Supplemental material for: Forecasting Multiple Time Series with One-Sided Dynamic Principal Components

Daniel Peña*
Department of Statistics and Institute of Financial Big Data
Universidad Carlos III de Madrid, Spain

and

Ezequiel Smucler
Department of Mathematics and Statistics
Universidad Torcuato Di Tella, Argentina

and

Victor J. Yohai
Instituto de Calculo and Department of Mathematics
School of Exact and Natural Sciences
Universidad de Buenos Aires - CONICET, Argentina

*D. P. has been supported by Grant ECO2015-66593-P of MINECO/FEDER/UE. E.S. was partially funded by a CONICET Ph.D fellowship and by grant PIP 112-201101-00339 from CONICET. Part of this work was conducted while E.S. was a Postdoctoral Research Fellow at the Department of Statistics, and at the Department of Computer Science at the University of British Columbia. The authors would like to thank Stefano Soccorsi for help with the programs to run the FHLZ procedure. They also acknowledge the help of Stefano Soccorsi for help with the programs to run the FHLZ procedure.
1 Proofs

This section includes the proofs of all the results stated in the main paper.

Proof of Proposition \[\text{2}.\] We prove the result for \(i = 1\), the other cases are proven similarly.

Fix \(a\) with \(\|a\| = 1\) and \(B\). Note that

\[
m_{\text{MSE}}^1(a, B) = E \|z_t - z_{R,1}^R(a, B)\|^2 = E \|z_{R,1}^R(a, B)\|^2 + E \|z_t\|^2 - 2E z_t z_{R,1}^R(a, B).
\]

Let \(f_t' = f_t(a) = (a'x_t, a'x_{t-1}, \ldots, a'x_{t-k_2}).\) Note that \(z_{R,1}^R(a, B) = B f_t'.\) Hence

\[
E \|z_{R,1}^R(a, B)\|^2 = Ef_t'BB'f_t = \text{Tr}(BB'_{k_1,k_2}(a)).
\]

Since \(BB'\) and \(S_{k_1,k_2}(a) - I_{k_2+1}\) are symmetric and semi-positive definite we have that

\[
\text{Tr}(BB'_{k_1,k_2}(a)) \geq \text{Tr}(BB' \inf_{\|a\| = 1} \lambda_{\text{min}}(S_{k_1,k_2}(a))) = \|B\|_F^2 \inf_{\|a\| = 1} \lambda_{\text{min}}(S_{k_1,k_2}(a)).
\]

On the other hand, by the Cauchy-Schwartz inequality

\[
E |z_t' z_{R,1}^R(a, B)| \leq (E \|z_t\|^2)^{1/2} (E \|z_{R,1}^R(a, B)\|^2)^{1/2}.
\]

Since \(BB'\) and \(I_{k_2+1}\) are symmetric and semi-positive definite we have that

\[
\text{Tr}(BB'_{k_1,k_2}(a)) \leq \text{Tr}(BB' \sup_{\|a\| = 1} \lambda_{\text{max}}(S_{k_1,k_2}(a))) = \|B\|_F^2 \sup_{\|a\| = 1} \lambda_{\text{max}}(S_{k_1,k_2}(a)).
\]

Hence,

\[
(E \|z_{R,1}^R(a, B)\|^2)^{1/2} \leq \|B\|_F \left( \sup_{\|a\| = 1} \lambda_{\text{max}}(S_{k_1,k_2}(a)) \right)^{1/2}.
\]

Provided by five anonymous referees whose useful comments and suggestions led to important improvements in the paper in both presentation and content.
Then
\[
\inf_{\|a\|=1} \|z_t - z_t^R(a, B)\|^2 \geq \|B\|_F^2 \inf_{\|a\|=1} \lambda_{\min}(S_{k_1^1, k_2^1}(a)) + \mathbb{E}\|z_t\|^2
\]
\[
- 2(\mathbb{E}\|z_t\|^2)^{1/2}\|B\|_F (\sup_{\|a\|=1} \lambda_{\max}(S_{k_1^1, k_2^1}(a)))^{1/2},
\]
from which the proposition follows immediately.

\[\square\]

**Lemma 1.** Fix \(k_1, k_2 \geq 0\). Then
\[
\sup_{\|a\|=1} \left\| \frac{F'_{k_1, k_2} F_{k_1, k_2}}{T - (k_1 + k_2)} - S_{k_1, k_2}(a) \right\|_F \xrightarrow{a.s.} 0.
\]

**Proof of Lemma 1.** Fix \(a\) with \(\|a\| = 1\). Then
\[
\frac{F'_{k_1, k_2} F_{k_1, k_2}}{T - (k_1 + k_2)} = \frac{1}{T - (k_1 + k_2)} \begin{pmatrix}
a'Z'_{k_1+k_2, k_1}
a'Z'_{k_1+k_2-1, k_1}
\vdots
a'Z'_{k_1, k_1}
\end{pmatrix}
\begin{pmatrix}
a Z_{k_1+k_2, k_1}
a Z_{k_1+k_2-1, k_1}
\vdots
a Z_{k_1, k_1}
\end{pmatrix}
\]

Fix \(k_1 \leq i, j \leq k_1 + k_2\). Then
\[
Z'_{i,k_1} Z_{j,k_1} = \begin{pmatrix}
Z'_{i}
\vdots
Z'_{i-k_1}
\end{pmatrix}
\begin{pmatrix}
Z_{i-1}
\vdots
Z_{i-k_1}
\end{pmatrix},
\]
\[
= \begin{pmatrix}
Z'_{i}Z_{j}
Z'_{i}Z_{j-1}
\vdots
Z'_{i}Z_{j-k_1}
\end{pmatrix}
\begin{pmatrix}
Z_{i-1}
\vdots
Z_{i-k_1}
\end{pmatrix}.
\]
Note that
\[ Z'_i Z_j = \sum_{r=1}^{T-k_1-k_2} z_{i+r} z'_{j+r}. \]

By the Ergodic Theorem
\[ \frac{Z'_i Z_j}{T-(k_1+k_2)} \xrightarrow{a.s.} \Sigma(i-j). \]

Hence
\[ \frac{Z'_{i,k_1} Z_{j,k_1}}{T-(k_1+k_2)} \xrightarrow{a.s.} \begin{pmatrix} \Sigma(i-j) & \Sigma(i-j+1) & \ldots & \Sigma(i-j+k_1) \\ \vdots & \vdots & \ddots & \vdots \\ \Sigma(i-j-k_1) & \Sigma(i-j-k_1+1) & \ldots & \Sigma(i-j) \end{pmatrix} = V_{k_1}(i-j). \]

We have shown that
\[ \frac{F'_{k_1,k_2} F_{k_1,k_2}}{T-(k_1+k_2)} \xrightarrow{a.s.} \begin{pmatrix} a' V_{k_1}(0) a \\ \vdots \\ a' V_{k_1}(-k_2) a \end{pmatrix} = S_{k_1,k_2}(a). \]

To prove that the convergence holds uniformly, it suffices to show that, for \( k_1 \leq i, j \leq k_1+k_2, \)
\[ \sup_{\|a\|=1} \left| \frac{a'Z'_{i,k_1} Z_{j,k_1} a}{T-(k_1+k_2)} - a' V_{k_1}(i-j) a \right| \xrightarrow{a.s.} 0. \]

