APPENDIX

A. *Proofs of technical results*

Proof of Theorem 1

Under Assumptions A to C, Lemma 3.1 of [1] implies that that

$$\sqrt{T_i}\left(\tilde{\boldsymbol{\vartheta}}_i - \boldsymbol{\vartheta}_i\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Gamma}(\boldsymbol{\vartheta}_i)\right),$$

conditionality on $\boldsymbol{\vartheta}_i$, where $\boldsymbol{\Gamma}(\boldsymbol{\vartheta})$ is a continuous function in $\boldsymbol{\vartheta}$. Thus, we have

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}}^{\circ}\right] = \sum_{i=1}^{K} w_i \mathbb{E}\left[\tilde{\boldsymbol{\vartheta}}_i\right] = \mathbb{E}\left[\tilde{\boldsymbol{\vartheta}}_i\right] = \mathbb{E}\left[\mathbb{E}\left[\tilde{\boldsymbol{\vartheta}}_i | \boldsymbol{\vartheta}_i\right]\right] = \mathbb{E}\left[\boldsymbol{\vartheta}_i\right] + o(T^{-1/2}) = \boldsymbol{\theta}_0^* + o(T^{-1/2}),$$

where $\boldsymbol{\theta}_0^* = \mathbb{E}\left[\boldsymbol{\vartheta}_i\right]$ and where the last equality is due to the asymptotic normality of $\tilde{\boldsymbol{\vartheta}}_i$ and the compactness of $\boldsymbol{\Theta}$. Moreover, using the expectation of quadratic forms, we have

$$\begin{aligned} \mathbb{E}\left[||\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*||_2^2\right] &= \operatorname{tr}\left\{\mathbb{E}\left[\left(\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*\right)\left(\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*\right)^{\top}\right]\right\} \\ &= \operatorname{tr}\left\{\operatorname{var}\left(\hat{\boldsymbol{\theta}}^{\circ}\right)\right\} + \operatorname{tr}\left\{\mathbb{E}\left[\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*\right]\mathbb{E}^{\top}\left[\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*\right]\right\}. \end{aligned} \tag{A.1}$$

Focusing on the first term in (A.1) and using the law of total variance, the variance of $\hat{\boldsymbol{\theta}}^{\circ}$ is of the following order elementwise:

$$\begin{aligned} \operatorname{var}\left(\hat{\boldsymbol{\theta}}^{\circ}\right) &= \sum_{i=1}^{K} w_i^2 \operatorname{var}\left(\tilde{\boldsymbol{\vartheta}}_i\right) = \sum_{i=1}^{K} w_i^2 \left\{\mathbb{E}\left[\operatorname{var}\left(\tilde{\boldsymbol{\vartheta}}_i | \boldsymbol{\vartheta}_i\right)\right] + \operatorname{var}\left(\mathbb{E}\left[\tilde{\boldsymbol{\vartheta}}_i | \boldsymbol{\vartheta}_i\right]\right)\right\} \\ &= \sum_{i=1}^{K} w_i^2 \left\{\operatorname{var}\left(\boldsymbol{\vartheta}_i\right) + o(T^{-1/2})\right\} = \mathcal{O}(K^{-1}). \end{aligned}$$

The intermediate results in the above development are based on the asymptotic normality of $\tilde{\boldsymbol{\vartheta}}_i$ (which provides orders of convergence for the conditional expectations and variances), the continuity of $\boldsymbol{\Gamma}(\boldsymbol{\vartheta})$, the assumption on the weights (w_i) as well as the compactness of $\boldsymbol{\Theta}$. As for the second term in (A.1), elementwise we have the order

$$\mathbb{E}\left[\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*\right]\mathbb{E}\left[\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*\right]^{\top} = o(T^{-1}).$$

Combining the above results back within (A.1) we obtain

$$\mathbb{E}\left[||\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*||_2^2\right] = o(T^{-1}) + \mathcal{O}(K^{-1}) = o(1).$$

By Markov's inequality, for any $M > 0$, we have

$$\Pr\left\{||\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*||_2 \geq M\right\} \leq \frac{1}{M^2}\mathbb{E}\left[||\hat{\boldsymbol{\theta}}^{\circ} - \boldsymbol{\theta}_0^*||_2^2\right] = o(1).$$

Thus, we have $\hat{\boldsymbol{\theta}}^{\circ} \xrightarrow{P} \boldsymbol{\theta}_0^*$ as $K, T \to \infty$ proving consistency of $\hat{\boldsymbol{\theta}}^{\circ}$ for a target value represented by $\boldsymbol{\theta}_0^*$.

To demonstrate that, in general, $\boldsymbol{\theta}_0^* = \mathbb{E}\left[\boldsymbol{\vartheta}_i\right] \neq \boldsymbol{\theta}_0$, we provide the following example. Consider a first-order moving-average process defined as

$$X_t^{(i)} = \vartheta_i \varepsilon_{t-1}^{(i)} + \varepsilon_t^{(i)},$$

where $\vartheta_i \overset{iid}{\sim} \mathcal{U}(0, 1/4)$, i.e. a uniform distribution between 0 and 1/4, and $\varepsilon_t^{(i)}$ is an *iid* sequence such that $\mathbb{E}[\varepsilon_t^{(i)}] = 0$ and $\text{var}[\varepsilon_t^{(i)}] = 1$. The theoretical WV at the first scale of such a process (conditionally on ϑ_i) is given by $\nu_1(\theta) = \frac{\theta^2 - \theta + 1}{2}$ and, thus, we have $\mu_1 := \mathbb{E}\left[\nu_1(\vartheta)\right] = \frac{43}{96}$. Considering only the first scale, we have

$$\theta_0 = \underset{\theta}{\text{argmin}}\ \mathbb{E}\left[\{\nu_1(\vartheta) - \nu_1(\theta)\}^2\right] = \underset{\theta}{\text{argmin}}\ \{\mu_1 - \nu_1(\theta)\}^2 = \frac{6 - \sqrt{21}}{12} < \mathbb{E}[\vartheta_i] = \frac{1}{8},$$

thus providing proof that there exist circumstances where $\boldsymbol{\theta}_0^* \neq \boldsymbol{\theta}_0$.

□

Proof of Theorem 2

Let $\bar{\nu}_j$ denote the j^{th} element (for $j \in \{1, \ldots, J\}$) of the vector $\bar{\boldsymbol{\nu}} := \sum_{i=1}^K w_i \hat{\boldsymbol{\nu}}_i$. Similarly, we let $\hat{\nu}_{i,j}$ denote the j^{th} element of $\hat{\boldsymbol{\nu}}_i$. Since $\hat{\nu}_{i,j}$ is an unbiased estimator and recalling the properties of the weights (w_i), we have

$$\mathbb{E}\left[\bar{\nu}_j\right] = \sum_{i=1}^K w_i \mathbb{E}\left[\hat{\nu}_{i,j}\right] = \sum_{i=1}^K w_i \mathbb{E}\left[\mathbb{E}\left[\hat{\nu}_{i,j}|\boldsymbol{\vartheta}_i\right]\right] = \sum_{i=1}^K w_i \mathbb{E}\left[\nu_j(\boldsymbol{\vartheta}_i)\right] = \mathbb{E}\left[\nu_j(\boldsymbol{\vartheta}_i)\right] =: \nu_{0,j}, \quad \text{(A.2)}$$

