APPENDIX B

In Sections B.1, B.2, B.3 and B.4 we prove Lemma A.1, Lemma A.2, Corollary A.3 and Lemma A.4, respectively, used in Appendices A.1-A.4 of the paper. These proofs require technical Lemmas B.1-B.5, which are proved in Sections B.5-B.9.

B.1 Proof of Lemma A.1

Write:

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right)-E_0\left[\varphi\left(\mu_t(\beta)\right)\right] \\
= & \frac{1}{T}\sum_{t=1}^{T}\varphi\left(\mu_t\left(\beta\right)\right)-E_0\left[\varphi\left(\mu_t\left(\beta\right)\right)\right] \\
& +\frac{1}{T}\sum_{t=1}^{T}\left\{\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},f_t(\beta),\beta)\right)-\varphi\left(\mu_t\left(\beta\right)\right)\right\} \\
+ & \frac{1}{T}\sum_{t=1}^{T}\left\{\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right)-\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},f_t(\beta),\beta)\right)\right\} \\
=: & \quad J_{1,T}(\beta)+J_{2,n,T}(\beta)+J_{3,n,T}(\beta).
\end{aligned}
$$

Term $J_{1,T}(\beta)$ is the time series average of a nonlinear transformation of process $\mu_t\left(\beta\right)$. We show in Section B.1.1 that $\sup_{\beta\in\mathcal{B}}|J_{1,T}(\beta)|=o_p(1)$. Term $J_{2,n,T}(\beta)$ accounts for the discrepancy between the cross-sectional average $\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},f_t(\beta),\beta)$ and the conditional expectation $\mu_t\left(\beta\right)$. We prove that $\sup_{\beta\in\mathcal{B}}|J_{2,n,T}(\beta)|=o_p(1)$ in Section B.1.2. Finally, term $J_{3,n,T}(\beta)$ is induced by the approximation of the pseudo-true factor value $f_t(\beta)$ with estimator $\hat{f}_{n,t}(\beta)$, and we show that $\sup_{\beta\in\mathcal{B}}|J_{3,n,T}(\beta)|=o_p(1)$ in Section B.1.3.

B.1.1 Proof that $\sup_{\beta\in\mathcal{B}}|J_{1,T}(\beta)|=o_p(1)$

From Theorem 21.9 in Davidson (1994) we have $\sup_{\beta\in\mathcal{B}}|J_{1,T}(\beta)|\xrightarrow{p}0$ if and only if:

(i) $|J_{1,T}(\beta)|\xrightarrow{p}0$, for any β in a dense subset of $\mathcal{B}$,

(ii) $J_{1,T}(\beta)$ is stochastically equicontinuous, that is, for any $\varepsilon > 0$ there exists $\delta > 0$ such that:

$$\limsup_{T\to\infty} P\left[\sup_{\beta,\beta'\in\mathcal{B}:\left\|\beta'-\beta\right\|\leq\delta}\left|J_{1,T}(\beta)-J_{1,T}(\beta')\right|>\varepsilon\right]<\varepsilon.$$

We prove points (i) and (ii) in Sections B.1.1.1 and B.1.1.2, respectively.

B.1.1.1 Pointwise convergence

Let $\beta \in \mathcal{B}$ be given. Let us first show that $\mu_t(\beta)$ is Near-Epoch Dependent in L^2-norm (L^2-NED) on $\{f_t : t \in \mathbb{Z}\}$ with size $-\alpha$. [1] Indeed, for $\mathcal{F}_{t-m}^{t+m} = \sigma\left(f_{t+\tau} : \tau = 0, \pm 1, ..., \pm m\right)$ we have:

$$\begin{aligned}\left\|\mu_t(\beta) - E\left[\mu_t(\beta) \mid \mathcal{F}_{t-m}^{t+m}\right]\right\|_2 &\leq \left\|E\left[a(Y_{i,t}, f_t(\beta), \beta) \mid \underline{f_t}\right] - E\left[a(Y_{i,t}, f_t(\beta), \beta) \mid f_t, ..., f_{t-m}\right]\right\|_2 \\ &= O\left(m^{-\alpha}\right),\end{aligned}$$

by Condition (1) (iv), where $\|Z\|_2 = E\left[\|Z\|_2\right]^{1/2}$. Since $\mu_t(\beta)$ is L^2-NED of size $-\alpha$ on $\{f_t : t \in \mathbb{Z}\}$, and φ is Lipshitz by condition (2), it follows that $\varphi\left(\mu_t\left(\beta\right)\right)$ is also L^2-NED of size $-\alpha$ on $\{f_t : t \in \mathbb{Z}\}$ by Theorem 17.12 of Davidson (1994). Moreover, from condition (2) we have that $\varphi\left(\mu_t\left(\beta\right)\right)$ is L^τ-bounded, $\tau > 2$. Then, since (f_t) is geometric strong mixing (Assumption A.4), it follows that $\varphi\left(\mu_t\left(\beta\right)\right) - E_0\left[\varphi\left(\mu_t\left(\beta\right)\right)\right]$ is a L^2-mixingale of size $-\alpha$ w.r.t. the filtration $\mathcal{F}_t = \sigma\left(f_\tau : \tau \leq t\right)$ by Theorem 17.5 in Davidson (1994). [2] Moreover, $\varphi\left(\mu_t\left(\beta\right)\right)$ is also uniformly integrable by Theorem 18.5 in Davidson (1994). It follows by the WLLN in Andrews (1988), Theorem 1, that $\frac{1}{T}\sum_{t=1}^{T}\left\{\varphi\left(\mu_t\left(\beta\right)\right) - E_0\left[\varphi\left(\mu_t\left(\beta\right)\right)\right]\right\} = o_p(1)$.

