{\boldsymbol{\nu}}$.

Theorem 2.1. Under Conditions (C1) to (C4) and assuming that the function $\psi(\cdot)$ is bounded, we have that the M-estimator $\hat{\boldsymbol{\nu}} := [\hat{\nu}^2_j]_{j=1,\dots,J}$ has the following asymptotic distribution:

$$\sqrt{T}\left(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{V}\right),$$

where $\boldsymbol{\nu} := [\nu^2_j]_{j=1,\dots,J}$ and $\boldsymbol{V} = \boldsymbol{M}\,\mathbb{E}[\boldsymbol{D}_0 \boldsymbol{D}_0^\top]\,\boldsymbol{M}^\top$.

The proof of this theorem can be found in Appendix B where the proofs for Condition (C2) (for Huber and Tukey Biweight functions in the Gaussian setting) and consistency of $\hat{\nu}^2_j$ can also be found. An extension of this result beyond intrinsically stationary processes, such as for example to fractional stochastic processes (see, e.g., Percival and Walden 2000), can be considered using the uniform functional dependence measure suggested in Zhou and Wu (2009) and Wu and Zhou (2011) but is left for future research.

Remark 2.4. The asymptotic covariance matrix $\boldsymbol{V}$ is a long-run covariance matrix which can be estimated via different methods. For example, the moving block bootstrap can be used applying the robust estimator $\hat{\boldsymbol{\nu}}$ in (3) to the bootstrapped samples thereby ensuring that the influence of potential outliers is limited. Alternatively, $\hat{\boldsymbol{V}}$ can be chosen as being the empirical version of $\boldsymbol{V}$ given in Theorem 2.1 which can be robustly estimated by the robust version of batched-mean estimator (see Zhang and Wu 2017) or the progressive batched-mean method generalized from the idea in Kim, Lahiri, and Nordman (2013). Moreover, when a parametric family $F_{\boldsymbol{\theta}}$ for (X_t) is assumed (as in the setting of this work), the parametric bootstrap could also be considered.

Aside from studying the behavior of the proposed M-estimator as a first-step estimator for the RGMWM (Section 4), as an additional exercise in Appendix C we compare the behaviour of $\hat{\nu}^2_j$ with the standard WV estimator $\tilde{\nu}^2_j$ and the median-type robust estimator in Mondal and Percival (2012b). In the latter simulation, it appears clearly that the proposed M-estimator is the best alternative to the standard estimator in the uncontaminated setting and the best overall in the contaminated setting. Based on the robustness properties of the proposed estimator $\hat{\boldsymbol{\nu}}$ (for bounded ψ-functions) and its asymptotic properties, this new estimator provides a suitable tool to perform robust scale-based analysis of variance for time series (see, e.g., Percival and Walden 2006, and references therein). More importantly, it delivers a first-step estimator with adequate properties based on which it is possible to perform robust parametric inference for a wide range of time series models as well as for large datasets as discussed in the following section.

3. Robust GMWM

The properties of the proposed M-estimator of WV can be directly carried over to the GMWM framework (see Guerrier et al. 2013). Indeed, as suggested in Guerrier, Molinari, and Victoria-Feser (2014), one can replace the standard estimator used in the GMWM with a robust estimator which, in this case, is the proposed M-estimator allowing us to deliver the RGMWM defined as

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\, (\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}))^{\top}\boldsymbol{\Omega}(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})), \tag{4}$$

where $\boldsymbol{\nu}(\boldsymbol{\theta}) := [\nu_j^2(\boldsymbol{\theta})]_{j=1,\dots,J}$ and $\boldsymbol{\Omega}$ is any symmetric positive-definite weighting matrix (for example, one could choose a non-random matrix $\boldsymbol{\Omega}_0$ or choose an estimator $\hat{\boldsymbol{\Omega}}$ for it, for example, $\boldsymbol{\Omega} := \hat{\boldsymbol{\Omega}} = \hat{\boldsymbol{V}}^{-1}$, where $\hat{\boldsymbol{V}}$ is a suitable estimator of $\boldsymbol{V}$, see Guerrier et al. 2013). A generally reasonable choice for $\hat{\boldsymbol{\Omega}}$ is to consider a diagonal matrix with elements equal to the diagonal elements of $\hat{\boldsymbol{V}}^{-1}$, where $\hat{\boldsymbol{V}}$ is obtained using one of the methods proposed in Remark 2.4. Moreover, the robustness (bounded IF) of the RGMWM estimator is inherited from the robustness of $\hat{\boldsymbol{\nu}}$ as shown in Genton and Ronchetti (2003) in the indirect inference framework.

With this in mind, in the following paragraphs we list and discuss the conditions for the consistency and asymptotic normality of the RGMWM which summarize and reduce those in Guerrier et al. (2013) for the standard GMWM. Denoting $||\cdot||_S$ as the matrix spectral norm, the conditions are as follows:

(C5) $\boldsymbol{\Theta}$ is compact.

(C6) $\boldsymbol{\nu}(\boldsymbol{\theta})$ is continuous and differentiable for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}$.

(C7) For $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$, $\boldsymbol{\nu}(\boldsymbol{\theta}_1) = \boldsymbol{\nu}(\boldsymbol{\theta}_2)$ if and only if $\boldsymbol{\theta}_1 = \boldsymbol{\theta}_2$.

(C8) $||\hat{\boldsymbol{\Omega}} - \boldsymbol{\Omega}_0||_S \xrightarrow{p} 0$.

Condition (C5) is commonly assumed but could be replaced by imposing other technical constraints if deemed more appropriate with respect to the parametric setting of reference (as proposed, for example, in Theorem 2.7 of Newey and McFadden 1994). Condition (C6) is easy to verify and is respected for most intrinsically stationary processes. However, Condition (C7) is an essential one which is often hard to verify. In this case, with respect to the Haar wavelet filter (which is one of the most commonly used wavelet filters), the discussion in Guerrier et al. (2013) and the results in Greenhall (1998) support the identifiability of a large class of (latent) time series models and are extended in supplementary material B as well as in Guerrier and Molinari (2016) thereby suggesting that Condition (C7) can hold in various settings. Finally, Condition (C8) addresses the choice of the weighting matrix $\boldsymbol{\Omega}$ and places conditions on an estimator, denoted $\hat{\boldsymbol{\Omega}}$, if this were chosen to define this matrix. In this perspective, the RGMWM is consistent for any matrix $\boldsymbol{\Omega}$ that is symmetric positive definite and, therefore, one needs to select an estimator $\hat{\boldsymbol{\Omega}}$ that converges to the non-random symmetric positive-definite matrix $\boldsymbol{\Omega}_0$ chosen for $\boldsymbol{\Omega}$. A final condition which has not been stated is the consistency of the WV estimator $\hat{\boldsymbol{\nu}}$ which is implied by Conditions (C2) to (C4) as seen in Section 2. With these conditions we can now state the consistency of the RGMWM estimator $\hat{\boldsymbol{\theta}}$.

Proposition 3.1. Under Conditions (C2) to (C8) we have that $\hat{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}_0$.

The proof of this proposition can be found in supplementary material C.1. With this result, we can finally give the conditions for the asymptotic normality of $\hat{\boldsymbol{\theta}}$. For this reason, let us define

$$\boldsymbol{A}(\boldsymbol{\theta}_0) := \frac{\partial}{\partial \boldsymbol{\theta}^{\top}}\, \boldsymbol{\nu}(\boldsymbol{\theta})\Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0},$$

which, if using a Haar wavelet filter for example, exists for a wide class of time series models (see, e.g., Zhang 2008). Using this definition, we can state these final conditions.

(C9) $\boldsymbol{\theta}_0 \in \operatorname{Int}(\boldsymbol{\Theta})$.

(C10) $\mathbf{H}(\boldsymbol{\theta}_0) := \boldsymbol{A}(\boldsymbol{\theta}_0)^{\top}\boldsymbol{\Omega}\boldsymbol{A}(\boldsymbol{\theta}_0)$ exists and is non-singular.

Condition (C9) is a standard regularity condition while Condition (C10) is also usually assumed since it depends on the specific parametric model $F_{\boldsymbol{\theta}}$ from which the time series (X_t) is generated and cannot therefore be verified in general. Since $\boldsymbol{\Omega}$ is non-singular by definition, we have that this condition relies mainly on the non-singularity of (the first p rows of) $\mathbf{A}(\boldsymbol{\theta}_0)$ which is used, for example, to discuss Condition (C7) in supplementary material B. Finally, as for the results of consistency in Proposition 3.1, an additional condition for the asymptotic normality of the RGMWM is the asymptotic normality of $\hat{\boldsymbol{\nu}}$ which is stated in Theorem 2.1. Having discussed these conditions, we can use them to state the following lemma.

Lemma 3.1. Under Conditions (C1) to (C10), the estimator $\hat{\boldsymbol{\theta}}$ has the following asymptotic distribution:

$$\sqrt{T}\left(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \mathbf{BVB}^{\top}\right),$$

where $\mathbf{B} = \mathbf{H}(\boldsymbol{\theta}_0)^{-1}\mathbf{A}(\boldsymbol{\theta}_0)^{\top}\boldsymbol{\Omega}$.

The proof of Lemma 3.1 is provided in supplemental material C.2. With the above results on the asymptotic properties of the RGMWM, the next paragraphs discuss some practical and theoretical advantages of the proposed robust framework.

3.1. Discussion: Practical Properties and Extensions

The RGMWM delivers various advantages that are mainly due to its two-step nature which allows it to benefit from the generality of the M-estimation framework presented in Section 2.1. A first advantage resides in the fact that the correction term needed to ensure Fisher consistency of the estimator defined in (3) only depends on the marginal distribution of the wavelet coefficients $(W_{j,t})$ which, if assumed to be Gaussian, can have an explicit form (for common ψ-functions) and only needs to be computed once for all levels j. A main advantage however resides in the fact that the dimension of the auxiliary parameter vector is always reasonable since in general $J < \log_2(T)$ which allows to make use of all the scales of WV (the auxiliary moments for the GMWM) without the need to select specific moments. This is not the case, for example, for Generalized Method of Moments (GMM) estimators where auxiliary moment-selection is an important issue since, according to the model that is being estimated, the choice should fall on all moments (which can be highly impractical) or on moments that are more "informative" than others (see, e.g., Andrews 1999). The RGMWM on the other hand can make use of all the possible scales of WV even for extremely large sample sizes, allowing it to preserve its statistical efficiency while gaining in terms of computational speed which is approximately of order $\mathcal{O}(T\log_2(T))$ while for the MLE, for example, it is roughly $\mathcal{O}(T^3)$. Supplementary material D shows results on the computational time required to estimate the parameters of some (latent) models for sample sizes up to 10 million, confirming the considerable computational advantage of the RGMWM over both standard and robust alternatives. Moreover, the wavelet decomposition (and consequent variance estimation) is computationally efficient based on well-known algorithms (see, e.g., Rioul and Duhamel 1992) and more recent approaches (see, e.g., Stocchi and Marchesi 2018) allowing the RGMWM to be scalable. Nevertheless, a possible limitation of this approach is that it requires a large enough sample size to estimate more complex models although, for example, it can already estimate four-parameter models with a sample size of $T = 20$ if using a Haar wavelet filter (without however claiming that such an estimate would be highly accurate).