It is easy to show that
\[ \frac{a'Z'_{i,k_1} Z_{j,k_1} a}{T-(k_1+k_2)} = \frac{\sum_{h=0}^{k_1} \sum_{r=0}^{k_1} a'_h Z'_{i-r} Z_{j-h} a_h}{T-(k_1+k_2)}. \]

Note that for any \( v, w \in \mathbb{R}^m \)
\[ v' Z'_i Z_j w = v' \left( \sum_{r=1}^{T-(k_1+k_2)} z_{i+r} z'_{j+r} \right) w. \]
Thus, to prove the lemma it will be enough to prove that

\[
\sup_{\|v\| \leq 1, \|w\| \leq 1} \left| \frac{T^{-(k_1+k_2)} v' \left( \sum_{r=1}^n z_{i+r} z_{j+r}' \right) w}{T - (k_1 + k_2)} - v' \Sigma(i-j) w \right| \xrightarrow{a.s.} 0.
\]

This follows immediately from

\[
\sup_{\|v\| \leq 1, \|w\| \leq 1} \left| \frac{T^{-(k_1+k_2)} v' \left( \sum_{r=1}^n z_{i+r} z_{j+r}' \right) w}{T - (k_1 + k_2)} - v' \Sigma(i-j) w \right| \leq \left\| \sum_{r=1}^n z_{i+r} z_{j+r}' \right\|_F \xrightarrow{a.s.} 0,
\]

and the Ergodic Theorem.

\[\square\]

**Lemma 2.** Fix $k_1, k_2 \geq 0$. Then

\[
\liminf_{T} \inf_{\|a\| = 1} \lambda_{\min} \left( \frac{F'_{k_1,k_2} F_{k_1,k_2}}{T - (k_1 + k_2)} \right) \geq \inf_{\|a\| = 1} \lambda_{\min} (S_{k_1,k_2}(a)),
\]

with probability one.

**Proof of Lemma 2.** It suffices to show that

\[
\sup_{\|a\| = 1} \lambda_{\min} \left( \frac{F'_{k_1,k_2} F_{k_1,k_2}}{T - (k_1 + k_2)} \right) - \lambda_{\min} (S_{k_1,k_2}(a)) \xrightarrow{a.s.} 0
\]

and this follows from Theorem 3.3.16 of [Horn and Johnson (1994)] and Lemma 1.

\[\square\]

For $i \geq 1$ let

\[
z^R_{t,i}(a^1, B^1, \ldots, a^i, B^i) = \sum_{c=1}^i r^R_t(a^c, B^c), \quad z^{*,i}_t = \sum_{c=1}^i r^R_t(a^{*,c}, B^{*,c}).
\]

Define

\[
A^0 = \frac{\mathbb{E}\|z_t\|^2}{m}, \quad A^1 = \frac{\mathbb{E}\|z_t - r^R_t(a^{*,1}, B^{*,1})\|^2}{m}.
\]
and for $i \geq 2$

$$A^i = 2^{i-1}A^1 + \sum_{c=2}^i 2^{(i+1-c)}E\|r_t^R(a^{*,c},B^{*,c})\|^2.$$ 

Let

$$k_{\text{max}}^0 = 0, \quad T^{*,0} = T, \quad (\hat{a}^0, \hat{B}^0) = (0,0), \quad \text{MSE}^0(\hat{a}^0, \hat{B}^0) = \frac{1}{(T-k_{\text{max}}^0)m} \sum_{t=k_{\text{max}}^0+1}^T \|z_t\|^2,$$

and $T^0 = \{(0,0)\}$ and $(a^{*,0}, B^{*,0}) = (0,0)$.

**Lemma 3.** Assume Condition 1 holds for $(k_i^1, k_i^2)$, $i = 1, \ldots, q$. Fix $0 \leq i \leq q$. Assume that, with probability one, for large enough $T$, $(\hat{a}^0, \hat{B}^0), \ldots, (\hat{a}^i, \hat{B}^i)$ exist. Then we have that, with probability one,

$$\limsup_T \text{MSE}^i(\hat{a}^i, \hat{B}^i) \leq A^i.$$

**Proof of Lemma 3.** The proof goes by induction. For $i = 0$ the result follows immediately from the Ergodic Theorem. Take $i = 1$. By definition of $(\hat{a}^1, \hat{B}^1)$ we have that

$$\text{MSE}^1(\hat{a}^1, \hat{B}^1) \leq \text{MSE}^1(a^{*,1}, B^{*,1}).$$

By the Ergodic Theorem $\text{MSE}^1(a^{*,1}, B^{*,1})$ converges to $A_1$ almost surely. Now assume the result has been proven for $i$. We will show that it holds for $i + 1$. By definition of $(\hat{a}^{i+1}, \hat{B}^{i+1})$ we have that

$$\text{MSE}^{i+1}(\hat{a}^{i+1}, \hat{B}^{i+1}) \leq \text{MSE}^{i+1}(a^{*,i+1}, B^{*,i+1}) = \frac{1}{T^{*,i+1}m} \sum_{t=k_{\text{max}}^{i+1}+1}^T \|z_t - \hat{z}_t - r_t^R(a^{*,i+1}, B^{*,i+1})\|^2.$$ 

Now

$$\frac{1}{T^{*,i+1}m} \sum_{t=k_{\text{max}}^{i+1}+1}^T \|z_t - \hat{z}_t - r_t^R(a^{*,i+1}, B^{*,i+1})\|^2 \leq \frac{2}{T^{*,i+1}m} \sum_{t=k_{\text{max}}^{i+1}+1}^T \|z_t - \hat{z}_t\|^2$$

$$+ \frac{2}{T^{*,i+1}m} \sum_{t=k_{\text{max}}^{i+1}+1}^T \|r_t^R(a^{*,i+1}, B^{*,i+1})\|^2.$$
For the first term, since $k_{i+1}^{i+1} \geq k_{i_{max}}^i$, 

$$\frac{2}{T_{*,i+1}^* m} \sum_{t=k_{i_{max}}^i+1}^{T} \|z_t - \hat{z}_t^i\|^2 \leq \frac{2T_{*,i}^{i+1}}{T_{*,i+1}^* m} \sum_{t=k_{i_{max}}^i+1}^{T} \|z_t - \hat{z}_t^i\|^2 = \frac{2T_{*,i}^{i+1}}{T_{*,i+1}^* m} \text{MSE}^i(\hat{a}^i, \hat{B}^i).$$

Using the inductive hypothesis, we get that with probability one

$$\limsup_T \frac{2}{T_{*,i+1}^* m} \sum_{t=k_{i_{max}}^i+1}^{T} \|z_t - \hat{z}_t^i\|^2 \leq 2A^i.$$

For the second term, using the Ergodic Theorem again, we get that with probability one,

$$\limsup_T \frac{2}{T_{*,i+1}^* m} \sum_{t=k_{i_{max}}^i+1}^{T} \|r^R_{t}(a^{*,i+1}, B^{*,i+1})\|^2 \leq 2 \mathbb{E} \|r^R_t(a^{*,i+1}, B^{*,i+1})\|^2 / m.$$

Hence, with probability one,

$$\limsup_T \text{MSE}^{i+1}(\hat{a}^{i+1}, \hat{B}^{i+1}) \leq 2A^i + 2 \frac{\mathbb{E} \|r^R_t(a^{*,i+1}, B^{*,i+1})\|^2}{m} = A^{i+1}.$$