where $\nu_j(\boldsymbol{\vartheta}_i)$ and $\nu_{0,j}$ denote the j^{th} element of the vectors $\boldsymbol{\nu}(\boldsymbol{\vartheta}_i)$ and $\boldsymbol{\nu}_0 := \mathbb{E}[\boldsymbol{\nu}(\boldsymbol{\vartheta}_i)]$, respectively. Next, using the law of total variance, we have that the variance of $\bar{\nu}_j$ is of the following order

$$\text{var}\left(\bar{\nu}_j\right) = \sum_{i=1}^K w_i^2 \,\text{var}\left(\hat{\nu}_{i,j}\right) = \sum_{i=1}^K w_i^2 \left\{\mathbb{E}\left[\text{var}\left(\hat{\nu}_{i,j}|\boldsymbol{\vartheta}_i\right)\right] + \text{var}\left(\mathbb{E}\left[\hat{\nu}_{i,j}|\boldsymbol{\vartheta}_i\right]\right)\right\} = \mathcal{O}(K^{-1}),$$

based on the compactness of $\boldsymbol{\Theta}$ (Assumption A), the continuity of $\boldsymbol{\nu}(\boldsymbol{\vartheta}_i)$ (Assumption B), the asymptotic normality of $\hat{\nu}_{i,j}$, the conditions on the weights (w_i) as well as the continuity of the asymptotic variance (Assumption C). Thus, by Chebyshev's inequality we have that $\bar{\boldsymbol{\nu}} \xrightarrow{P} \boldsymbol{\nu}_0$. Next, we study the expected value of the objective function of $\hat{\boldsymbol{\theta}}^{\dagger}$. Again by the expectation of quadratic forms, we have

$$\mathbb{E}[\hat{Q}^{\dagger}(\boldsymbol{\theta})] = \mathbb{E}\left[\|\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\|_{\boldsymbol{\Omega}}^2\right] = \|\boldsymbol{\nu}_0 - \boldsymbol{\nu}(\boldsymbol{\theta})\|_{\boldsymbol{\Omega}}^2 + \operatorname{tr}\left\{\boldsymbol{\Omega} \operatorname{var}(\bar{\boldsymbol{\nu}})\right\},$$

using (A.2). Since the term $\operatorname{tr}\left\{\boldsymbol{\Omega} \operatorname{var}(\bar{\boldsymbol{\nu}})\right\}$ is constant with respect to $\boldsymbol{\theta}$, we have

$$\boldsymbol{\theta}_0 := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ \mathbb{E}\left[\|\boldsymbol{\nu}(\boldsymbol{\vartheta}_i) - \boldsymbol{\nu}(\boldsymbol{\theta})\|_{\boldsymbol{\Omega}}^2\right] = \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\operatorname{argmin}}\ Q^*(\boldsymbol{\theta}),$$

where $Q^*(\boldsymbol{\theta}) := \|\boldsymbol{\nu}_0 - \boldsymbol{\nu}(\boldsymbol{\theta})\|_{\boldsymbol{\Omega}}^2$. Using the consistency of $\bar{\boldsymbol{\nu}}$ and the continuity of $\hat{Q}^{\dagger}(\boldsymbol{\theta})$ in $\boldsymbol{\theta}$, we have that for each $\boldsymbol{\theta} \in \boldsymbol{\Theta}$, $|\hat{Q}^{\dagger}(\boldsymbol{\theta}) - Q^*(\boldsymbol{\theta})| \xrightarrow{P} 0$. Moreover, for all $\boldsymbol{\theta}_1, \boldsymbol{\theta}_2 \in \boldsymbol{\Theta}$ we have

$$\begin{aligned}
\left|\hat{Q}^{\dagger}(\boldsymbol{\theta}_1) - \hat{Q}^{\dagger}(\boldsymbol{\theta}_2)\right| &= \left|\|\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}_1)\|_{\boldsymbol{\Omega}}^2 - \|\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}_2)\|_{\boldsymbol{\Omega}}^2\right| \\
&= \left|\|\boldsymbol{\nu}(\boldsymbol{\theta}_2) - \boldsymbol{\nu}(\boldsymbol{\theta}_1)\|_{\boldsymbol{\Omega}}^2 - 2\left\{\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}_2)\right\}^{\top} \boldsymbol{\Omega} \left\{\boldsymbol{\nu}(\boldsymbol{\theta}_2) - \boldsymbol{\nu}(\boldsymbol{\theta}_1)\right\}\right| \\
&\leq \|\boldsymbol{\nu}(\boldsymbol{\theta}_2) - \boldsymbol{\nu}(\boldsymbol{\theta}_1)\|_{\boldsymbol{\Omega}}^2 + 2\left|\|\left\{\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}_2)\right\}^{\top} \boldsymbol{\Omega} \left\{\boldsymbol{\nu}(\boldsymbol{\theta}_2) - \boldsymbol{\nu}(\boldsymbol{\theta}_1)\right\}\right| \\
&\leq \|\boldsymbol{\nu}(\boldsymbol{\theta}_2) - \boldsymbol{\nu}(\boldsymbol{\theta}_1)\|_2^2 \|\boldsymbol{\Omega}\|_F + 2\|\boldsymbol{\nu}(\boldsymbol{\theta}_2) - \boldsymbol{\nu}(\boldsymbol{\theta}_1)\|_2 \|\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}_2)\|_2 \|\boldsymbol{\Omega}\|_F
\end{aligned} \tag{A.3}$$

where the first inequality is due to the triangular inequality and the second to the Cauchy-Schwarz inequality. The norms $\|\cdot\|_2$ and $\|\cdot\|_F$ denote the L_2 and the Frobenius norms, respectively. Since $\boldsymbol{\nu}(\boldsymbol{\theta})$ is continuously differentiable (Assumption B) and $\boldsymbol{\Theta}$ is compact (Assumption A) we have that for all $j \in \{1, \ldots, J\}$ there exists a constant C_j such that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \nu_j(\boldsymbol{\vartheta})\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\theta}} \right\|_2^2 \leq C_j.$$

Next, we define the matrix $\mathbf{A}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J)$, which is Jacobian matrix of $\boldsymbol{\nu}(\boldsymbol{\theta})$ whose j^{th} row is evaluated in $\boldsymbol{\theta}_j$. Based on Assumption A we have that

$$\sup_{\boldsymbol{\theta}_1,\ldots,\boldsymbol{\theta}_J\in\boldsymbol{\Theta}} \|\mathbf{A}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J)\|_F = \sqrt{\sum_{j=1}^{J} \sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}} \left\| \frac{\partial}{\partial \boldsymbol{\theta}} \nu_j(\boldsymbol{\vartheta})\Big|_{\boldsymbol{\vartheta}=\boldsymbol{\theta}} \right\|_2^2} \leq \sqrt{\sum_{j=1}^{J} C_j} =: C. \tag{A.4}$$

