Cross Validation, Monte Carlo Cross Validation, and Accumulated Prediction Errors: Asymptotic Properties

Ching-Kang Ing

Institute of Statistics, National Tsing Hua University, Hsinchu, Taiwan

Outline

Cross Validation (CV)

- Consider

$$y_i = \boldsymbol{x}_i'\boldsymbol{\beta} + e_i = \sum_{j=1}^{p} x_{ij}\beta_j + e_i,$$

  where $\boldsymbol{\beta}' = (\beta_1, \ldots, \beta_p)$ and e_i are i.i.d. r.v.s with $E(e_i) = 0$ and $E(e_i^2) = \sigma^2 > 0$.
- Suppose $\beta_i = 0$ for some $i \in \mathbb{P} = \{1, \ldots, p\}$.
- We are interested in choosing the smallest true model

$$y_i = \boldsymbol{x}_{i,\alpha^*}'\boldsymbol{\beta}_{\alpha^*} + e_i,$$

  where $\alpha^* = \{i : \beta_i \neq 0\}$, $\boldsymbol{\beta}_{\alpha^*}' = (\beta_i, i \in \alpha^*)$ and $\boldsymbol{x}_{i,\alpha^*}' = (x_{ij}, j \in \alpha^*)$, among the candidate models

$$E(\boldsymbol{y}) \in C(\boldsymbol{X}_\alpha),$$

  where $\boldsymbol{y}' = (y_1, \ldots, y_n)$, $\boldsymbol{X}_\alpha' = (\boldsymbol{x}_{i,\alpha}, \ldots, \boldsymbol{x}_{n,\alpha})$ with $\boldsymbol{x}_{i,\alpha}' = (x_{ij}, j \in \alpha)$, and $\alpha \in 2^{\mathbb{P}}$.

Delete-n_v-out CV (CV(n_v))

- Let $(y_i, \boldsymbol{x}_i)$, $i = 1, \ldots, n$, be observations.
- CV(n_v) splits the data into two parts:

$$\{(y_i, \boldsymbol{x}_i), i \in s\} \quad \text{and} \quad \{(y_i, \boldsymbol{x}_i), i \in s^c\},$$

  where s and s^c are subsets of $\mathbb{N} = \{1, \ldots, n\}$ with $s \cap s^c = \emptyset$ and $s \cup s^c = \mathbb{N}$.
  - The first part is referred to as the validation (testing) sample, whereas the second part is referred to as the training sample.
- Denote $\sharp(s)$ and $\sharp(s^c)$ by n_v and n_c, respectively, noting that $n_v + n_c = n$.
- The CV evaluates the performance of α using

$$\begin{aligned} \hat{\Gamma}_{\alpha,n}(n_v) &= \frac{1}{n_v \binom{n}{n_v}} \sum_{\text{All } (s,s^c) \text{ combinations}} \sum_{i \in s} (y_i - \boldsymbol{x}'_{i,\alpha} \hat{\boldsymbol{\beta}}_{\alpha,s^c})^2 \\ &= \frac{1}{n_v \binom{n}{n_v}} \sum_{\text{All } (s,s^c) \text{ combinations}} \|\boldsymbol{y}_s - \hat{\boldsymbol{y}}_{\alpha,s^c}\|^2, \qquad (\hat{\boldsymbol{y}}_{\alpha,s^c} = \boldsymbol{X}_{\alpha,s} \hat{\boldsymbol{\beta}}_{\alpha,s^c}) \end{aligned}$$

  where $\boldsymbol{y}'_A = (y_i, i \in A)$, $A \subseteq \mathbb{N}$ and $\hat{\boldsymbol{\beta}}_{\alpha,s^c} = (\boldsymbol{X}'_{\alpha,s^c} \boldsymbol{X}_{\alpha,s^c})^{-1} \boldsymbol{X}'_{\alpha,s^c} \boldsymbol{y}_{s^c}$ with $\boldsymbol{X}'_{\alpha,A} = (\boldsymbol{x}_{i,\alpha}, i \in A)$.

Some statistical properties of CV

Define $\boldsymbol{P}_\alpha = \boldsymbol{X}_\alpha(\boldsymbol{X}'_\alpha \boldsymbol{X}_\alpha)^{-1}\boldsymbol{X}'_\alpha$ and $\boldsymbol{Q}_{\alpha,A} = \boldsymbol{X}_{\alpha,A}(\boldsymbol{X}'_\alpha \boldsymbol{X}_\alpha)^{-1}\boldsymbol{X}'_{\alpha,A}$.

Fact 1

$$n_v^{-1}\|\boldsymbol{y}_s - \hat{\boldsymbol{y}}_{\alpha,s^c}\|^2 = n_v^{-1}\|(\boldsymbol{I}_{n_v} - \boldsymbol{Q}_{\alpha,s})^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{\beta}}_\alpha)\|^2,$$

where $\boldsymbol{I}_{n_v}$ denotes the n_v-dimensional identity matrix and

$$\hat{\boldsymbol{\beta}}_\alpha = (\boldsymbol{X}'_\alpha \boldsymbol{X}_\alpha)^{-1}\boldsymbol{X}'_\alpha \boldsymbol{y},$$

which is the least squares estimate of $\boldsymbol{\beta}_\alpha$ based on the "full" data.