[1] A strictly stationary process X_t is L^2-LED on $\{f_t : t \in \mathbb{Z}\}$ with size $-\alpha$ if:

$$\left\|X_t - E\left[X_t \mid \mathcal{F}_{t-m}^{t+m}\right]\right\|_2 = O\left(m^{-\alpha}\right),\ m \to \infty,$$

where $\mathcal{F}_{t-m}^{t+m} = \sigma\left(f_{t+\tau} : \tau = 0, \pm 1, ..., \pm m\right)$ [see Davidson (1994), Definition 17.1].

[2] A strictly stationary, $\mathcal{F}_t$-adapted, zero-mean process X_t is a L^2-mixingale of size $-\alpha$ w.r.t. $\mathcal{F}_t$ if:

$$\left\|X_t - E\left[X_t \mid \mathcal{F}_{t-m}\right]\right\|_2 = O\left(m^{-\alpha}\right),\ m \to \infty,$$

[see Definition 1 in Andrews (1988) and Definition 16.1 in Davidson (1994)].

B.1.1.2 Stochastic equicontinuity

From Theorem 21.11 in Davidson (1994), $J_{1,T}(\beta) = \frac{1}{T}\sum_{t=1}^{T} q_t(\beta)$ is stochastically equicontinuous if:

$$\left|q_t\left(\beta^{'}\right) - q_t(\beta)\right| \leq B_t \left\|\beta^{'} - \beta\right\|, \quad \forall \beta, \beta \in \mathcal{B}, \tag{B.1}$$

and B_t is a stochastic sequence independent of β such that $\frac{1}{T}\sum_{t=1}^{T} B_t = O_p(1)$. Now, $q_t(\beta) = \varphi(\mu_t(\beta)) - E_0[\varphi(\mu_t(\beta))]$. Using that φ is Lipshitz, and denoting by L the Lipshitz constant, we get:

$$\left|q_t\left(\beta^{'}\right) - q_t(\beta)\right| \leq L\left\|\mu_t\left(\beta^{'}\right) - \mu_t(\beta)\right\| + LE\left[\left\|\mu_t\left(\beta^{'}\right) - \mu_t(\beta)\right\|\right].$$

Using $\left\|\mu_t\left(\beta^{'}\right) - \mu_t(\beta)\right\| \leq \sup_{\beta\in\mathcal{B}} E\left[\left\|\frac{\partial a(Y_{i,t}, f_t(\beta), \beta)}{\partial \beta'}\right\| |\underline{f_t}\right] \left\|\beta^{'} - \beta\right\|$, we see that (B.1) is satisfied with $B_t = \sup_{\beta\in\mathcal{B}} E\left[\left\|\frac{\partial a(Y_{i,t}, f_t(\beta), \beta)}{\partial \beta'}\right\| |\underline{f_t}\right] + E\left[\sup_{\beta\in\mathcal{B}} E\left[\left\|\frac{\partial a(Y_{i,t}, f_t(\beta), \beta)}{\partial \beta'}\right\| |\underline{f_t}\right]\right]$. Thus $\frac{1}{T}\sum_{t=1}^{T} B_t = O_p(1)$ follows from $E\left[\sup_{\beta\in\mathcal{B}} E\left[\left\|\frac{\partial a(Y_{i,t}, f_t(\beta), \beta)}{\partial \beta'}\right\| |\underline{f_t}\right]\right] < \infty$, which is implied by condition (1) (iii).

B.1.2 Proof that $\sup_{\beta\in\mathcal{B}} |J_{2,n,T}(\beta)| = o_p(1)$

Let $\varepsilon > 0$. Using that φ is Lipshitz, and denoting by L the Lipshitz constant, we have that $\|x - y\| \leq L/\varepsilon$ implies $|\varphi(x) - \varphi(y)| \leq \varepsilon$. Thus, we get:

$$\begin{aligned} & P\left[\sup_{\beta\in\mathcal{B}} \left|\frac{1}{T}\sum_{t=1}^{T}\left\{\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, f_t(\beta), \beta)\right) - \varphi(\mu_t(\beta))\right\}\right| \geq \varepsilon\right] \\ \leq \; & P\left[\sup_{\beta\in\mathcal{B}} \sup_{1\leq t\leq T} \left\|\frac{1}{n}\sum_{i=1}^{n}[a(Y_{i,t}, f_t(\beta), \beta) - \mu_t(\beta)]\right\| \geq L/\varepsilon\right] = P_{1,\varepsilon}. \end{aligned}$$

To bound $P_{1,\varepsilon}$, define for any $\delta > 0$ the event:

$$\Omega_{1,n,T}(\delta) = \left\{\sup_{\beta\in\mathcal{B}} \sup_{1\leq t\leq T} \left\|\frac{1}{n}\sum_{i=1}^{n}[a(Y_{i,t}, f_t(\beta), \beta) - \mu_t(\beta)]\right\| \leq \delta\right\}.$$

Lemma B.1: *Under conditions (1) (i)-(iii), (v)-(vi) and (4) of Lemma A.1:* $P[\Omega_{1,n,T}(\delta)] \to 1$ *as* $n, T \to \infty$, *for any* $\delta > 0$.