By the mean value theorem, we have

$$\begin{aligned}\|\boldsymbol{\nu}(\boldsymbol{\theta}_2)-\boldsymbol{\nu}(\boldsymbol{\theta}_1)\|_2^2 &= \|\mathbf{A}\left(\boldsymbol{\theta}_1^*,\ldots,\boldsymbol{\theta}_J^*\right)(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)\|_2^2 \\ &= (\boldsymbol{\theta}_1-\boldsymbol{\theta}_2)^\top \mathbf{A}^\top\left(\boldsymbol{\theta}_1^*,\ldots,\boldsymbol{\theta}_J^*\right)\mathbf{A}\left(\boldsymbol{\theta}_1^*,\ldots,\boldsymbol{\theta}_J^*\right)(\boldsymbol{\theta}_1-\boldsymbol{\theta}_2) \\ &\leq \|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2^2 \left\|\mathbf{A}^\top\left(\boldsymbol{\theta}_1^*,\ldots,\boldsymbol{\theta}_J^*\right)\mathbf{A}\left(\boldsymbol{\theta}_1^*,\ldots,\boldsymbol{\theta}_J^*\right)\right\|_F \\ &\leq \|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2^2 \|\mathbf{A}\left(\boldsymbol{\theta}_1^*,\ldots,\boldsymbol{\theta}_J^*\right)\|_F^2 \leq C^2\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2^2,\end{aligned}$$

where the first two inequalities are consequences of Cauchy-Schwarz inequality and the last one is due to (A.4). Returning to (A.3), we have

$$\left|\hat{Q}^\dagger(\boldsymbol{\theta}_1)-\hat{Q}^\dagger(\boldsymbol{\theta}_2)\right| \leq B_n\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2,$$

where

$$B_n := C\|\boldsymbol{\Omega}\|_F\left\{C\|\boldsymbol{\theta}_1-\boldsymbol{\theta}_2\|_2+2\|\bar{\boldsymbol{\nu}}-\boldsymbol{\nu}(\boldsymbol{\theta}_2)\|_2\right\}.$$

Since $\boldsymbol{\Theta}$ is compact (Assumption A), $\|\boldsymbol{\Omega}\|_F$ is bounded (Assumption C) and $\bar{\boldsymbol{\nu}}$ is consistent, we have that B_n is bounded in probability. Thus, Corollary 2.2 of [2] implies that

$$\sup_{\boldsymbol{\theta}\in\boldsymbol{\Theta}}|\hat{Q}^\dagger(\boldsymbol{\theta})-Q^*(\boldsymbol{\theta})| \xrightarrow{P} 0.$$

Finally, Theorem 2.1 of [3] implies that $\|\hat{\boldsymbol{\theta}}^\dagger-\boldsymbol{\theta}_0\| \xrightarrow{P} 0$, which concludes the proof.

□

Proof of Proposition 1

Under Assumptions A and B, the AWV and MS-GMWM estimators can be defined in terms of their derivatives, i.e.,

$$\hat{\boldsymbol{\theta}}^\dagger := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argzero}}\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{Q}^\dagger(\boldsymbol{\theta}),$$

and

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta}\in\boldsymbol{\Theta}}{\text{argzero}}\frac{\partial}{\partial\boldsymbol{\theta}}\widehat{Q}(\boldsymbol{\theta}),$$

respectively, where argzero stands for the value of $\boldsymbol{\theta}$ that allows the expression to be zero.

Considering this, the derivative of $\widehat{Q}(\boldsymbol{\theta})$ is given by

$$\frac{\partial}{\partial \boldsymbol{\theta}}\widehat{Q}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\sum_{i=1}^{K} w_i \|\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}(\boldsymbol{\theta})\|_{\boldsymbol{\Omega}}^2 = -\sum_{i=1}^{K} 2w_i \frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\nu}(\boldsymbol{\theta})\boldsymbol{\Omega}(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}(\boldsymbol{\theta}))$$
$$= -2\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\nu}(\boldsymbol{\theta})\boldsymbol{\Omega}\sum_{i=1}^{K} w_i(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}(\boldsymbol{\theta})).$$

Knowing that $\sum_{i=1}^{K} w_i = 1$, we finally have that

$$\frac{\partial}{\partial \boldsymbol{\theta}}\widehat{Q}(\boldsymbol{\theta}) = -2\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\nu}(\boldsymbol{\theta})\boldsymbol{\Omega}\left(\sum_{i=1}^{K} w_i\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}(\boldsymbol{\theta})\right).$$

If we take the derivative of $\widehat{Q}^{\dagger}(\boldsymbol{\theta})$, we obtain

$$\frac{\partial}{\partial \boldsymbol{\theta}}\widehat{Q}^{\dagger}(\boldsymbol{\theta}) = \frac{\partial}{\partial \boldsymbol{\theta}}\|\sum_{i=1}^{K} w_i\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}(\boldsymbol{\theta})\|_{\boldsymbol{\Omega}}^2 = -2\frac{\partial}{\partial \boldsymbol{\theta}}\boldsymbol{\nu}(\boldsymbol{\theta})\boldsymbol{\Omega}\left(\sum_{i=1}^{K} w_i\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}(\boldsymbol{\theta})\right).$$

Since $\widehat{Q}(\boldsymbol{\theta})$ and $\widehat{Q}^{\dagger}(\boldsymbol{\theta})$ have the same derivative, under Assumptions A and B they have the same solution in zero and, consequently, we have that $\hat{\boldsymbol{\theta}}^{\dagger} = \hat{\boldsymbol{\theta}}$ thus concluding the proof. □

Proof of Proposition 2

This proof is adapted and closely follows the proof of Lemma 3.1 in [1]. More specifically, given the results on the consistency in Theorem 2, the proof of asymptotic normality of $\hat{\boldsymbol{\theta}}^{\dagger}$ naturally follows the standard proof of asymptotic normality for extremum estimators (see e.g. [3]). Indeed, using again the notation $\bar{\boldsymbol{\nu}} := \sum_{i=1}^{K} w_i\hat{\boldsymbol{\nu}}_i$, we start by studying the asymptotic distribution of this quantity. To do so, for $\mathbf{s} \in \mathcal{S} := \left\{\mathbf{x} \in \mathbb{R}^J \mid \|\mathbf{x}\|_2 = 1\right\}$, we define $X_i(\mathbf{s}) = \mathbf{s}^{\top}(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_0)$. Unconditionally, we have that $X_i(\mathbf{s})$, for $i = 1, \ldots, K$, are mean-zero iid random variables for which we define $\boldsymbol{\Upsilon} := \mathrm{var}(\hat{\boldsymbol{\nu}}_i)$, that allows us to define $\mathrm{var}\left[X_i(\mathbf{s})\right] = \mathbf{s}^{\top}\boldsymbol{\Upsilon}\mathbf{s} =: \sigma^2(\mathbf{s})$. Moreover, for sufficiently large K, there exists a constant M such that

$$\frac{\max_{i=1,\ldots,K} w_i^2}{\sum_{i=1}^{K} w_i^2} \leq \frac{\max_{i=1,\ldots,K} w_i^2}{\sum_{i=1}^{K} w_i} = \max_{i=1,\ldots,K} w_i^2 \leq M^2K^{-2} = \mathcal{O}(K^{-2}).$$

Considering the above definitions and properties, by Hajek-Sidak's central limit theorem we consequently have that

$$\frac{\sum_{i=1}^{K} w_i\mathbf{s}^{\top}(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_0)}{\sigma(\mathbf{s})\sqrt{\sum_{i=1}^{K} w_i^2}} = \frac{\sum_{i=1}^{K} w_i X_i(\mathbf{s})}{\sigma(\mathbf{s})\sqrt{\sum_{i=1}^{K} w_i^2}} \xrightarrow{\mathcal{D}} \mathcal{N}(0,1).$$

Recalling that by definition the weights are such that $K\sum_{i=1}^{K} w_i^2 \to c$ (where c is a positive constant), by Slutsky's theorem we have that

$$\sqrt{K}\sum_{i=1}^{K} w_i \mathbf{s}^\top \left(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, c\,\sigma^2(\mathbf{s})\right).$$

This allows us to use the Cramer-Wold device to finally obtain

$$\sqrt{K}\sum_{i=1}^{K} w_i \left(\hat{\boldsymbol{\nu}}_i - \boldsymbol{\nu}_0\right) \xrightarrow{\mathcal{D}} \mathcal{N}\left(\mathbf{0}, \boldsymbol{\Sigma}\right). \tag{A.5}$$

