Appendix A

Here we define the processes listed in Sec. II. For each process we define $\nu_{\tau_j}(\boldsymbol{\theta})$ which is the Haar WV implied by it, where $\tau_j = 2^j, j = 1, \ldots, J$ are the dyadic scales at which the WV is computed. The processes are the following:

(P1) *Gaussian White Noise* (WN) with parameter $\sigma^2 \in \mathbb{R}^+$. This process is defined as

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

and has the following implied Haar WV at scale τ_j:

$$\nu_{\tau_j}\left(\sigma^2\right) = \frac{\sigma^2}{\tau_j}.$$

(P2) *Quantization Noise* (QN) (see e.g. [5]) with parameter $Q^2 \in \mathbb{R}^+$. This process has a PSD of the form

$$S_X(f) = 4Q^2 \sin^2\left(\frac{\pi f}{\Delta t}\right)\Delta t, \;\; f < \frac{\Delta t}{2}$$

and has the following implied Haar WV at scale τ_j:

$$\nu_{\tau_j}\left(Q^2\right) = \frac{6Q^2}{\tau_j^2}.$$

(P3) *Drift* with parameter $\omega \in \Omega$ where Ω is either $\mathbb{R}^+$ or $\mathbb{R}^-$. This process is defined as:

$$X_t = \omega t$$

and has the following implied Haar WV at scale τ_j:

$$\nu_{\tau_j}\left(\omega\right) = \frac{\tau_j^2\omega^2}{16}.$$

(P4) *Random walk* (RW) with parameter $\gamma^2 \in \mathbb{R}^+$. This process is defined as:

$$X_t = \sum_{t=1}^{T} \gamma Z_t$$

where $Z_t \overset{iid}{\sim} \mathcal{N}(0,1)$ and has the following implied Haar WV at scale τ_j:

$$\nu_{\tau_j}\left(\gamma^2\right) = \frac{\left(\tau_j^2 + 2\right)\gamma^2}{12\tau_j}.$$

(P5) *Auto-Regressive* AR(1) process with parameter $\rho \in (-1, +1)$ and $\upsilon^2 \in \mathbb{R}^+$. This process is defined as:

$$X_t = \rho X_{t-1} + \upsilon Z_t \;\; \text{where} \;\; Z_t \overset{iid}{\sim} \mathcal{N}(0,1)$$

and it has the following implied Haar WV at scale τ_j:

$$\nu_{\tau_j}\left(\rho, \upsilon^2\right) = \frac{\left(\frac{\tau_j}{2} - 3\rho - \frac{\tau_j\rho^2}{2} + 4\rho^{\frac{\tau_j}{2}+1} - \rho^{\tau_j+1}\right)\upsilon^2}{\frac{\tau_j^2}{2}\left(1-\rho\right)^2\left(1-\rho^2\right)}.$$

Appendix B

Proof of Lemma 1: Under the Gaussian assumption for processes (P1) and (P5) we know that these processes have finite moments as does (P4) after first-differencing. Process (P1) and the first-differencing of process (P4) have independent observations and process (P5) is strongly mixing (see [1]). Moreover, process (P2) follows a triangular distribution with finite moments (see App. K) and, together with the first difference of process (P3), they have no dependence structure on previous observations. Therefore all processes from (P1) to (P5) fulfill the criteria of Lemma 1 by definition. ■

Appendix C

Let us define the AV and, for this purpose, let $\left(\overline{Y}_t\left(\kappa_j\right)\right)$ be the sample average of $\kappa_j = 2^{j-1} = \tau_j/2$ consecutive observations of (Y_t), i.e.

$$\overline{Y}_t\left(\kappa_j\right) = \frac{1}{\kappa_j}\sum_{i=0}^{\kappa_j - 1} Y_{t-i}\,.$$

The AV at scale τ_j, denoted as $\phi^2_{\tau_j}$, aims to measure the amount of variability contained in the process $\left(\overline{Y}_t\left(\kappa_j\right)\right)$. This quantity is defined as half of the expectation of squared differences between adjacent non-overlapping $\overline{Y}_t\left(\kappa_j\right)$

$$\phi^2_{\tau_j} = \frac{1}{2}\mathbb{E}\left[\left(\overline{Y}_t\left(\kappa_j\right) - \overline{Y}_{t-\kappa_j}\left(\kappa_j\right)\right)^2\right].$$