Since $P_{1,\varepsilon} = 1 - P[\Omega_{1,n,T}(L/\varepsilon)]$, by Lemma B.1 we get that $P_{1,\varepsilon} \to 0$ as $n, T \to \infty$, for any $\varepsilon > 0$. It follows that $\sup\limits_{\beta \in \mathcal{B}} |J_{2,n,T}(\beta)| = o_p(1)$.

B.1.3 Proof that $\sup\limits_{\beta \in \mathcal{B}} |J_{3,n,T}(\beta)| = o_p(1)$

Let $\varepsilon > 0$ be given. Then:

$$\begin{aligned} & P\left[\sup_{\beta \in \mathcal{B}} \left| \frac{1}{T} \sum_{t=1}^{T} \left\{ \varphi\left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) - \varphi\left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, f_t(\beta), \beta) \right) \right\} \right| \geq \varepsilon \right] \\ \leq \; & P\left[\sup_{\beta \in \mathcal{B}} \sup_{1 \leq t \leq T} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) - a(Y_{i,t}, f_t(\beta), \beta) \right] \right\| \geq L/\varepsilon \right] = P_{2,\varepsilon}. \end{aligned}$$

To bound $P_{2,\varepsilon}$, define for any $\delta > 0$ the event:

$$\Omega_{2,n,T}(\delta) = \left\{ \sup_{\beta \in \mathcal{B}} \sup_{1 \leq t \leq T} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) - a(Y_{i,t}, f_t(\beta), \beta) \right] \right\| \leq \delta \right\}.$$

Lemma B.2: *Under conditions (1) (i)-(iii), (vii)-(viii) and (4) of Lemma A.1:* $P[\Omega_{2,n,T}(\delta)] \to 1$ *as* $n, T \to \infty$, *for any* $\delta > 0$.

Since $P_{2,\varepsilon} = 1 - P[\Omega_{2,n,T}(L/\varepsilon)]$, by Lemma B.2 we get that $P_{2,\varepsilon} \to 0$ as $n, T \to \infty$, for any $\varepsilon > 0$. It follows $\sup\limits_{\beta \in \mathcal{B}} |J_{3,n,T}(\beta)| = o_p(1)$.

B.2 Proof of Lemma A.2

Let us first show that $\frac{1}{n} \sum\limits_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{U}$ for any $1 \leq t \leq T$ and $\beta \in \mathcal{B}$, with probability approaching 1. We need the following Lemma B.3.

Lemma B.3: *Under conditions (1) (ii) and (ix) of Lemma A.2, for any $\eta > 0$ there exists a compact set $\mathcal{K} \subset \mathcal{U}$ such that $P[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}] \geq 1 - \eta$.*

Now, let $\eta > 0$ be given and, by Lemma B.3, let $\mathcal{K}_1$ be a compact subset of $\mathcal{U}$ such that $P[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_1] \geq 1 - \eta$. Let further $\delta > 0$ be such that $\{x \in \mathbb{SR}^{r\times r} : dist(x, \mathcal{K}_1) \leq \delta\} \subset \mathcal{U}$, where $\mathbb{SR}^{r\times r}$ is the subset of (r, r) symmetric matrices. Then:

$$\begin{aligned} P_{n,T} &:= P\left[\left\{\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta), 1 \leq t \leq T, \beta \in \mathcal{B}\right\} \subset \mathcal{U}\right] \\ &\geq P[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_1] + P[\Omega_{1,n,T}(\delta)] - 1 \geq P[\Omega_{1,n,T}(\delta)] - \eta. \end{aligned}$$

From Lemma B.1, it follows that $\limsup_{n,T\to\infty} P_{n,T} \geq 1 - \eta$. Since η can be chosen arbitrarily small, we get $\lim_{n,T\to\infty} P_{n,T} = 1$. Therefore we can focus on the event $\left\{\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta), 1 \leq t \leq T, \beta \in \mathcal{B}\right\} \subset \mathcal{U}$.

Let us now prove Lemma A.2. Let $\varepsilon, \eta > 0$ be given. We have to prove that:

$$\limsup_{n,T\to\infty} P\left[\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - E_0[\varphi(\mu_t(\beta))]\right| \geq \varepsilon\right] \leq \eta. \quad \text{(B.2)}$$

Let us introduce an approximation for φ that is globally Lipshitz. More precisely, let $\mathcal{K} \subset \mathcal{U}$ be a compact set and let $\tilde{\varphi}$ be a Lipshitz function on $\mathcal{U}$ such that

$$\tilde{\varphi} = \varphi \text{ on } \mathcal{K} \text{ and } |\tilde{\varphi}| \leq |\varphi| \text{ on } \mathcal{U}. \quad \text{(B.3)}$$