Proof

Since

$$(\boldsymbol{X}'_{\alpha,s^c}\boldsymbol{X}_{\alpha,s^c})^{-1} = (\boldsymbol{X}'_\alpha \boldsymbol{X}_\alpha - \boldsymbol{X}'_{\alpha,s}\boldsymbol{X}_{\alpha,s})^{-1},$$

by making use of

$$(\boldsymbol{A} - \boldsymbol{B}'\boldsymbol{B})^{-1} = \boldsymbol{A}^{-1} + \boldsymbol{A}^{-1}\boldsymbol{B}'(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{A}^{-1}\boldsymbol{B}')^{-1}\boldsymbol{B}\boldsymbol{A}^{-1}, \quad (1)$$

($\boldsymbol{I}$: identity matrix of suitable dimension)

we obtain

$$\begin{aligned}\hat{\boldsymbol{y}}_{\alpha,s^c} &= \boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{\beta}}_{\alpha,s^c} \\ &= \boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{\beta}}_\alpha + \boldsymbol{Q}_{\alpha,s}(\boldsymbol{I} - \boldsymbol{Q}_{\alpha,s})^{-1}\boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{Q}_{\alpha,s}\boldsymbol{y}_s - \boldsymbol{Q}_{\alpha,s}(\boldsymbol{I} - \boldsymbol{Q}_{\alpha,s})^{-1}\boldsymbol{Q}_{\alpha,s}\boldsymbol{y}_s,\end{aligned}$$

and hence

$$\boldsymbol{y}_s - \boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{\beta}}_{\alpha,s^c} = (\boldsymbol{I} - \boldsymbol{Q}_{\alpha,s})^{-1}(\boldsymbol{y}_s - \boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{\beta}}_\alpha).$$

This completes the proof.

Remark

By Fact 1, we have

$$\hat{\Gamma}_{\alpha,n}(1) = n^{-1}\sum_{i=1}^{n}\{(1 - w_{i\alpha})^{-1}(y_i - \boldsymbol{x}'_{i,\alpha}\hat{\boldsymbol{\beta}}_\alpha)\}^2, \quad \rightarrow \text{ Delete-1-out CV (conventional CV)}$$

where $w_{i\alpha} = (\boldsymbol{P}_\alpha)_{ii}$, the ith diagonal element of $\boldsymbol{P}_\alpha$.

Theorem 1

Assume

$$\frac{\boldsymbol{X}'\boldsymbol{X}}{n} \underset{n\to\infty}{\longrightarrow} \boldsymbol{R}, \text{ a positive definite matrix,} \tag{2}$$

where $\boldsymbol{X}' = (\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n)$, and

$$\max_{1\le i\le n} w_{i\alpha} \underset{n\to\infty}{\longrightarrow} 0 \text{ for any } \alpha \in 2^{\mathbb{P}}. \tag{3}$$

Then

(1)

$$\hat{\Gamma}_{\alpha,n}(1) = \sigma^2 + \frac{1}{n}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} + o_p(1),$$

($\boldsymbol{I}$: n-dimensional identity matrix)

if α is an incorrect model (namely $\alpha^* - \alpha \neq \emptyset$),

(2)

$$\hat{\Gamma}_{\alpha,n}(1) = n^{-1}\boldsymbol{e}'\boldsymbol{e} + \frac{2d_\alpha\sigma^2}{n} - \frac{1}{n}\boldsymbol{e}'\boldsymbol{P}_\alpha\boldsymbol{e} + o_p(n^{-1}),$$

if α is a correct model ($\alpha^* - \alpha = \emptyset$), where $d_\alpha = \sharp(\alpha)$ and $\boldsymbol{e}' = (e_1, \ldots, e_n)$.

Proof

- Since

$$(1 - w_{i\alpha})^2 = 1 + 2w_{i\alpha} + O(w_{i\alpha}^2),$$

where $O(\cdot)$ denotes a bound uniform over $1 \leq i \leq n$, it holds that

$$\hat{\Gamma}_{\alpha,n}(1) = \frac{1}{n}\sum_{i=1}^{n}\gamma_{i\alpha}^2 + \frac{1}{n}\sum_{i=1}^{n}(2w_{i\alpha} + O(w_{i\alpha}^2))\gamma_{i,\alpha}^2 \equiv (\mathrm{I}) + (\mathrm{II}),$$

where $\gamma_{i,\alpha} = y_i - \boldsymbol{x}_{i,\alpha}'\hat{\boldsymbol{\beta}}_\alpha$.

- If α is incorrect, then

$$\begin{aligned}
(\mathrm{I}) &= n^{-1}\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{y} \\
&= n^{-1}(\boldsymbol{e}' + \boldsymbol{\beta}'\boldsymbol{X}')(\boldsymbol{I} - \boldsymbol{P}_\alpha)(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{e}) \\
&= n^{-1}\boldsymbol{e}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{e} + 2n^{-1}\boldsymbol{e}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} + n^{-1}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} \\
&\overset{\text{why?}}{=} \sigma^2 + n^{-1}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} + o_p(1),
\end{aligned}$$

and $(\mathrm{II}) = o_p(1)$. Then (1) follows.

If α is correct, then

$$(\mathrm{I}) = n^{-1}\boldsymbol{e}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{e},$$

and

$$\begin{aligned}
(\mathrm{II}) &= n^{-1}\sum_{i=1}^{n}(2w_{i\alpha} + O(w_{i\alpha}^2))(e_i - \boldsymbol{x}'_{i,\alpha}(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha))^2 \\
&\overset{\text{why?}}{=} \frac{2}{n}\sum_{i=1}^{n} w_{i\alpha}e_i^2 + o_p(n^{-1}) \\
&= \frac{2}{n}\sum_{i=1}^{n} w_{i\alpha}\sigma^2 + \frac{2}{n}\sum_{i=1}^{n} w_{i\alpha}(e_i^2 - \sigma^2) + o_p(n^{-1}) \\
&\overset{\text{why?}}{=} \frac{2d_\alpha\sigma^2}{n} + o_p(1).
\end{aligned}$$

This completes the proof.

Corollary 1

Let

$$\hat{\alpha} = \mathrm{argmin}_{\alpha\in 2^{\mathbb{P}}}\hat{\Gamma}_{\alpha,n}(1).$$

Then $\lim_{n\to\infty} P(\alpha^* - \hat{\alpha} = \emptyset) = 1$, but $\lim_{n\to\infty} P(\hat{\alpha} = \alpha^*)$ is in general less than 1.

