
Mallows Model Averaging Estimator for the MIDAS Model with Almon Polynomial Weight

Hsin-Chieh Wong[a] and Wen-Jen Tsay[b]

National Central University,[a] National Chengchi University,[a]

and Academia Sinica[b]

Supplementary Material

The Supplementary Material shows that the proposed MMA estimator has the property of asymptotic optimality. The Supplementary Material is structured as follows. In S1-S3 we deal with the properties of the OLS estimator, the penalty term, and the proposed MMA, respectively. We then prove Theorem 1 by extending the method of Wan et al. (2010) in the last section.

S1 Proof of Lemma 1

Proof of Lemma 1. To start with, it is known that:

$$\text{rank}\,[A'A] = \text{rank}\,[AA'] = \text{rank}\,[A] = \text{rank}\,[A']\,,$$

for any matrix A. From this, for all $q \leq Q$ we have:

$$\operatorname{rank}\left[\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q\right] = \operatorname{rank}\left[\tilde{\mathbf{X}}_q\right] = \operatorname{rank}\left[\mathbf{X}_q\mathbf{U}_q\right], \tag{S1.1}$$

where the second step follows from equation (3.7). Since $\mathbf{X}_Q$ have full rank, $\mathbf{X}_q$ have full rank for all $q \leq Q$. Hence, we observe:

$$\operatorname{rank}\left[\mathbf{X}_q\mathbf{U}_q\right] = \operatorname{rank}\left[\mathbf{U}_q\right], \tag{S1.2}$$

whereby $\mathbf{X}_q$ is an $n \times (1 + \sum_{i=1}^{k_q} K_i)$ matrix with rank $(1 + \sum_{i=1}^{k_q} K_i)$. By combining (S1.1) and (S1.2), we have:

$$\operatorname{rank}\left[\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q\right] = \operatorname{rank}\left[\mathbf{U}_q\right].$$

Remember that the matrix $\mathbf{V}(K, p)$ defined in (2.7) stands for a Vandermonde matrix. It is also widely known that:

$$\det\left(\mathbf{V}(N, N)\right) = \prod_{1 \leq i < j \leq N} (j - i),$$

which means that the N rows of $\mathbf{V}(N, N)$ are linearly independent. Recall from the assumption that $p_i < K_i$ for $i \in \mathbb{N}$. Since K_i is a finite number, we have $\operatorname{rank}\left[\mathbf{V}(K_i, p_i)\right] = p_i$ for $i \in \mathbb{N}$. Thus, we obtain:

$$\operatorname{rank}\left[\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q\right] = \operatorname{rank}\left[\mathbf{U}_q\right] = 1 + \sum_{i=1}^{k_q} p_i.$$

Note that $\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q$ is an $\left(1 + \sum_{i=1}^{k_q} p_i\right) \times \left(1 + \sum_{i=1}^{k_q} p_i\right)$ matrix. Therefore, we conclude that the matrix $\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q$ has full rank and is invertible for all $q \leq Q$. $\square$

S2 Proof of Lemma 2

Proof of Lemma 2. For item (i), note that we have the invertible property of $\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q$ by Lemma 1. Thus, we obtain that:

$$\begin{aligned}\mathrm{tr}\left(\mathbf{P}_q\right) =& \mathrm{tr}\left(\tilde{\mathbf{X}}_q\left(\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q\right)^{-1}\tilde{\mathbf{X}}_q'\right)\\ =& \mathrm{tr}\left(\left(\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q\right)^{-1}\left(\tilde{\mathbf{X}}_q'\tilde{\mathbf{X}}_q\right)\right)\\ =& \mathrm{tr}\left(\mathbf{I}_{(1+\sum_{i=1}^{k_q}p_i)}\right) = 1+\sum_{i=1}^{k_q}p_i,\end{aligned}$$

and $\mathrm{tr}\left(\mathbf{P}(\mathbf{w})\right) = \mathrm{tr}\left(\sum_{q=1}^{Q} w_q\mathbf{P}_q\right) = \sum_{q=1}^{Q} w_q\mathrm{tr}\left(\mathbf{P}_q\right) = 1+\sum_{q=1}^{Q}\sum_{i=1}^{k_q} w_q p_i$.