The result is proven. \hfill \Box

We will note for $i = 1, \ldots, q$

$$g^i(M) = \left( M \left( \inf_{\|a\|=1} \lambda_{min} \left( S_{k^i_1,k^i_2}(a) \right) \right)^{1/2} \right)^{2} - \left( A^{i-1} \right)^{1/2}.$$ 

If Condition 1 holds for $k^i_1, k^i_2$, clearly $g^i(M) \to +\infty$ when $M \to +\infty$. For $i = 1, \ldots, q$ let

$$s^i = \left( \frac{m A^{i-1}}{\inf_{\|a\|=1} \lambda_{min} \left( S_{k^i_1,k^i_2}(a) \right) } \right)^{1/2}.$$

The following Lemma is a key result.
Lemma 4. Assume Condition \([7]\) holds for \((k_1^i, k_2^i), i = 1, \ldots, q\). Fix \(1 \leq i \leq q\). Assume that, with probability one, for large enough \(T\), \((\hat{a}^0, \hat{B}^0), \ldots, (\hat{a}^{i-1}, \hat{B}^{i-1})\) exist. Then if \(M > s_i\), with probability one,

\[
\liminf_T \inf_{\|a\|=1, \|B\|_F \geq M} \text{MSE}^i(a, B) \geq g^i(M).
\]

Proof of Lemma \([4]\). The triangle inequality implies that

\[
m^{1/2} \text{MSE}^i(a, B)^{1/2} \geq \left( \frac{1}{T^{*,i}} \left\| F_{k_1^i, k_2^i} B \right\|_F^2 \right)^{1/2} - \left( \frac{1}{T^{*,i}} \sum_{t=k_{\max}+1}^T \| z_t - \hat{z}_t^{-1} \|_2 \right)^{1/2}
\]

We will bound the right hand side of the last inequality. Note that

\[
\frac{1}{T^{*,i}} \sum_{t=k_{\max}+1}^T \| z_t - \hat{z}_t^{-1} \|_2^2 \leq \frac{T^{*,i-1}}{T^{*,i-1}} \frac{1}{t=k_{\max}+1} \sum_{t=k_{\max}+1}^T \| z_t - \hat{z}_t^{-1} \|_2^2 = \frac{m T^{*,i-1}}{T^{*,i}} \text{MSE}^{i-1}(\hat{a}^{i-1}, \hat{B}^{i-1}).
\]

By Lemma \([3]\) with probability one

\[
\limsup_T \text{MSE}^{i-1}(\hat{a}^{i-1}, \hat{B}^{i-1}) \leq A^{i-1}.
\]

On the other hand

\[
\inf_{\|a\|=1, \|B\|_F \geq M} \left\| F_{k_1^i, k_2^i} \frac{B}{\sqrt{T^{*,i}}} \right\|_F \geq M \inf_{\|a\|=1} \inf_{\|B\|_F=1} \left\| \frac{F_{k_1^i, k_2^i} B}{\sqrt{T^{*,i}}} \right\|_F.
\]

Note that

\[
\inf_{\|B\|_F=1} \left\| \frac{F_{k_1^i, k_2^i} B}{\sqrt{T^{*,i}}} \right\|_F = \inf_{\|B\|_F=1} \left\| \text{vec} \left( \frac{F_{k_1^i, k_2^i} B}{\sqrt{T^{*,i}}} \right) \right\|_F
\]

\[
= \inf_{\|B\|_F=1} \left\| \left( I_m \otimes \frac{F_{k_1^i, k_2^i}}{\sqrt{T^{*,i}}} \right) \text{vec}(B) \right\|
\]

\[
= \inf_{\|B\|_F=1} \left\| \left( I_m \otimes \frac{F_{k_1^i, k_2^i}}{\sqrt{T^{*,i}}} \right) b \right\|
\]

\[
= \lambda_{\min}^{1/2} \left( \left( I_m \otimes \frac{F_{k_1^i, k_2^i}}{\sqrt{T^{*,i}}} \right) \left( I_m \otimes \frac{F_{k_1^i, k_2^i}}{\sqrt{T^{*,i}}} \right) \right)
\]

\[
= \lambda_{\min}^{1/2} \left( I_m \otimes \frac{F_{k_1^i, k_2^i} F_{k_1^i, k_2^i}}{T^{*,i}} \right) = \lambda_{\min}^{1/2} \left( \frac{F_{k_1^i, k_2^i} F_{k_1^i, k_2^i}}{T^{*,i}} \right).
\]
By Lemma 2 with probability one

\[
\liminf_T \inf_{\|a\| = 1, \|B\|_F \geq M} \left\| \frac{F_{k_1^i, k_2^i}}{\sqrt{T^{*,i}}} B \right\|_F \geq M \left( \inf_{\|a\| = 1} \lambda_{\min} \left( S_{k_1^i, k_2^i}(a) \right) \right)^{1/2}.
\]

It follows that if \( M > s^i \)

\[
\liminf_T \inf_{\|a\| = 1, \|B\|_F \geq M} \text{MSE}(a, B) \geq g^i(M).
\]

Fix \( 1 \leq i \leq q \). Let \( \mathbb{P}_{T^{*,i}} \) be the empirical probability measure that places mass \( 1/T^{*,i} \) at \( y_1 = (z_1, \ldots, z_{k_{\max}^i} + 1), \ldots, y_{T^{*,i}} = (z_{T^{*,i}}, \ldots, z_T) \). The process \( (y_t)_t \) is strictly stationary and ergodic. Let

\[
L_{a_1, B_1, \ldots, a_i, B_i}(y_t) = \frac{1}{m} \| z_{t+k_{\max}^i} - z_{t+k_{\max}^i}^R \|_2.
\]

**Lemma 5.** Fix \( 1 \leq i \leq q \). For each \( M > 0 \)

\[
\sup \{ \| \mathbb{P}_{T^{*,i}} L_{a_1, B_1, \ldots, a_i, B_i} - \mathbb{P} L_{a_1, B_1, \ldots, a_i, B_i} \| : \|a_1\| = \cdots = \|a_i\| = 1, \|B_1\|, \ldots, \|B_i\| \leq M \}^{a.s.} \rightarrow 0.
\]

**Proof of Lemma 5.** Let

\[
\mathcal{L} = \{ L_{a_1, B_1, \ldots, a_i, B_i} : \|a_1\| = \cdots = \|a_i\| = 1, \|B_1\|, \ldots, \|B_i\| \leq M \}.
\]