Notation

- $b = \binom{n}{n_v}$
- $\mathcal{B}_{n_v} = \{s : s \subseteq \mathbb{N} \text{ and } \sharp(s) = n_v\}$
- $\boldsymbol{P}_{\alpha,s} = \boldsymbol{X}_{\alpha,s}(\boldsymbol{X}'_{\alpha,s}\boldsymbol{X}_{\alpha,s})^{-1}\boldsymbol{X}'_{\alpha,s}$
- $\hat{\boldsymbol{R}}_{\alpha,s} = \frac{1}{n_v}\sum_{t\in s}\boldsymbol{x}_{t,\alpha}\boldsymbol{x}'_{t,\alpha}$
- $\hat{\boldsymbol{R}}_{\alpha,s^c} = \frac{1}{n_c}\sum_{t\in s^c}\boldsymbol{x}_{t,\alpha}\boldsymbol{x}'_{t,\alpha}$
- $\boldsymbol{\gamma}_{\alpha,s} = \boldsymbol{y}_s - \boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{\beta}}_\alpha$

Then, Fact 1 yields

$$\hat{\Gamma}_{\alpha,n}(n_v) = \frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}(\boldsymbol{I}_{n_v}-\boldsymbol{Q}_{\alpha,s})^{-1}(\boldsymbol{I}_{n_v}-\boldsymbol{Q}_{\alpha,s})^{-1}\boldsymbol{\gamma}_{\alpha,s} = \text{(I)}+\text{(II)}+\cdots+\text{(VI)}, \tag{4}$$

where

$$\text{(I)} = \frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}, \qquad \text{(II)} = \frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{P}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s},$$

$$\text{(III)} = \frac{1}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}(\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}-\hat{\boldsymbol{R}}^{-1}_{\alpha,s})\boldsymbol{X}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}(\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}-\hat{\boldsymbol{R}}^{-1}_{\alpha,s})\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s},$$

$$\text{(IV)} = 2\frac{n_v}{n_c}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{P}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s},$$

$$\text{(V)} = \frac{2}{n_c}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}(\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}-\hat{\boldsymbol{R}}^{-1}_{\alpha,s})\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s},$$

$$\text{(VI)} = \frac{2n_v}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{P}_{\alpha,s}\boldsymbol{X}_{\alpha,s}(\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}-\hat{\boldsymbol{R}}^{-1}_{\alpha,s})\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}. \qquad (\boldsymbol{P}_{\alpha,s}\boldsymbol{X}_{\alpha,s}=\boldsymbol{X}_{\alpha,s})$$

Hint

Using (1), one gets

$$(\boldsymbol{I}_{n_v}-\boldsymbol{Q}_{\alpha,s})^{-1} = \boldsymbol{I}_{n_v} + \frac{n_v}{n_c}\boldsymbol{P}_{\alpha,s} + \frac{1}{n_c}\boldsymbol{X}_{\alpha,s}(\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}-\hat{\boldsymbol{R}}^{-1}_{\alpha,s})\boldsymbol{X}'_{\alpha,s}.$$

Theorem 2

Assume the same assumptions as in Theorem 1. Assume also that

$$\lim_{n\to\infty} \max_{s\in\mathcal{B}_{n_v}} \|\hat{\boldsymbol{R}}_s - \hat{\boldsymbol{R}}_{s^c}\| = 0, \tag{5}$$

where $\hat{\boldsymbol{R}}_A = \frac{1}{\sharp(A)} \sum_{t\in A} \boldsymbol{x}_t \boldsymbol{x}_t'$ with $A \subseteq \mathbb{N}$. Suppose

$$\frac{n_v}{n} \to 1 \quad \text{and} \quad n_c = n - n_v \to \infty. \tag{6}$$

Then

(i) For $\alpha^* - \alpha \neq \emptyset$,

$$\hat{\Gamma}_{\alpha,n}(n_v) = n^{-1}\boldsymbol{e}'\boldsymbol{e} + n^{-1}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} + o_p(1) + R_n, \tag{7}$$

where R_n is a non-negative random variable.

(ii) For $\alpha^* - \alpha = \emptyset$,

$$\hat{\Gamma}_{\alpha,n}(n_v) = n^{-1}\boldsymbol{e}'\boldsymbol{e} + \frac{d_\alpha \sigma^2}{n_c} + o_p(n_c^{-1}). \tag{8}$$

(iii) $\lim_{n\to\infty} P(\hat{\alpha}_{n_v} = \alpha^*) = 1$, where $\hat{\alpha}_{n_v} = \operatorname{argmin}_{\alpha\in 2^{\mathbb{P}}} \hat{\Gamma}_{\alpha,n}(n_v)$.

Proof

- We begin by considering the case of $\alpha^* - \alpha = \emptyset$.
- We will first show that

$$\hat{\Gamma}_{\alpha,n}(n_v) = (\mathrm{I}) + (\mathrm{II})(1 + o(1)), \tag{9}$$

via proving

$$|(K)| = (\mathrm{II})o(1), \tag{10}$$

where $K = \mathrm{III}, \mathrm{IV}, \mathrm{V}$, and VI.