It should be noted that $\phi^2_{\tau_j}$ is assumed not to depend on time. The condition for this property to hold is that the first-order difference of the series (Y_t) is stationary (see [6]), which is indeed the case for the processes listed in Sec. II. Several estimators of the AV have been proposed in the literature (see e.g. [6] and the references therein for a detailed discussion). The asymptotically most efficient among them is based on the maximum-overlap estimator proposed in [3] and can be computed from a realization (y_t) using

$$\hat{\phi}^2_{\tau_j} = \frac{1}{2\left(T - \tau_j + 1\right)}\sum_{t=\tau_j}^{T}\left(\overline{y}_t(\kappa_j) - \overline{y}_{t-\kappa_j}(\kappa_j)\right)^2. \quad \text{(D-1)}$$

On the other hand, the WV can be defined as

$$\nu_{\tau_j} \equiv \text{var}\left(W_{j,t}\right)$$

where $(W_{j,t})$ are the wavelet coefficients issued from a *Maximal Overlap* (MO-) *Discrete Wavelet Transform* (DWT) obtained by filtering (Y_t) with the corresponding wavelet filter at level j (see [3, 7]), i.e.

$$W_{j,t} = \sum_{l=0}^{\tau_j - 1}\tilde{h}_{j,l} Y_{t-l}, \; t \in \mathbb{Z}.$$

The MODWT wavelet filters $\{\tilde{h}_{j,l}\}$ are actually a rescaled version of the DWT filter $h_{j,l} = 1/2^{j/2}$, i.e. $\tilde{h}_{j,l} = h_{j,l}/2^{j/2} = 1/\tau_j$. As in the case of the AV, the WV are assumed not to depend on time and, when using the Haar filter, the condition for this property to hold is the same as for the AV (i.e. the first order difference of the series (Y_t) is stationary). Based on this, the WV estimator is defined as

$$\hat{\nu}_{\tau_j} = \frac{1}{M_j}\sum_{t=\tau_j}^{T} W^2_{j,t}.$$

Appendix D

Proof of Corollary 1: Let us consider the Haar-based WV estimator and re-express it as

$$\begin{aligned}\hat{\nu}_{\tau_j} &= \frac{1}{M_j}\sum_{t=\tau_j}^{T} W_{j,t}^2 \\ &= \frac{1}{M_j}\sum_{t=\tau_j}^{T}\left(\sum_{t=1}^{\kappa_j}\frac{1}{\tau_j}y_t - \sum_{t=1}^{\kappa_j}\frac{1}{\tau_j}y_{t-\kappa_j}\right)^2 \\ &= \frac{1}{M_j\tau_j^2}\sum_{t=\tau_j}^{T}\left(\sum_{t=1}^{\kappa_j}y_t - \sum_{t=1}^{\kappa_j}y_{t-\kappa_j}\right)^2\end{aligned}$$

By the same logic, we can also re-express the AV estimator is as follows

$$\begin{aligned}\hat{\phi}_{\tau_j}^2 &= \frac{1}{2\left(T-\tau_j+1\right)}\sum_{t=\tau_j}^{T}\left(\overline{y}_t(\kappa_j) - \overline{y}_{t-\kappa_j}(\kappa_j)\right)^2 \\ &= \frac{1}{2M_j}\sum_{t=\tau_j}^{T}\left(\frac{2}{\tau_j}\sum_{t=1}^{\kappa_j}y_t - \frac{2}{\tau_j}\sum_{t=1}^{\kappa_j}y_{t-\kappa_j}\right)^2 \\ &= \frac{2}{M_j\tau_j^2}\sum_{t=\tau_j}^{T}\left(\sum_{t=1}^{\kappa_j}y_t - \sum_{t=1}^{\kappa_j}y_{t-\kappa_j}\right)^2\end{aligned}$$

since $M_j = (T-\tau_j+1)$. By comparing the two expressions we can see that $\hat{\phi}_{\tau_j}^2 = 2\hat{\nu}_{\tau_j}$. Under the conditions of Lemma 1 we know that $\hat{\nu}_{\tau_j}$ is a consistent estimator (i.e. $\hat{\nu}_{\tau_j} \xmapsto{\mathcal{P}} \nu_{\tau_j}$). Given the continuous mapping theorem we have that $2\hat{\nu}_{\tau_j} \xmapsto{\mathcal{P}} 2\nu_{\tau_j}$ and, since $\phi_{\tau_j}^2 = 2\nu_{\tau_j}$, we finally have $\hat{\phi}_{\tau_j}^2 = 2\hat{\nu}_{\tau_j} \xmapsto{\mathcal{P}} 2\nu_{\tau_j} = \phi_{\tau_j}^2$. Therefore, $\hat{\phi}_{\tau_j}^2$ is a consistent estimator (i.e. $\hat{\phi}_{\tau_j}^2 \xmapsto{\mathcal{P}} \phi_{\tau_j}^2$). ■