\( \mathcal{L} \) is VC-major, since it is formed by polynomials of bounded degree. It has an integrable envelope, since \( \mathbb{E} \| z \|^2 < +\infty \). Moreover, if we take \( \mathcal{L}^0 \) to be the subset of \( \mathcal{L} \) formed by taking only \( a_c \) and \( B_c, c = 1, \ldots, i \), with rational entries it follows that: \( \mathcal{L}^0 \) is countable, and each element of \( \mathcal{L} \) is the pointwise limit of elements of \( \mathcal{L}^0 \). Then the lemma follows from Proposition 1 of Adams and Nobel (2010).  \( \square \)
Proof of Proposition 1. The proof goes by induction. Take \( i = 1 \). Take \( M > 0 \) such that
\[
g^{i}(M) > 2A^{i-1}, \quad M > s_{i}.
\]
Since \( \text{MSE}^{i}(a, B) \) is continuous, it attains its minimum over the compact set
\[
K = \{(a, B) : \|a\| = 1, \|B\|_{F} \leq M\}.
\]
Let
\[
A = \left\{ \lim \inf_{\|a\|=1, \|B\|_{F} \geq M} \inf_{M > s_{i}} \text{MSE}^{i}(a, B) \geq g^{i}(M) \right\},
\]
\[
B = \left\{ \limsup_{T} \text{MSE}^{i}(0, 0) \leq A^{i-1} \right\},
\]
\[
C = \{ \text{For large enough } T, (\hat{a}^{0}, \hat{B}^{0}), \ldots, (\hat{a}^{i-1}, \hat{B}^{i-1}) \text{ exist.} \}.
\]
Note that since \( k_{\max}^{i-1} \leq k_{\max}^{i} \)
\[
\text{MSE}^{i}(0, 0) = \frac{1}{T^{*}^{i}m} \sum_{t=k_{\max}^{i}+1}^{T} \|z_{t} - \hat{z}_{t}^{i-1}\|^{2} \leq \frac{T^{*}^{i-1}}{T^{*}^{i}m} \frac{1}{T^{*}^{i-1}m} \sum_{t=k_{\max}^{i-1}+1}^{T} \|z_{t} - \hat{z}_{t}^{i-1}\|^{2} = \frac{T^{*}^{i-1}}{T^{*}^{i}} \text{MSE}^{i-1}(\hat{a}^{i-1}, \hat{B}^{i-1}).
\]
Hence, by Lemma 3, \( \mathbb{P}(B) = 1 \). By Lemma 4, \( \mathbb{P}(A) = 1 \). \( \mathbb{P}(C) = 1 \) holds trivially, because \( (\hat{a}^{0}, \hat{B}^{0}) \) is just \( (0, 0) \). Then \( \mathbb{P}(A \cap B \cap C) = 1 \).

Assume we are working in the set \( A \cap B \cap C \) in what follows. Then for all sufficiently large \( T \) we have that
\[
\text{MSE}^{i}(0, 0) \leq 2A^{i-1}
\]
and
\[
\inf_{\|a\|=1, \|B\|_{F} \geq M} \text{MSE}^{i}(a, B) > 2A^{i-1}.
\]
Hence

\[ \min_{K} \text{MSE}^i(a, B) \leq 2A^{i-1} \]

and

\[ \inf_{\|a\|=1,\|B\|_F \geq M} \text{MSE}^i(a, B) > 2A^{i-1}. \]

It follows that there exists a global minimizer of \( \text{MSE}^i \) in \( K \), in particular there exists at least one \((\hat{a}^i, \hat{B}^i)\)

Now for the inductive step. Assume the result holds for \( i - 1 \), we want to show it then holds for \( i \). Then the proof goes like above, word by word, the only difference being that the inductive hypothesis must be used to claim that with probability one, for large enough \( T, (\hat{a}^0, \hat{B}^0), \ldots, (\hat{a}^i, \hat{B}^i) \) exist.

\[ \square \]

To keep the notation light, in what follows we always assume we are working on the set with probability one where for large enough \( T, (\hat{a}^1, \hat{B}^1), \ldots, (\hat{a}^q, \hat{B}^q) \) exist. Let

\[ Q^i(a_1, B_1, \ldots, a_i, B_i) = \mathbb{E}\|z_t - z_{R,i}^R(a_1, B_1, \ldots, a_i, B_i)\|^2. \]

Note that

\[ Q^i(a_1, B_1, \ldots, a_i, B_i) = \mathbb{P}L_{a_1, B_1, \ldots, a_i, B_i}. \]

**Lemma 6.** Assume Condition \( \square \) holds for \((k^i_1, k^i_2), i = 1, \ldots, q\). Fix \( 1 < i \leq q \). Assume \( d((\hat{a}^c, \hat{B}^c), \mathcal{I}^c) \overset{a.s.}{\to} 0 \) for \( c = 1, \ldots, i - 1 \). Then if \((a^{*c}, B^{*c}) \in \mathcal{I}^c, c = 1, \ldots, i, \)

\[ Q^i(\hat{a}^1, \hat{B}^1, \hat{a}^2, \hat{B}^2, \ldots, a^{*i}, B^{*i}) \overset{a.s.}{\to} Q^i(a^{*1}, B^{*1}, a^{*2}, B^{*2}, \ldots, a^{*i}, B^{*i}) \]

and

\[ |Q^i(\hat{a}^1, \hat{B}^1, \hat{a}^2, \hat{B}^2, \ldots, \hat{a}^i, \hat{B}^i) - Q^i(a^{*1}, B^{*1}, a^{*2}, B^{*2}, \ldots, a^{*i}, B^{*i})| \overset{a.s.}{\to} 0. \]
Proof of Lemma. Fix $\varepsilon > 0$. Since for $c = 1, \ldots, i - 1$

$$d((\hat{a}^c, \hat{B}^c), T^c) \xrightarrow{a.s.} 0,$$

with probability one, for all sufficiently large $T$ and for all $c = 1, \ldots, i - 1$ we can find $(a^{*c}, B^{*c}) \in T^c$ such that $\|\hat{a}^c - a^{*c}\| + \|\hat{B}^c - B^{*c}\| < \varepsilon$. Note that $(a^{*c}, B^{*c})$ can depend on $T$, even if this dependence is not made explicit in the notation. It follows from the proof of Proposition that there exists an $M > 0$ such that $\|b_{h}^{*c}\| \leq M$ for all $c$ and $h$. Take $c$ with $1 \leq c \leq i - 1$. Since $\|\hat{a}^c - a^{*c}\| < \varepsilon$ if we let $\eta_h^c = \sum_{h=0}^{k_2} \|z_{t-h}\|$

$$|\hat{c}_t - f_t(a^{*c})| = \left| \sum_{h=0}^{k_2} z_{t-h} (\hat{a}_h^c - a_h^{*c}) \right| \leq \sum_{h=0}^{k_2} \|z_{t-h}\| \|\hat{a}_h^c - a_h^{*c}\| \leq \varepsilon \sum_{h=0}^{k_2} \|z_{t-h}\| = \varepsilon \eta_h^c. $$

Using the triangle inequality

$$\|r_t^R(\hat{a}^c, \hat{B}^c) - r_t^R(a^{*c}, B^{*c})\| \leq \|r_t^R(a^{*c}, B^{*c}) - r_t^R(\hat{a}^c, \hat{B}^c)\| + \|r_t^R(\hat{a}^c, \hat{B}^c) - r_t^R(a^{*c}, B^{*c})\|.$$ 

The first term in the inequality above is

$$\|r_t^R(a^{*c}, B^{*c}) - r_t^R(\hat{a}^c, \hat{B}^c)\| = \left\| \sum_{h=0}^{k_2} b_h^{*c} \left( \hat{c}_t - f_t(a^{*c}) \right) \right\| \leq \varepsilon \sum_{h=0}^{k_2} \|b_h^{*c}\| \eta_{t-h}^c \leq \varepsilon M \sum_{h=0}^{k_2} \eta_{t-h}^c.$$ 