To show (10) holds with $K = \mathrm{VI}$, note that

$$(\mathrm{VI}) \leq 2\{(A) + (B)\}, \tag{11}$$

where

$$\begin{aligned} (A) &= \frac{n_v}{n_c^2}\frac{1}{n_v b}\left|\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}(\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c} - \hat{\boldsymbol{R}}^{-1}_{\alpha,s})(\hat{\boldsymbol{R}}_{\alpha,s} - \hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s}\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}\right|, \\ (B) &= \frac{n_v}{n_c^2}\frac{1}{n_v b}\left|\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{R}}^{-1}_{\alpha,s}(\hat{\boldsymbol{R}}_{\alpha,s} - \hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s}\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}\right|. \end{aligned}$$

In addition,

$$(B) \overset{\text{why?}}{\leq} \frac{n_v^2}{n_c^2}\frac{1}{n_v b}\max_{s\in\mathcal{B}_{n_v}}\|\hat{\boldsymbol{R}}_{\alpha,s} - \hat{\boldsymbol{R}}_{\alpha,s^c}\|\max_{s\in\mathcal{B}_{n_v}}\|\hat{\boldsymbol{R}}^{-1}_{\alpha,s}\|\sum_{s\in\mathcal{B}_{n_v}}\boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{P}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}, \tag{12}$$

and

$$(A) \overset{\text{why?}}{\leq} \frac{n_v^2}{n_c^2}\frac{1}{n_v b}\max_{s\in\mathcal{B}_{n_v}}\|\hat{\boldsymbol{R}}_{\alpha,s} - \hat{\boldsymbol{R}}_{\alpha,s^c}\|^2\max_{s\in\mathcal{B}_{n_v}}\|\hat{\boldsymbol{R}}^{-1}_{\alpha,s}\|\max_{s\in\mathcal{B}_{n_v}}\|\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}\|\sum_{s\in\mathcal{B}_{n_v}}\boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{P}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}. \tag{13}$$

- By (2), we have

$$\max_{s\in\mathcal{B}_{n_v}} \|\hat{\boldsymbol{R}}_{\alpha,s}^{-1}\| = O(1),$$

which, together with (5), and (11)–(13), yields

$$|(\mathrm{VI})| = (\mathrm{II})o(1). \tag{14}$$

- Similarly, it can be shown that

$$|(\mathrm{V})| = (\mathrm{II})o(1). \tag{15}$$

- Since $\frac{n_v}{n_c} \to \infty$, it is easy to see that

$$|(\mathrm{IV})| = (\mathrm{IV}) = (\mathrm{II})o(1). \qquad \text{((IV) is non-negative)} \tag{16}$$

Now, for (III), we have

$$|(\mathrm{III})| \le \frac{n_v}{n_c^2}\frac{1}{n_v b}\left|\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}(\hat{\boldsymbol{R}}_{\alpha,s}-\hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}(\hat{\boldsymbol{R}}_{\alpha,s}-\hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s}\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}\right|$$

$$\le \frac{n_v}{n_c^2}\frac{1}{n_v b}\left|\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{R}}^{-1}_{\alpha,s}(\hat{\boldsymbol{R}}_{\alpha,s}-\hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}(\hat{\boldsymbol{R}}_{\alpha,s}-\hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s}\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}\right|$$

$$+ \frac{n_v}{n_c^2}\frac{1}{n_v b}\left|\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{X}_{\alpha,s}\hat{\boldsymbol{R}}^{-1}_{\alpha,s}(\hat{\boldsymbol{R}}_{\alpha,s}-\hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c} (\hat{\boldsymbol{R}}_{\alpha,s}-\hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s^c}(\hat{\boldsymbol{R}}_{\alpha,s}-\hat{\boldsymbol{R}}_{\alpha,s^c})\hat{\boldsymbol{R}}^{-1}_{\alpha,s}\boldsymbol{X}'_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}\right|.$$

This and an argument similar to that used to prove (14) give

$$|(\mathrm{III})| = (\mathrm{II})o(1). \tag{17}$$

Hence, (10) follows from (14)–(17). Now, by an argument similar to that use to prove (14) again, we have

$$(\mathrm{II}) = \frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{Q}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s} + (\mathrm{II})o(1). \tag{18}$$

Moreover,

$$
\begin{aligned}
&\frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{Q}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s}\\
=\ &\frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}}\sum_{(i,j)\in s\times s} P_{\alpha,ij}\gamma_{i,\alpha}\gamma_{j,\alpha} \qquad ([P_{\alpha,ij}]_{i,j\in s}=\boldsymbol{P}_\alpha)\\
=\ &\frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}}\sum_{i\in s} P_{\alpha,ii}\gamma_{i,\alpha}^2 + \frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{s\in\mathcal{B}_{n_v}}\sum_{(i,j)\in s\times s, i\neq j} P_{\alpha,ij}\gamma_{i,\alpha}\gamma_{j,\alpha} \qquad (P_{\alpha,ii}=w_{i\alpha})\\
\overset{\text{why?}}{=}\ &\frac{n_v^2}{n_c^2}\frac{1}{n}\sum_{i=1}^{n} w_{i\alpha}\gamma_{i,\alpha}^2 - \frac{n_v^2}{n_c^2}\frac{n_v-1}{n-1}\frac{1}{n}\sum_{i=1}^{n} w_{i\alpha}\gamma_{i,\alpha}^2\\
=\ &\frac{n_v^2}{n_c^2}\frac{1}{n}\frac{n_c}{n-1}\sum_{i=1}^{n} w_{i\alpha}\sigma^2 + O\left(\frac{1}{n_c}\sum_{i=1}^{n} w_{i\alpha}(\gamma_{i,\alpha}^2-\sigma^2)\right)\\
\overset{\text{why?}}{=}\ &\frac{d_\alpha\sigma^2}{n_c} + o_p(n_c^{-1}). \qquad (19)
\end{aligned}
$$

- By (18) and (19), we have

$$(\mathrm{II}) = \frac{d_\alpha \sigma^2}{n_c} + o_p(n_c^{-1}). \tag{20}$$

- Moreover,

$$(\mathrm{I}) = \frac{1}{n}\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{y} = \frac{1}{n}\boldsymbol{e}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{e} = n^{-1}\boldsymbol{e}'\boldsymbol{e} + O_p(n^{-1}). \tag{21}$$

- In view of (20), (21) and (9), the desired conclusion (8) follows.