Such a function exists by condition (2) (i). Then (B.2) follows if:

$$A_{1,\varepsilon} = \limsup_{n,T\to\infty} P\left[\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - E_0[\tilde{\varphi}(\mu_t(\beta))]\right| \geq \varepsilon/3\right] \leq \eta/2,$$

$$\begin{aligned} & A_{2,\varepsilon} \\ = \; & \limsup_{n,T\to\infty} P\left[\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - \tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right)\right]\right| \geq \varepsilon/3\right] \\ & \leq \eta/2, \end{aligned} \quad \text{(B.4)}$$

and:

$$\sup_{\beta\in\mathcal{B}} \left|E_0\left[\tilde{\varphi}\left(\mu_t(\beta)\right)\right] - E_0\left[\varphi\left(\mu_t(\beta)\right)\right]\right| \leq \varepsilon/3. \tag{B.5}$$

From (B.3) and conditions (1) (ii) and (2) (ii), function $\tilde{\varphi}$ is Lipshitz and such that $E_0\left[|\tilde{\varphi}(\mu_t(\beta))|^\delta\right] < \infty$ for some $\delta > 2$. Thus we get $A_{1,\varepsilon} = 0$ by Lemma A.1. Let us now show that the inequalities (B.4) and (B.5) hold for an appropriate choice of the approximating function $\tilde{\varphi}$.

Let us first consider $A_{2,\varepsilon}$. Since $\tilde{\varphi} = \varphi$ on $\mathcal{K}$, in the event that defines $A_{2,\varepsilon}$ only the terms with $\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{K}^c$ contribute to the sum. Let $\mathcal{K}_0 \subset \mathcal{K}$ be a compact set and $\delta > 0$ such that:

$$dist\left(\mathcal{K}_0, \mathcal{K}^c\right) > 2\delta. \tag{B.6}$$

Define for $\delta > 0$ the events:

$$\Omega_{3,n,T}(\delta) = \left\{\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T} \frac{1}{\lambda_t(\beta)} \left\| \frac{1}{n}\sum_{i=1}^{n} \left[a(Y_{i,t}, f_t(\beta), \beta) - \mu_t(\beta)\right] \right\| \leq \delta \right\}.$$

and:

$$\Omega_{4,n,T}(\delta) = \left\{\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T} \frac{1}{\lambda_t(\beta)} \left\| \frac{1}{n}\sum_{i=1}^{n} \left[a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) - a(Y_{i,t}, f_t(\beta), \beta)\right] \right\| \leq \delta \right\}.$$

Lemma B.4: *Under conditions (1) (i)-(iii), (v)-(vi), (ix)-(x) and (4) of Lemma A.2:* $P\left[\Omega_{3,n,T}(\delta)\right] \to 1$ *as* $n, T \to \infty$, *for any* $\delta > 0$.

Lemma B.5: *Under conditions (1) (i)-(iii), (vii)-(x) and (4) of Lemma A.2:* $P\left[\Omega_{4,n,T}(\delta)\right] \to 1$ *as* $n, T \to \infty$, *for any* $\delta > 0$.

When $\Omega_{n,T}(\delta) = \cap_{j=1}^{4}\Omega_{j,n,T}(\delta)$ occurs, we have that $\|A - D\| \leq 2\delta$ and $A = (1+\Delta)\, D$ with $\|\Delta\| \leq 2\delta$, P-a.s., for $A = \frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)$ and $D = \mu_t(\beta)$, for

$t = 1, ..., T$ and $\beta \in \mathcal{B}$. Then, by (B.3) and condition (2) (ii) we have:

$$\begin{aligned}\left|\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - \tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right)\right| &\leq 2\left|\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right)\right| \\ &\leq C\left\|\mu_t(\beta)\right\|^{\tau}\psi\left(\mu_t(\beta)\right),\end{aligned}$$

for a constant C and $\delta < 1/4$, and by (B.6):

$$\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{K}^c \Rightarrow \mu_t(\beta) \in \mathcal{K}_0^c,$$

P-a.s., for any $t = 1, ..., T$ and $\beta \in \mathcal{B}$. Thus, when $\Omega_{n,T}(\delta)$ occurs, we have:

$$\begin{aligned}&\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - \tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right)\right]\right| \\ \leq\ & C\sup_{\beta\in\mathcal{B}}\frac{1}{T}\sum_{t=1}^{T} 1\left\{\mu_t(\beta) \in \mathcal{K}_0^c\right\}\left\|\mu_t(\beta)\right\|^{\tau}\psi\left(\mu_t(\beta)\right) \\ \leq\ & \frac{C}{T}\sum_{t=1}^{T}\left(1 - 1\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_0\right]\right)\sup_{\beta\in\mathcal{B}}\left\|\mu_t(\beta)\right\|^{\tau}\psi\left(\mu_t(\beta)\right).\end{aligned}$$

It follows that:

$$\begin{aligned}&P\left[\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - \tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right)\right]\right| \geq \varepsilon/3\right] \\ \leq\ & \sum_{j=1}^{4} P\left[\Omega_{j,n,T}(\delta)^c\right] \\ & + P\left[\frac{C}{T}\sum_{t=1}^{T}\left(1 - 1\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_0\right]\right)\sup_{\beta\in\mathcal{B}}\left\|\mu_t(\beta)\right\|^{\tau}\psi\left(\mu_t(\beta)\right) \geq \varepsilon/3\right].\end{aligned}$$