Item (ii) follows from the fact that $\mathrm{tr}\left(\mathbf{P}_m\mathbf{P}_l\right) = 1+\sum_{i=1}^{k_{\min\{m,l\}}} p_i$ and simple algebra. For item (iii), we have:

$$\lambda_{\max}\left(\mathbf{P}(\mathbf{w})\right) = \max_{\boldsymbol{\eta}}\frac{\boldsymbol{\eta}'\mathbf{P}(\mathbf{w})\boldsymbol{\eta}}{\boldsymbol{\eta}'\boldsymbol{\eta}} \leq \sum_{q=1}^{Q} w_q\max_{\boldsymbol{\eta}}\frac{\boldsymbol{\eta}'\mathbf{P}_q\boldsymbol{\eta}}{\boldsymbol{\eta}'\boldsymbol{\eta}} = 1,$$

because $\mathbf{P}_q$ is an idempotent matrix. □

S3 Proof of Proposition 1

Proof of Proposition 1. Observe that:

$$C_n(\mathbf{w}) - L_n(\mathbf{w}) = \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + 2\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu} - 2\left(\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon} - \sigma^2 k^{(p)}(\mathbf{w})\right) \tag{S3.3}$$

and:

$$\begin{aligned} E[\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}|\mathbf{X}^{(m)}] &= E\left[\operatorname{tr}\left(\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}\right)\middle|\mathbf{X}^{(m)}\right] \\ &= E\left[\operatorname{tr}\left(\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}\boldsymbol{\epsilon}'\right)\middle|\mathbf{X}^{(m)}\right] \\ &= \operatorname{tr}\left(\mathbf{P}(\mathbf{w})\cdot E\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}^{(m)}\right]\right). \end{aligned}$$

Recall that $E\left[\boldsymbol{\epsilon}\boldsymbol{\epsilon}'|\mathbf{X}^{(m)}\right] = \sigma^2\mathbf{I}$. Thus, we have:

$$E[\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}|\mathbf{X}^{(m)}] = \sigma^2\operatorname{tr}\left(\mathbf{P}(\mathbf{w})\cdot\mathbf{I}\right) = \sigma^2 k^{(p)}(\mathbf{w}).$$

This result therefore follows by taking the expectation of (S3.3). □

S4 Proof of Theorem 1

We now shall prove Theorem 1. The proof is adapted from Wan et al. (2010, Theorems 1 and 2). Let us now outline the way to prove Theorem 1, which is basically split into Step I and Step II.

For Step I, we extend the idea of the proof by Wan et al. (2010), which is identical to that of Li (1987). Indeed, Theorem 2.1 of Li (1987) presents asymptotic optimality for a broad class of linear estimators. More specifically, we recall from (3.22) that:

$$C_n(\mathbf{w}) = L_n(\mathbf{w}) + \boldsymbol{\epsilon}'\boldsymbol{\epsilon} + 2\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu} - 2\left(\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon} - \sigma^2 k^{(p)}(\mathbf{w})\right). \tag{S4.4}$$

The first step is to thus show that as $n \to \infty$:

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \frac{\left|2\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}\right|}{R_n(\mathbf{w})} \xrightarrow{p} 0, \tag{S4.5}$$

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \frac{\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}-\sigma^2 k^{(p)}(\mathbf{w})\right|}{R_n(\mathbf{w})} \xrightarrow{p} 0, \tag{S4.6}$$

and:

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \left|\frac{L_n(\mathbf{w})}{R_n(\mathbf{w})}-1\right| \xrightarrow{p} 0. \tag{S4.7}$$

Therefore, the proof of the first step will be split into three parts. Nevertheless, a similar strategy is adopted to deal with each part, respectively. In other words, we first provide an upper bound for the case of non-stochastic $\mathbf{X}^{(m)}$. We then complete the proof by removing this constraint for the case of non-stochastic $\mathbf{X}^{(m)}$.