- To show (7), note that (4) implies (why?)

$$\hat{\Gamma}_{\alpha,n}(n_v) \geq \frac{1}{n_v b} \sum_{s \in \mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s} \boldsymbol{\gamma}_{\alpha,s} = \frac{1}{n} \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{y}. \tag{22}$$

- Since $\alpha^* - \alpha \neq \emptyset$,

$$\begin{aligned} \frac{1}{n}\boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{y} &= n^{-1}\boldsymbol{e}'\boldsymbol{e} + n^{-1}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} + 2n^{-1}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{e} \\ &= n^{-1}\boldsymbol{e}'\boldsymbol{e} + n^{-1}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} + o_p(1), \end{aligned} \tag{23}$$

  and hence (7) is ensured by (22) and (23).
- Finally, (iii) is an immediate consequence of (i) and (ii).

Monte Carlo Cross Validation (MCCV)

- Let $s_i \overset{\text{i.i.d.}}{\sim} U(\mathcal{B}_{n_v})$, where $\mathcal{B}_{n_v}$ is defined in the note for CV.
- Define MCCV as follows:

$$\hat{\Gamma}_{\alpha,n}^{\mathrm{MCCV}} = \frac{1}{n_v b} \sum_{i=1}^{b} \|\boldsymbol{y}_{s_i} - \hat{\boldsymbol{y}}_{\alpha,s_i^c}\|^2,$$

where b is the number of Monte Carlo simulations used to calculate CV.

Theorem 3

Assume the same assumptions as in Theorem 1. Suppose

$$\max_{1\le j\le b}\left\|\frac{1}{n_v}\sum_{i\in s_j}\boldsymbol{x}_i\boldsymbol{x}_i' - \frac{1}{n_c}\sum_{i\notin s_j}\boldsymbol{x}_i\boldsymbol{x}_i'\right\| = o_p(1), \tag{24}$$

and

$$Ee_1^4<\infty,\ \frac{n^2}{bn_c^2}\to 0,\ \frac{n_v}{n}\to 1, \text{ and } n_c\to\infty. \tag{25}$$

Then,

(i) for $\alpha^*-\alpha\neq\emptyset$,

$$\hat{\Gamma}_{\alpha,n}^{\mathrm{MCCV}} = \frac{1}{n_v b}\sum_{i=1}^{b}\boldsymbol{e}_{s_i}'\boldsymbol{e}_{s_i} + \frac{1}{n}\boldsymbol{\beta}'\boldsymbol{X}'(\boldsymbol{I}-\boldsymbol{P}_\alpha)\boldsymbol{X}\boldsymbol{\beta} + o_p(1) + R_n,$$

where $\boldsymbol{e}_{s_i}' = (e_j, j\in s_i)$ and R_n is some positive r.v.,

(ii) for $\alpha^*-\alpha=\emptyset$,

$$\hat{\Gamma}_{\alpha,n}^{\mathrm{MCCV}} = \frac{1}{n_v b}\sum_{i=1}^{b}\boldsymbol{e}_{s_i}'\boldsymbol{e}_{s_i} + n_c^{-1}d_\alpha\sigma^2 + o_p(n_c^{-1}),$$

(iii) $\lim_{n\to\infty}P(\hat{\alpha}_{\mathrm{MCCV}}=\alpha^*)=1$, where $\hat{\alpha}_{\mathrm{MCCV}} = \mathrm{argmin}_{\alpha\in 2^{\mathbb{P}}}\hat{\Gamma}_{\alpha,n}^{\mathrm{MCCV}}$.

Proof

By an argument similar to that used to prove (9) and (18) in the note for CV, one has for $\alpha^* - \alpha = \emptyset$,

$$\hat{\Gamma}_{\alpha,n}^{\mathrm{MCCV}} = \frac{1}{n_v b}\sum_{i=1}^{b}\boldsymbol{\gamma}'_{\alpha,s_i}\boldsymbol{\gamma}_{\alpha,s_i} + \frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{i=1}^{b}\boldsymbol{\gamma}'_{\alpha,s_i}\boldsymbol{P}_{\alpha,s_i}\boldsymbol{\gamma}_{\alpha,s_i}(1+o_p(1)), \tag{26}$$

and

$$\frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{i=1}^{b}\boldsymbol{\gamma}'_{\alpha,s_i}\boldsymbol{P}_{\alpha,s_i}\boldsymbol{\gamma}_{\alpha,s_i}(1+o_p(1)) = \frac{n_v^2}{n_c^2}\frac{1}{n_v b}\sum_{i=1}^{b}\boldsymbol{\gamma}'_{\alpha,s_i}\boldsymbol{Q}_{\alpha,s_i}\boldsymbol{\gamma}_{\alpha,s_i}. \tag{27}$$

In addition, by (19) in the note for CV,

$$
\begin{aligned}
& E\left[\frac{n_v^2}{n_c^2} \frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{\gamma}'_{\alpha,s_i} \boldsymbol{Q}_{\alpha,s_i} \boldsymbol{\gamma}_{\alpha,s_i} \middle| (y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n) \right] \\
= \quad & \frac{n_v^2}{n_c^2} \frac{1}{n_v} E[\boldsymbol{\gamma}'_{\alpha,s_1} \boldsymbol{Q}_{\alpha,s_1} \boldsymbol{\gamma}_{\alpha,s_1} | (y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)] \\
\overset{\text{why?}}{=} \quad & \frac{n_v^2}{n_c^2} \frac{1}{n_v} \frac{1}{\binom{n}{n_v}} \sum_{s \in \mathcal{B}_{n_v}} \boldsymbol{\gamma}'_{\alpha,s} \boldsymbol{Q}_{\alpha,s} \boldsymbol{\gamma}_{\alpha,s} \\
= \quad & \frac{d_\alpha \sigma^2}{n_c} + o_p(n_c^{-1}),
\end{aligned}
$$

implying

$$
E[n_c V_n | \mathcal{F}_n] \xrightarrow{pr.} d_\alpha \sigma^2, \tag{28}
$$

where

$$
V_n = \frac{n_v^2}{n_c^2} \frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{\gamma}'_{\alpha,s_i} \boldsymbol{Q}_{\alpha,s_i} \boldsymbol{\gamma}_{\alpha,s_i}
$$

and $\mathcal{F}_n$ is the σ-field generated by $(y_1, \boldsymbol{x}_1), \ldots, (y_n, \boldsymbol{x}_n)$.