For model comparison purposes, using its "model-independent" nature based on the robust weights of the proposed M-estimator in (3), the latter allows to graphically compare potential candidate models on the basis of the estimated WV as is routinely done, for example, with error characterization in the field of signal processing (see, e.g., El-Sheimy, Hou, and Niu 2008). Indeed, decreasing linear trends in the log–log plot of the WV can indicate the presence of white noise or rounding-error models while increasing linear trends can indicate the presence of non-stationary components such as drifts and random walks whereas slight "bumps" in the plot can indicate the presence of ARMA components. In Section 5, for example, the graphical display of the WV is used to detect and check the fit of the model when analysing a real dataset on personal saving rates. Moreover, continuing with its model comparison advantages, the RGMWM estimator also delivers a general framework for robust goodness-of-fit tests and model selection. Indeed, the objective function in (4) can be used as a statistic for a goodness-of-fit test (Sargan-Hansen test or J-test) as proposed by Hansen (1982), where the asymptotic distribution under the null hypothesis is chi-squared with $J - p$ degrees of freedom. Moreover, model selection criteria can also be built based on the (penalized) GMM objective function which in the RGMWM setting would be given by

$$T(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}_k))^\top \hat{\boldsymbol{\Omega}}_k(\hat{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}_k)) + \Lambda\left(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\Omega}}_k\right), \tag{5}$$

where $\hat{\boldsymbol{\theta}}_k$ and $\hat{\boldsymbol{\Omega}}_k$ denote respectively the estimated parameter vector and weighting matrix for the kth model within a set of K candidate models, while $\Lambda\left(\hat{\boldsymbol{\theta}}_k, \hat{\boldsymbol{\Omega}}_k\right)$ is a possible penalty term. Penalized objective functions have been proposed for model and moment selection for GMM estimators such as, for example, Andrews (1999) and successively Andrews and Lu (2001) who proposed penalty terms that reflect the number of moment conditions or Zhang and Guerrier (2020) who derived a penalty term using the covariance penalty criterion of Efron (2004). Nevertheless, since these model selection criteria only rely on a consistent estimator for the model's parameters $\boldsymbol{\theta}_k$, a robust version is (almost) readily available by using the RGMWM framework. Moreover, the estimator $\hat{\boldsymbol{\Omega}}_k$ in (5) can be made model-independent by choosing $\hat{\boldsymbol{\Omega}}_k := \hat{\boldsymbol{\Omega}}$ (e.g., based on $\hat{V}^{-1}$ as proposed in Remark 2.4) making model comparison computationally more efficient. While the J-test statistic will be used in the analysis of real data in Section 5, the study of a possible implementation of a robust model selection criterion is left for further research.

Finally, aside from providing the basis for model-independent outlier-detection (which can be of great importance for fault-detection algorithms, see, for example, Guerrier et al. 2012, and references therein), another advantage of the RGMWM estimator is that it can easily be extended to more complex settings such as multivariate time series (see, e.g., Xu et al. 2019), locally stationary time series (see, e.g., Nason, Von Sachs, and Kroisandt 2000) or to random fields (see, e.g., Mondal and Percival 2012b,a) thereby delivering a computationally efficient framework for robust inference in these settings as well.

4. Simulation Studies

The aim of this section is to show that the RGMWM has a reasonable performance in settings where there is no data contamination and has a better performance than the classical (and possibly robust) alternatives when the data are contaminated (as mentioned in the previous sections, a simulation study regarding the performance of the proposed robust WV can also be found in Appendix C). In order to compare the proposed framework with existing approaches, in this section we limit ourselves to relatively short time series and stationary models while we explore large and/or intrinsically stationary time series in the applied examples in Section 5 and in the supplemental material. Concerning the robust alternatives, there is a lack of implemented and generally available (or usable) robust methods for parametric inference on time series models. For this reason, we were only able to successfully implement two robust estimators with which to compare the proposed RGMWM estimator: the Yule-Walker estimator based on the robust autocovariance estimator (YW) (as used for example in Sarnaglia, Reisen, and Lévy-Leduc 2010b) and the indirect inference estimator based on the YW estimator (as proposed in Genton and Ronchetti 2003) using a Tukey Biweight function with tuning constant $c = 2.2$ (chosen based on preliminary simulation studies in order to be highly robust). In the latter case, AR(p^*) models were used as auxiliary models with $p^* = p + 1$ where, in this case, p represents the number of parameters in the models of interest (excluding the innovation variance parameter) and the number of simulated samples was 100. Since the YW estimator is appropriate for autoregressive models while the indirect inference estimator is used for all other models, we will denote both with a common acronym, that is, R-YW (for Robust Yule-Walker based estimators). On the other hand, the GMWM and RGMWM estimators are made available through the open-source R package gmwm[2] where the default options are the Haar wavelet filter (at dyadic scales $\tau_j = 2^j$) and a diagonal matrix for the weighting matrix $\hat{\boldsymbol{\Omega}}$ with elements proportional to the estimated variance of $\hat{\nu}_j^2$ (which was used for the simulations also to make more reasonable comparisons with the indirect inference estimators which were based on the identity matrix). More details on this choice can be found in Appendix C. The Tukey Biweight was used also for the RGMWM with tuning constant based on an asymptotic efficiency of 60% thereby guaranteeing high robustness of the resulting estimator.

For the simulation studies different types of contamination were used, going from scale-contamination to additive and replacement outliers as well as patchy outliers and level-shifts. Innovation-type contamination was not considered since it does not affect the estimators much (see Maronna, Martin, and Yohai 2006; Maronna et al. 2019, for an overview of different contamination settings). We denote the proportion of contaminated observations with ϵ and the size of contamination (i.e., the variance of the observations which are added to the uncontaminated observations) with σ_ϵ^2. Finally, when dealing with level-shifts, we denote μ_{ϵ_i} as the size of the ith shift in level.

The performance of these estimators is investigated on the following models and contamination settings:

- AR(1): a zero-mean first-order autoregressive model with parameter vector $[\rho_1 \; \upsilon^2]^\top = [0.9 \; 1]^\top$, scale-based contamination at level $j = 3$, $\epsilon = 0.01$ and $\sigma_\epsilon^2 = 100$;
- AR(2): a zero-mean second-order autoregressive model with parameter vector $[\rho_1 \; \rho_2 \; \upsilon^2]^\top = [0.5 \; -0.3 \; 1]^\top$, replacement isolated outliers, $\epsilon = 0.05$ and $\sigma_\epsilon^2 = 9$;
- ARMA(1,2): a zero-mean autoregressive-moving average model with parameter vector $[\rho \; \varrho_1 \; \varrho_2 \; \upsilon^2]^\top = [0.5 \; -0.1 \; 0.5 \; 1]^\top$, and level-shift contamination with $\epsilon = 0.05$, $\mu_{\epsilon_1} = 5$ and $\mu_{\epsilon_2} = -3$;
- ARMA(3,1): a zero-mean autoregressive-moving average model with parameter vector $[\rho_1 \; \rho_2 \; \rho_3 \; \varrho_1 \; \upsilon^2]^\top = [0.7 \; 0.3 \; -0.2 \; 0.5 \; 2]^\top$, patchy outliers, $\epsilon = 0.01$ and $\sigma_\epsilon^2 = 100$;
- SSM: a state-space model (X_t) interpreted as a composite (latent) process in certain engineering applications. This model is defined as

$$Y_t^{(i)} = \rho_{(i)} Y_{t-1}^{(i)} + W_t^{(i)}, \quad W_t^{(i)} \overset{\text{iid}}{\sim} \mathcal{N}(0, \upsilon_{(i)}^2)$$
$$X_t = \sum_{i=1}^{2} Y_t^{(i)} + Z_t, \quad Z_t \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma^2)$$

with parameter vector

$$[\rho_{(1)} \; \upsilon_{(1)}^2 \; \rho_{(2)} \; \upsilon_{(2)}^2 \; \sigma^2]^\top = [0.99 \; 0.1 \; 0.6 \; 2 \; 3]^\top,$$

additive isolated outliers, $\epsilon = 0.05$ and $\sigma_\epsilon^2 = 9$.

To measure the statistical performance of the estimators we choose to use a robust and relative version of the root mean squared error (RMSE) defined as follows:

$$\text{RMSE}^* := \sqrt{\text{med}\left(\frac{\hat{\theta}_i - \theta_{i,0}}{\theta_{i,0}}\right)^2 + \text{mad}\left(\frac{\hat{\theta}_i}{\theta_{i,0}}\right)^2},$$

with med$(\cdot)$ representing the median, mad$(\cdot)$ the median absolute deviation and $\hat{\theta}_i$ and $\theta_{i,0}$ representing the ith element of the estimated and true parameter vectors, respectively. Finally, for each simulation, the number of simulated samples is 500 while the sample size is $T = 10^3$ which delivers $J = 9$ scales for the GMWM-type estimators.

Figure 1 displays the (logarithm of the) RMSE* of the estimators in both uncontaminated and contaminated settings for all the models presented above. When considering the AR(1) and AR(2) and ARMA(3,1) models, the RGMWM does not lose much in uncontaminated settings while it performs generally as well or better than the R-YW estimator in contaminated ones, especially concerning the variance parameter υ^2 in the ARMA(3,1). As for the ARMA(1,2) model, the RGMWM is not as efficient as the others in the uncontaminated case while it adequately bounds the influence of outliers in contaminated ones, performing generally better than the R-YW estimator, especially for the variance parameter υ^2 as in the case of the ARMA(3,1) model. For the SSM model, standard estimators in their default implementation did not appear to be numerically stable while, to the best of our knowledge, robust alternatives have never been implemented. From Figure 1, it can be seen how the RGMWM is extremely close to the GMWM in uncontaminated settings while it remains more stable than the GMWM in the contaminated ones. When considered jointly with the additional simulation study in supplementary material D, this simulation exercise shows that the RGMWM provides a computationally efficient and numerically stable method to robustly estimate the parameters of many linear state-space models which, to date, has been almost unfeasible in practice with alternative robust estimators. Indeed, supplementary material D illustrates the computational efficiency of the RGMWM for time series models with a moderately high number of parameters for sample sizes of $T = 10^7$ for which the RGMWM is computed in just over a minute, denoting the added value of this approach since these models and sample sizes are extremely common, for example, in the natural sciences and engineering. As a final note, since this simulation study does not use a full weighting matrix for $\hat{\boldsymbol{\Omega}}$, the efficiency of the (R)GMWM could be improved by choosing an alternative matrix.

5. Application: Personal Saving Rates

Having highlighted the properties of the RGMWM in a controlled simulated setting, in this section, we conclude this work by presenting the results when using the RGMWM for an analysis on real data concerning personal saving rates. In addition, in supplemental material E we present the results of an analysis on the measurement error issued from an inertial sensor based on a calibration sample of size $T = 9 \cdot 10^5$ that requires the

[2]The gmwm package can be downloaded from *https://github.com/SMAC-Group/gmwm*.