For the second term,

$$\|r_t^R(\hat{a}^c, \hat{B}^c) - r_t^R(a^{*c}, B^{*c})\| = \left\| \sum_{h=0}^{k_2} \hat{c}_t - f_t(a^{*c}) \right\| \leq \varepsilon \sum_{h=0}^{k_2} \eta_{t-h}^c.$$ 

Let

$$q(a_1, B_1, \ldots, a_{i-1}, B_{i-1}, a_1^2, B_1^2, \ldots, a_{i-1}^2, B_{i-1}^2) = \mathbb{E} \|z_t^{R,i-1}(a_1, B_1, \ldots, a_{i-1}, B_{i-1}) - z_t^{R,i-1}(a_1^2, B_1^2, \ldots, a_{i-1}^2, B_{i-1}^2)\|^2.$$ 

12
and
\[ s(a_1, B_1, a_2, B_2) = \mathbb{E}[r^R_t(a_1, B_1) - r^R_t(a_2, B_2)]^2. \]

Hence
\[ s(\hat{a}^c, \hat{B}^c, a^*, B^*) \leq \varepsilon^2(1 + M)^2 \mathbb{E}\left( \sum_{h=0}^{k_2} \eta_{t-h}^c \right)^2. \]

and
\[ q(\hat{a}^1, \hat{B}^1, \ldots, \hat{a}^{i-1}, \hat{B}^{i-1}, a^*, B^*) \leq \varepsilon^2(1 + M)^2 \mathbb{E}\left( \sum_{c=1}^{i-1} \sum_{h=0}^{k_2} \eta_{t-h}^c \right)^2. \]

Using the triangle inequality we get
\[
\left| \left( Q_i(\hat{a}^i, \hat{B}^i, \hat{a}^i, \hat{B}^i, \ldots, a^*, B^*) \right)^{1/2} - \left( Q_i(a^*, B^*, a^*, B^*, \ldots, a^*, B^*) \right)^{1/2} \right|
\]
\[ \leq \left( q(\hat{a}^i, \hat{B}^i, \ldots, \hat{a}^{i-1}, \hat{B}^{i-1}, a^*, B^*) \right)^{1/2} \]

and
\[
\left| \left( Q_i(\hat{a}^i, \hat{B}^i, \hat{a}^i, \hat{B}^i, \ldots, \hat{a}^i, \hat{B}^i) \right)^{1/2} - \left( Q_i(a^*, B^*, a^*, B^*, \ldots, a^*, B^*) \right)^{1/2} \right|
\]
\[ \leq \left( q(\hat{a}^i, \hat{B}^i, \ldots, \hat{a}^{i-1}, \hat{B}^{i-1}, a^*, B^*) \right)^{1/2} \]

The desired result now follows easily. \( \square \)

**Lemma 7.** Assume Condition 7 holds for \((k_1, k_2)\), \(i = 1, \ldots, q\). Fix \(1 \leq i \leq q\). Assume \(d((\hat{a}^c, \hat{B}^c), T^c) \xrightarrow{a.s.} 0\) for \(c = 0, \ldots, i - 1\). Then, with probability one,

\[ \limsup_T \text{MSE}_i(\hat{a}^i, \hat{B}^i) \leq \text{MSE}_i(a^*, B^*) \]

**Proof of Lemma 7.** For \(i = 1\) the result follows immediately from the definition of \((\hat{a}^i, \hat{B}^i)\) and the Ergodic Theorem. Take \(i > 1\). From the proofs of Proposition 1 and 2 we know
that there exists an $M > 0$ such that with probability one

$$\limsup_T \|\hat{B}^c\|_F \leq \frac{M}{2}, \ c = 1, \ldots, i,$$

and moreover for all $(a^{*,i}, B^{*,i}) \in \mathcal{I}^i, \ \|B^{*,i}\|_F \leq M/2$. Then if

$$A = \left\{ \limsup_T \|\hat{B}^c\|_F \leq \frac{M}{2}, \ c = 1, \ldots, i \right\},$$

$\mathbb{P}(A) = 1$. Define the sets

$$B = \left\{ \sup \left\{ \mathbb{P}_T L_{a_1, a_2, \ldots, a_i, B} - \mathbb{P}_L a_1, a_2, \ldots, a_i, B : \|a_c\| = 1, \|B_c\| \leq M, c \leq i \right\} \to 0 \right\}$$

$$C = \left\{ Q'(\hat{a}^1, \hat{B}^1, \hat{a}^2, \hat{B}^2, \ldots, a^{*,i}, B^{*,i}) \to Q'(a^{*,1}, B^{*,1}, a^{*,2}, B^{*,2}, \ldots, a^{*,i}, B^{*,i}) \right\}$$

By Lemma 5, $\mathbb{P}(B) = 1$. By Lemma 6, $\mathbb{P}(C) = 1$. Assume in what follows that we are working in the set $A \cap B \cap C$.

Fix $\zeta > 0$. By definition of $(\hat{a}^i, \hat{B}^i)$ we have that

$$\text{MSE}^i(\hat{a}^i, \hat{B}^i) \leq \text{MSE}^i(a^{*,i}, B^{*,i}) = \mathbb{P}_T L_{\hat{a}^1, \hat{B}^1, a^{*,i}, B^{*,i}}.$$

Then, for sufficiently large $T$,

$$\mathbb{P}_T L_{\hat{a}^1, \hat{B}^1, a^{*,i}, B^{*,i}} \leq \mathbb{P}_L a_1, a_2, \ldots, a^{*,i}, B^{*,i} + \frac{\zeta}{2} = Q'(a^{*,1}, B^{*,1}, a^{*,2}, B^{*,2}, \ldots, a^{*,i}, B^{*,i}) + \frac{\zeta}{2} \leq Q'(a^{*,1}, B^{*,1}, a^{*,2}, B^{*,2}, \ldots, a^{*,i}, B^{*,i}) + \zeta$$

$$= \text{MSE}_i(a^{*,i}, B^{*,i}) + \zeta.$$?

The result is proven.

Lemma 8. Assume Condition 7 holds for $(k^i_1, k^i_2)$, $i = 1, \ldots, q$. Fix $1 \leq i \leq q$. Assume $d((\hat{a}^c, \hat{B}^c), \mathcal{I}^c) \xrightarrow{a.s.} 0$ for $c = 0, \ldots, i - 1$. Fix $\varepsilon > 0$. Let

$$A = \left\{ \limsup_T d((\hat{a}^i, \hat{B}^i), \mathcal{I}^i) \geq \varepsilon \right\},$$

14
and assume $\mathbb{P}(A) > 0$. Then there exists $M_0$ such that if $M > M_0$, there exists a set $E(M)$ with probability one, such that in the set $A \cap E(M)$ there exists a subsequence $(T_k)_k$ such that

$$\liminf_k \mathbb{P}_{T_k} L_{a_k^1 \hat{B}_k^1, \ldots, a_k^i \hat{B}_k^i} \geq \inf \left\{ \text{MSE}^i_*(a, B) : \|a\| = 1, \|B\|_F \leq M, d((a, B), I^1) \geq \varepsilon/2 \right\},$$

where $(\hat{a}_k^i, \hat{B}_k^i)$ are the values of $(\hat{a}^i, \hat{B}^i)$ over the subsequence $(T_k)_k$.