- We also have

$$\begin{aligned} Var(n_c V_n|\mathcal{F}_n) &= \frac{n_v^2}{n_c^2}\frac{1}{b^2}\sum_{i=1}^{b} Var(\boldsymbol{\gamma}'_{\alpha,s_i}\boldsymbol{Q}_{\alpha,s_i}\boldsymbol{\gamma}_{\alpha,s_i}|\mathcal{F}_n) \\ &\le \frac{n_v^2}{n_c^2}\frac{1}{b}E((\boldsymbol{\gamma}'_{\alpha,s_1}\boldsymbol{Q}_{\alpha,s_1}\boldsymbol{\gamma}_{\alpha,s_1})^2|\mathcal{F}_n) \\ &= \frac{n_v^2}{n_c^2}\frac{1}{b}\frac{1}{\binom{n}{n_v}}\sum_{s\in\mathcal{B}_{n_v}}(\boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{Q}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s})^2. \end{aligned}$$

- Since $Ee_1^4 < \infty$, we have

$$E((\boldsymbol{\gamma}'_{\alpha,s}\boldsymbol{Q}_{\alpha,s}\boldsymbol{\gamma}_{\alpha,s})^2) \le C < \infty \text{ for all } s \in \mathcal{B}_{n_v}, \qquad (C: \text{ some positive constant})$$

which, together with $\frac{n_v^2}{n_c^2}\frac{1}{b} \to 0$, yields

$$Var[n_c V_n|\mathcal{F}_n] \xrightarrow{pr.} 0. \qquad (29)$$

- It is shown in the Appendix that

$$(28) + (29) \quad \text{implies} \quad n_c V_n \xrightarrow{pr.} d_\alpha \sigma^2. \tag{30}$$

- In view of (26), (27) and (30), (ii) is proved once

$$\frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{\gamma}'_{\alpha,s_i} \boldsymbol{\gamma}_{\alpha,s_i} = \frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{e}'_{s_i} \boldsymbol{e}_{s_i} + o_p(n_c^{-1}). \tag{31}$$

- To show (31), note first that

$$\boldsymbol{\gamma}_{\alpha,s_i} = \boldsymbol{e}_{s_i} - \boldsymbol{X}_{\alpha,s_i}(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha),$$

yielding

$$\begin{aligned} \frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{\gamma}'_{\alpha,s_i} \boldsymbol{\gamma}_{\alpha,s_i} \quad &= \quad \frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{e}'_{s_i} \boldsymbol{e}_{s_i} - \frac{2}{n_v b} \sum_{i=1}^{b} (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha)' \boldsymbol{X}'_{\alpha,s_i} \boldsymbol{e}_{s_i} \\ &+ (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha)' \left(\frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{X}'_{\alpha,s_i} \boldsymbol{X}_{\alpha,s_i} \right) (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha). \end{aligned} \tag{32}$$

Since

$$\begin{aligned}
\left| \frac{1}{n_v b} \sum_{i=1}^{b} (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha)' \boldsymbol{X}'_{\alpha,s_i} \boldsymbol{e}_{s_i} \right| &\leq \|\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha\| \left\| \frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{X}'_{\alpha,s_i} \boldsymbol{e}_{s_i} \right\|, \\
E\left(\left\| \frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{X}'_{\alpha,s_i} \boldsymbol{e}_{s_i} \right\| \right) &\leq \frac{1}{n_v b} \sum_{i=1}^{b} E(\|\boldsymbol{X}'_{\alpha,s_i} \boldsymbol{e}_{s_i}\|) \\
&= \frac{1}{n_v} E[E(\|\boldsymbol{X}'_{\alpha,s_1} \boldsymbol{e}_{s_1}\| | \mathcal{F}_n)] \\
&= \frac{1}{n_v} E\left(\frac{1}{\binom{n}{n_v}} \sum_{s \in \mathcal{B}_{n_v}} \|\boldsymbol{X}'_{\alpha,s} \boldsymbol{e}_s\| \right) \\
&\overset{\text{why?}}{=} O(n_v^{-\frac{1}{2}}),
\end{aligned}$$

and

$$\|\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha\| = O_p(n^{-\frac{1}{2}}),$$

one has

$$\frac{2}{n_v b} \sum_{i=1}^{b} (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha)' \boldsymbol{X}'_{\alpha,s_i} \boldsymbol{e}_{s_i} \overset{\text{why?}}{=} O_p(n^{-1}). \tag{33}$$

- Similarly, it can be shown that

$$(\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha)' \left(\frac{1}{n_v b} \sum_{i=1}^{b} \boldsymbol{X}'_{\alpha,s_i} \boldsymbol{X}_{\alpha,s_i} \right) (\hat{\boldsymbol{\beta}}_\alpha - \boldsymbol{\beta}_\alpha) = O_p(n^{-1}). \tag{34}$$

- Combining (32)–(34), we obtain the desired conclusion (31). Thus, (ii) is proved.
- The proof of (i) is left as an exercise.
- Finally, we note that (iii) is an immediate consequence of (i) & (ii).

Appendix: the proof of (30)

To prove (30), we need the so-called conditional Chebyshev's inequality, which is stated as follows.