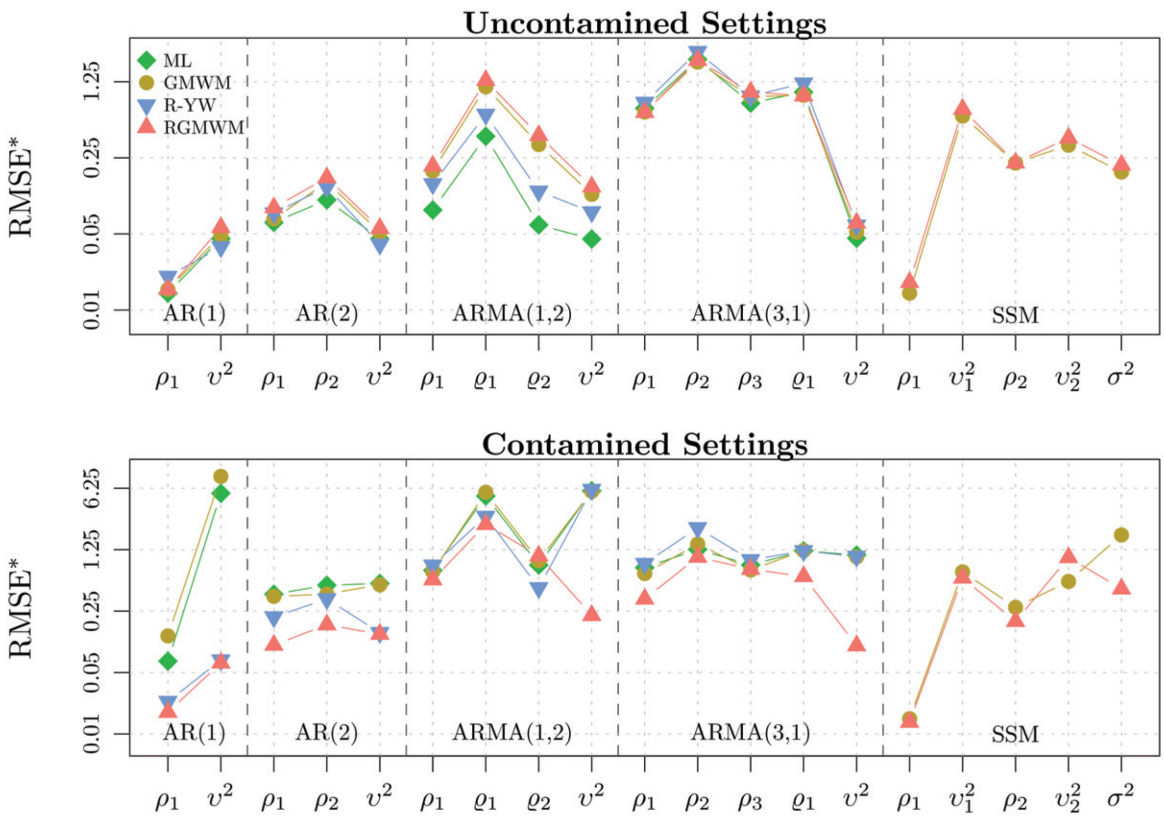

Figure 1. Top row: logarithm of the RMSE* of the estimators in an uncontaminated setting. Bottom row: logarithm of the RMSE* of the estimators in a contaminated setting. R-YW represents the YW estimator for the AR(1) and AR(2) models while it represents the indirect inference estimator for the other models.

estimation of a state-space model with 6 parameters. Indeed, the wide class of intrinsically stationary models for which the RGMWM can be used allows it to be employed in a large variety of applications where outliers and other types of data contamination can often occur. As mentioned above, in this section, the RGMWM will be used to analyse data consisting in the monthly seasonally adjusted Personal Saving Rates (PSR) from January 1959 to May 2015 provided by the Federal Reserve Bank of St. Louis.[3] The study of PSR is an essential part of the overall investigation on the health of national and international economies since, within more general economic models, PSR can greatly impact the funds available for investment which in turn determine the productive capacity of an economy. Understanding the behaviour of PSR is therefore an important step in correct economic policy decision making. In this sense, Slacalek and Sommer (2012) study the factors behind saving rates and investigate different models which, among others, are compared to the random-walk-plus-noise (local level) model (RWN). As opposed to the latter model, various time-varying models are proposed in the literature to explain precautionary PSR together with risk aversion in the light of different factors such as financial shocks or others (see, for example, Owen and Wu 2007; Brunnermeier and Nagel 2008). Nevertheless, as emphasized in Pankratz (2012), modelling the time series with a stationary model, or a dth-order non-stationary model such as an ARIMA, can be useful under many aspects such as, for example, to understand if a dynamic model is needed for forecasting and, if so, what kind of model is appropriate.

In this example, we consider the RWN model and we use the WV log-log plot and a J-test (see Section 3.1) to understand what kind of model could fit the time series. Taking into account the analysis in Slacalek and Sommer (2012) who, among others, consider *two* main drivers to PSR (unemployment and interest rates) in addition to fluctuating target wealth, we consider the sum of *two* AR(1) processes (i.e., ARMA(2,1)) for the noise component of the RWN model. Based on both the visual fit of the WV of the candidate models in the plots as well as the p-values of the J-tests (for which the null hypothesis is that the model fits the data well), we find that a random walk plus an ARMA(2,1) indeed appears to be a good fit (J-test p-values are 0.109 and 0.279 for GMWM and RGMWM, respectively). This can be seen in Figure 2 where, in the top part, the saving rate time series is represented along with the identified outliers and, in the bottom part, we see the log–log representation of the classic and robust estimated and model-implied WV respectively. Indeed, for the bottom part, the diagonal plots show the classic and robust estimations, respectively, each with the estimated WV and the WV implied by the estimated RWN model; when the difference in the two lines lies within their confidence intervals, the chosen model can be considered as adequate. The off-diagonal plots compare the classic and robust estimated WV (upper off-diagonal) and the WV implied by the GMWM and RGMWM model parameter estimates (lower off-diagonal).

[3] U.S. Bureau of Economic Analysis, Personal Saving Rate [PSAVERT], retrieved from FRED, Federal Reserve Bank of St. Louis; *https://fred.stlouisfed.org/series/PSAVERT*

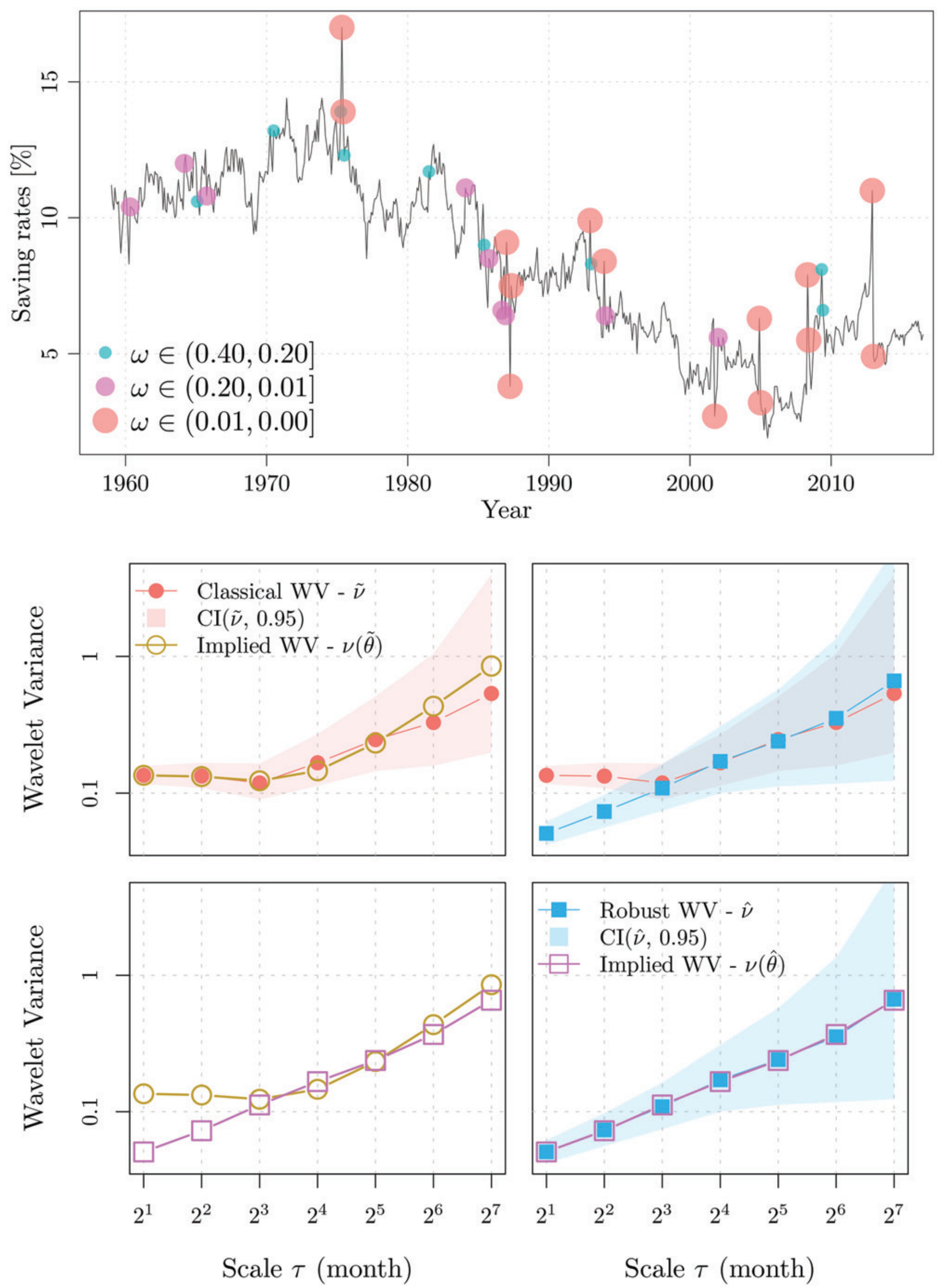

Figure 2. Top figure: Saving rates time series with different types of points indicating outliers identified through the weights of the proposed M-estimator. Bottom figure: log-log scale WV plots for saving rates series with classic estimated WV superposed with model-implied WV based on the parameters estimated through the GMWM (top left); classic and robust estimated WV with respective confidence intervals superposed (top right); classic and robust model-implied WV based on the GMWM and RGMWM estimates respectively (bottom left); robust estimated WV superposed with model-implied WV based on the parameters estimated through the RGMWM estimator (bottom right).

Table 1. Random Walk plus ARMA(2,1) model estimates for the PSR data.

	GMWM		RGMWM	
	Estimate	CI(·, 95%)	Estimate	CI(·, 95%)
γ^2	7.95×10^{-2}	$(3.67 \times 10^{-2}; 1.11 \cdot 10^{-1})$	5.85×10^{-2}	$(1.54 \times 10^{-2}; 9.97 \times 10^{-2})$
ρ_1	1.64×10^{-1}	$(5.93 \times 10^{-2}; 2.89 \times 10^{-1})$	6.00×10^{-1}	$(4.48 \times 10^{-1}; 7.55 \times 10^{-1})$
ρ_2	3.06×10^{-3}	$(-1.31 \times 10^{-1}; 1.48 \times 10^{-1})$	1.84×10^{-1}	$(3.10 \times 10^{-2}; 2.46 \times 10^{-1})$
ϱ	2.43×10^{-1}	$(2.02 \times 10^{-1}; 2.81 \times 10^{-1})$	2.92×10^{-1}	$(2.28 \times 10^{-1}; 3.45 \times 10^{-1})$
σ^2	3.14×10^{-1}	$(2.59 \times 10^{-1}; 3.85 \times 10^{-1})$	1.32×10^{-1}	$(8.59 \times 10^{-2}; 1.80 \times 10^{-1})$

NOTE: Estimated parameters with GMWM and RGMWM estimators with γ^2 being the random walk parameter, ρ_i the ith autoregressive parameter, ϱ the moving average parameter and σ^2 the innovation variance of the ARMA(2,1) model. Confidence intervals (CI) based on the approach used in Guerrier et al. (2013).

It can be seen how significant the difference is between the standard and robust WV estimates, especially at the first scales where the confidence intervals of the estimated WV do not overlap (upper off-diagonal plot). This leads to a difference in the model-implied WV whose parameters have been estimated through the GMWM and RGMWM (lower off-diagonal plot). It can also be noticed how the confidence intervals of the robust WV estimator $\hat{\boldsymbol{\nu}}$ are wider than those of the classical estimator due to the trade-off between (bias) robustness and variance.