**Proof of Lemma 8.** From the proofs of Proposition 1 and 2 we know that there exists an $M > 0$ such that with probability one

$$\limsup_T \|\hat{B}_c\|_F \leq M_0^2, c = 1, \ldots, i.$$

Take $M > M_0$. Then if

$$B(M) = \left\{ \limsup_T \|\hat{B}_c\|_F \leq M/2, c = 1, \ldots, i \right\},$$

$$\mathbb{P}(B(M)) = 1.$$ Let

$$C = \left\{ \sup \{\mathbb{P}_{T^*} L_{a_1, a_2, \ldots, a_i, B_i, \ldots, a_i, B_i} - \mathbb{P} L_{a_1, a_2, \ldots, a_i, B_i, \ldots, a_i, B_i} : \|a_c\| = 1, \|B_c\| \leq M, c \leq i \} \to 0 \right\}.$$

By Lemma 5, $\mathbb{P}(C) = 1$.

Assume now that $i = 1$. Take $E(M) = B(M) \cap C$. Assume in what follows that we are working in the set $A \cap E(M)$. Then we can find a subsequence $(T_k)_k$ such that $\|\hat{B}_k^1\|_F \leq M$, and $d((\hat{a}_k^1, \hat{B}_k^1), I^1) \geq \varepsilon/2$ for all $k$. Fix $\zeta > 0$. Then for sufficiently large $k$

$$\mathbb{P}_{T_k^*} L_{\hat{a}_k^1, \hat{B}_k^1} \geq \mathbb{P} L_{\hat{a}_k^1, \hat{B}_k^1} - \zeta = \text{MSE}^1_*(\hat{a}_k^1, \hat{B}_k^1) - \zeta.$$

Moreover,

$$\text{MSE}^1_*(\hat{a}_k^1, \hat{B}_k^1) \geq \inf \left\{ \text{MSE}^1_*(a, B) : \|a\| = 1, \|B\|_F \leq M, d((a, B), I^1) \geq \varepsilon/2 \right\}.$$
This finishes the proof for \( i = 1 \).

Now for the case \( i > 1 \). Let

\[
D = \left\{ |Q^i(\hat{a}^1, \hat{B}^1, \hat{a}^2, \hat{B}^2, \ldots, \hat{a}^i, \hat{B}^i) - Q^i(a^*1, B^*1, a^*2, B^*2, \ldots, \hat{a}^i, \hat{B}^i)| \to 0 \right\},
\]

By Lemma 6, \( \mathbb{P}(D) = 1 \). Take \( E(M) = B(M) \cap C \cap D \). Assume in what follows that we are working in the set \( A \cap E(M) \). Then we can find a subsequence \( (T_k)_k \) such that

\[
\|\hat{B}_k\|_F \leq M, 1 \leq c \leq i, \text{ and } d((\hat{a}^i_k, \hat{B}^i_k), T^i) \geq \varepsilon/2 \text{ for all } k. \]

Fix \( \zeta > 0 \). Then for sufficiently large \( k \)

\[
\mathbb{P}_{T_k^*} L_{\hat{a}^i_k, \hat{B}^i_k} \geq \mathbb{P} L_{\hat{a}^i_k, \hat{B}^i_k} - \frac{\zeta}{2} = Q^i(\hat{a}^1_k, \hat{B}^1_k, \hat{a}^2_k, \hat{B}^2_k, \ldots, \hat{a}^i_k, \hat{B}^i_k) - \frac{\zeta}{2} \geq Q^i(a^{*1}, B^{*1}, a^{*2}, B^{*2}, \ldots, \hat{a}^i_k, \hat{B}^i_k) - \zeta = \text{MSE}^i(a^*_k, B^*_k) - \zeta.
\]

Moreover,

\[
\text{MSE}^i(a^*_k, B^*_k) \geq \inf \left\{ \text{MSE}^i(a, B) : \|a\| = 1, \|B\|_F \leq M, d((a, B), T^i) \geq \varepsilon/2 \right\}.
\]

The result is proven.

\[\square\]

**Proof of Theorem 7** The proof goes by induction. Take \( i = 1 \). Fix \( \varepsilon > 0 \). Let

\[
A = \left\{ \limsup_T d((\hat{a}^i, \hat{B}^i), T^i) \geq \varepsilon \right\}.
\]

We will show that \( \mathbb{P}(A) = 0 \). Assume \( \mathbb{P}(A) > 0 \). Define the sets

\[
B = \left\{ \lim_T d((\hat{a}^c, \hat{B}^c), T^c) = 0, c = 0, \ldots, i - 1 \right\},
\]

\[
C = \left\{ \limsup_T \text{MSE}^i(\hat{a}^i, \hat{B}^i) \leq \text{MSE}^i(a^{*i}, B^{*i}) \right\},
\]

\[16\]
Clearly $\mathbb{P}(B) = 1$. By Lemma $7$, $\mathbb{P}(C) = 1$. Take the $M_0$ given by Lemma $8$. Let $M$ be such that
\[ M > M_0, \quad \|B^*\|_F \leq M, c = 0, \ldots, i \]
Let $E(M)$ be the set provided by Lemma $8$. Then $\mathbb{P}(A \cap B \cap C \cap E(M)) > 0$. Assume in what follows that we are working in the set $A \cap B \cap C \cap E(M)$. Since in particular we are in set $A \cap E(M)$, there exists a subsequence $(T_k)_k$ such that
\[ \liminf_k \mathbb{P}_{T_k^*} L_{\tilde{a}_k^1, \tilde{B}_k^1, \ldots, \tilde{a}_k^i, \tilde{B}_k^i} \geq \inf \{ \text{MSE}^*_i(a, B) : \|a\| = 1, \|B\|_F \leq M, d((a, B), I^i) \geq \varepsilon/2 \} \, , \]
Moreover, since we are in set $B \cap C$
\[ \limsup_k \mathbb{P}_{T_k^*} L_{\tilde{a}_k^1, \tilde{B}_k^1, \ldots, \tilde{a}_k^i, \tilde{B}_k^i} \leq \text{MSE}^*_i(a^{*,i}, B^{*,i}) \, . \]
Then
\[ \text{MSE}^*_i(a^{*,i}, B^{*,i}) \geq \liminf_k \mathbb{P}_{T_k^*} L_{\tilde{a}_k^1, \tilde{B}_k^1, \ldots, \tilde{a}_k^i, \tilde{B}_k^i} \geq \inf \{ \text{MSE}^*_i(a, B) : \|a\| = 1, \|B\|_F \leq M, d((a, B), I^i) \geq \varepsilon/2 \} > \text{MSE}^*_i(a^{*,i}, B^{*,i}) \, , \]
a contradiction. It must be that $\mathbb{P}(A) = 0$.

For the inductive step, the proof goes exactly as above, only that the inductive hypothesis has to be invoked to claim that $\mathbb{P}(B) = 1$. \qed
Since the following involves only the first ODPC, to keep the notation light, we will drop the superscript indicating the component number from $z_{R,1}^R, (a^{*,1}, B^{*,1}), \text{MSE}^1, \text{MSE}^1_*, \text{etc.}$ For $h = 0, \ldots, k_2$ let $b_h = (b_{h,1}, \ldots, b_{h,m})'$. For $a \in \mathbb{R}^{m(k_1+1)}$ let $f_t'(a) = (f_t(a), \ldots, f_{t-k_2}(a))$.