Having obtained this result, we proceed with the main proof. By Assumption B and by the definition of $\hat{\boldsymbol{\theta}}^\dagger$, we have

$$\left.\frac{\partial \widehat{Q}^\dagger(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^\dagger} = \mathbf{0}_{p\times 1} \iff \left.\frac{\partial}{\partial \boldsymbol{\theta}}\left[\left(\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)^\intercal \boldsymbol{\Omega}\left(\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)\right]\right|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^\dagger} = \mathbf{0}_{p\times 1},$$

which, up to a constant, yields

$$\underbrace{\left(\frac{\partial}{\partial \boldsymbol{\theta}}\left(\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta})\right)^\intercal\Big|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}^\dagger}\right)}_{\mathbf{B}(\hat{\boldsymbol{\theta}}^\dagger)} \boldsymbol{\Omega}\left(\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}^\dagger)\right) = \mathbf{0}_{p\times 1}. \tag{A.6}$$

The mean value theorem ensures that, based on Assumption A, there exists a matrix $\mathbf{A}(\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_J)$, as defined in the proof of Theorem 2, that can be used to expand $\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}^\dagger)$ around $\boldsymbol{\theta}_0$ in the following way

$$\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\hat{\boldsymbol{\theta}}^\dagger) = \bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}_0) - \mathbf{A}(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_J^*)\left(\hat{\boldsymbol{\theta}}^\dagger - \boldsymbol{\theta}_0\right). \tag{A.7}$$

Based on the derivatives in the proofs of Proposition 1 (whose solutions for zero occur when $\bar{\boldsymbol{\nu}} = \boldsymbol{\nu}(\boldsymbol{\theta})$) and using Theorem 2, we have that $\bar{\boldsymbol{\nu}} \xrightarrow{p} \boldsymbol{\nu}(\boldsymbol{\theta}_0)$ and $\hat{\boldsymbol{\theta}}^\dagger \xrightarrow{p} \boldsymbol{\theta}_0$ such that the multivariate mean value theorem also guarantees that the matrix $\mathbf{A}(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_J^*)$ has the following property

$$\mathbf{A}(\boldsymbol{\theta}_1^*, \ldots, \boldsymbol{\theta}_J^*) \xrightarrow{p} -\left.\frac{\partial}{\partial \boldsymbol{\theta}^\intercal}\left(\boldsymbol{\nu}(\boldsymbol{\theta}_0) - \boldsymbol{\nu}(\boldsymbol{\theta})\right)\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \left.\frac{\partial}{\partial \boldsymbol{\theta}^\intercal}\boldsymbol{\nu}(\boldsymbol{\theta})\right|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0} = \mathbf{A}(\boldsymbol{\theta}_0),$$

given that $\partial/\partial\boldsymbol{\theta}\,\boldsymbol{\nu}(\boldsymbol{\theta}^\intercal)$ is continuous. Plugging (A.7) in the third factor of (A.6), multiplying by $\sqrt{K}$ and using Assumption B allows us to state that $\sqrt{K}\left(\hat{\boldsymbol{\theta}}^\dagger - \boldsymbol{\theta}_0\right)$ is equal to

$$-\left[\mathbf{B}(\hat{\boldsymbol{\theta}}^\dagger)\,\boldsymbol{\Omega}\,\mathbf{A}(\hat{\boldsymbol{\theta}}^\dagger, \boldsymbol{\theta}_0)\right]^{-1}\mathbf{B}(\hat{\boldsymbol{\theta}}^\dagger)\,\boldsymbol{\Omega}\,\sqrt{K}\left(\bar{\boldsymbol{\nu}} - \boldsymbol{\nu}(\boldsymbol{\theta}_0)\right). \tag{A.8}$$

Knowing that $\mathbf{B}(\hat{\boldsymbol{\theta}}^\dagger) \xrightarrow{p} -\mathbf{A}(\boldsymbol{\theta}_0)^\intercal$ by the continuous mapping theorem, and that $\hat{\boldsymbol{\theta}}^\dagger \xrightarrow{p} \boldsymbol{\theta}_0$ from

Theorem 2, by continuous mapping theorem (and the continuity of matrix inverse) we have that

$$-\left[\mathbf{B}(\hat{\boldsymbol{\theta}}^{\dagger})\,\boldsymbol{\Omega}\,\mathbf{A}(\hat{\boldsymbol{\theta}}^{\dagger},\boldsymbol{\theta}_0)\right]^{-1}\mathbf{B}(\hat{\boldsymbol{\theta}}^{\dagger})\,\boldsymbol{\Omega}\xrightarrow{p}\left[\mathbf{A}(\boldsymbol{\theta}_0)^{\intercal}\boldsymbol{\Omega}\mathbf{A}(\boldsymbol{\theta}_0)\right]^{-1}\mathbf{A}(\boldsymbol{\theta}_0)^{\intercal}\boldsymbol{\Omega}=\mathbf{H}(\boldsymbol{\theta}_0)^{-1}\mathbf{A}(\boldsymbol{\theta}_0)^{\intercal}\boldsymbol{\Omega}.$$

By again using Slutsky's theorem with Theorem 1 of [4] (in conjunction with Assumption A) as well as (A.5), we have that (A.8) has the following asymptotic distribution

$$\sqrt{K}\left(\hat{\boldsymbol{\theta}}^{\dagger}-\boldsymbol{\theta}_0\right)\xrightarrow{\mathcal{D}}\mathcal{N}(\mathbf{0},\boldsymbol{\Lambda}_0),$$

with $\boldsymbol{\Lambda}_0$ being given by

$$\boldsymbol{\Lambda}_0:=\mathbf{H}(\boldsymbol{\theta}_0)^{-1}\mathbf{A}(\boldsymbol{\theta}_0)^{\intercal}\boldsymbol{\Omega}\boldsymbol{\Sigma}\boldsymbol{\Omega}\mathbf{A}(\boldsymbol{\theta}_0)\mathbf{H}(\boldsymbol{\theta}_0)^{-1},$$

where $\boldsymbol{\Sigma}$ is defined in (A.5), thus concluding the proof. □

Proof of Proposition 3

When $\boldsymbol{\nu}(\boldsymbol{\theta})=\mathbf{W}\boldsymbol{\theta}$, as shown in [5], the GMWM has an explicit solution given by

$$\bar{\boldsymbol{\vartheta}}=(\mathbf{W}^{\intercal}\boldsymbol{\Omega}\mathbf{W})^{-1}\mathbf{W}^{\intercal}\boldsymbol{\Omega}\hat{\boldsymbol{\nu}}.$$

Hence, in this case the AGMWM estimator can be expressed as

$$\hat{\boldsymbol{\theta}}^{\circ}=\sum_{i=1}^{K}w_i\tilde{\boldsymbol{\vartheta}}_i=\sum_{i=1}^{K}w_i(\mathbf{W}^{\intercal}\boldsymbol{\Omega}\mathbf{W})^{-1}\mathbf{W}^{\intercal}\boldsymbol{\Omega}\hat{\boldsymbol{\nu}}_i=(\mathbf{W}^{\intercal}\boldsymbol{\Omega}\mathbf{W})^{-1}\mathbf{W}^{\intercal}\boldsymbol{\Omega}\underbrace{\sum_{i=1}^{K}w_i\hat{\boldsymbol{\nu}}_i}_{\bar{\boldsymbol{\nu}}}.$$