The estimated parameters of the RWN model using the GMWM and RGMWM are given in Table 1 along with their respective confidence intervals. There are two main differences between the two estimations: (i) the estimates of the first autoregressive parameter ρ_1 and innovation variance σ^2 are significantly different; (ii) the second autoregressive parameter ρ_2 is not significant using the GMWM. These differences highlight how the conclusions concerning parameter values and model selection can considerably change when outliers are present in the data. Indeed, the choice of the model would then affect the decisions taken toward the selection of appropriate causal and dynamic models to better explain the behaviour of saving rates. The selected model based on the robust fit can in fact be interpreted as a sum of latent models along the lines given in Slacalek and Sommer (2012) where the ARMA(2,1) can be seen as a sum of two AR(1) models (see Granger and Morris 1976) where each of them represents, for example, the reaction of PSR to changes in uncertainty (affected by unemployment) and interest rates, respectively, while the random walk describes the continuous fluctuations of target wealth which also drives PSR.

The additional benefit of the RGMWM, and more specifically of the proposed M-estimator of WV, is also to deliver weights that allow to identify outliers which may not be visible simply by looking at the time series. As shown in the top part of Figure 2, the outliers identified by the RGMWM can be interpreted in the light of the national and global economic and political events. Limiting ourselves to the major identified outliers, the first one corresponds to a rise in the precautionary savings in the aftermath of the OPEC oil crisis and the 1974 stock market crash. In the months following October 1987 we can see an instability in the PSR with a rise and sudden fall linked to the "Black Monday" stock market crash which added to the savings and loans crisis which lasted to the early 1990s. This period also saw an economic recession where a rise in the saving rates, highlighted by the presence of high outliers, led to a drop in aggregate demand and bankruptcies. Finally, the various financial crises of the 21st century led to sudden and isolated rises in PSR as indicated again by the outliers.

Norm Notation

Given the use of different types of norm considered in these appendices, as a reference below we provide their notations and definitions:

- $|\boldsymbol{a}|_q := \left(\sum_{j=1}^{p} |a_j|^q\right)^{1/q}$ represents the l_q-norm where $\boldsymbol{a} = (a_1, \dots, a_p)^\top \in \mathbb{R}^p$.
- $\|X\|_q := (\mathbb{E}[|X|^q])^{1/q}$.
- $\|\boldsymbol{X}\|_S$ denotes the spectral norm of a matrix $\boldsymbol{X}$.

Appendix A: Short Literature Review

A detailed discussion on robust estimation and inference methods for time series models can be found in Maronna, Martin, and Yohai (2006), Chapter 8. An important part of the literature in this domain has dealt with time series models such as autoregressive and/or moving average models. For example, Künsch (1984b) proposes optimal robust M-estimators of the parameters of autoregressive processes by studying the properties of their influence function (see also Martin and Yohai 1986). Denby and Martin (1979) develop a generalized M-estimator for the parameter of a first-order autoregressive process whereas Bustos and Yohai (1986), Allende and Heiler (1992) and de Luna and Genton (2001); Genton and Ronchetti (2003) extend the research to include moving average models using generalized M-estimation theory and/or indirect inference (see, e.g., Gourieroux, Monfort, and Renault 1993). Bianco et al. (1996) proposed a class of robust estimators for regression models with ARIMA errors based on τ-estimators of scale (Yohai and Zamar 1988). Ronchetti and Trojani (2001) developed a robust version of the generalized method of moments (proposed by Hansen 1982) for estimating the parameters of time series models in economics, while Ortelli and Trojani (2005) further developed a robust efficient method of moments and Cizek (2016) proposed a generalized method of trimmed moments. Mancini, Ronchetti, and Trojani (2005) developed optimal bias-robust estimators for a class of conditional location and scale time series models while La Vecchia and Trojani (2010) developed conditionally unbiased optimal robust estimators for general diffusion processes, for which approximation methods for computing integrals are proposed. Cizek (2008) studies the properties of a two-step least weighted squares robust time-series regression estimator and Agostinelli and Bisaglia (2010) proposed a weighted MLE for ARFIMA processes.

Two-step robust approaches can be built upon robust (Kalman) filtering or robust moment estimation. Robust estimators of moments, such as autocovariances, include Ma and Genton (2000), Lévy-Leduc et al. (2011), and Chang and Politis (2016) (see also Rousseeuw and Croux 1993), and for a review, see, for example, Dürre, Fried, and Liboschik (2015). They have been used by for example, Molinares, Reisen, and Cribari-Neto (2009) as plugin estimators for ARFIMA models (see also Reisen and Molinares 2012), by Sarnaglia, Reisen, and Lévy-Leduc (2010a) for the parameters of the periodic AR model with the Yule–Walker equation and by Bahamonde and Veiga (2016) for the GARCH(1,1). The idea of making the Kalman filter robust was originated with Masreliez and Martin (1977) and Cipra (1992) who proposed robust modifications of exponential smoothing (see also Cipra and Hanzak 2011 and Croux, Gelper, and Mahieu 2010 for a multivariate version). For a robust version of the Holt-Winters smoother, see Gelper, Fried, and Croux (2010), and other proposals can be found in for example, Ruckdeschel, Spangl, and Pupashenko (2014) and Calvet, Czellar, and Ronchetti (2015). Muler, Peña, and Yohai (2009) developed a class of robust estimates for ARMA models that are closely related to robust filtering. Robustness properties of wavelet filtering have been studied for the identically and independently distributed (iid) case by Renaud (2002). Several robust local filters have been proposed so far since the median filter proposal from Tukey (1977): Bruce et al. (1994) pre-process the estimation of the wavelet coefficients via a "fast and robust smooth/cleaner;" Krim and Schick (1999) derived a robust estimator of the wavelet coefficients based on minimax description length; Härdle and Gasser (1984) develop a locally weighted smoothing using M-estimation and Fried, Einbeck, and Gather (2007) propose a nonparametric, weighted repeated median filter. Sardy, Tseng, and Bruce (2001) proposed a robust wavelet-based estimator using a robust loss-penalized function, for which appropriately choosing the smoothing parameter is an important robustness issue as revealed, for example, by Cantoni and Ronchetti (2001).

Appendix B: Proof of Theorem 2.1

In this appendix we discuss the asymptotic normality of the proposed WV estimator $\hat{\boldsymbol{\nu}}$. Before proving Theorem 2.1, we need two additional results which are namely, (i) the consistency of $\hat{\nu}_j^2$ as well as (ii) the Bouligand differentiability of the Huber ψ-function if chosen for the estimator $\hat{\boldsymbol{\nu}}$ (which is needed for a Taylor expansion in the proof). We start with the consistency of $\hat{\nu}_j^2$ which is stated in the following proposition (followed by its proof).

Proposition B.1. Under Conditions (C2) to (C4), we have that

$$\hat{\nu}_j^2 \xrightarrow{p} \nu_j^2.$$

Proof. We firstly verify the point-wise convergence

$$\frac{1}{M_j}\sum_{t=1}^{M_j}\psi(W_{j,t},\nu_j^2) \xrightarrow{p} \mathbb{E}[\psi(W_{j,t},\nu_j^2)]$$

for any $\nu_j^2 > 0$. Since $\left(\psi(W_{j,t},\nu_j^2)\right)$ is a time-invariant function of $\left(W_{j,t}\right)$, it is also a stationary process (see Wooldridge 1994) based on Condition (C3). Recalling the notation $\|Z\|_q := (\mathbb{E}[|Z|^q])^{1/q}$ and denoting $A := 1/M_j \sum_{t=1}^{M_j} \psi(W_{j,t},\nu_j^2) - \mathbb{E}[\psi(W_{j,t},\nu_j^2)]$, by Markov inequality we have

$$\mathbb{P}\left(|A| \geq \epsilon\right) \leq \frac{\|A\|_2^2}{\epsilon^2}. \tag{B.6}$$

Applying the definition of the projection operator $\mathcal{P}_t\cdot$, computations show that

$$\psi(W_{j,t},\nu_j^2) - \mathbb{E}[\psi(W_{j,t}.\nu_j^2)] = \sum_{l=0}^{\infty}\mathcal{P}_{t-l}\psi(W_{j,t},\nu_j^2),$$

hence the numerator on the right side of the inequality in (B.6) can be written as

$$\left\|\frac{1}{M_j}\sum_{t=1}^{M_j}\psi(W_{j,t},\nu_j^2) - \mathbb{E}[\psi(W_{j,t},\nu_j^2)]\right\|_2 = \frac{1}{M_j}\left\|\sum_{t=1}^{M_j}\sum_{l=0}^{\infty}\mathcal{P}_{t-l}\psi(W_{j,t},\nu_j^2)\right\|_2.$$

Noticing that $(\mathcal{P}_{t-l}\psi(W_{j,t},\nu_j^2))$ (for $t = 1,\ldots,M_j$) forms a martingale difference sequence, by first applying the triangle inequality and then Burkholder's moment inequality for martingale differences (Burkholder 1988), we have

$$\begin{aligned}\frac{1}{M_j}\left\|\sum_{t=1}^{M_j}\sum_{l=0}^{\infty}\mathcal{P}_{t-l}\psi(W_{j,t},\nu_j^2)\right\|_2 &\leq \frac{1}{M_j}\sum_{l=0}^{\infty}\left\|\sum_{t=1}^{M_j}\mathcal{P}_{t-l}\psi(W_{j,t},\nu_j^2)\right\|_2\\ &\leq \frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\left\|\mathcal{P}_0\psi(W_{j,l},\nu_j^2)\right\|_2.\end{aligned}$$

We would therefore want to show that the latter term tends to zero to prove consistency. For this reason, following the proof in Wu (2011), we now write

$$\mathcal{P}_0\psi(W_{j,t},\nu_j^2) = \mathbb{E}[\psi(W_{j,t},\nu_j^2)|\mathcal{F}_0] - \mathbb{E}[\psi(W_{j,t},\nu_j^2)|\mathcal{F}_{-1}],$$

where, recalling that $W_{j,t}^\star$ is a coupled version of $W_{j,t}$, we can notice that

$$\mathbb{E}[\psi(W_{j,t},\nu_j^2)|\mathcal{F}_{-1}] = \mathbb{E}[\psi(W_{j,t}^\star,\nu_j^2)|\mathcal{F}_{-1}],$$

since the filtrations $\mathcal{F}_t$ and $\mathcal{F}_t^\star$ are the same up to $t = -1$ (and are different at $t = 0$). This implies that

$$\mathbb{E}[\psi(W_{j,t},\nu_j^2)|\mathcal{F}_{-1}] = \mathbb{E}[\psi(W_{j,t}^\star,\nu_j^2)|\mathcal{F}_{-1}] = \mathbb{E}[\psi(W_{j,t}^\star,\nu_j^2)|\mathcal{F}_0].$$

This allows us to rewrite

$$\left\|\mathcal{P}_0\psi(W_{j,t},\nu_j^2)\right\|_2 = \left\|\mathbb{E}[\psi(W_{j,t},\nu_j^2) - \psi(W_{j,t}^\star,\nu_j^2)|\mathcal{F}_0]\right\|_2,$$

which, by Jensen's inequality, gives us

$$\left\|\mathbb{E}[\psi(W_{j,t},\nu_j^2) - \psi(W_{j,t}^\star,\nu_j^2)|\mathcal{F}_0]\right\|_2 \leq \left\|\psi(W_{j,t},\nu_j^2) - \psi(W_{j,t}^\star,\nu_j^2)\right\|_2.$$