Appendix E

Proof of Corollary 2: Consider the vector $\hat{\boldsymbol{\phi}} = [\hat{\phi}_{\tau_j}]_{\tau_j\in\boldsymbol{\eta}}$. Since $\hat{\boldsymbol{\phi}}_{\tau_j}$ is consistent by Corollary 1, i.e. $\hat{\boldsymbol{\phi}} \xmapsto{\mathcal{P}} \boldsymbol{\phi}$, by the continuous mapping theorem we also have that $h(\hat{\boldsymbol{\phi}}) \xmapsto{\mathcal{P}} h\left(\boldsymbol{\phi}\right)$ if $h\left(\cdot\right)$ is a continuous function (see e.g. Theorem 1.14 of [2]). For $\hat{\theta}_{AV}$ defined by (3), we may write $\hat{\theta}_{AV} = k(\hat{\boldsymbol{\phi}})$. Then for (P1), (P2) or (P3), $k\left(\cdot\right)$ is continuous (see Sec. III), such that $\hat{\theta}_{AV} = k(\hat{\boldsymbol{\phi}}) \xmapsto{\mathcal{P}} k\left(\boldsymbol{\phi}\right)$. Given Corollary 1 and since there is a direct relationship between AV and WV (i.e. $\phi_{\tau_j}^2 = 2\nu_{\tau_j}$), we can use the expressions for the WV of these processes and express them in the form

$$\phi_{\tau_j}^2(\boldsymbol{\theta}) = c\boldsymbol{\theta}^{\alpha}\tau_j^{\lambda}$$

where for process (P1), for example, $c=2$, $\alpha=1$ and $\lambda=-1$. Taking the logarithm we have

$$\log(\phi_{\tau_j}(\boldsymbol{\theta})) = \frac{1}{2}\log(c) + \frac{\alpha}{2}\log(\boldsymbol{\theta}) + \frac{\lambda}{2}\log(\tau_j). \quad \text{(D-2)}$$

Based on the estimator defined in (2), we solve

$$\frac{\partial}{\partial\boldsymbol{\theta}}\sum_{\tau_j\in\boldsymbol{\eta}}\left(\log(\hat{\phi}_{\tau_j}) - \frac{1}{2}\log(c) - \frac{\alpha}{2}\log(\boldsymbol{\theta}) - \frac{\lambda}{2}\log(\tau_j)\right)^2 = 0$$

for $\boldsymbol{\theta}$ which gives

$$\hat{\boldsymbol{\theta}}_{AV} = c^{-1/\alpha}\exp\left(\frac{2}{|\boldsymbol{\tau}|\alpha}\sum_{\tau_j\in\boldsymbol{\eta}}\left(\log(\hat{\phi}_{\tau_j}) - \frac{\lambda}{2}\log(\tau_j)\right)\right)$$

which takes the general form given in (3) that we have defined as $k(\hat{\boldsymbol{\phi}})$ where $\boldsymbol{\phi} \equiv \boldsymbol{\phi}(\boldsymbol{\theta})$. If we replace $\boldsymbol{\phi}$ instead of $\hat{\boldsymbol{\phi}}$ in $k(\hat{\boldsymbol{\phi}})$ we get exactly $\boldsymbol{\theta}$ which shows that $k(\boldsymbol{\phi}) = \boldsymbol{\theta}$. Hence, we have $\hat{\theta}_{AV} = k(\hat{\boldsymbol{\phi}}) \xmapsto{\mathcal{P}} k(\boldsymbol{\phi}) = \boldsymbol{\theta}$ thus concluding the proof. ■