Now, based on the proof in Proposition 1, we have that $\hat{\boldsymbol{\theta}}^{\dagger}$ is the solution in $\boldsymbol{\theta}$ of the following equation

$$2\frac{\partial}{\partial\boldsymbol{\theta}}\boldsymbol{\nu}(\boldsymbol{\theta})\boldsymbol{\Omega}\left(\underbrace{\sum_{i=1}^{K}w_i\hat{\boldsymbol{\nu}}_i}_{\bar{\boldsymbol{\nu}}}-\boldsymbol{\nu}(\boldsymbol{\theta})\right)=0,$$

which, in the case where $\boldsymbol{\nu}(\boldsymbol{\theta})=\boldsymbol{W}(\boldsymbol{\theta})$, delivers

$$2\mathbf{W}^{\intercal}\boldsymbol{\Omega}\left(\bar{\boldsymbol{\nu}}-\mathbf{W}\boldsymbol{\theta}\right)=0.$$

The solution is therefore given by

$$\hat{\boldsymbol{\theta}}^{\dagger}=(\mathbf{W}^{\intercal}\boldsymbol{\Omega}\mathbf{W})^{-1}\mathbf{W}^{\intercal}\boldsymbol{\Omega}\bar{\boldsymbol{\nu}},$$

which is the same as for the AGMWM and, based on Proposition 1, the same as for the MS-GMWM. □

B. *Individual fits*

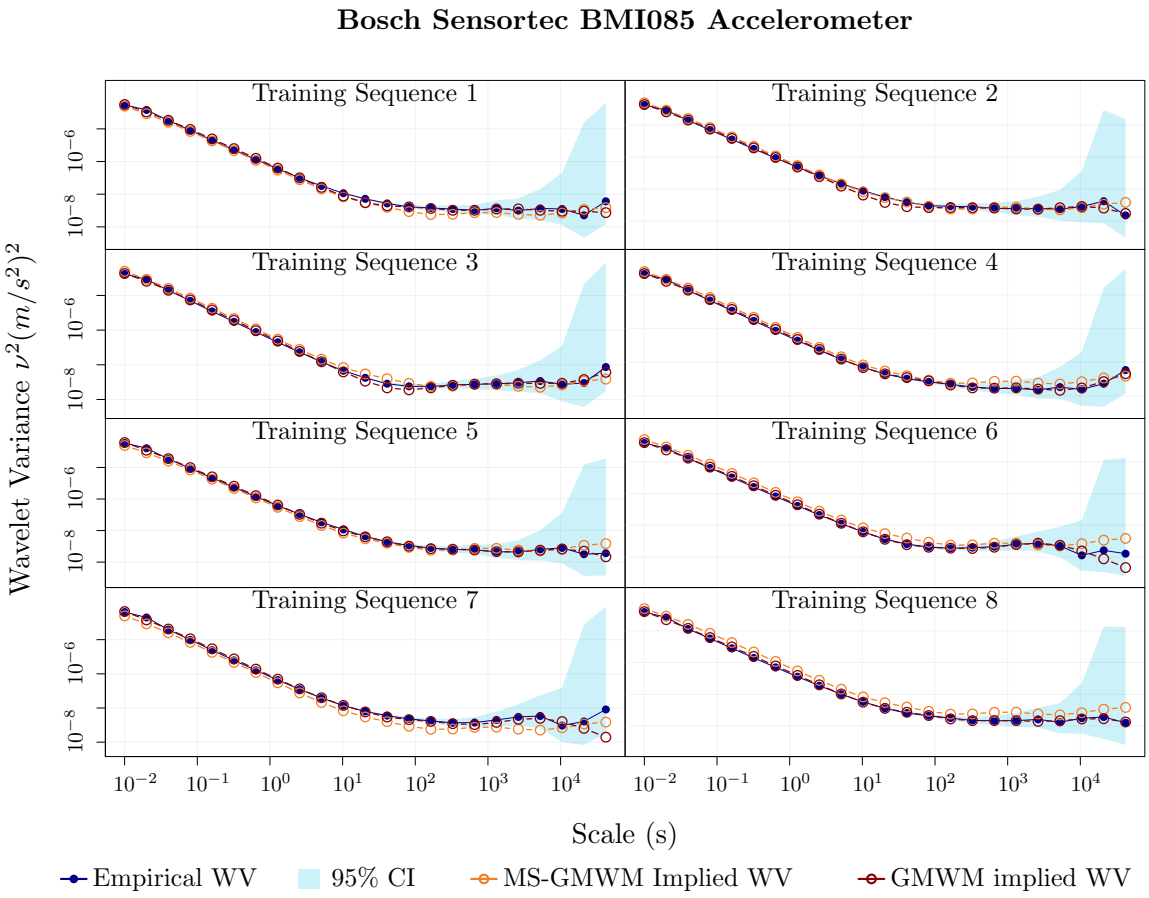

Fig. B.1. Empirical WV (blue doted line) for the training sequences of accelerometer of a Bosch Sensortec BMI085 6-Axis IMU and their respective 95% confidence intervals (blue shaded area). Red dotted line represent the implied WV from the individual solution of the GMWM on sequence $1, \dots 8$, while orange dotted line the implied WV from the MS-GMWM computed trough the AWV.