Moreover, recall that

$$\psi(W_{j,t},\nu_j^2) = \omega^2\left(r_{j,t};\nu_j^2,c\right) r_{j,t}^2 - a(\nu_j^2,c),$$

where $\omega(\cdot) \in [0,1]$ are weights (given for example by the Huber or Tukey biweight functions) and $r_{j,t} = W_{j,t}/\nu_j$. Given this, we can denote $\phi(W_{j,t}/\nu_j,c) := \omega\left(r_{j,t};\nu_j^2,c\right) r_{j,t}$ and, combining the above notations and expansions we have

$$\begin{aligned}&\frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\left\|\mathcal{P}_0\psi(W_{j,l},\nu_j^2)\right\|_2\\ &\leq \frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\left\|\psi(W_{j,l},\nu_j^2) - \psi(W_{j,l}^\star,\nu_j^2)\right\|_2\\ &= \frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\left\|[\phi(W_{j,l}/\nu_j,c)]^2 - [\phi(W_{j,l}^\star/\nu_j,c)]^2\right\|_2\\ &= \frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\Big\|[\phi(W_{j,l}/\nu_j,c) + \phi(W_{j,l}^\star/\nu_j,c)]\\ &\qquad [\phi(W_{j,l}/\nu_j,c) - \phi(W_{j,l}^\star/\nu_j,c)]\Big\|_2\\ &= \frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\mathbb{E}\Big\{[\phi(W_{j,l}/\nu_j,c) + \phi(W_{j,l}^\star/\nu_j,c)]^2\\ &\qquad [\phi(W_{j,l}/\nu_j,c) - \phi(W_{j,l}^\star/\nu_j,c)]^2\Big\}^{1/2}.\end{aligned}$$

By Hölder's inequality we have that the last term is smaller or equal to

$$\frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}2\left\|\phi(W_{j,l}/\nu_j,c)\right\|_4\left\|\phi(W_{j,l}/\nu_j,c) - \phi(W_{j,l}^\star/\nu_j,c)\right\|_4,$$

and, noticing that

$$|\phi(W_{j,t}/\nu_j,c) - \phi(W_{j,t}^\star/\nu_j,c)| \leq \left|\frac{W_{j,t} - W_{j,t}^\star}{\nu_j}\right|,$$

we can finally write

$$\begin{aligned}&\frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\left\|\mathcal{P}_0\psi(W_{j,l},\nu_j^2)\right\|_2\\ &\quad\leq \frac{1}{\sqrt{M_j}\nu_j}\sum_{l=0}^{\infty}2\left\|\phi(W_{j,l}/\nu_j,c)\right\|_4\left\|W_{j,l} - W_{j,l}^\star\right\|_4.\end{aligned}$$

At this point, we can also underline that $|\phi(W_{j,t}/\nu_j,c)| \leq c$ implying that, for all $p > 0$, $\left\|\phi(W_{j,t}/\nu_j,c)\right\|_p \leq k < \infty$. Using Condition (C4) we finally have

$$\begin{aligned}\frac{1}{\sqrt{M_j}}\sum_{l=0}^{\infty}\left\|\mathcal{P}_0\psi(W_{j,l},\nu_j^2)\right\|_2 &\leq \frac{2k}{\sqrt{M_j}\nu_j}\sum_{l=0}^{\infty}\delta_{l,4}^j\\ &= \mathcal{O}_p\left(\frac{1}{\sqrt{M_j}}\right).\end{aligned}$$

Therefore, using these results in (B.6), we have

$$\frac{1}{M_j}\sum_{t=1}^{M_j}\psi(W_{j,t},\nu_j^2) \xrightarrow{p} \mathbb{E}[\psi(W_{j,t},\nu_j^2)], \tag{B.7}$$

for every ν_j^2. Based on the requirements of Lemma 5.10 of Van der Vaart (2000), knowing that $\mathbf{N} \subset \mathbb{R}^+$ and using Condition (C2) we have that

$$\hat{\nu}_j^2 \xrightarrow{p} \nu_j^2,$$

thus concluding the proof. □

Having proved consistency, we now deliver an additional technical result that allows to perform an expansion in the case where the Huber ψ-function is chosen. This result is provided in the following lemma (followed again by its proof).

Lemma B.1. Assuming the wavelet coefficient process $(W_{j,t})$ is Gaussian, the function $\psi(W_{j,t},\nu_j^2)$ using Huber weights is Bouligand-differentiable at $\nu_{j,0}^2$ as follows:

$$\psi'(W_{j,t},\nu_{j,0}^2) = \begin{cases} -\frac{W_{j,t}^2}{\nu_{j,0}^4} & \text{if } |r_{j,t}| \le c \\ 0 & \text{if } |r_{j,t}| > c \end{cases}$$

The proof of this lemma is given below.

Proof. Let us define $r_0 := W_{j,t}/\sqrt{\nu_{j,0}^2}$ and $r := W_{j,t}/\sqrt{\nu^2}$ where $\nu^2 = \nu_{j,0}^2 + h$. Let $\mathcal{X} \subseteq \mathbb{R}^l$ be an open set and let $f : \mathcal{X} \to \mathbb{R}^m$ be a function. By the definition in Christmann and Van Messem (2008), f is Bouligand-differentiable (B-differentiable) at point $x_0 \in \mathcal{X}$ if there exists a positive homogeneous function[4] $f'(x_0;\cdot)$ such that $f(x_0+h) = f(x_0)+f'(x_0;h)+o(h)$. Below are the computations of the B-derivatives for the five cases of the Huber weight function, consider h be a value close enough to 0:

1. Setting $r_0 = c$ we have:
   - If $h \ge 0$ $(r \le c)$:

$$\begin{aligned}
&\psi'(W_{j,t},\nu_{j,0}^2)(h) + o(h)\\
&= \psi(W_{j,t},\nu_{j,0}^2+h) - \psi(W_{j,t},\nu_{j,0}^2)\\
&= r^2 - a_\psi(c) - r_0^2 + a_\psi(c)\\
&= \frac{W_{j,t}^2}{\nu_{j,0}^2+h} - \frac{W_{j,t}^2}{\nu_{j,0}^2}\\
&= \frac{W_{j,t}^2}{\nu_{j,0}^2}\left(\frac{-h}{\nu_{j,0}^2+h}\right)\\
&= -\frac{W_{j,t}^2}{\nu_{j,0}^2}\left(\frac{h}{\nu_{j,0}^2} - \frac{h^2}{\nu_{j,0}^2(\nu_{j,0}^2+h)}\right)\\
&= -\frac{W_{j,t}^2}{\nu_{j,0}^4}h + \underbrace{\frac{W_{j,t}^2 h^2}{\nu_{j,0}^2(\nu_{j,0}^2+h)}}_{o(h)} := \Delta
\end{aligned}$$

   - If $h < 0$ $(r > c)$:

$$\begin{aligned}
&\psi'(W_{j,t},\nu_{j,0}^2)(h) + o(h)\\
&\psi(W_{j,t},\nu_{j,0}^2+h) - \psi(W_{j,t},\nu_{j,0}^2)\\
&c^2 - a_\psi(c) - r_0^2 + a_\psi(c) = c^2 - c^2 = 0
\end{aligned}$$

2. Setting $r_0 = -c$ we have:
   - If $h < 0$ $(r < -c)$:

$$\begin{aligned}
&\psi'(W_{j,t},\nu_{j,0}^2)(h) + o(h)\\
&= \psi(W_{j,t},\nu_{j,0}^2+h) - \psi(W_{j,t},\nu_{j,0}^2)\\
&c^2 - a_\psi(c) - r_0^2 + a_\psi(c) = 0
\end{aligned}$$

   - If $h \ge 0$ $(r \ge -c)$:

$$\begin{aligned}
&\psi'(W_{j,t},\nu_{j,0}^2)(h) + o(h)\\
&\psi(W_{j,t},\nu_{j,0}^2+h) - \psi(W_{j,t},\nu_{j,0}^2)\\
&r^2 - a_\psi(c) - r_0^2 + a_\psi(c)\\
&\ldots = \Delta
\end{aligned}$$

3. Setting $r_0 > c$ and h is small enough such that $r \ge c$, we have

$$\begin{aligned}
&\psi'(W_{j,t},\nu_{j,0}^2)(h) + o(h)\\
&\psi(W_{j,t},\nu_{j,0}^2+h) - \psi(W_{j,t},\nu_{j,0}^2)\\
&c^2 - a_\psi(c) - r_0^2 + a_\psi(c) = 0
\end{aligned}$$

4. Setting $r_0 < -c$ and h is small enough such that $r \le -c$, we have

$$\begin{aligned}
\psi'(W_{j,t},\nu_{j,0}^2)(h) + o(h) &= \psi(W_{j,t},\nu_{j,0}^2+h) - \psi(W_{j,t},\nu_{j,0}^2)\\
&= c^2 - a_\psi(c) - r_0^2 + a_\psi(c) = 0
\end{aligned}$$

5. Setting $-c < r_0 < c$ and h is small enough such that $-c \le r \le c$, we have

$$\begin{aligned}
\psi'(W_{j,t},\nu_{j,0}^2)(h) + o(h) &= \psi(W_{j,t},\nu_{j,0}^2+h) - \psi(W_{j,t},\nu_{j,0}^2)\\
&= r^2 - a_\psi(c) - r_0^2 + a_\psi(c)\\
&= \ldots = \Delta
\end{aligned}$$

We therefore have that the first B-derivative of the function $\psi(W_{j,t},\nu_j^2)$ at $\nu_{j,0}^2$ is given by

$$\psi'(W_{j,t},\nu_{j,0}^2)(h) = \psi'(W_{j,t},\nu_{j,0}^2)h = \begin{cases} -\frac{W_{j,t}^2}{\nu_{j,0}^4}h & \text{if } |r_{j,t}| \le c \\ 0 & \text{if } |r_{j,t}| > c, \end{cases}$$

and it is easy to see that $\psi(W_{j,t},\nu_{j,0}^2)(\cdot)$ is positive homogeneous.

The approach used in this proof can be used to obtain expressions for the B-derivatives of other piecewise differentiable weight functions (see Scholtes 2012). It can be seen how it extends the classic derivative for $|r_0| < c$ also to the points ν_0^2 such that $|r_0| = c$. However, the Frechet differentiability of this function has also been discussed in Clarke (1986). □

[4] $f : \mathcal{X} \to Z$ is called positive homogeneous if $f(\alpha x) = \alpha f(x)$, for all $\alpha \ge 0$ and $x \in \mathcal{X}$

As mentioned earlier, Lemma B.1 is useful for the results on asymptotic normality of the proposed estimator to hold in case the choice of the ψ-function corresponds to the Huber ψ-function, which is not continuous differentiable with respect to ν_j^2. Without such smoothness condition, the classical proof based on the mean value theorem would fail. However, in this case, we could apply Theorem 5.21 in Van der Vaart (2000), which avoid the smoothness condition. However, we need to verify two additional conditions. Firstly, the function $\psi(W_{j,t}, \nu_j^2)$ is locally Lipschitz with respect to ν_j^2. This is true since $\left|\psi(W_{j,t}, \nu_j^2) - \psi(W_{j,t}, \nu_j^{2\,\prime})\right| \leq \frac{W_{j,t}^2}{\nu_j^2 \nu_j^{2\,\prime}} |\nu_j^2 - \nu_j^{2\,\prime}| \leq \frac{W_{j,t}^2}{C^2} |\nu_j^2 - \nu_j^{2\,\prime}|$ for all $W_{j,t} \in \mathbb{R}$ and $\nu_j^2 \geq C > 0$. Secondly, we need to verify that $\mathbb{E}\psi(W_{j,t}, \nu_j^2)$ is differentiable at $\nu_{j,0}^2$, which is implied by Condition (C1). And Lemma B.1 leads directly to Condition (C1). Therefore, in the following proof we only focus on the case when $\psi(\cdot, \nu_j^2)$ is continuous differentiable with respect to ν_j^2. Since otherwise, we can apply Theorem 5.21 in Van der Vaart (2000) by replacing A_j by m_j and removing the steps of showing $A_j \xrightarrow{p} m_j$, the rest are similar.