**Proof of Theorem 3** For all $a \in \mathbb{R}^{m(k_1+1)}$,

\[
\begin{align*}
    f_t(a) &= \begin{pmatrix}
        z_t' & \cdots & z_{t-k_1}' \\
        \vdots & \ddots & \vdots \\
        z_{t-k_2}' & \cdots & z_{t-k_1-k_2}'
    \end{pmatrix} a \\
    &= \begin{pmatrix}
        u_t' B & \cdots & u_{t-k_1}' B \\
        \vdots & \ddots & \vdots \\
        u_{t-k_2}' B & \cdots & u_{t-k_1-k_2}' B
    \end{pmatrix} a + \begin{pmatrix}
        e_t' & \cdots & e_{t-k_1}' \\
        \vdots & \ddots & \vdots \\
        e_{t-k_2}' & \cdots & e_{t-k_1-k_2}'
    \end{pmatrix} a \\
    &= \begin{pmatrix}
        u_t' & \cdots & u_{t-k_1}' \\
        \vdots & \ddots & \vdots \\
        u_{t-k_2}' & \cdots & u_{t-k_1-k_2}'
    \end{pmatrix} (I_{k_1+1} \otimes B) a + \begin{pmatrix}
        e_t' & \cdots & e_{t-k_1}' \\
        \vdots & \ddots & \vdots \\
        e_{t-k_2}' & \cdots & e_{t-k_1-k_2}'
    \end{pmatrix} a \\
    &= U_t (I_{k_1+1} \otimes B) a + E_t a,
\end{align*}
\]
By Condition 2a), for sufficiently large \( m \), the rows of \( B \) are linearly independent. By the Hahn-Banach Theorem, there exists \( \tilde{a}_0 \in \mathbb{R}^m \) with \( \|\tilde{a}_0\| = 1 \) such that

\[
\tilde{a}_0' b = \|b_0 - P(b_0|b_1, \ldots, b_{k_2})\| \quad \text{and} \quad \tilde{a}_0' b_i = 0 \quad \text{for} \quad i = 1, \ldots, k_2,
\]

where \( P(b_0|b_1, \ldots, b_{k_2}) \) is the projection of \( b_0 \) onto the space spanned by \( b_1, \ldots, b_{k_2} \). Let \( B_{-(0)} \) be the matrix obtained by removing the first row and column of \( B \). Then

\[
P(b_0|b_1, \ldots, b_{k_2}) = B'_{-(0)} \left( B_{-(0)} B'_{-(0)} \right)^{-1} B_{-(0)} b_0.
\]

Hence using Condition 2a)

\[
\frac{\|P(b_0|b_1, \ldots, b_{k_2})\|^2}{m} = \frac{b_0' B'_{-(0)} \left( \frac{B_{-(0)} B'_{-(0)}}{m} \right)^{-1} B_{-(0)} b_0}{m} \to 0.
\]

It follows that there exists a fixed constant \( c > 0 \) such that for all sufficiently large \( m \)

\[
n(b_0) = \|b_0 - P(b_0|b_1, \ldots, b_{k_2})\| \geq c\sqrt{m}.
\]

Let \( \tilde{a} \in \mathbb{R}^{m(k_1+1)} \) be defined by \( \tilde{a} = (\overline{a}_0, 0_m, \ldots, 0_m)' \). Since \( (a^*, B^*) \in I \)

\[
mMSE_s(a^*, B^*) \leq mMSE_s(\tilde{a}, B/n(b_0))
\]

\[
= \mathbb{E} \|B'u_t + e_t - (B'/n(b_0))U_t(I_{k_1+1} \otimes B)\tilde{a} - (B'/n(b_0))E_t\tilde{a}\|^2
\]

\[
= \mathbb{E} \|B'u_t - B'U_t(I_{k_1+1} \otimes (B/n(b_0)))\tilde{a}\|^2 + \mathbb{E} \|e_t - (B'/n(b_0))E_t\tilde{a}\|^2
\]

\[
+ 2\mathbb{E} (B'u_t - B'U_t(I_{k_1+1} \otimes (B/n(b_0))))\tilde{a}'(e_t - (B'/n(b_0))E_t\tilde{a}) = 0.
\]

By Condition 2b) and c), \( \mathbb{E} (B'u_t - B'U_t(I_{k_1+1} \otimes (B/n(b_0))))\tilde{a}'(e_t - (B'/n(b_0))E_t\tilde{a}) = 0. \)

By construction of \( \tilde{a} \)

\[
(I_{k_1+1} \otimes (B/n(b_0)))\tilde{a} = (1, 0, \ldots, 0)' \in \mathbb{R}^{(k_1+1)(k_2+1)}
\]
which implies that $U_t(I_{k_1+1} \otimes (B/n(b_0)))\tilde{a} = u_t$. Hence

$$\frac{1}{m} \mathbb{E} \|B'u_t - B'U_t(I_{k_1+1} \otimes (B/n(b_0)))\tilde{a}\|^2 = 0.$$  

Note that

$$\mathbb{E}\|e_t - (B'/\|b_0\|)E_t\tilde{a}\|^2 = \mathbb{E}\|e_t\|^2 + \mathbb{E}\|(B'/n(b_0))E_t\tilde{a}\|^2 - 2\mathbb{E}e'_t(B'/n(b_0))E_t\tilde{a}.$$  

Now $E_t\tilde{a} = (e'_t\tilde{a}_0, \ldots, e'_{t-k_2}\tilde{a}_0)'$ and hence

$$\|E_t\tilde{a}\|^2 = \sum_{h=0}^{k_2} (e'_{t-h}\tilde{a}_0)^2$$  

It follows that $\mathbb{E}\|E_t\tilde{a}\|^2 = (k_2 + 1) (\tilde{a}'_0\Sigma^e(0)\tilde{a}_0)$. Hence, using Condition 2a) and c)

$$\frac{1}{m} \mathbb{E}\|(B'/n(b_0))E_t\tilde{a}\|^2 \leq \frac{\|B'\|^2}{m} (k_2 + 1) (\tilde{a}'_0\Sigma^e(0)\tilde{a}_0) \frac{1}{n(b_0)^2} = O \left( \frac{1}{m} \right).$$  

Finally,

$$\frac{1}{mn(b_0)} \mathbb{E}\|e'B'E_t\tilde{a}\| \leq \left( \mathbb{E}\|e_t\|^2/m \right)^{1/2} \left( \mathbb{E}\|(B'/n(b_0))E_t\tilde{a}\|^2/m \right)^{1/2} = O \left( \frac{1}{\sqrt{m}} \right).$$  

Since $\mathbb{E}\|e_t\|^2 = \sum_{j=1}^{m} e_{t,j}^2$, the Theorem is proven. \hfill \Box

We will need the following general results on linear predictors.