For Step II, we prove that as $n \to \infty$:

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \frac{\mathrm{tr}\left(\mathbf{P}(\mathbf{w})\right)\left(\hat{\sigma}_Q^2-\sigma^2\right)}{R_n(\mathbf{w})} \xrightarrow{p} 0, \tag{S4.8}$$

because:

$$\hat{C}_n(\mathbf{w}) = C_n(\mathbf{w}) + 2\mathrm{tr}\left(\mathbf{P}(\mathbf{w})\right)\left(\hat{\sigma}_Q^2-\sigma^2\right). \tag{S4.9}$$

Proof of Theorem 1. **Part I of Step I.** First, for any $\delta > 0$ and the case of non-stochastic $\mathbf{X}^{(m)}$, we observe:

$$
\begin{aligned}
P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\frac{|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}|}{R_n(\mathbf{w})}>\delta\right] &\leq P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}|>\delta\Xi_n\right]\\
&\leq P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\sum_{q=1}^{Q}w_q|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}|>\delta\Xi_n\right]\\
&\leq P\left[\bigcup_{q=1}^{Q}\left\{|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}|>\delta\Xi_n\right\}\right]\\
&\leq \sum_{q=1}^{Q}P\left[|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}|>\delta\Xi_n\right]\\
&\leq \sum_{q=1}^{Q}\frac{E|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}|^{2N}}{\delta^{2N}\Xi_n^{2N}}, \qquad \text{(S4.10)}
\end{aligned}
$$

in which the last step follows by Chebyshev's inequality. Through Theorem 2 of Whittle (1960), we next obtain:

$$
\begin{aligned}
&\sum_{q=1}^{Q}\frac{E|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}|^{2N}}{\delta^{2N}\Xi_n^{2N}}\\
&\leq C_1\delta^{-2N}\Xi_n^{-2N}\sum_{q=1}^{Q}\left|\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right|^{N}, \qquad \text{(S4.11)}
\end{aligned}
$$

for some constant $C_1 > 0$. Recall from (3.19) that:

$$
R_n(\mathbf{w}) = |\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}| + \sigma^2\mathrm{tr}\left(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\right),
$$

and so we have $R_n(\mathbf{w}) \geq |\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}|$. Therefore, we obtain:

$$
P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\frac{|\boldsymbol{\epsilon}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}|}{R_n(\mathbf{w})}>\delta\right] \leq C_1\delta^{-2N}\Xi_n^{-2N}\sum_{q=1}^{Q}\left[R_n(\mathbf{w}_q^0)\right]^N. \qquad \text{(S4.12)}
$$

Finally, when $\mathbf{X}^{(m)}$ is random, with the dominated convergence theorem (DCT), the result (S4.5) is claimed by combining (3.29) and (S4.12).

Part II of Step I. For the case of non-stochastic $\mathbf{X}^{(m)}$, by Chebyshev's inequality, Theorem 2 of Whittle (1960), and the fact (3.21) that $R_n(\mathbf{w}) > \sigma^2 \mathrm{tr}\left(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\right)$, we similarly have:

$$
\begin{aligned}
& P\left[\sup_{\mathbf{w}\in\mathcal{H}_n} \frac{\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon} - \sigma^2 k^{(p)}(\mathbf{w})\right|}{R_n(\mathbf{w})} > \delta\right] \\
& \leq P\left[\sup_{\mathbf{w}\in\mathcal{H}_n} \left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon} - \sigma^2 k^{(p)}(\mathbf{w})\right| > \delta\Xi_n\right] \\
& \leq \sum_{q=1}^{Q} \frac{E\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_q^0)\boldsymbol{\epsilon} - \sigma^2 k^{(p)}(\mathbf{w}_q^0)\right|^{2N}}{\delta^{2N}\Xi_n^{2N}} \\
& \leq C_2 \delta^{-2N}\Xi_n^{-2N} \sum_{q=1}^{Q} \left[\mathrm{tr}\left(\mathbf{P}(\mathbf{w}_q^0)\mathbf{P}(\mathbf{w}_q^0)\right)\right]^N \\
& \leq C_2' \delta^{-2N}\Xi_n^{-2N} \sum_{q=1}^{Q} \left[R_n(\mathbf{w}_q^0)\right]^N, \qquad \text{(S4.13)}
\end{aligned}
$$

where $C_2 > 0$ and $C_2' > 0$ are constants. When $\mathbf{X}^{(m)}$ is random, by combining (3.29) and (S4.13), the result (S4.6) is thus proved by DCT.