Conditional Chebyshev's inequality

Let X be a positive r.v., $\epsilon > 0$, and $\mathcal{F}$ be a σ-field. Then

$$P(X > \epsilon|\mathcal{F}) \le \frac{E(X|\mathcal{F})}{\epsilon} \text{ a.s. (almost surely).} \tag{35}$$

Proof

Let $D \in \mathcal{F}$. Then, by the definition of conditional expectation,

$$\begin{aligned} \int_D P(X > \epsilon|\mathcal{F})\, \mathrm{d}\mathcal{P} = \int_D E(I_{X>\epsilon}|\mathcal{F})\, \mathrm{d}\mathcal{P} = \int_D I_{X>\epsilon}\, \mathrm{d}\mathcal{P} &\le \int_D \frac{X}{\epsilon} I_{X>\epsilon}\, \mathrm{d}\mathcal{P} \\ &\le \int_D \frac{X}{\epsilon}\, \mathrm{d}\mathcal{P} \\ &= \int_D \frac{E(X|\mathcal{F})}{\epsilon}\, \mathrm{d}\mathcal{P}. \end{aligned}$$

Hence (35) follows.

Now, (30) is ensured by the following fact.

Fact

Let $\{X_n\}$ and $\{\mathcal{F}_n\}$ be sequences of r.v.s and σ-field, respectively. If

$$Var(X_n|\mathcal{F}_n) \xrightarrow{pr.} 0, \tag{36}$$

and

$$E(X_n|\mathcal{F}_n) \xrightarrow{pr.} C \text{ (a constant)}, \tag{37}$$

then

$$X_n \xrightarrow{pr.} C. \tag{38}$$

Proof

By the conditional Chebyshev's inequality and (36), it holds that for any $\epsilon > 0$

$$P(|X_n - E(X_n|\mathcal{F}_n)| > \epsilon|\mathcal{F}_n) \leq \frac{Var(X_n|\mathcal{F}_n)}{\epsilon^2} \xrightarrow{pr.} 0,$$

which, together with the dominated convergence theorem, yields

$$X_n - E(X_n|\mathcal{F}_n) \xrightarrow{pr.} 0. \quad \text{(why?)}$$

This and (37) give (38).

Accumulated Prediction Errors (APE)

Consider the following stochastic regression model

$$y_t = \sum_{i=1}^{p} \beta_i x_{ti} + \epsilon_t, \tag{39}$$

where $\epsilon_t \overset{\text{i.i.d.}}{\sim} (0, \sigma^2)$ and $\boldsymbol{x}_t = (x_{t1}, \ldots, x_{tp})'$ is $\mathcal{F}_{t-1}$-measurable, meaning that $\boldsymbol{x}_t$ can be completely decided by the information collected at time $t-1$.

Example 1: Autoregressive (AR) models

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \epsilon_t,$$

where $A(z) = 1 - a_1 z - a_2 z^2 - \cdots - a_p z^p \neq 0$, $\forall |z| \leq 1$.

Example 2: Autoregressive exogenous models

$$y_t = a_1 y_{t-1} + \cdots + a_p y_{t-p} + \eta_1 Z_{t1} + \cdots + \eta_q Z_{tq} + \epsilon_t,$$

where $A(z) \neq 0$, $\forall |z| \leq 1$, and $(\epsilon_t, Z_{t1}, \ldots, Z_{tq})$ are i.i.d..

- Let α and α^* be defined as in the note for CV.
- We are interested in choosing α^* based on APE, which assigns α to a positive value:

$$\mathrm{APE}_\alpha = \sum_{t=M+1}^{n} (y_t - \boldsymbol{x}'_{t,\alpha}\hat{\boldsymbol{\beta}}_{t-1,\alpha})^2,$$

where

$$\hat{\boldsymbol{\beta}}_{t,\alpha} = \left(\sum_{j=1}^{t} \boldsymbol{x}_{j,\alpha}\boldsymbol{x}'_{j,\alpha}\right)^{-1} \left(\sum_{j=1}^{t} \boldsymbol{x}_{j,\alpha} y_j\right)$$

and M is some integer to be specified later.

Theorem 4

Assume

$$\frac{1}{n}\sum_{t=1}^{n} \boldsymbol{x}_t \boldsymbol{x}_t' \xrightarrow{pr.} \boldsymbol{R} \text{ (p.d.)}, \quad E|\epsilon_1|^{q_1} < \infty \text{ for some large } q_1, \tag{40}$$

and

$$\max_{1\le t\le n, 1\le i\le p} E|x_{ti}|^{q_2} < \infty \text{ for some large } q_2, \quad E\|\sqrt{t}(\hat{\boldsymbol{\beta}}_t - \boldsymbol{\beta})\|^{q_3} < \bar{k} < \infty, \tag{41}$$

for all $t \geq M$ and some large q_3.

(i) For $\alpha^* - \alpha = \emptyset$,

$$\mathrm{APE}_\alpha = \sum_{t=M+1}^{n} \epsilon_t^2 + \sigma^2 d_\alpha \log n + o_p(\log n),$$

(ii) For $\alpha^* - \alpha \neq \emptyset$,

$$\mathrm{APE}_\alpha \geq \sum_{t=M+1}^{n} \epsilon_t^2 + \Delta_n(\alpha) + O_p(1),$$

where $\frac{1}{n}\Delta_n(\alpha) \xrightarrow{pr.} \gamma > 0$.

(iii) $\lim_{n\to\infty} P(\hat{\alpha}_{\mathrm{APE}} = \alpha^*) = 1$, where $\hat{\alpha}_{\mathrm{APE}} = \mathrm{argmin}_{\alpha\in 2^{\mathbb{P}}} \mathrm{APE}_\alpha$.