From Lemmas B.1-B.2 and B.4-B.5 we get:

$$A_{2,\varepsilon} \leq P\left[\frac{C}{T}\sum_{t=1}^{T}\left(1 - 1\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_0\right]\right)\sup_{\beta\in\mathcal{B}}\left\|\mu_t(\beta)\right\|^{\tau}\psi\left(\mu_t(\beta)\right) \geq \varepsilon/3\right] = P_{2,\varepsilon}.$$

By the Markov inequality, the Minkowsky inequality and the Cauchy-Schwartz inequality, we have:

$$
\begin{aligned}
P_{2,\varepsilon} &\leq \frac{3C}{\varepsilon} E\left[(1-1\left[\{\mu_t(\beta), \beta\in\mathcal{B}\}\subset\mathcal{K}_0\right])\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)\|^{\tau}\psi\left(\mu_t(\beta)\right)\right] \\
&\leq \frac{3C}{\varepsilon}\left(1-P\left[\{\mu_t(\beta), \beta\in\mathcal{B}\}\subset\mathcal{K}_0\right]\right)^{1/p} E\left[\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)\|^{\tau q}\psi\left(\mu_t(\beta)\right)^q\right]^{1/q} \\
&\leq \frac{3C}{\varepsilon}\left(1-P\left[\{\mu_t(\beta), \beta\in\mathcal{B}\}\subset\mathcal{K}_0\right]\right)^{1/p} E\left[\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)\|^{\tau q p'}\right]^{1/(p'q)} \\
&\quad E\left[\sup_{\beta\in\mathcal{B}}\psi\left(\mu_t(\beta)\right)^{qq'}\right]^{1/(qq')},
\end{aligned}
$$

with $p, q, p', q' > 1$ such that $1/p + 1/q = 1$, $1/p' + 1/q' = 1$. Fix $q \in (1, \frac{4}{1+\tau})$ and $p' = 4/(\tau q)$. We get:

$$
A_{2,\varepsilon} \leq \frac{3C}{\varepsilon}\left(1-P\left[\{\mu_t(\beta), \beta\in\mathcal{B}\}\subset\mathcal{K}_0\right]\right)^{1/p} c_1,
$$

where $c_1 = E\left[\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)\|^4\right]^{\tau/4} E\left[\sup_{\beta\in\mathcal{B}}\psi\left(\mu_t(\beta)\right)^4\right]^{1/q-\tau/4} < \infty$ by conditions (1) (ii) and (2) (ii).

Let us now bound $\sup_{\beta\in\mathcal{B}}\left|E_0\left[\tilde{\varphi}\left(\mu_t(\beta)\right)\right] - E_0\left[\varphi\left(\mu_t(\beta)\right)\right]\right|$. By similar arguments as above:

$$
\sup_{\beta\in\mathcal{B}}\left|E_0\left[\tilde{\varphi}\left(\mu_t(\beta)\right)\right] - E_0\left[\varphi\left(\mu_t(\beta)\right)\right]\right| \leq C\left(1-P\left[\{\mu_t(\beta), \beta\in\mathcal{B}\}\subset\mathcal{K}_0\right]\right)^{1/p} c_1.
$$

From Lemma B.3, we can fix $\mathcal{K}_0$, $\mathcal{K}$ and δ such that $P\left[\{\mu_t(\beta), \beta\in\mathcal{B}\}\subset\mathcal{K}_0\right] \geq 1 - \min\left\{\left(\frac{\varepsilon\eta}{6Cc_1}\right)^p, \left(\frac{\varepsilon}{3Cc_1}\right)^p\right\}$ and (B.6) hold. Then (B.4) and (B.5) follow, and the proof is concluded.

B.3 Proof of Corollary A.3

Let us first consider Case (A) and check condition (2) of Lemma A.2. (i) Let $\mathcal{K}\subset\mathcal{U}$ be compact, and let $w, z \in \mathcal{K}$. Using that $w^{-1} - z^{-1} = -z^{-1}(w-z)w^{-1}$ we deduce that φ is Lipshitz on $\mathcal{K}$ with Lipshitz constant $L = \sup_{z\in\mathcal{K}}\left\|z^{-1}\right\|^2 < \infty$. (ii) Let $w, z \in \mathcal{U}$, $w = (1+\Delta)z$, $\|\Delta\| \leq 1/2$. Then $1+\Delta$ is a nonsingular matrix. From

$w^{-1} = z^{-1}(1+\Delta)^{-1}$ we see that condition (2) (ii) is satisfied with $C = \left\|(1+\Delta)^{-1}\right\|$, $\tau = 0$ and $\psi(z) = \|z^{-1}\|$.