Appendix F

Proof of Corollary 3: First, we apply the same approach used in the proof of Theorem 2. Since $\hat{\gamma}_{AV}^2$ as defined in (4) can be written as $\hat{\gamma}_{AV}^2 = k(\hat{\boldsymbol{\phi}})$, with $k(\cdot)$ a continuous function, and since $\hat{\boldsymbol{\phi}} \xmapsto{\mathcal{P}} \boldsymbol{\phi} = \left[\phi_{\tau_j}\right]_{\tau_j\in\boldsymbol{\tau}}$, with $\phi_{\tau_j} = \sqrt{\gamma^2\left(\frac{\tau_j^2+2}{6\tau_j}\right)}$, then by using (4) we have

$$\begin{aligned}\hat{\gamma}_{AV}^2 &\xmapsto{\mathcal{P}} 6\exp\left\{\frac{2}{|\boldsymbol{\tau}|}\sum_{\tau_j\in\boldsymbol{\tau}}\left[\log\left(\phi_{\tau_j}\right) - \frac{1}{2}\log\left(\tau_j\right)\right]\right\} \\ &= 6\exp\left\{\frac{2}{|\boldsymbol{\tau}|}\sum_{\tau_j\in\boldsymbol{\tau}}\left[\log\left(\sqrt{\gamma^2\left(\frac{\tau_j^2+2}{6\tau_j^2}\right)}\right)\right]\right\} \\ &= \tilde{\gamma}^2.\end{aligned}$$

Now, one can always write

$$\begin{aligned}\gamma^2 &= 6\exp\left\{\frac{2}{|\boldsymbol{\tau}|}\sum_{\tau_j\in\boldsymbol{\tau}}\left[\log\left(\frac{\gamma}{\sqrt{6}}\right)\right]\right\} \\ &= 6\exp\left\{\frac{2}{|\boldsymbol{\tau}|}\sum_{\tau_j\in\boldsymbol{\tau}}\left[\log\left(\sqrt{\gamma^2\left(\frac{\tau_j^2+2}{6\tau_j^2}\right)}\right) - b_{AV}(\tau_j)\right]\right\}\end{aligned}$$

and

$$\begin{aligned}b_{AV}(\tau_j) &= \log\left(\sqrt{\gamma^2\left(\frac{\tau_j^2+2}{6\tau_j^2}\right)}\right) - \log\left(\frac{\gamma}{\sqrt{6}}\right) \\ &= \log\left(\sqrt{1+\frac{2}{\tau_j^2}}\right) > 0\end{aligned}$$

since $\tau_J < \infty$. Hence $\tilde{\gamma}^2 > \gamma_0^2$ which concludes the proof. ■

Appendix G

Proof of Corollary 4: The proof is straightforward since any combination of a process chosen among (P1), (P2), (P3) and (P4) with any other different process(es) satisfying assumption (A3) does not have a linear representation of the form defined in (1). So we may apply the same rational as in Corollary 3 to show the inconsistency of the estimator for θ. Therefore, $\hat{\theta}_l$ cannot be a consistent estimator of θ. ■

Appendix H

Proof of Lemma 2: The proof is straightforward. Indeed, for processes (P1), (P2) or (P3), we have that

$$\begin{aligned}\hat{\theta}_{AV} &= \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{\tau_j\in\boldsymbol{\tau}} \varepsilon_{\tau_j}^2 \\ &= \underset{\theta\in\Theta}{\operatorname{argmin}} \sum_{\tau_j\in\boldsymbol{\tau}} \left[\log\left(\hat{\phi}_{\tau_j}\right) - \log\left(\phi_{\tau_j}(\theta)\right)\right]^2 \\ &= \underset{\theta\in\Theta}{\operatorname{argmin}} \left(\hat{\boldsymbol{\nu}}_{\boldsymbol{\tau}} - \boldsymbol{\nu}_{\boldsymbol{\tau}}(\theta)\right)^T \boldsymbol{\Omega}^\star \left(\hat{\boldsymbol{\nu}}_{\boldsymbol{\tau}} - \boldsymbol{\nu}_{\boldsymbol{\tau}}(\theta)\right)\end{aligned}$$

where the elements of $\boldsymbol{\Omega}^\star$ are given by (8). Therefore, we have that $\hat{\theta}_{AV} = \hat{\theta}_{\boldsymbol{\nu}}$ as they are the solutions of the same problem. ■