Part III of Step I. To start with, we note that the result (S4.7) is equivalent to:

$$
\sup_{\mathbf{w}\in\mathcal{H}_n} \left| \frac{|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}|^2 - \sigma^2 \mathrm{tr}\left(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\right) - 2\left(\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\left(\mathbf{I} - \mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}\right)}{R_n(\mathbf{w})} \right| \xrightarrow{p} 0. \qquad \text{(S4.14)}
$$

Therefore, to prove (S4.7), it suffices to show that as $n \to \infty$:

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \left| \frac{\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}}{R_n(\mathbf{w})} \right| \xrightarrow{p} 0, \tag{S4.15}$$

and:

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \left| \frac{|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}|^2 - \sigma^2 \mathrm{tr}\left(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\right)}{R_n(\mathbf{w})} \right| \xrightarrow{p} 0. \tag{S4.16}$$

We now shall prove (S4.15). For the case of non-stochastic $\mathbf{X}^{(m)}$, by Chebyshev's inequality and Theorem 2 of Whittle (1960), we observe that for any $\delta > 0$:

$$\begin{aligned}
&P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\left|\frac{\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}}{R_n(\mathbf{w})}\right| > \delta\right] \\
&\quad\le P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}\right| > \delta\Xi_n\right] \\
&\quad\le P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\left|\sum_{t=1}^{Q} w_t\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\sum_{q=1}^{Q} w_q\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right| > \delta\Xi_n\right] \\
&\quad\le P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\sum_{t=1}^{Q}\sum_{q=1}^{Q} w_t w_q\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right| > \delta\Xi_n\right] \\
&\quad\le P\left[\bigcup_{t=1}^{Q}\bigcup_{q=1}^{Q}\left\{\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right| > \delta\Xi_n\right\}\right] \\
&\quad\le \sum_{t=1}^{Q}\sum_{q=1}^{Q} P\left[\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right| > \delta\Xi_n\right] \\
&\quad\le \sum_{t=1}^{Q}\sum_{q=1}^{Q}\frac{E\left|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right|^{2N}}{\delta^{2N}\Xi_n^{2N}} \\
&\quad\le C_3\delta^{-2N}\Xi_n^{-2N}\sum_{t=1}^{Q}\sum_{q=1}^{Q}\left|\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_t^0)\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right|^{N},
\end{aligned} \tag{S4.17}$$

where $C_3 > 0$ is a constant. We further note that:

$$\begin{aligned}
&\left|\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}\right| \\
&\le \lambda_{\max}^2\left(\mathbf{P}(\mathbf{w})\right)\cdot\left|\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}\right| \\
&\le \left|\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\left(\mathbf{I}-\mathbf{P}(\mathbf{w})\right)\boldsymbol{\mu}\right|,
\end{aligned} \tag{S4.18}$$

in which the second step holds by item (iii) of Lemma 2. By the fact of

(3.20) that $R_n(\mathbf{w}) \geq |\boldsymbol{\mu}'(\mathbf{I}-\mathbf{P}(\mathbf{w}))(\mathbf{I}-\mathbf{P}(\mathbf{w}))\boldsymbol{\mu}|$, we thus have:

$$\begin{aligned} &P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\left|\frac{\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})(\mathbf{I}-\mathbf{P}(\mathbf{w}))\boldsymbol{\mu}}{R_n(\mathbf{w})}\right| > \delta\right] \\ &\quad \leq C_3\delta^{-2N}\Xi_n^{-2N}\sum_{t=1}^{Q}\sum_{q=1}^{Q}\left|\boldsymbol{\mu}'\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\left(\mathbf{I}-\mathbf{P}(\mathbf{w}_q^0)\right)\boldsymbol{\mu}\right|^N \\ &\quad \leq C_3\delta^{-2N}\Xi_n^{-2N}Q\sum_{q=1}^{Q}\left[R_n(\mathbf{w}_q^0)\right]^N, \end{aligned} \tag{S4.19}$$

for some constant $C_3 > 0$. When $\mathbf{X}^{(m)}$ is random, this completes the proof of the result (S4.15) by using (3.29) and DCT.