Let us now consider Case (B). Use the block notation:

$$x = \begin{pmatrix} x_{11} & x_{12} \\ x_{21} & x_{22} \end{pmatrix}.$$

Then $\varphi(x) = x_{11} - x_{12}x_{22}^{-1}x_{21} = \varphi_1(x) - \varphi_2(x)$, where $\varphi_1(x) = x_{11}$ and $\varphi_2(x) = x_{12}x_{22}^{-1}x_{21}$. The convergence of the transformation with φ_1 follows from Lemma A.1. Let us focus on φ_2 and apply Lemma A.2. Condition (2) (i) is satisfied, since φ_2 is the product of three mappings that are Lipshitz on compact sets. To check condition (2) (ii), let $w, z \in \mathcal{U}$, $w = (1+\Delta)z$, $\|\Delta\| \leq 1/2$. Then:

$$\|\varphi_2(w)\| \leq \|w_{12}\| \left\|w_{22}^{-1}\right\| \|w_{21}\| \leq \|w\|^2 \left\|w_{22}^{-1}\right\| \leq \|1+\Delta\|^2 \|z\|^2 \left\|w_{22}^{-1}\right\|.$$

Denote by $d = r - s$ the dimension of w_{22}. Since w is positive definite, and norms on $\mathbb{R}^{d\times d}$ are equivalent, we have:

$$\begin{aligned} \left\|w_{22}^{-1}\right\| &\leq C \sup_{x\in\mathbb{R}^d:\|x\|=1} x' w_{22}^{-1} x = C\left(\inf_{x\in\mathbb{R}^d:\|x\|=1} x' w_{22} x\right)^{-1} \leq C\left(\inf_{x\in\mathbb{R}^r:\|x\|=1} x' w x\right)^{-1} \\ &= C \sup_{x\in\mathbb{R}^r:\|x\|=1} x' w^{-1} x \leq CC' \left\|w^{-1}\right\|, \end{aligned}$$

where C, C' are constants. From the argument above for the inverse mapping, we know that $\|w^{-1}\| \leq \left\|(1+\Delta)^{-1}\right\| \|z^{-1}\|$. We get that $\|\varphi_2(w)\| \leq CC' \left\|(1+\Delta)^{-1}\right\| \|1+\Delta\|^2 \|z\|^2 \|z^{-1}\|$ and condition (2) (ii) is satisfied with $\tau = 2$ and $\psi(z) = \|z^{-1}\|$.

B.4 Proof of Lemma A.4

The proof consists of several steps. a) First, we prove that $\hat{\beta}$ is consistent. For any $\varepsilon > 0$ we have:

$$\begin{aligned} P\left[|\hat{\beta} - \beta_0| \geq \varepsilon\right] &\leq P\left[\sup_{\beta\in\mathcal{B}:|\beta-\beta_0|\geq\varepsilon} \mathcal{L}_{nT}(\beta,\hat{\theta}) \geq \mathcal{L}_{nT}(\hat{\beta},\hat{\theta})\right] \\ &\leq P\left[\sup_{\beta\in\mathcal{B}:|\beta-\beta_0|\geq\varepsilon} \mathcal{L}_{nT}(\beta,\hat{\theta}) \geq \mathcal{L}_{nT}(\beta_0,\theta_0)\right]. \end{aligned}$$

By using (1i), (1ii) and (5i), we get:

$$P\left[|\hat{\beta}-\beta_0| \geq \varepsilon\right] \leq P\left[\sup_{\beta\in\mathcal{B}:|\beta-\beta_0|\geq\varepsilon} \mathcal{L}^*(\beta)-\mathcal{L}^*(\beta_0) \geq o_p(1)\right].$$

The probability in the RHS is $o(1)$, since $\sup_{\beta\in\mathcal{B}:|\beta-\beta_0|\geq\varepsilon} \mathcal{L}^*(\beta)-\mathcal{L}^*(\beta_0) < 0$ by identification condition (2i) and compactness of calB.

b) Second, we prove that $\hat{\beta}-\beta_0 = O_p\left(\frac{1}{\sqrt{nT}}\right)$. Estimator $\hat{\beta}$ satisfies with probability approaching (w.p.a.) 1 the first-order condition:

$$0 = \frac{\partial \mathcal{L}_{nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right) = \frac{\partial \mathcal{L}^*_{nT}}{\partial \beta}\left(\hat{\beta}\right) + \frac{1}{n}\frac{\partial \mathcal{L}_{1,nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right) + \frac{1}{n^2}\frac{\partial \mathcal{L}_{2,nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right).$$

By using the mean value Theorem, (4ii) and (5ii), we get:

$$0 = \frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} + \frac{\partial \mathcal{L}^*_{nT}\left(\tilde{\beta}\right)}{\partial \beta \partial \beta'}\left(\hat{\beta}-\beta_0\right) + O_p\left(\frac{1}{n}\right),$$

where $\tilde{\beta}$ is a mean value. From (3i) and the consistency of $\hat{\beta}$, matrix $\left[-\frac{\partial \mathcal{L}^*_{nT}\left(\tilde{\beta}\right)}{\partial \beta \partial \beta'}\right]^{-1}$ exists w.p.a. and is $O_p(1)$. Thus, from (4i) and since $T/n \to 0$, we get:

$$\hat{\beta}-\beta_0 = \left[-\frac{\partial \mathcal{L}^*_{nT}\left(\tilde{\beta}\right)}{\partial \beta \partial \beta'}\right]^{-1}\left(\frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} + O_p\left(\frac{1}{n}\right)\right) = O_p\left(\frac{1}{\sqrt{nT}}\right).$$

c) Third, $\hat{\beta}^* - \beta_0 = O_p\left(\frac{1}{\sqrt{nT}}\right)$ follows from similar arguments as above, by setting functions $\mathcal{L}_{1,nT}(\beta,\theta)$ and $\mathcal{L}_{2,nT}(\beta,\theta)$ equal to zero.