Appendix I

Proof of Corollary 5: Since $\hat{\theta}_{AV} = \hat{\theta}_{\boldsymbol{\nu}}^\star$ with $\hat{\theta}_{\boldsymbol{\nu}}^\star$ being the GMWM estimator based on $\boldsymbol{\Omega}^\star$, and since the GMWM estimator with minimal asymptotic variance is $\hat{\theta}_{\boldsymbol{\nu}}^\circ$ where $\boldsymbol{\Omega}^\circ = \mathbf{V}^{-1} \neq \boldsymbol{\Omega}^\star$, we verify the result of Corollary 5. ■

Appendix J

In [4] it was shown that the WV estimator $\hat{\boldsymbol{\nu}}$ is asymptotically normally distributed with covariance matrix $\mathbf{V}$ for which an estimator was also given. The latter is however extremely computationally demanding and a parametric bootstrap estimator is often preferred. Since $\hat{\boldsymbol{\nu}}$ is a consistent estimator, it is possible to construct an asymptotically optimal GMWM estimator based on $\boldsymbol{\Omega}^\circ$ as follows:

1) Compute the GMWM estimator, say $\hat{\theta}_{\boldsymbol{\nu}}$, based on any positive definite $\boldsymbol{\Omega}$;
2) Compute a parametric bootstrap estimator of $\mathbf{V}$, say $\widehat{\mathbf{V}}$, based on $\hat{\theta}_{\boldsymbol{\nu}}$ whose consistency follows from the consistency of $\hat{\theta}_{\boldsymbol{\nu}}$;
3) Compute the GMWM estimator, say $\hat{\theta}_{\boldsymbol{\nu}}^\circ$, based on $\boldsymbol{\Omega}^\circ = \widehat{\mathbf{V}}^{-1}$. By the continuity of matrix inversion we have that $\boldsymbol{\Omega}^\circ \to \mathbf{V}^{-1}$ and therefore we have that

$$\lim_{T\to\infty} \frac{\operatorname{var}\left(\hat{\theta}_{\boldsymbol{\nu}}^\circ\right)}{\operatorname{var}\left(\hat{\theta}_{AV}\right)} < 1.$$

Appendix K

Triangular distribution of quantization noise: A quantization noise process (X_t) is defined as follows:

$$X_t = \sqrt{Q_0^2}\dot{U}_t\delta t$$

where δt is the time difference between consecutive observations and

$$\dot{U}_t = \frac{U_{t+1} - U_t}{\delta t}$$

with $U_t = \sqrt{12}U_t^*$ and $U_t^* \overset{i.i.d.}{\sim} \mathcal{U}(0,1)$ (i.e. U_t^* follows a standard uniform distribution).

Since δt is supposed to be the same for any time t (i.e. the sampling frequency is the same for all observations), we have

$$\begin{aligned}X_t &= \sqrt{Q_0^2}\frac{U_{t+1} - U_t}{\delta t}\delta t \\ &= \sqrt{Q_0^2}\sqrt{12}(U_{t+1}^* - U_t^*) = \sqrt{12Q_0^2}R_t\end{aligned}$$

where $R_t \equiv U_{t+1}^* - U_t^*$ is the difference between two standard uniform variables which by definition follows a standard triangular distribution. Hence, (X_t) follows a rescaled standard triangular distribution which has finite moments since $Q_0^2 < \infty$. ■

References

[1] Andrews, D.W.: First order autoregressive processes and strong mixing. Yale University (1983)
[2] DasGupta, A.: Asymptotic Theory of Statistics and Probability. Springer (2008)
[3] Greenhall, C.A.: Recipes for Degrees of Freedom of Frequency Stability Estimators. IEEE Transactions on Instrumentation and Measurement **40**(6), 994–999 (1991)
[4] Guerrier, S., Skaloud, J., Stebler, Y., Victoria-Feser, M.P.: Wavelet-variance-based estimation for composite stochastic processes. Journal of the American Statistical Association **108**(503), 1021–1030 (2013)
[5] Papoulis, A.: Probability, Random Variables, and Stochastic Process, third edition edn. McGraw-Hill, Inc. (1991)
[6] Percival, D.B., Guttorp, P.: Long-Memory Processes, the Allan Variance and Wavelets. Wavelets in geophysics **4**, 325–344 (1994)
[7] Percival, D.B., Walden, A.T.: Wavelet Methods for Time Series Analysis. Cambridge Series in Statistical and Probabilistic Mathematics. Cambridge University Press, 32 Avenue of the Americas, New York, NY 10013-2473, USA (2000)