By Chebyshev's inequality, Theorem 2 of Whittle (1960), and the fact that $R_n(\mathbf{w}) > \sigma^2\text{tr}(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w}))$, we thus obtain:

$$\begin{aligned} &P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\left|\frac{|\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\boldsymbol{\epsilon}|^2-\sigma^2\text{tr}(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w}))}{R_n(\mathbf{w})}\right| > \delta\right] \\ &\quad \leq P\left[\sup_{\mathbf{w}\in\mathcal{H}_n}\sum_{t=1}^{Q}\sum_{q=1}^{Q}w_tw_q\left||\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\boldsymbol{\epsilon}|^2-\sigma^2\text{tr}\left(\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\right)\right| > \delta\Xi_n\right] \\ &\quad \leq P\left[\bigcup_{t=1}^{Q}\bigcup_{q=1}^{Q}\left\{\left||\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\boldsymbol{\epsilon}|^2-\sigma^2\text{tr}\left(\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\right)\right| > \delta\Xi_n\right\}\right] \\ &\quad \leq \sum_{t=1}^{Q}\sum_{q=1}^{Q}P\left[\left||\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\boldsymbol{\epsilon}|^2-\sigma^2\text{tr}\left(\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\right)\right| > \delta\Xi_n\right] \\ &\quad \leq \sum_{t=1}^{Q}\sum_{q=1}^{Q}\frac{E\left||\boldsymbol{\epsilon}'\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\boldsymbol{\epsilon}|^2-\sigma^2\text{tr}\left(\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\right)\right|^{2N}}{\delta^{2N}\Xi_n^{2N}} \\ &\quad \leq C_3'\delta^{-2N}\Xi_n^{-2N}\sum_{\mathbf{w}\in\mathcal{H}_n}\left[\text{tr}\left(\mathbf{P}(\mathbf{w}_q^0)\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_t^0)\mathbf{P}(\mathbf{w}_q^0)\right)\right]^N, \end{aligned} \tag{S4.20}$$

where $C_3' > 0$ is a constant. Again, we note that:

$$\begin{aligned}\operatorname{tr}\left(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\right) &\le \lambda_{\max}^2\left(\mathbf{P}(\mathbf{w})\right)\cdot \operatorname{tr}\left(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\right)\\ &\le \operatorname{tr}\left(\mathbf{P}(\mathbf{w})\mathbf{P}(\mathbf{w})\right), \end{aligned} \tag{S4.21}$$

in which the second step holds by item (iii) of Lemma 2. When $\mathbf{X}^{(m)}$ is random, the result (S4.16) is thus obtained by adopting (3.29) and DCT. This proof is completed by combining Parts I-III.

Step II. The Mallows criterion can next be written as:

$$\hat{C}_n(\mathbf{w}) = C_n(\mathbf{w}) + 2\operatorname{tr}\left(\mathbf{P}(\mathbf{w})\right)\left(\hat{\sigma}_Q^2 - \sigma^2\right), \tag{S4.22}$$

where σ^2 is replaced by $\hat{\sigma}_Q^2$. Hence, in this step, it suffices to prove that as $n \to \infty$ then:

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \frac{\operatorname{tr}\left(\mathbf{P}(\mathbf{w})\right)\left(\hat{\sigma}_Q^2 - \sigma^2\right)}{R_n(\mathbf{w})} \xrightarrow{p} 0. \tag{S4.23}$$

We observe that:

$$\sup_{\mathbf{w}\in\mathcal{H}_n} \frac{\operatorname{tr}\left(\mathbf{P}(\mathbf{w})\right)\left(\hat{\sigma}_Q^2 - \sigma^2\right)}{R_n(\mathbf{w})} \le \frac{\left(1 + \sum_{i=1}^{k_Q} p_i\right)}{\Xi_n}\left|\hat{\sigma}_Q^2 - \sigma^2\right|. \tag{S4.24}$$