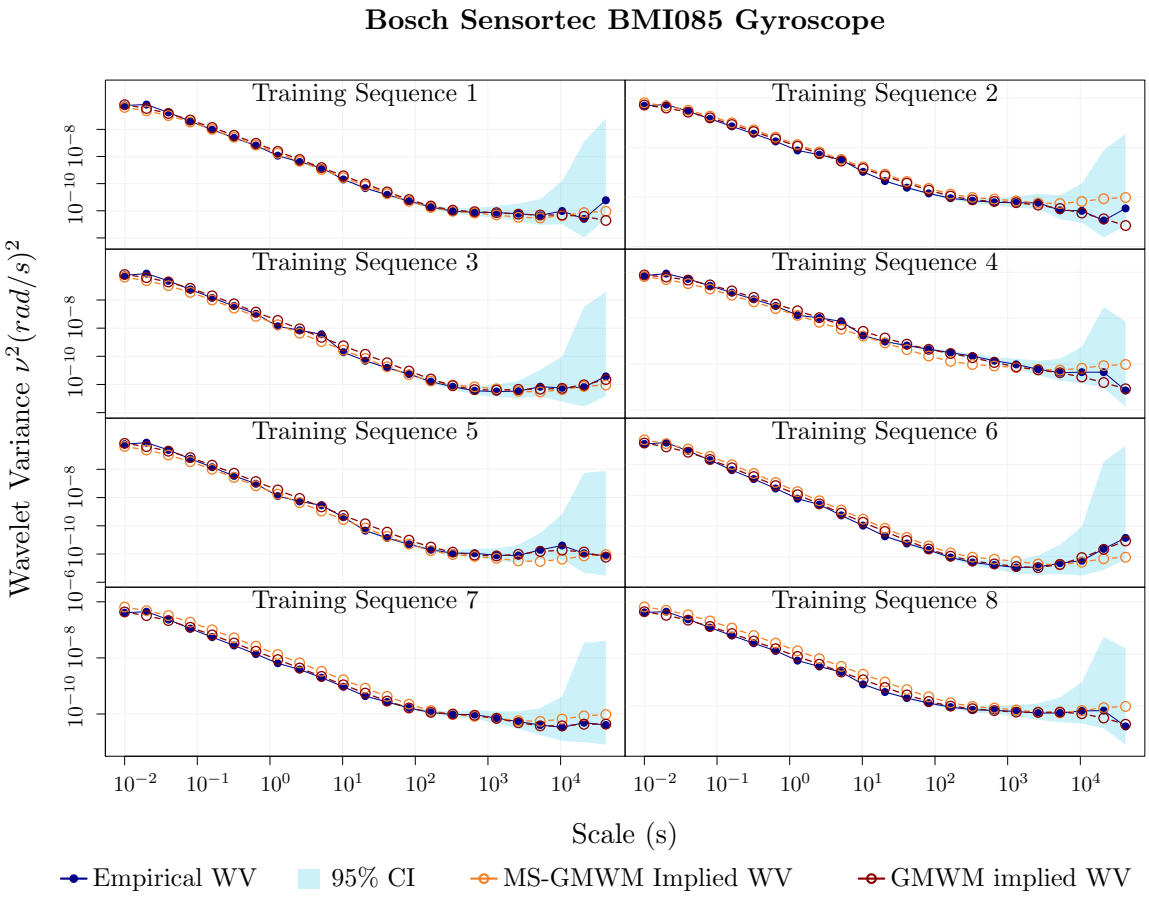

Fig. B.2. Empirical WV (blue doted line) for the training sequences of gyroscope of a Bosch Sensortec BMI085 6-Axis IMU and their respective 95% confidence intervals (blue shaded area). Red dotted line represent the implied WV from the individual solution of the GMWM on sequence $1, \dots 8$, while orange dotted line the implied WV from the MS-GMWM computed trough the AWV.

C. Parameter distribution accelerometer and gyroscope

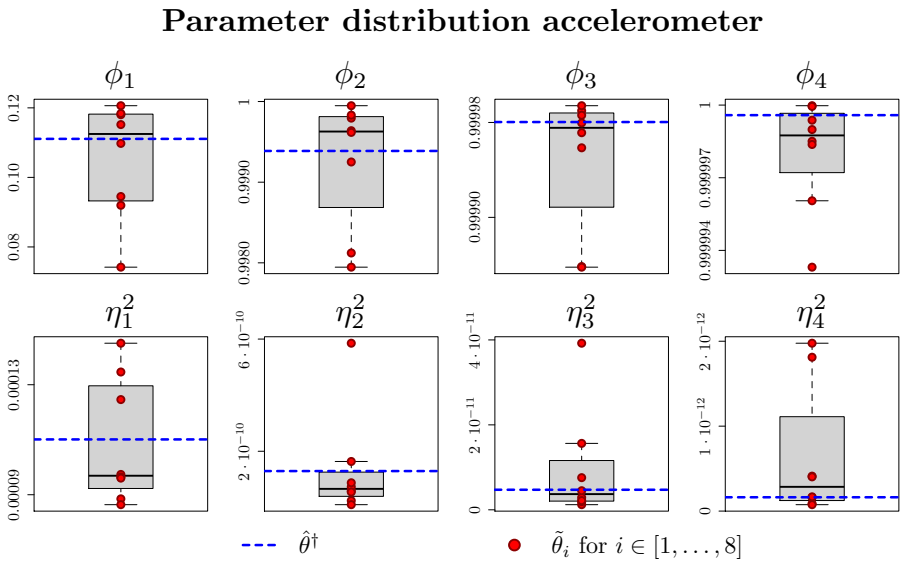

Fig. C.1. Empirical parameter distribution of $\tilde{\boldsymbol{\theta}}_i$ for model $\mathcal{M}_i, i \in [1, ..., 8]$, with each individual estimates i represented by a red dot, with the MS estimates $\hat{\boldsymbol{\theta}}^\dagger$ (blue dashed line) for model $\mathcal{M}_{\text{MS}}$ for the accelerometers.

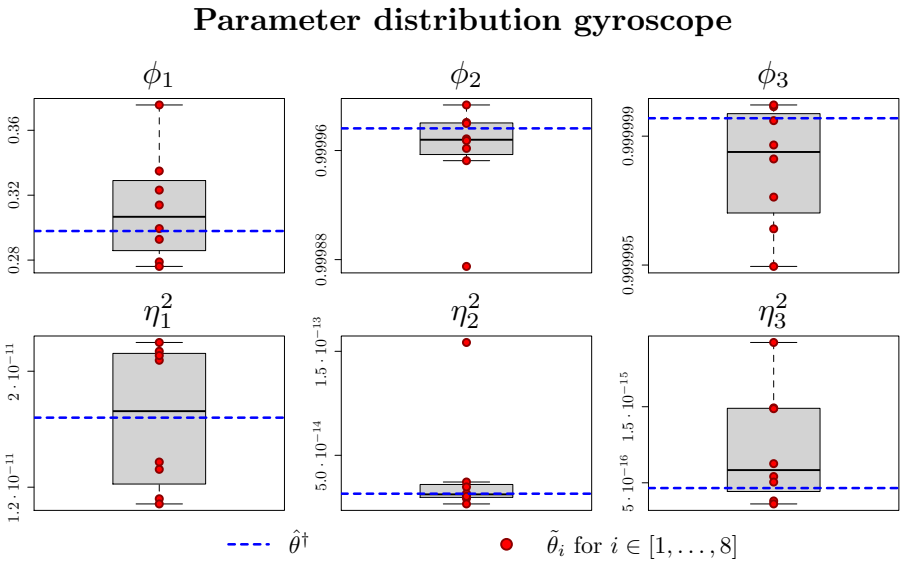

Fig. C.2. Empirical parameter distribution of $\tilde{\boldsymbol{\theta}}_i$ for model $\mathcal{M}_i, i \in [1, ..., 8]$, with each individual estimates i represented by a red dot, with the MS estimates $\hat{\boldsymbol{\theta}}^\dagger$ (blue dashed line) for model $\mathcal{M}_{\text{MS}}$ for the gyroscopes.

D. Case Study II - Impact on Navigation

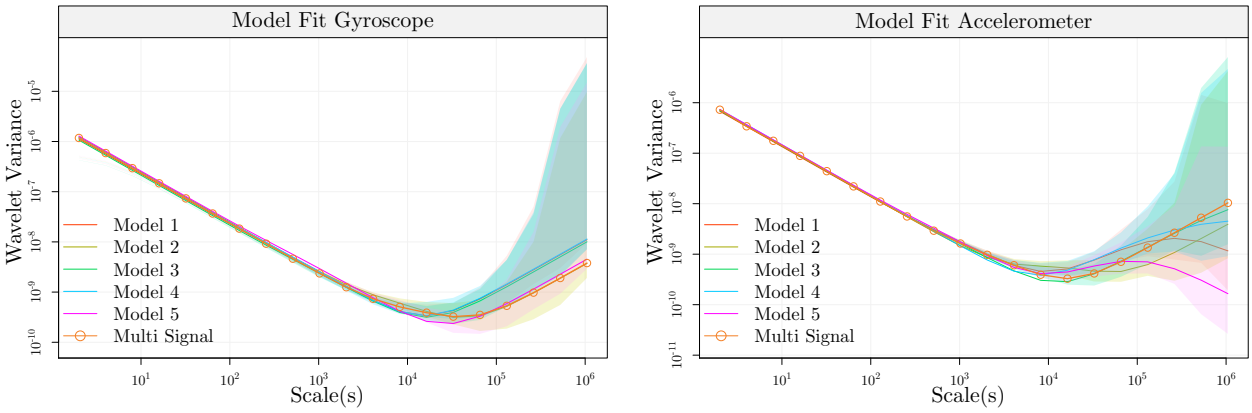

Fig. D.1. WV implied by respective estimated models for the x-axis gyroscope (left plot) and x-axis accelerometer (right plot) for the NavChip™ 6-Axis Intersense IMU based on the model $WN + AR1 + RW$ for the gyroscope and $AR1 + AR1 + WN$ for the accelerometer. For each plot the plain colored lines represent the WV implied by the model estimated via the GMWM on each individual replicate (e.g. Model 1 fitted based on the first replicate), while the orange line with circles represents the AWV fit on all replicates $1, \ldots, 5$ with the shaded areas consisting in the 95% confidence intervals for the WV for each replicate.