Proof

- We first prove (i). Define

$$\boldsymbol{V}_{t,\alpha} = \left(\sum_{i=1}^{t} \boldsymbol{x}_{i,\alpha}\boldsymbol{x}'_{i,\alpha}\right)^{-1}, \quad Q_{t,\alpha} = \left(\sum_{j=1}^{t} \boldsymbol{x}_{j,\alpha}\epsilon_j\right)' \boldsymbol{V}_{t,\alpha} \left(\sum_{j=1}^{t} \boldsymbol{x}_{j,\alpha}\epsilon_j\right),$$

  and $d_{t,\alpha} = \boldsymbol{x}'_{t,\alpha}\boldsymbol{V}_{t,\alpha}\boldsymbol{x}_{t,\alpha}$.
- By making use of

$$\boldsymbol{V}_{t,\alpha} = \boldsymbol{V}_{t-1,\alpha} - \frac{\boldsymbol{V}_{t-1,\alpha}\boldsymbol{x}_{t,\alpha}\boldsymbol{x}'_{t,\alpha}\boldsymbol{V}_{t-1,\alpha}}{1 + \boldsymbol{x}'_{t,\alpha}\boldsymbol{V}_{t-1,\alpha}\boldsymbol{x}_{t,\alpha}}$$

  and

$$\frac{1}{1 + \boldsymbol{x}'_{t,\alpha}\boldsymbol{V}_{t-1,\alpha}\boldsymbol{x}_{t,\alpha}} = 1 - d_{t,\alpha},$$

  one gets

$$\sum_{t=M+1}^{n} \left(\boldsymbol{x}'_{t,\alpha}\boldsymbol{V}_{t-1,\alpha}\sum_{j=1}^{t-1} \boldsymbol{x}_{j,\alpha}\epsilon_j\right)^2 (1 - d_{t,\alpha}) + Q_{n,\alpha} - Q_{M,\alpha}$$

$$= \sum_{t=M+1}^{n} d_{t,\alpha}\epsilon_t^2 + 2\sum_{t=M+1}^{n} \boldsymbol{x}'_{t,\alpha}(\hat{\boldsymbol{\beta}}_{t-1,\alpha} - \boldsymbol{\beta}_\alpha)\epsilon_t(1 - d_{t,\alpha}). \quad (42)$$

- Moreover, by (40) and (41), it can be shown that

$$\sum_{t=M+1}^{n}\left(\boldsymbol{x}'_{t,\alpha}\boldsymbol{V}_{t-1,\alpha}\sum_{j=1}^{t-1}\boldsymbol{x}_{j,\alpha}\epsilon_j\right)^2 d_{t,\alpha}=O_p(1), \quad (43)$$

$$Q_n=O_p(1), \quad Q_M=O_p(1), \quad (44)$$

$$\sum_{t=M+1}^{n}\boldsymbol{x}'_{t,\alpha}(\hat{\boldsymbol{\beta}}_{t-1,\alpha}-\boldsymbol{\beta}_\alpha)\epsilon_t(1-d_{t,\alpha})=o_p(\log n). \quad (45)$$

- Also, we have

$$\begin{aligned}\sum_{t=M+1}^{n} d_{t,\alpha}\epsilon_t^2 &= \sum_{t=M+1}^{n} d_{t,\alpha}\sigma^2+\sum_{t=M+1}^{n} d_{t,\alpha}(\epsilon_t^2-\sigma^2)\\ &= \sigma^2 d_\alpha\log n+o_p(\log n)+O_p(1). \quad (46)\end{aligned}$$

- Consequently, (i) follows from (42)–(46), and the fact that

$$\begin{aligned}
\mathrm{APE}_\alpha &= \sum_{t=M+1}^{n} (\epsilon_t - \boldsymbol{x}'_{t,\alpha}(\hat{\boldsymbol{\beta}}_{t-1,\alpha} - \boldsymbol{\beta}_\alpha))^2 \\
&= \sum_{t=M+1}^{n} \epsilon_t^2 - 2 \sum_{t=M+1}^{n} (\hat{\boldsymbol{\beta}}_{t-1,\alpha} - \boldsymbol{\beta}_\alpha)\boldsymbol{x}_{t,\alpha}\epsilon_t + \sum_{t=M+1}^{n} [\boldsymbol{x}'_{t,\alpha}(\hat{\boldsymbol{\beta}}_{t-1,\alpha} - \boldsymbol{\beta}_\alpha)]^2 \\
&= \sum_{t=M+1}^{n} \epsilon_t^2 + \sum_{t=M+1}^{n} [\boldsymbol{x}'_{t,\alpha}(\hat{\boldsymbol{\beta}}_{t-1,\alpha} - \boldsymbol{\beta}_\alpha)]^2 + o_p(\log n).
\end{aligned}$$

- To show (ii), note that by Theorem 2.1 of Wei (1992), we have

$$\begin{aligned}
\sum_{t=M+1}^{n} (y_t - \boldsymbol{x}'_{t,\alpha}\hat{\boldsymbol{\beta}}_{t-1,\alpha})^2 &\geq \sum_{t=1}^{n}(y_t - \boldsymbol{x}'_{t,\alpha}\hat{\boldsymbol{\beta}}_{n,\alpha})^2 - \sum_{t=1}^{M}(y_t - \boldsymbol{x}'_{t,\alpha}\hat{\boldsymbol{\beta}}_{M,\alpha})^2 \\
&= \boldsymbol{y}'(\boldsymbol{I} - \boldsymbol{P}_\alpha)\boldsymbol{y} + O_p(1),
\end{aligned}$$

  which implies (ii) (why?).
- Now, (iii) follows directly from (i) and (ii).

References

1. Shao, J. (1993). Linear model selection by cross-validation. *Journal of the American Statistical Association*, **88**, 486–494.
2. Wei, C. Z. (1992). On predictive least squares principles. *The Annals of Statistics*, **20**, 1–42.