Recall from (3.26) that:

$$\hat{\sigma}_Q^2 = \frac{(Y - \tilde{\mathbf{X}}_Q\hat{\Gamma}_Q)'(Y - \tilde{\mathbf{X}}_Q\hat{\Gamma}_Q)}{n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)} = \frac{Y'(\mathbf{I} - \mathbf{P}_Q)Y}{n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)},$$

where $\mathbf{I}$ is the $n \times n$ identity matrix, and Q corresponds to the largest approximating model. Thus, we have:

$$\begin{aligned}
&\frac{\left(1+\sum_{i=1}^{k_Q} p_i\right)}{\Xi_n}\left|\hat{\sigma}_Q^2-\sigma^2\right| \\
=&\frac{\left(1+\sum_{i=1}^{k_Q} p_i\right)}{\Xi_n}\left|\frac{Y'(\mathbf{I}-\mathbf{P}_Q)Y}{n-\left(1+\sum_{i=1}^{k_Q} p_i\right)}-\sigma^2\right| \\
\leq&\frac{\left(1+\sum_{i=1}^{k_Q} p_i\right)}{n-\left(1+\sum_{i=1}^{k_Q} p_i\right)}\frac{|\boldsymbol{\mu}'(\mathbf{I}-\mathbf{P}_Q)\boldsymbol{\mu}|}{\Xi_n}+\frac{2\left(1+\sum_{i=1}^{k_Q} p_i\right)|\boldsymbol{\mu}'(\mathbf{I}-\mathbf{P}_Q)\boldsymbol{\epsilon}|}{\Xi_n\left(n-\left(1+\sum_{i=1}^{k_Q} p_i\right)\right)} \\
&+\frac{\left(1+\sum_{i=1}^{k_Q} p_i\right)\left|\boldsymbol{\epsilon}'(\mathbf{I}-\mathbf{P}_Q)\boldsymbol{\epsilon}-\sigma^2\left(1+\sum_{i=1}^{k_Q} p_i\right)\right|}{\Xi_n\left(n-\left(1+\sum_{i=1}^{k_Q} p_i\right)\right)} \\
=:&J_1+J_2+J_3. \qquad \text{(S4.25)}
\end{aligned}$$

Therefore, the remaining proof is to bound J_1, J_2, and J_3. Here, $\mathbf{w}_Q^0$ is an $Q\times 1$ vector in which the Qth element is one, and the others are zeros, so that:

$$\mathbf{P}_Q=\mathbf{P}(\mathbf{w}_Q^0). \qquad \text{(S4.26)}$$

By (S4.26), we have that as $n\to\infty$:

$$\begin{aligned}
\frac{\boldsymbol{\mu}'(\mathbf{I}-\mathbf{P}_Q)\boldsymbol{\mu}}{\Xi_n^2}&=\frac{\boldsymbol{\mu}'(\mathbf{I}-\mathbf{P}(\mathbf{w}_Q^0))\boldsymbol{\mu}}{\Xi_n^2} \\
&=\frac{R_n(\mathbf{w}_Q^0)+2\boldsymbol{\mu}'(\mathbf{I}-\mathbf{P}_Q)\mathbf{P}_Q\boldsymbol{\epsilon}+\boldsymbol{\epsilon}'\mathbf{P}_Q\boldsymbol{\epsilon}}{\Xi_n^2} \\
&=\frac{R_n(\mathbf{w}_Q^0)+\boldsymbol{\epsilon}'\mathbf{P}_Q\boldsymbol{\epsilon}}{\Xi_n^2}\to 0, \qquad \text{(S4.27)}
\end{aligned}$$

in which the last step holds by utilizing (3.29). By Assumption 2, we get that as $n \to \infty$:

$$\frac{\boldsymbol{\mu}'\boldsymbol{\mu}}{n} = \frac{\sum_{t=0}^{n-1} \mu_t^2}{n} = O(1). \tag{S4.28}$$

Combining (3.31), (S4.27), and (S4.28), we obtain that as $n \to \infty$:

$$J_1 \leq \left(\frac{\left(1 + \sum_{i=1}^{k_Q} p_i\right)^2}{n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)} \frac{\boldsymbol{\mu}'(\mathbf{I} - \mathbf{P}_Q)\boldsymbol{\mu}}{\Xi_n^2} \frac{\boldsymbol{\mu}'\boldsymbol{\mu}}{n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)} \right)^{1/2} \to 0. \tag{S4.29}$$