This section present a further evaluation similar to the one presented in Section. IV. This time, we consider the NavChip™ 6-Axis IMU supplied by Intersense[1], a high-quality MEMs IMU for navigation applications, e.g., in UAVs [6]. Such an inertial module combines a 3-axis gyroscope and a 3-axis accelerometer. We collect $K = 5$ replicates of sensor data in static conditions, each one lasting 1.5 hours.

To identify the error process we visually analyse the empirical WV of the replicates and find that the sum of a WN, an AR1 and a RW appears well suited for the gyroscope signal while the sum of two AR1 processes and a WN fits the accelerometer WV well as shown in Fig. D.1. More specifically: (i) the plain colored lines represent the WV implied by the processes fit to the individual replicates fitted through the GMWM ($\mathcal{M}_i$ is the model fit on replicate $i = 1, \ldots, 5$) and (ii) the orange line with circles represents the model fitted through the AWV ($\mathcal{M}_{\text{MS}}$) on all 5 replicates with the shaded area representing the 95% confidence intervals for the WV. Each fitted model lies within the confidence intervals of the empirical WV which, to avoid visual cluttering, was omitted from the plot. Considering these representations, it is straightforward to detect differences in the models fit to the signals via the individual and joint approaches.

[1]The Intersense NavChip™ family: https://www.intersense.com/navchip

Given this, in order to confirm whether to use a single replicate or a multi-signal approach we perform the near-stationarity test put forward in [7] by simulating 100 bootstrap replicates under the estimated $F_{\hat{\boldsymbol{\theta}}^\dagger}$ which, keeping in mind the discrete nature of the bootstrapped test statistic, gives us a zero p-value thereby allowing us to reject the null hypothesis that all replicates are issued from the same data-generating process with $\boldsymbol{\vartheta}_i = \theta_0$ for all i (i.e. G is a Dirac point mass distribution).

With the goal of producing further evidence for the argument that some (or all) fitted models are different from each other (in terms of parameter values) we perform a Principal Component Analysis (PCA) on the parameter vectors $\tilde{\boldsymbol{\vartheta}}_i$ as estimated by GMWM for all the models and on $\hat{\boldsymbol{\theta}}^\dagger$ (AWV). The first two dimensions of the PCA are shown in Fig. D.2, where we can identify three clusters of models, highlighting the differences in behaviour of those replicates. Based on this result we can consider $\mathcal{M}_2$ (the model estimated by the GMWM applied to the second replicate) as the replicate which is closest to the fit given by $\mathcal{M}_{\text{MS}}$.

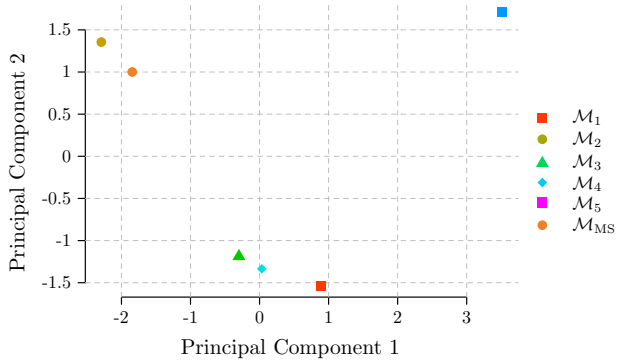

Fig. D.2. First two principal components on the parameter vectors $\tilde{\boldsymbol{\vartheta}}_i$ for models $i = 1, \ldots, 5$ ($\mathcal{M}_i$) and $\hat{\boldsymbol{\theta}}^\dagger$ on replicates 1 to 5 ($\mathcal{M}_{\text{MS}}$).

Given the similarities of the parameter vector $\tilde{\boldsymbol{\vartheta}}_2$ of $\mathcal{M}_2$ with $\hat{\boldsymbol{\theta}}^\dagger$ of $\mathcal{M}_{\text{MS}}$, we compute the 95% confidence intervals of their parameter values for both the gyroscope and accelerometer signals. These are shown in Tab I and Tab II. The 95% confidence intervals for $\mathcal{M}_2$ are constructed by parametric bootstrap, i.e., by simulating new samples under the stationary estimated model, while the ones for $\mathcal{M}_{\text{MS}}$ are obtained by re-sampling the different replicates $X_t^{(i)}$ (for $i = 1, \ldots, 5$) as described for example in [7]. Looking at these tables, while the parameter values can be considered fairly close, we can clearly see how the AWV confidence intervals better capture the uncertainty in the parameters due to the near-stationary setting.

Parameter	$\mathcal{M}_2$	$\mathcal{M}_{\text{MS}}$
ϕ_1	$9.99 \cdot 10^{-1}$	$9.99 \cdot 10^{-1}$
	$(9.99 \cdot 10^{-1}, 9.99 \cdot 10^{-1})$	$(9.99 \cdot 10^{-1}, 9.99 \cdot 10^{-1})$
η_1^2	$4.76 \cdot 10^{-14}$	$2.82 \cdot 10^{-13}$
	$(2.00 \cdot 10^{-14}, 7.53 \cdot 10^{-14})$	$(1.82 \cdot 10^{-13}, 4.09 \cdot 10^{-13})$
ϕ_2	$9.99 \cdot 10^{-1}$	$9.98 \cdot 10^{-1}$
	$(9.99 \cdot 10^{-1}, 9.99 \cdot 10^{-1})$	$(9.97 \cdot 10^{-1}, 9.99 \cdot 10^{-1})$
η_2^2	$1.28 \cdot 10^{-12}$	$4.72 \cdot 10^{-12}$
	$(1.08 \cdot 10^{-12}, 1.50 \cdot 10^{-12})$	$(2.40 \cdot 10^{-12}, 7.07 \cdot 10^{-12})$
σ^2	$1.39 \cdot 10^{-6}$	$1.39 \cdot 10^{-6}$
	$(1.39 \cdot 10^{-6}, 1.40 \cdot 10^{-6})$	$(1.37 \cdot 10^{-6}, 1.43 \cdot 10^{-6})$

TABLE I

PARAMETER ESTIMATES AND 95% CONFIDENCE INTERVALS FOR THE AR1 + AR1 + WN MODEL FOR THE NAVCHIP™ 6-AXIS IMU ACCELEROMETER BASED ON SOLELY ON THE SECOND REPLICATE ($\mathcal{M}_2$) AND ON THE AWV ($\mathcal{M}_{\text{MS}}$).

Parameter	$\mathcal{M}_2$	$\mathcal{M}_{\text{MS}}$
ϕ	$9.99 \cdot 10^{-1}$	$9.99 \cdot 10^{-1}$
	$(9.99 \cdot 10^{-1}, 9.99 \cdot 10^{-1})$	$(9.99 \cdot 10^{-1}, 9.99 \cdot 10^{-1})$
η^2	$1.79 \cdot 10^{-12}$	$6.99 \cdot 10^{-13}$
	$(1.37 \cdot 10^{-12}, 2.22 \cdot 10^{-12})$	$(2.60 \cdot 10^{-13}, 1.13 \cdot 10^{-12})$
γ^2	$4.25 \cdot 10^{-14}$	$4.42 \cdot 10^{-14}$
	$(2.00 \cdot 10^{-14}, 6.50 \cdot 10^{-14})$	$(2.55 \cdot 10^{-14}, 8.03 \cdot 10^{-14})$
σ^2	$1.29 \cdot 10^{-6}$	$1.22 \cdot 10^{-6}$
	$(1.29 \cdot 10^{-6}, 1.29 \cdot 10^{-6})$	$(1.13 \cdot 10^{-6}, 1.33 \cdot 10^{-6})$

TABLE II

PARAMETER ESTIMATES AND 95% CONFIDENCE INTERVALS FOR THE WN + AR1 + RW MODEL FOR THE NAVCHIP™ 6-AXIS IMU GYROSCOPE BASED ON SOLELY ON THE SECOND REPLICATE ($\mathcal{M}_2$) AND ON THE AWV ($\mathcal{M}_{\text{MS}}$).