d) Fourth, let us show that $\hat{\theta}$ is consistent. For any $\varepsilon > 0$ we have:

$$\begin{aligned} P\left[|\hat{\theta}-\theta_0| \geq \varepsilon\right] &\leq P\left[\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \mathcal{L}_{nT}(\hat{\beta},\theta) \geq \mathcal{L}_{nT}(\hat{\beta},\hat{\theta})\right] \\ &\leq P\left[\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \mathcal{L}_{nT}(\hat{\beta},\theta) \geq \mathcal{L}_{nT}(\beta_0,\theta_0)\right]. \end{aligned}$$

Using (1ii), (5i) and the consistency of $\hat{\beta}$, the RHS probability is such that:

$$\begin{aligned} & P\left[\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \mathcal{L}_{nT}(\hat{\beta},\theta) \geq \mathcal{L}_{nT}(\beta_0,\theta_0)\right] \\ = & P\left[\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \frac{1}{n}\mathcal{L}_{1,nT}(\hat{\beta},\theta) - \frac{1}{n}\mathcal{L}_{1,nT}(\beta_0,\theta_0) \geq \mathcal{L}^*_{nT}\left(\beta_0\right) - \mathcal{L}^*_{nT}\left(\hat{\beta}\right) + O_p\left(\frac{1}{n^2}\right)\right] \\ = & P\left[\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \mathcal{L}_1(\beta_0,\theta) - \mathcal{L}_1(\beta_0,\theta_0) \geq n\left[\mathcal{L}^*_{nT}\left(\beta_0\right) - \mathcal{L}^*_{nT}\left(\hat{\beta}\right)\right] + o_p\left(1\right)\right]. \end{aligned}$$

Thus:

$$P\left[|\hat{\theta}-\theta_0| \geq \varepsilon\right] \leq P\left[\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \mathcal{L}_1(\beta_0,\theta) - \mathcal{L}_1(\beta_0,\theta_0) \geq n\left[\mathcal{L}^*_{nT}\left(\beta_0\right) - \mathcal{L}^*_{nT}\left(\hat{\beta}\right)\right] + o_p\left(1\right)\right].$$

To prove that $n\left[\mathcal{L}^*_{nT}\left(\beta_0\right) - \mathcal{L}^*_{nT}\left(\hat{\beta}\right)\right] = o_p(1)$, we can expand $\mathcal{L}^*_{nT}\left(\beta_0\right)$ and $\mathcal{L}^*_{nT}\left(\hat{\beta}\right)$ around $\hat{\beta}^*$ at order 2 and use that $\dfrac{\partial\mathcal{L}^*_{nT}\left(\hat{\beta}^*\right)}{\partial\beta} = 0$. More precisely, we get:

$$\mathcal{L}^*_{nT}\left(\beta_0\right) = \mathcal{L}^*_{nT}\left(\hat{\beta}^*\right) + \frac{1}{2}\left(\beta_0 - \hat{\beta}^*\right)' \frac{\partial^2\mathcal{L}^*_{nT}\left(\dot{\beta}\right)}{\partial\beta\partial\beta'}\left(\beta_0 - \hat{\beta}^*\right),$$

and:

$$\mathcal{L}^*_{nT}\left(\hat{\beta}\right) = \mathcal{L}^*_{nT}\left(\hat{\beta}^*\right) + \frac{1}{2}\left(\hat{\beta} - \hat{\beta}^*\right)' \frac{\partial^2\mathcal{L}^*_{nT}\left(\ddot{\beta}\right)}{\partial\beta\partial\beta'}\left(\hat{\beta} - \hat{\beta}^*\right),$$

where $\dot{\beta}$ and $\ddot{\beta}$ are mean values. From b) and c) above, we have $\beta_0 - \hat{\beta}^* = O_p\left(1/\sqrt{nT}\right)$ and $\hat{\beta} - \hat{\beta}^* = O_p\left(1/\sqrt{nT}\right)$, and from (3i) we get $n\left[\mathcal{L}^*_{nT}\left(\beta_0\right) - \mathcal{L}^*_{nT}\left(\hat{\beta}\right)\right] = O_p\left(1/T\right) = o_p(1)$. Therefore we get:

$$P\left[|\hat{\theta}-\theta_0| \geq \varepsilon\right] \leq P\left[\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \mathcal{L}_1(\beta_0,\theta) - \mathcal{L}_1(\beta_0,\theta_0) \geq o_p\left(1\right)\right].$$

The RHS probability is $o(1)$, since $\sup_{\theta\in\Theta:|\theta-\theta_0|\geq\varepsilon} \mathcal{L}_1(\beta_0,\theta) - \mathcal{L}_1(\beta_0,\theta_0) < 0$ from (2ii) and the compactness of Θ.