Proof of Theorem 2.1. Given Condition (C1), let us denote $\psi'(W_{j,t}, \nu_j{}^2) = \partial/\partial\nu_j{}^2\, \psi(W_{j,t}, \nu_j{}^2)$ and apply the mean value theorem to $\sum_{t=1}^{M_j} \psi(W_{j,t}, \hat{\nu}_j^2)$ around ν_j^2 obtaining

$$\sum_{t=1}^{M_j} \psi(W_{j,t}, \hat{\nu}_j^2) = \sum_{t=1}^{M_j} \psi(W_{j,t}, \nu_j^2) + \sum_{t=1}^{M_j} \psi'(W_{j,t}, \nu_j^{*2})(\hat{\nu}_j^2 - \nu_j^2) = 0$$

where

$$|\nu_j^{*2} - \nu_j^2| \leq |\hat{\nu}_j^2 - \nu_j^2|. \tag{B.8}$$

Rearranging the expansion and multiplying by $\sqrt{T}$ yields

$$\sqrt{T}(\hat{\nu}_j^2 - \nu_j^2) = \sqrt{\frac{T}{M_j}} \left[\underbrace{-\frac{1}{M_j} \sum_{t=1}^{M_j} \psi'(W_{j,t}, \nu_j^{*2})}_{A_j} \right]^{-1} \underbrace{\frac{1}{\sqrt{M_j}} \sum_{t=1}^{M_j} \psi(W_{j,t}, \nu_j^2)}_{B_j}. \tag{B.9}$$

Let us start from term A_j. We can rewrite this term as

$$-\frac{1}{M_j} \sum_{t=1}^{M_j} \psi'(W_{j,t}, \nu_j^{*2}) = -\frac{1}{M_j} \sum_{t=1}^{M_j} \psi'(W_{j,t}, \nu_j^2) - \frac{1}{M_j} \underbrace{\sum_{t=1}^{M_j} [\psi'(W_{j,t}, \nu_j^{*2}) - \psi'(W_{j,t}, \nu_j^2)]}_{C_j}.$$

Since $\left(\psi'(W_{j,t}, \nu_j^2)\right)$ is a time-invariant function of $\left(W_{j,t}\right)$, it is also a stationary process (see Wooldridge 1994) based on Condition (C3). Let us start from the first term on the right side of the above equality and define $m_j = \mathbb{E}[-\psi'(W_{j,t}, \nu_j^2)]$. Then by Markov inequality, we have

$$\mathbb{P}\left(\left| \frac{1}{M_j} \sum_{t=1}^{M_j} -\psi'(W_{j,t}, \nu_j^2) - m_j \right| \geq \epsilon \right) \leq \frac{\left\| \frac{1}{M_j} \sum_{t=1}^{M_j} -\psi'(W_{j,t}, \nu_j^2) - m_j \right\|_2^2}{\epsilon^2},$$

where, following the same reasoning as for the proof of Proposition B.1, $\left\| \frac{1}{M_j} \sum_{t=1}^{M_j} -\psi'(W_{j,t}, \nu_j^2) - m_j \right\|_2^2$ can be bounded by following inequalities:

$$\begin{aligned}
\left\| \frac{1}{M_j} \sum_{t=1}^{M_j} -\psi'(W_{j,t}, \nu_j^2) - m_j \right\|_2^2 &= \frac{1}{M_j^2} \left\| \sum_{t=1}^{M_j} \sum_{l=0}^{\infty} \mathcal{P}_{t-l} \psi'(W_{j,t}, \nu_j^2) \right\|_2^2 \\
&\leq \frac{1}{M_j^2} \sum_{l=0}^{\infty} \left\| \sum_{t=1}^{M_j} \mathcal{P}_{t-l} \psi'(W_{j,t}, \nu_j^2) \right\|_2^2 \\
&\leq \frac{1}{M_j} \sum_{l=0}^{\infty} \left\| \mathcal{P}_0 \psi'(W_{j,l}, \nu_j^2) \right\|_2^2.
\end{aligned}$$

Under Conditions (C1) to (C3), we can follow the same steps as the proof of Proposition B.1 to deliver

$$\frac{1}{M_j} \sum_{t=1}^{M_j} -\psi'(W_{j,t}, \nu_j^2) \xrightarrow{p} \mathbb{E}[-\psi'(W_{j,t}, \nu_j^2)] = m_j.$$

If the Bouligand derivative $\psi'(\cdot)$ has discontinuous points (continuous almost everywhere), then Theorem 5.21 in Van der Vaart (2000) can be directly applied to the ψ-function under the same conditions and deliver the desired result.

As for term C_j, since $\hat{\nu}_j^2$ is a consistent estimator of ν_j^2, by (B.8) so is ν_j^{*2}. Moreover, since $\psi'(\cdot)$ is continuous almost everywhere by Condition (C1), using the continuous mapping theorem, we have

$$\frac{1}{M_j} \sum_{t=1}^{M_j} [\psi'(W_{j,t}, \nu_j^{*2}) - \psi'(W_{j,t}, \nu_j^2)] \xrightarrow{p} 0,$$

which finally yields

$$A_j \xrightarrow{p} m_j.$$

Let us now focus on term B_J for which we intend to show convergence to a normal distribution in order to make use of Slutsky's theorem. For this reason we verify the requirements of Theorem 7 in Wu (2011) most of which have already been verified in the proof of Proposition B.1. Indeed, based Condition (C3) we have that $\left(\psi(W_{j,t}, \nu_j^2)\right)$ is a stationary process which can be represented as

$$\psi(W_{j,t}, \nu_j^2) = \sum_{l=0}^{\infty} \mathcal{P}_{t-l} \psi(W_{j,t}, \nu_j^2),$$

where $\left(\mathcal{P}_{t-l} \psi(W_{j,t}, \nu_j^2)\right)_{l=0,\ldots,\infty}$ is a martingale difference sequence. Based on this, we verify the conditions of Theorem 3 in Wu (2011) such that the martingale central limit theorem can be applied. Firstly we need to show that

$$\sum_{l=0}^{\infty} \left\| \mathcal{P}_0 \psi(W_{j,l}, \nu_j^2) \right\|_2^2 < \infty.$$

This requirement was verified in the proof of Proposition B.1 since it was shown that, with $k < \infty$, we have

$$\sum_{l=0}^{\infty} \left\| \mathcal{P}_0 \psi(W_{j,l}, \nu_j^2) \right\|_2 \leq \frac{2\,k}{\nu_j} \sum_{l=0}^{\infty} \delta_{l,4}^{j},$$

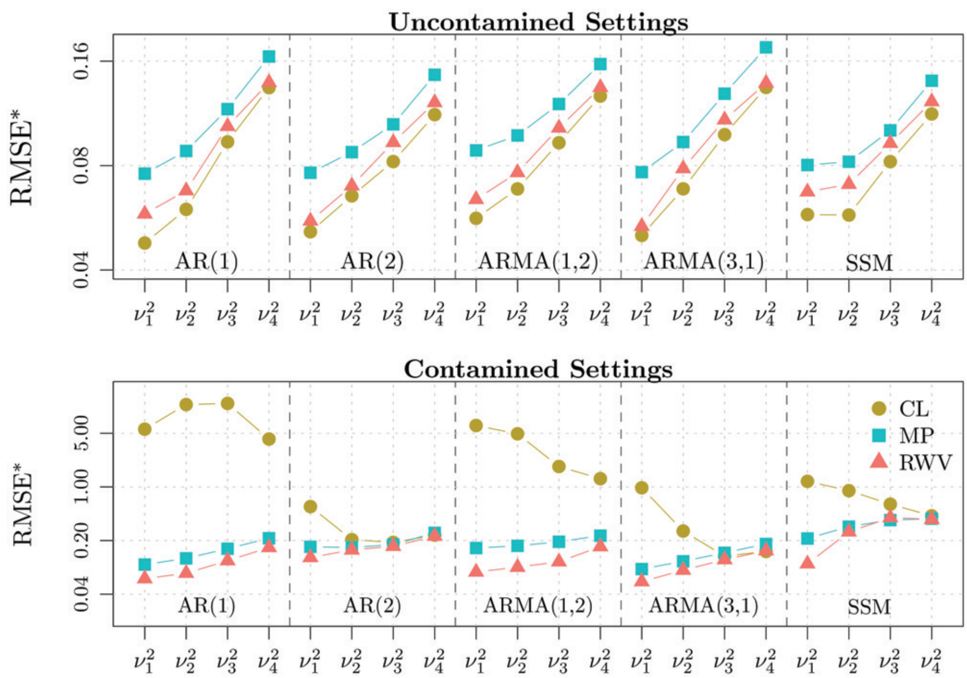

Figure C.1. Top row: RMSE* of the estimators of WV in an uncontaminated setting. Bottom row: RMSE* of the estimators of WV in a contaminated setting.

thereby, based on Condition (C4), verifying the above requirement. Hence, based on Theorem 3 in Wu (2011) we have that the term B_j has the following asymptotic distribution

$$\frac{1}{\sqrt{M_j}}\sum_{t=1}^{M_j}\psi(W_{j,t},\nu_j^2)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\mathbb{E}[D_0^2]\right),$$

where $D_0 := \sum_{t=0}^{\infty}\mathcal{P}_0\psi(W_{j,t},\nu_j^2)$. Since $M_j = \mathcal{O}(T)$, using all the above results we apply Slutsky's theorem to (B.9) to obtain

$$\sqrt{T}(\hat{\nu}_j^2-\nu_j^2)\xrightarrow{\mathcal{D}}\mathcal{N}\left(0,\frac{\mathbb{E}[D_0^2]}{m_j^2}\right).$$

Finally, employing the Crámer-Wold device we can deliver the final result:

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

where $\boldsymbol{V} = \boldsymbol{M}\,\mathbb{E}[\boldsymbol{D}_0\boldsymbol{D}_0^\top]\boldsymbol{M}^\top$ with $\boldsymbol{D}_0 := \sum_{t=0}^{\infty}\mathcal{P}_0\psi(\boldsymbol{W}_t,\boldsymbol{\nu})$ and $\boldsymbol{M} := \mathbb{E}[-\partial/\partial\boldsymbol{\nu}\,\psi(\boldsymbol{W}_t,\boldsymbol{\nu})]$. □

Appendix C: Additional Simulation Studies: WV Estimation

In this appendix we investigate the performance of the proposed M-estimator of WV in (3) which we denote as RWV in this section. For this purpose, we compare it with the standard estimator of WV, denoted as CL, and with the the robust estimator proposed in Mondal and Percival (2012a), denoted as MP. With respect to the latter estimator, we implement the median-type estimator for which most results were available and which was actually used in the simulation studies presented in Mondal and Percival (2012a) for which we specify in more detail below, the outlier processes:

- "Isolated:" outliers occur for an ϵ-proportion of randomly sampled variables from (X_t) by adding (additive) or replacing (replacement) white noise with variance σ_ϵ^2.
- "Scale-based:" the process (X_t) has outliers for certain scales of decomposition, meaning that their behaviour over certain periods of time suffers from contamination. More specifically, let $(W_{j,t})$ represent the wavelet coefficients from the level of decomposition j. Then scale-based outliers occur by multiplying an ϵ-proportion of random *consecutive* portions of $(W_{j,t})$ by σ_ϵ^2.
- "Level-shift:" a constant μ_ϵ is added to an ϵ-proportion of random *consecutive* portions of (X_t).
- "Patchy:" white noise with variance σ_ϵ^2 is added to an ϵ-proportion of random *consecutive* portions of (X_t).