Having compared the different approaches in terms of model fit and uncertainty evaluation, we investigate the navigation performance on the 6 different models. For this purpose, the estimated stochastic models are used in an Extended Kalman Filter for INS/GNSS navigation in order to compare the performance of these models in terms of position and orientation error as well as consistency of the confidence intervals for the navigation state within a realistic navigation scenario. The procedure is similar to the one presented in Section IV, with the difference that the number of available data sequences ($K = 5$) is not sufficient to split those into a training and validation set. Thus, the same data sequences used to fit models $\mathcal{M}_i$, with $i \in [1, ..., 5]$ and $\mathcal{M}_{\text{MS}}$ are also used to generate the noisy inertial readings during the Monte-Carlo navigation simulations. The results are presented in Fig. D.3 and Fig. D.4.

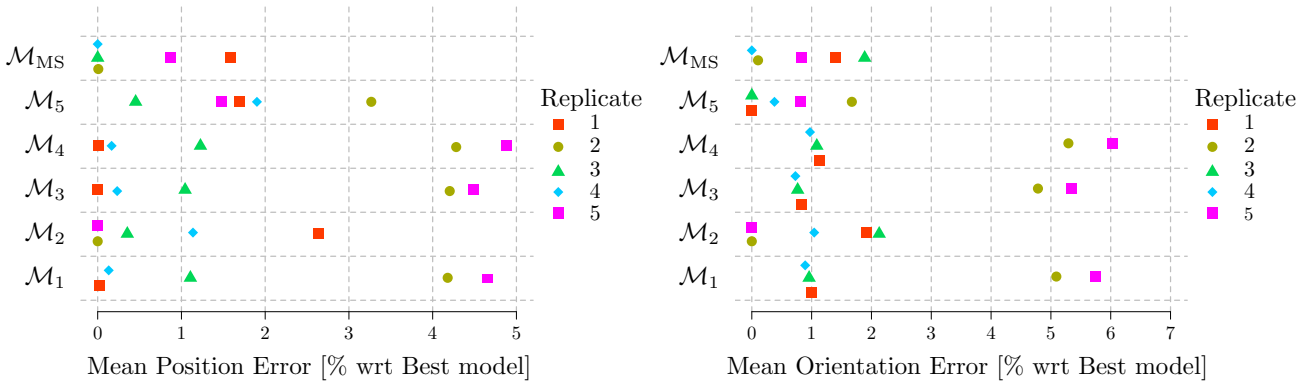

Fig. D.3. Mean position and orientation errors achieved by using each estimated model to predict the error on all replicates, $\mathcal{M}_i$ represents the model estimated on replicate i and $\mathcal{M}_{\text{MS}}$ represents the model estimated via the multi-signal AWV. The results are expressed in percentage with respect the best performing model on one specific static acquisition. For instance, the model based on the second replicate ($\mathcal{M}_2$) has the lowest position error on replicate 5 (0%) while, always referred to the position error for replicate 5, the model based on the fourth replicate ($\mathcal{M}_4$) performs on average 5% worse (results based on 500 Monte-Carlo runs).

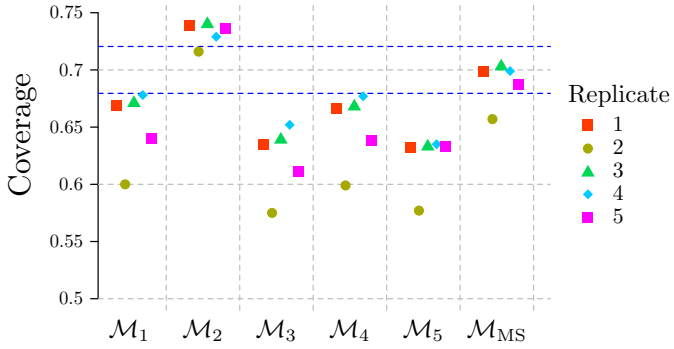

Fig. D.4. Empirical coverage of the 70% confidence intervals derived from the EKF covariance matrices for each model estimated on individual replicates ($\mathcal{M}_i$) and via the multi-signal AWV ($\mathcal{M}_{\text{MS}}$). Results represent empirical coverage of each individual model when applied to all replicates with the blue dashed lines representing confidence intervals for the true coverage. For instance, the model based on the second replicate ($\mathcal{M}_2$) tends to "over-cover" the navigation states for all the other replicates except for those of its reference signal which just lie within the confidence intervals of the required 70% coverage rate (results based on 500 Monte-Carlo runs).

As expected, it is possible to see that each model fitted on a single replicate performs best when employed against noise coming from the data replicate it was estimated on. However, when the noise comes from a different static acquisition, the performances may degrade up to 5% in position and 6% in orientation. This is related to the overfitting of the model on a single realisation of the stochastic error processes. On the contrary, the models estimated with the AWV technique perform well, if not optimally, on each of the static acquisition replicates. This shows that the proposed method is able to better capture the underlying nature of the data-generating

process. This is also visible in the coverage of the confidence intervals for the navigation states (position and orientation) as estimated by the EKF: the models estimated with the AWV have the correct coverage, within 68% and 72%, given the number of Monte-Carlo runs, while the coverage is often outside such bounds for the single-signal models.

REFERENCES

[1] S. Guerrier, R. Molinari, M.-P. Victoria-Feser, and H. Xu. Robust two-step wavelet-based inference for time series models. *Journal of the American Statistical Association*, pages 1–50, 2021.

[2] W. K. Newey. Uniform convergence in probability and stochastic equicontinuity. *Econometrica: Journal of the Econometric Society*, pages 1161–1167, 1991.

[3] K.W Newey and D. McFadden. Large sample estimation and hypothesis. In R.F. Engle and McFadden D.L., editors, *Handbook of Econometrics, Volume 4*, chapter 26, pages 2112–2245. Elsevier, 1994.

[4] M. Weber. A weighted central limit theorem. *Statistics & probability letters*, 76(14):1482–1487, 2006.

[5] S. Guerrier, J. Jurado, M. Khaghani, G. Bakalli, M. Karemera, R. Molinari, S. Orso, J. Raquet, C. Schubert, J. Skaloud, et al. Wavelet-based moment-matching techniques for inertial sensor calibration. *IEEE Transactions on Instrumentation and Measurement*, 2020.

[6] D. A. Cucci, M. Rehak, and J. Skaloud. Bundle adjustment with raw inertial observations in UAV applications. *ISPRS Journal of Photogrammetry and Remote Sensing*, 130:1–12, 2017.

[7] G. Bakalli, A. Radi, N. El-Sheimy, R. Molinari, and S. Guerrier. A computational multivariate-based technique for inertial sensor calibration. In *Proceedings of the 30th International Technical Meeting of The Satellite Division of the Institute of Navigation (ION GNSS+ 2017)*, pages 3028–3038, 2017.