To bound J_2, we first consider the case of non-stochastic $\mathbf{X}^{(m)}$ and then observe that for any $\delta > 0$ and some constant $C_4 > 0$,

$$\begin{aligned} & P\left[\frac{2\left(1 + \sum_{i=1}^{k_Q} p_i\right) |\boldsymbol{\mu}'(\mathbf{I} - \mathbf{P}_Q)\boldsymbol{\epsilon}|}{\Xi_n \left(n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)\right)} > \delta \right] \\ & \leq |\boldsymbol{\mu}'(\mathbf{I} - \mathbf{P}_Q)(\mathbf{I} - \mathbf{P}_Q)\boldsymbol{\mu}| \frac{4C_4 \left(1 + \sum_{i=1}^{k_Q} p_i\right)^2}{\delta^2 \Xi_n^2 \left(n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)\right)^2}, \end{aligned} \tag{S4.30}$$

in which this inequality holds by Chebyshev's inequality and Theorem 2 of Whittle (1960). When $\mathbf{X}^{(m)}$ is random, by combining (3.29) and (S4.27), we have that as $n \to \infty$:

$$J_2 \to 0 \tag{S4.31}$$

by DCT. Finally, because $E[\boldsymbol{\epsilon}'(\mathbf{I} - \mathbf{P}_Q)\boldsymbol{\epsilon}] = \sigma^2[n - (1 + \sum_{i=1}^{k_Q} p_i)]$, by Theorem 2 of Whittle (1960), we obtain that for any $\delta > 0$, there exist a constant

$C_5 > 0$ and $\kappa > 0$ such that:

$$E\left|\boldsymbol{\epsilon}'(\mathbf{I}-\mathbf{P}_Q)\boldsymbol{\epsilon} - \sigma^2\left(n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)\right)\right|^2$$
$$\leq C_5 \kappa^{1/(N+\delta)} \mathrm{tr}\left((\mathbf{I}-\mathbf{P}_Q)(\mathbf{I}-\mathbf{P}_Q)\right)$$
$$= C_5 \kappa^{1/(N+\delta)} \left(n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)\right). \quad \text{(S4.32)}$$

Thus, to bound J_3, for any $\delta > 0$, by Markov inequality we have that as $n \to \infty$:

$$P\left[\left|\frac{1}{n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)} \boldsymbol{\epsilon}'(\mathbf{I}-\mathbf{P}_Q)\boldsymbol{\epsilon} - \sigma^2\right| \geq \delta\right]$$
$$\leq \frac{E\left|\boldsymbol{\epsilon}'(\mathbf{I}-\mathbf{P}_Q)\boldsymbol{\epsilon} - \sigma^2\left(n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)\right)\right|^2}{\delta^2\left(n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)\right)^2}$$
$$\leq \frac{C_5 \kappa^{1/(N+\delta)}}{\delta^2\left(n - \left(1 + \sum_{i=1}^{k_Q} p_i\right)\right)} \to 0. \quad \text{(S4.33)}$$

Combining (S4.25), (S4.29), (S4.31), and (S4.33), we get that as $n \to \infty$:

$$\sup_{\mathbf{w} \in \mathcal{H}_n} \frac{\mathrm{tr}\left(\mathbf{P}(\mathbf{w})\right)\left(\hat{\sigma}_Q^2 - \sigma^2\right)}{R_n(\mathbf{w})} \xrightarrow{p} 0.$$

Combining Steps I and II, Theorem 1 is now obtained. □

Bibliography

Li, K. C. (1987). “Asymptotic optimality for C_p, C_L, cross-validation and generalized cross-validation: Discrete index set.” *Annals of Statistics.* 15, 958–975.

Wan, A. T. K., Zhang, X., and Zou, G. (2010). “Least squares model averaging by Mallows criterion.” *Journal of Econometrics.* 156, 277–283.

Whittle, P. (1960). “Bounds for the moments of linear and quadratic forms in independent variables.” *Theory of Probability and Its Applications.* 5, 302–305.