In order to assess the estimators' performance, we test them in the same settings as those used in Section 4 of the main manuscript and we make use of the same measure of statistical performance which, as a reminder, is the RMSE defined as follows:

$$\text{RMSE}^* := \sqrt{\text{med}\left(\frac{\hat{\nu}_j^2-\nu_{j,0}^2}{\nu_{j,0}^2}\right)^2+\text{mad}\left(\frac{\hat{\nu}_j^2}{\nu_{j,0}^2}\right)^2},$$

where $\nu_{j,0}^2$ represents the true model-implied WV for scale τ_j. Figure C.1 represents the logarithm of this measure for all the considered processes and estimators. As can be observed, in the uncontaminated settings the best estimator is obviously the standard estimator CL which is however, in many cases, closely followed by the proposed estimator RWV while the alternative robust estimator MP is generally less precise/efficient than the other two estimators. In the contaminated settings, however, the standard estimator becomes highly biased as expected while the two robust estimators are only marginally affected by the different forms of contamination. Between the latter two, the proposed estimator RWV is nevertheless the best since it reports a lower RMSE* for almost all considered scales of WV.

To conclude, the simulation study highlights how the proposed estimator RWV is the best alternative to the standard estimator CL in the uncontaminated settings while it is overall the best estimator in all the considered contaminated settings.

Remark C.1. In order to estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\nu}}$ (i.e., $\boldsymbol{V}$), the simulation studies in Section 4 use the parametric bootstrap since we assume to know the parametric model we want to estimate. More specifically, for each simulation we

1. estimate the parameters of the model of interest using $\boldsymbol{\Omega} = \mathbf{I}$, i.e. the identity matrix;
2. use the estimated parameters to simulate H time series on which we estimate the WV $\hat{\boldsymbol{\nu}}^{(h)}$;
3. compute the covariance matrix of $\hat{\boldsymbol{\nu}}^{(h)}$, for $h = 1, \dots, H$;
4. use the diagonal elements of this covariance matrix to deliver $\hat{\boldsymbol{V}}$ and define $\boldsymbol{\Omega} := \hat{\boldsymbol{V}}^{-1}$.

Supplementary Material

The supplementary material contains (i) the proof of Proposition 2.1, (ii) a discussion about the choice of the tuning constant, (iii) a formal investigation on the identifiability of wavelet variances, (iv) a formal verification of Conditions (C3) and (C4) for the Haar wavelet filter in causal ARMA models, (v) the asymptotic properties of the RGMWM (consistency and asymptotic normality), (vi) a simulation study to assess the computational efficiency of the RGMWM and (vii) an additional application to inertial sensor stochastic calibration.

Acknowledgments

We thank to Stephen Portnoy and Roger Koenker for their valuable inputs and suggestions throughout the development of this work as well as for their comments on the final draft. We would also like to thank Fabio Trojani for his interesting and helpful considerations. Moreover, we thank Donald Percival and Debashis Mondal for sharing WV code and some data for the case studies presented in the article and in its supplemental material. Moreover, we thank Charles Greenhall for his precious insight into the discussion on the injective relationship between the spectral density and the WV. Finally, we thank Mucyo Karemera, Marc-Olivier Boldi, Samuel Orso and Marco Avella for discussing the results of this work with us.

Funding

Stéphane Guerrier is partially supported by Swiss National Science Foundation grant #176843 and Innosuisse-Boomerang Grant 37308.1 IP-ENG. Roberto Molinari was partially supported by the NSF under Grants SES-1534433 and SES-1853209 and in part by the NCATS-NIH under Grant UL1 TR002014. Maria-Pia Victoria-Feser is partially supported by a Swiss National Science Foundation grant #182684.

References

Abry, P., and Didier, G. (2018), "Wavelet Estimation for Operator Fractional Brownian Motion," *Bernoulli*, 24, 895–928. [1996]

Agostinelli, C., and Bisaglia, L. (2010), "Arfima Processes and Outliers: A Weighted Likelihood Approach," *Journal of Applied Statistics*, 37, 1569–1584. [2006]

Allende, H., and Heiler, S. (1992), "Recursive Generalized m Estimates for Autoregressive Moving-average Models," *Journal of Time Series Analysis*, 13, 1–18. [2006]

Andrews, D. (1999), "Consistent Moment Selection Procedures for Generalized Method of Moments Estimation," *Econometrica*, 67, 543–563. [1997,2002]

Andrews, D. W. K., and Lu, B. (2001), "Consistent Model and Moment Selection Procedures for GMM Estimation with Application to Dynamic Panel Data Models," *Journal of Econometrics*, 101, 123–164. [2002]

Bahamonde, N., and Veiga, H. (2016), "A Robust Closed-form Estimator for the GARCH(1,1) Model," *Journal of Statistical Computation and Simulation*, 86, 1605–1619. [2006]

Bakalli, G., Radi, A., Nassar, S., Guerrier, S., Zhang, Y., and Molinari, R. (2018), "A Two-step Computationally Efficient Procedure for IMU Classification and Calibration," In *2018 IEEE/ION Position, Location and Navigation Symposium (PLANS)*, pp. 534–540. [1996]

Bianco, A. M., Garcia Ben, M., Martinez, E. J., and Yohai, V. J. (1996), "Robust Procedures for Regression Models With ARIMA Errors," In Albert Prat (ed) *Proceedings in Computational Statistics, COMPSTAT 96*, 27–38, Heidelberg: Physica-Verlag. [2006]

Bruce, A. G., Donoho, D. L., Gao, H. Y., and Martin, R. D. (1994), "Denoising and Robust Non-linear Wavelet Analysis," in *SPIE Proceedings, Wavelet Applications*, pp. 325–336. [2006]

Brunnermeier, M. K., and Nagel, S. (2008), "Do Wealth Fluctuations Generate Time-varying Risk Aversion? Micro-evidence on Individuals' Asset Allocation (Digest Summary)," *American Economic Review*, 98, 713–736. [2004]

Burkholder, D. L. (1988), "Sharp Inequalities for Martingales and Stochastic Integrals," *Astérisque*, 157, 75–94. [2007]

Bustos, O. H., and Yohai, V. J. (1986), "Robust Estimates for ARMA Models," *Journal of the American Statistical Association*, 81, 155–168. [2006]

Calvet, L. E., Czellar, V., and Ronchetti, E. (2015), "Robust Filtering," *Journal of the American Statistical Association*, 110, 1591–1606. [2006]

Cantoni, E., and Ronchetti, E. (2001), "Resistant Selection of the Smoothing Parameter for Smoothing Splines," *Statistics and Computing*, 11, 141–146. [2006]

Chang, C. C., and Politis, D. N. (2016), "Robust Autocorrelation Estimation," *Journal of Computational and Graphical Statistics*, 25, 144–166. [2006]

Christmann, A., and Van Messem, A. (2008), "Bouligand Derivatives and Robustness of Support Vector Machines for Regression," *The Journal of Machine Learning Research*, 9, 915–936. [2008]

Cipra, T. (1992), "Robust Exponential Smoothing," *Journal of Forecasting*, 11, 57–69. [2006]

Cipra, T., and Hanzak, T. (2011), "Exponential Smoothing for Time Series With Outliers," *Kybernetika*, 4, 165–178. [2006]

Cizek, P. (2008), "Efficient Robust Estimation of Time-series Regression Models," *Application of Mathematics*, 53, 267–279. [2006]

——— (2016), "Generalized Method of Trimmed Moments," *Journal of Statistical Planning and Inference*, 171, 63–78. [2006]

Clarke, B. R. (1986), "Nonsmooth Analysis and Fréchet Differentiability of m-functionals," *Probability Theory and Related Fields*, 73, 197–209. [2008]

Clausen, P., Skaloud, J., Molinari, R., Lee, J., and Guerrier, S. (2018), "Use of a New Online Calibration Platform With Applications to Inertial Sensors," *IEEE Aerospace and Electronic Systems Magazine*, 30–36. [1996]

Croux, C., X., Gelper, S., and Mahieu, K. (2010), " Robust Exponential Smoothing of Multivariate Time Series," *Journal of Computational Statistics & Data Analysis*, 54, 2999–3006. [2006]

de Luna, X., and Genton, M. G. (2001), "Robust Simulation-based Estimation of ARMA Models," *Journal of Computational and Graphical Statistics*, 10, 370–387. [2006]

Denby, L., and Martin, R. D. (1979), "Robust Estimation of the First-order Autoregressive Parameter," *Journal of the American Statistical Association*, 74, 140–146, [2006]

Dhrymes, J. (2005), "Moments of Truncated (Normal) Distributions," Unpublished note, 10, 95–107. [1999]

Duffie, D., and Singleton, K. J. (1993), "Simulated Moments Estimation of Markov Models of Asset Prices," *Econometrica: Journal of the Econometric Society*, 61, 929–952. [1997]

Durbin, J., and Koopman, S. J. (2012). *Time Series Analysis by State Space Methods*. Oxford: Oxford University Press. [1996]

Dürre, A., Fried, R., and Liboschik, T. . (2015), "Robust Estimation of (Partial) Autocorrelation," *WIREs Computational Statistics*, 7, 205–222. [2006]

Efron, B. (2004), "The Estimation of Prediction Error," *Journal of the American Statistical Association*, 99, 619–632. [2002]

El-Sheimy, N., Hou, H., and Niu, X. (2008), "Analysis and Modeling of Inertial Sensors Using Allan Variance," *IEEE Transactions on Instrumentation and Measurement*, 57, 140–149. [2002]

Fonseca, R. V., Mondal, D., and Zhang, L. (2019), "Wavelet Variances for Heavy-tailed Time Series," *Environmetrics*, 30, e2563. [2000]

Foufoula-Georgiou, E., and Kumar, P. (2014), *Wavelets in Geophysics* (Vol. 4). San Diego, California, USA: Academic Press. [1996]

Fried, R., Einbeck, J., and Gather, U. (2007), "Weighted Repeated Median Smoothing and Filtering," *Journal of the American Statistical Association*, 102, 1300–1308. [2006]

Gallant, A. R., and Tauchen, G. (1996), "Which Moments to Match?" *Econometric Theory*, 12, 657–681. [1997]

Gallegati, M. (2012), "A Wavelet-based Approach to Test for Financial Market Contagion," *Computational Statistics & Data Analysis*, 56, 3491–3497. [1996]

Gelper, S., Fried, R., and Croux, C. (2010), "Robust Forecasting With Exponential and Holt-Winters Smoothing," *Journal of Forecasting*, 29, 285–300. [2006]

Genton, M. G.. and Lucas, A. (2003), "Comprehensive Definitions of Breakdown Points for Independent and Dependent Observations," *Journal of the Royal Statistical Society*, Series B, 65, 81–94. [1999]

Genton, M. G., and Ronchetti, E. (2003), "Robust Indirect Inference," *Journal of the American Statistical Association*, 98, 1–10. [2001,2002,2006]

Gourieroux, C., Monfort, A., and Renault, E. (1993), "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118. [1997,2006]

Granger, C. W. J., and Morris, M. J. (1976), "Time Series Modelling and Interpretation," *Journal of the Royal Statistical Society*, Series A, 246–257. [2006]

Greenhall, C. A. (1998), "Spectral Ambiguity of Allan Variance," *IEEE Transactions on Instrumentation and Measurement*, 47, 623–627. [1998,2001]

Guerrier, S., and Molinari, R. (2016), "On the Identifiability of Latent Models for Dependent Data," *Submitted manuscript*, Available at: *http://arxiv.org/abs/1607.05884*. [2001]