**Lemma 9.** Let $X$ be a zero mean random variable and $W' = (W_1, \ldots, W_n)$ be a zero mean random vector such that $\text{cov}(W)$ is non-singular. Let $P(W_i|X)$ be the best linear predictor of $W_i$ based on $X$. Then

$$\sum_{i=1}^{n} \mathbb{E}P(W_i|X)^2 \leq \lambda_{\text{max}}(\text{cov}(W)).$$
Proof of Lemma 9. Let $P(X|W)$ be the best linear predictor of $X$ based on $W_1, \ldots W_n$. Then

$$P(X|W) = \mathbb{E}(XW')(\text{cov}(W))^{-1}W$$

Hence

$$\mathbb{E}X^2 \geq \mathbb{E}P(X|W)^2 = \mathbb{E}(\mathbb{E}(XW')(\text{cov}(W))^{-1}WW'(\text{cov}(W))^{-1}\mathbb{E}(XW))$$

$$= \mathbb{E}(XW')(\text{cov}(W))^{-1}\mathbb{E}(XW) \geq (\lambda_{\text{max}}(\text{cov}(W)))^{-1}\sum_{i=1}^{n}(E(XW_i))^2.$$ 

Now,

$$\sum_{i=1}^{n} \mathbb{E}P(W_i|X)^2 = \sum_{i=1}^{n} \mathbb{E}(\frac{\mathbb{E}XW_i}{\mathbb{E}X^2}X)^2 = \sum_{i=1}^{n} (\frac{\mathbb{E}XW_i}{\mathbb{E}X^2})^2 \leq \lambda_{\text{max}}(\text{cov}(W)).$$

\hfill \Box

Lemma 10. Let $W' = (W_1, \ldots, W_n)$ be a zero mean random vector such that $\text{cov}(W)$ is non-singular. Let $Y_1, \ldots, Y_k$ be zero mean random variables and $P(W_i|Y)$ be the best linear predictor of $W_i$ based on $Y_1, \ldots, Y_k$. Then

$$\frac{1}{n} \sum_{i=1}^{n} (\mathbb{E}(W_i - P(W_i|Y))^2 - \mathbb{E}W_i^2) \geq -\frac{k\lambda_{\text{max}}(\text{cov}(W))}{n}.$$ 

Proof of Lemma 10. Let $X_1, \ldots, X_k$ be zero mean, unit variance, uncorrelated random variables with the same linear span as $Y_1, \ldots, Y_k$. Let $P(W_i|X)$ be the best linear predictor of $W_i$ based on $X_1, \ldots, X_k$. Then, using Lemma 9

$$\sum_{i=1}^{n} \mathbb{E}(W_i - P(W_i|X))^2 = \sum_{i=1}^{n} \mathbb{E}(W_i - P(W_i|X))^2 = \sum_{i=1}^{n} (\mathbb{E}W_i^2 - \mathbb{E}P(W_i|X)^2)$$

$$= \sum_{i=1}^{n} \mathbb{E}W_i^2 - \sum_{i=1}^{n} \sum_{j=1}^{k} \mathbb{E}P(W_i|X_j)^2 \geq \sum_{i=1}^{n} \mathbb{E}W_i^2 - k\lambda_{\text{max}}(\text{cov}(W)).$$

21
Hence
\[
\frac{1}{n} \sum_{i=1}^{n} \left( \mathbb{E}(W_i - P(W_i|Y))^2 - \mathbb{E}W_i^2 \right) \geq -\frac{k\lambda_{\text{max}}(\text{cov}(W))}{n}.
\]

\[\square\]

**Proof of Theorem 3**

\[\text{MSE}_*(a^*, B^*) = \frac{1}{m} \mathbb{E}\|\mathbf{z}_t - \mathbf{z}_t^R(a^*, B^*)\|^2 = \frac{1}{m} \mathbb{E}\|\mathbf{B}'\mathbf{u}_t - \mathbf{z}_t^R(a^*, B^*) + \mathbf{e}_t\|^2 = \frac{1}{m} \mathbb{E}\|\mathbf{X}_t - \mathbf{z}_t^R(a^*, B^*)\|^2 + \frac{1}{m} \mathbb{E}\|\mathbf{e}_t\|^2 + \frac{2}{m} \mathbb{E}\text{tr}(\mathbf{B}'\mathbf{u}_t - \mathbf{z}_t^R(a^*, B^*)）。\]

By Condition 2c)

\[\mathbb{E}e_t' (\mathbf{B}'\mathbf{u}_t - \mathbf{z}_t^R(a^*, B^*)) = \mathbb{E}e_t' \mathbf{z}_t^R(a^*, B^*) = 0.\]

By Theorem 2, it suffices to show that

\[\frac{1}{m} \mathbb{E}e_t' \mathbf{z}_t^R(a^*, B^*) = O\left(\frac{1}{\sqrt{m}}\right).\]

Let \(\hat{e}_{t,j}\) be the best linear predictor of \(e_{t,j}\) based on \(z_{t,j}^R(a^*, B^*)\). Then

\[\mathbb{E} (e_{t,j} - \hat{e}_{t,j})^2 = \mathbb{E}e_{t,j}^2 - \frac{(\mathbb{E}z_{t,j}^R(a^*, B^*)e_{t,j})^2}{\mathbb{E}z_{t,j}^R(a^*, B^*)^2}.\]

By Condition 4a), \(\mathbb{E}z_{t,j}^R(a^*, B^*)^2 \leq \mathbb{E}z_{t,j}^2 \leq L\), since \(z_{t,j}^R(a^*, B^*)\) is obtained by projecting \(z_{t,j}\) on the space spanned by \(f_1(a^*), \ldots, f_{t-k_2}(a^*)\). Then

\[
\frac{1}{m} \sum_{j=1}^{m} \mathbb{E}(e_{t,j} - \hat{e}_{t,j})^2 \leq \frac{1}{m} \sum_{j=1}^{m} \left( \mathbb{E}e_{t,j}^2 - \frac{(\mathbb{E}z_{t,j}^R(a^*, B^*)e_{t,j})^2}{L} \right).
\]

Hence by Condition 4b), using Lemma 10 we get

\[
-\frac{(k_2 + 1)\lambda_{\text{max}}(\Sigma^e(0))}{m} \leq \frac{1}{m} \sum_{j=1}^{m} \left( \mathbb{E}(e_{t,j} - P(e_{t,j}|f_1(a^*), \ldots, f_{t-k_2}(a^*)))^2 - \mathbb{E}e_{t,j}^2 \right)
\]

\[
\leq \frac{1}{m} \sum_{j=1}^{m} (\mathbb{E}(e_{t,j} - \hat{e}_{t,j})^2 - \mathbb{E}e_{t,j}^2) \leq -\frac{1}{m} \sum_{j=1}^{m} \left( \mathbb{E}z_{t,j}^R(a^*, B^*)e_{t,j})^2 \right).
\]

22
and thus
\[
\frac{1}{m} \sum_{j=1}^{m} (\mathbb{E}z_{t,j}(a^*, B^*)e_{t,j})^2 \leq \frac{(k_2 + 1)L\lambda_{\max}(\Sigma^e(0))}{m}
\]

Applying the Cauchy-Schwartz inequality

\[
\frac{1}{m} \sum_{j=1}^{m} |\mathbb{E}z_{t,j}(a^*, B^*)e_{t,j}| \leq \frac{1}{m} \left( \sum_{j=1}^{m} (\mathbb{E}z_{t,j}(a^*, B^*)e_{t,j})^2 \right)^{1/2} m^{1/2}
\]

\[
= \left( \frac{1}{m} \sum_{j=1}^{m} (\mathbb{E}z_{t,j}(a^*, B^*)e_{t,j})^2 \right)^{1/2}
\]

\[
\leq \left( \frac{(k_2 + 1)L\lambda_{\max}(\Sigma^e(0))}{m} \right)^{1/2}
\]

The result now follows immediately. \(\square\)

**References**