Guerrier, S., Waegli, A., Skaloud, J., and Victoria-Feser, M. P. (2012), "Fault Detection and Isolation in Multiple Mems-imus Configurations," *IEEE Transactions on Aerospace and Electronic Systems*, 48, 2015–2031. [2002]

Guerrier, S., Skaloud, J., Stebler, Y., and Victoria-Feser, M. P. (2013), "Wavelet-variance-based Estimation for Composite Stochastic Processes," *Journal of the American Statistical Association*, 108, 1021–1030. [1996,1998,2001,2005]

Guerrier, S., Molinari, R., and Victoria-Feser, M. P. (2014), "Estimation of Time Series Models via Robust Wavelet Variance," *Austrian Journal of Statistics*, 267–277. [2001]

Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*. New York: John Wiley. [1997,1998]

Hansen, L. P. (1982), "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica: Journal of the Econometric Society*, 1029–1054. [2002,2006]

Härdle, W., and Gasser, T. (1984), " Robust Non-parametric Function Fitting," *Journal of the Royal Statistical Society*, Series B, 46, 42–51. [2006]

Huber, P. J. (1964), "Robust Estimation of a Location Parameter," *The Annals of Mathematical Statistics*, 35, 73–101. [1998]

Huber, P. J. (1981), *Robust Statistics*. New York: John Wiley. [1999]

Jia, X., An, H., Fang, W., Sun, X., and Huang, X. (2015), "How Do Correlations of Crude Oil Prices Co-move? A Grey Correlation-based Wavelet Perspective," *Energy Economics*, 49, 588–598. [1996]

Kim, Y., Lahiri, S. N., and Nordman, D. J. (2013), "A Progressive Block Empirical Likelihood Method for Time Series," *Journal of the American Statistical Association*, 108, 1506–1516. [2000]

Krim, H., and Schick, I. (1999), "Minimax Description Length for Signal Denoising and Optimized Representation," *IEEE Transactions on Information Theory*, 45, 898–908. [2006]

Künsch, H. (1984a), "Infinitesimal Robustness for Autoregressive Processes," *The Annals of Statistics*, 12, 843–863, [1997]

——— (1984b), "Infinitesimal Robustness for Autoregressive Processes," *The Annals of Statistics*, 843–863. [1998,2006]

La Vecchia, D., and Trojani, F. (2010), "Infinitesimal Robustness for Diffusions," *Journal of the American Statistical Association*, 105, 703–712. [2006]

Lévy-Leduc, C., Boistard, H., Moulines, E., Taqqu, M., and Reisen, V. A. (2011), "Robust Estimation of the Scale and the Autocovariance Function of Gaussian Short and Long Range Dependent Processes," *Journal of Time Series Analysis*, 32, 135–156. [2006]

Ma, Y., and Genton, M. (2000), "Highly Robust Estimation of the Autocovariance Function," *Journal of Time Series Analysis*, 21, 663–684. [2006]

Mancini, L., Ronchetti, E., and Trojani, F. (2005), "Optimal Conditionally Unbiased Bounded Influence Inference in Dynamic Location and Scale Models," *Journal of the American Statistical Association*, 100, 628–641. [2006]

Maronna, R. A., Martin, R. D., and Yohai, V. J. (2006), *Robust Statistics: Theory and Methods*. Chichester: Wiley. [1996,1997,1999,2003,2006]

Maronna, R. A., Martin, R. D., Yohai, V. J., and Salibián-Barrera, M. (2019), *Robust Statistics: Theory and Methods (with R)*. Hoboken, NJ Wiley. [1996,1997,2003]

Martin, R. D., and Yohai, V. J. (1986), "Influence Functionals for Time Series," *The Annals of Statistics*, 14, 781–818. [2006]

Masreliez, C. J., and Martin, R. D. (1977), "Robust Bayesian Estimation for the Linear Model and Robustifying the Kalman Filter," *IEEE Transactions on Automatic Control*, 22, 361–371. [2006]

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration," *Econometrica: Journal of the Econometric Society*, 57, 995–1026. [1997]

Molinares, F. F., Reisen, V. A., and Cribari-Neto, F. (2009), "Robust Estimation in Longmemory Processes Under Additive Outliers," *Journal of Statistical Planning and Inference*, 139, 2511–2525, [2006]

Mondal, D., and Percival, D. B. (2012a), "M-estimation of Wavelet Variance," *Annals of the Institute of Statistical Mathematics*, 64, 27–53. [1997,1998,2002,2010]

——— (2012b), "Wavelet Variance Analysis for Random Fields on a Regular Lattice," *IEEE Transactions on Image Processing*, 21, 537–549. [2000,2002]

Muler, N., Peña, D., and Yohai, V. J. (2009), "Robust Estimation for ARMA Models," *The Annals of Statistics*, 37, 816–840. [1997,2006]

Nason, G. P., Von Sachs, R., and Kroisandt, G. (2000), "Wavelet Processes and Adaptive Estimation of the Evolutionary Wavelet Spectrum," *Journal of the Royal Statistical Society*, Series B, 62, 271–292. [2002]

Newey, W. K., and McFadden, D. (1994), "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics*, 4, 2111–2245. [2001]

Ortelli, C., and Trojani, F. (2005), "Robust Efficient Method of Moments," *Journal of Econometrics*, 128, 69–97. [2006]

Owen, A. L., and Wu, S. (2007). Financial shocks and worry about the future. *Empirical Economics*, 33, 515–530. [2004]

Pankratz, A. (2012), *Forecasting with Dynamic Regression Models* (vol. 935). John Wiley & Sons. [2004]

Percival, D. B. (1995), "On Estimation of the Wavelet Variance," *Biometrika*, 82, 619–631. [1996,1997,1998]

——— (2016), "A Wavelet Perspective on the Allan Variance," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, 63, 538–554. [1999]

D. B. Percival and A. T. Walden. *Wavelet Methods for Time Series Analysis* (Vol. 4). Cambridge University Press, 2006. [1996,1998,2000]

——— (2000), *Wavelet Methods for Time Series Analysis*. Cambridge Series in Statistical and Probabilistic Mathematics. New York: Cambridge University Press. [2000]

Radi, A., Bakalli, G., Guerrier, S., El-Sheimy, N., Sesay, A. B., and Molinari, R. (2019), " A Multisignal Wavelet Variance-based Framework for Inertial Sensor Stochastic Error Modeling," *IEEE Transactions on Instrumentation and Measurement*, 68, 4924–4936. doi: 10.1109/TIM.2019.2899535. [1996]

Reisen, V. A., and Molinares, F. F. . "Robust Estimation in Time Series With Long and Short Memory Properties," *Annales Mathematicae et Informaticae*, 39, 207–224, 2012. [2006]

Renaud, O. (2002), "Sensitivity and Other Properties of Wavelet Regression and Density Estimators," *Statistica Sinica*, 12, 1275–1290. [2006]

Rioul, O., and Duhamel, P. (1992), "Fast Algorithms for Discrete and Continuous Wavelet Transforms," *IEEE Transactions on Information Theory*, 38, 569–586. [2002]

Ronchetti, E., and Trojani, F. (2001), "Robust Inference with GMM Estimators," *Journal of Econometrics*, 101, 37–69. [2006]

Rousseeuw, P. J., and Croux, C. (1993), "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association*, 88, 1273–1283. [2006]

Ruckdeschel, P., Spangl, B., and Pupashenko, D. (2014), "Robust Kalman Tracking and Smoothing With Propagating and Non-propagating Outliers," *Statistical Papers*, 55. [2006]

Sardy, S., Tseng, P., and Bruce, A. G. (2001), "Robust Wavelet Denoising," *IEEE Transactions on Signal Processing*, 49, 1146–1152. [2006]

Sarnaglia, A. J. Q., Reisen, V. A., and Lévy-Leduc, C. (2010a), "Robust Estimation of Periodic Autoregressive Processes in the Presence of Additive Outliers," *Journal of Multivariate Analysis*, 101, 2168–2183. [2006]

——— (2010b), "Robust Estimation of Periodic Autoregressive Processes in the Presence of Additive Outliers," *Journal of Multivariate Analysis*, 101, 2168–2183. [2002]

Scholtes, S. (2012), *Introduction to Piecewise Differentiable Equations*. Springer Science & Business Media. New York: Springer-Verlag. [2008]

Serroukh, A., Walden, A. T., and Percival, D. B. (2000), "Statistical Properties and Uses of the Wavelet Variance Estimator for the Scale Analysis of Time Series," *Journal of the American Statistical Association*, 95, 184–196. [1997,1998]

Shumway, R. H., and Stoffer, D. S. (2013), *Time Series Analysis and Its Applications*. Springer, New York, NY, USA: Springer Science & Business Media. [1996]

Slacalek, J., and Sommer, M. (2012), "What Drives Household Saving?" Unpublished note, 10, 95–107. [2004,2006]

Smith, A. A. (1993), "Estimating Nonlinear Time-series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84, [1997]

Stocchi, M., and Marchesi, M. (2018), "Fast Wavelet Transform Assisted Predictors of Streaming Time Series," *Digital Signal Processing*, 77, 5–12. [2002]

Tukey, J. W. (1977), *Exploratory Data Analysis*, Reading, MA: Addison-Wesley. Preliminary editions: 1970, 1971. [2006]

Van der Vaart, A. (2000), *Asymptotic Statistics* (Vol. 3), Cambridge, United Kingdom: Cambridge University Press. [2008,2009]

Wooldridge, J. M. (1994), "Estimation and Inference for Dependent Processes," *Handbook of Econometrics*, 4, 2639–2738. [2007,2009]

Wu, W. (2005), "Nonlinear System Theory: Another Look at Dependence," *Proceedings of the National Academy of Sciences*, 102, 14150–14154. [1999]

——— (2011), "Asymptotic Theory for Stationary Processes," *Statistics and its Interface*, 4, 207–226. [1999,2000,2007,2009,2010]

Wu, W. B., and Zhou, Z. (2011), "Gaussian Approximations for Non-stationary Multiple Time Series," *Statistica Sinica*, 21, 1397–1413. [2000]

Xie, S., and Krishnan, S. (2013), "Wavelet-based Sparse Functional Linear Model With Applications to Eegs Seizure Detection and Epilepsy Diagnosis," *Medical & Biological Engineering & Computing*, 51, 49–60. [1996]

Xu, H., Guerrier, S., Molinari, R., and Karemera, M. (2019), "Multivariate Signal Modelling With Applications to Inertial Sensor Calibration," *IEEE Transactions on Signal Processing*, 67, 5143–5152. [2002]

Yohai, V. J., and Zamar, R. H. (1988), "High Breakdown Point Estimates of Regression by Means of the Minimization of an Efficient Scale," *Journal of the American Statistical Association*, 83, 406–413. [2006]

Zhang, N. F. (2008), "Allan Variance of Time Series Models for Measurement Data," *Metrologia*, 45, 549–561. [2001]

Zhang, X., and Guerrier, S. (2020), "Covariance Penalty Based Model Selection Criterion for Indirect Estimators," Submitted Working Paper, pp. 1–30. [2002]

Zhou, Z., and Wu, W. B. (2009), "Local Linear Quantile Estimation for Nonstationary Time Series," *The Annals of Statistics*, 37, 2696–2729. [2000]

Zhang, D., and Wu, W. (2017), "Gaussian Approximation for High Dimensional Time Series," *The Annals of Statistics*, 45, 1895–1919. [2000]

Ziaja, A., Antoniadou, I., Barszcz, T., Staszewski, W. J., and Worden, K. (2016), "Fault Detection in Rolling Element Bearings Using Wavelet-Based Variance Analysis and Novelty Detection," *Journal of Vibration and Control*, 22, 396–411. [1996]