e) Fifth, let us show the asymptotic normality. The first-order conditions for $\hat{\beta}, \hat{\theta}$

are:

$$
\begin{aligned}
0 &= \frac{\partial \mathcal{L}_{nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right) = \frac{\partial \mathcal{L}^*_{nT}}{\partial \beta}\left(\hat{\beta}\right) + \frac{1}{n}\frac{\partial \mathcal{L}_{1,nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right) + \frac{1}{n^2}\frac{\partial \mathcal{L}_{2,nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right),\\
0 &= \frac{\partial \mathcal{L}_{nT}}{\partial \theta}\left(\hat{\beta},\hat{\theta}\right) \Leftrightarrow 0 = \frac{\partial \mathcal{L}_{1,nT}}{\partial \theta}\left(\hat{\beta},\hat{\theta}\right) + \frac{1}{n}\frac{\partial \mathcal{L}_{2,nT}}{\partial \theta}\left(\hat{\beta},\hat{\theta}\right),
\end{aligned}
$$

where a factor $1/n$ in the second equation cancels. Let us multiply the first equation by $\sqrt{nT}$, the second equation by $\sqrt{T}$, and use the mean-value Theorem to get:

$$
\begin{aligned}
0 &= \sqrt{nT}\frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} + \frac{\partial^2 \mathcal{L}^*_{nT}(\tilde{\beta})}{\partial \beta \partial \beta'}\sqrt{nT}\left(\hat{\beta}-\beta_0\right)\\
&\quad + \sqrt{\frac{T}{n}}\frac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial \beta} + \frac{1}{n}\frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \beta \partial \beta'}\sqrt{nT}\left(\hat{\beta}-\beta_0\right) + \frac{1}{\sqrt{n}}\frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \beta \partial \theta'}\sqrt{T}\left(\hat{\theta}-\theta_0\right)\\
&\quad + \frac{1}{n}\sqrt{\frac{T}{n}}\frac{\partial \mathcal{L}_{2,nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right),
\end{aligned}
$$

and:

$$
\begin{aligned}
0 &= \sqrt{T}\frac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial \theta} + \frac{1}{\sqrt{n}}\frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \theta \partial \beta'}\sqrt{nT}\left(\hat{\beta}-\beta_0\right) + \frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \theta \partial \theta'}\sqrt{T}\left(\hat{\theta}-\theta_0\right)\\
&\quad + \frac{\sqrt{T}}{n}\frac{\partial \mathcal{L}_{2,nT}}{\partial \beta}\left(\hat{\beta},\hat{\theta}\right),
\end{aligned}
$$

where $\tilde{\beta}$ and $\tilde{\theta}$ are mean values. Thus, we get from (4ii), (5ii) and $T/n \to 0$:

$$
-\begin{bmatrix} \frac{\partial^2 \mathcal{L}^*_{nT}(\tilde{\beta})}{\partial \beta \partial \beta'} + \frac{1}{n}\frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \beta \partial \beta'} & \frac{1}{\sqrt{n}}\frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \beta \partial \theta'} \\ \frac{1}{\sqrt{n}}\frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \theta \partial \beta'} & \frac{\partial^2 \mathcal{L}_{1,nT}(\tilde{\beta},\tilde{\theta})}{\partial \theta \partial \theta'} \end{bmatrix} \begin{bmatrix} \sqrt{nT}\left(\hat{\beta}-\beta_0\right) \\ \sqrt{T}\left(\hat{\theta}-\theta_0\right) \end{bmatrix} = \begin{bmatrix} \sqrt{nT}\frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} \\ \sqrt{T}\frac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial \theta} \end{bmatrix} + o_p(1).
$$

Then, from (3i), (3ii), (3iii),

$$
\begin{aligned}
\sqrt{nT}\left(\hat{\beta}-\beta_0\right) &= \left(I_0^*\right)^{-1}\sqrt{nT}\frac{\partial \mathcal{L}^*_{nT}(\beta_0)}{\partial \beta} + o_p(1),\\
\sqrt{T}\left(\hat{\theta}-\theta_0\right) &= I_{1,\theta\theta}^{-1}\sqrt{T}\frac{\partial \mathcal{L}_{1,nT}(\beta_0,\theta_0)}{\partial \theta} + o_p(1).
\end{aligned}
$$

The joint asymptotic normality follows from (4i).

f) Finally, the asymptotic expansion of $\hat{\beta}^*$ is the same as that of $\hat{\beta}$. The conclusion follows.

B.5 Proof of Lemma B.1

The proof of Lemma B.1 is obtained by a similar argument as in the proof of Lemma B.4 (Section B.8) replacing $\lambda_t(\beta)$ with 1.

B.6 Proof of Lemma B.2

The proof of Lemma B.2 is obtained by a similar argument as in the proof of Lemma B.5 (Section B.9) replacing $\lambda_t(\beta)$ with 1.

B.7 Proof of Lemma B.3

Let $\lambda(x)$ and $\Lambda(x)$ denote the smallest and respectively the largest eigenvalues of the symmetric matrix $x \in \mathbb{SR}^{r\times r}$, and let $\lambda_t(\beta) = \lambda\left(\mu_t(\beta)\right)$ and $\Lambda_t(\beta) = \Lambda\left(\mu_t(\beta)\right)$. For any constants c, C such that $0 < c \leq C < \infty$ define the compact set $\mathcal{K}_{c,C} = \left\{x \in \mathbb{SR}^{r\times r} : c \leq \lambda(x) \leq \Lambda(x) \leq C\right\} \subset \mathcal{U}$. Then:

$$\begin{aligned}
P\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_{c,C}\right] &= P\left[\inf_{\beta\in\mathcal{B}}\lambda_t(\beta) \geq c,\ \sup_{\beta\in\mathcal{B}}\Lambda_t(\beta) \leq C\right] \\
&\geq 1 - P\left[\inf_{\beta\in\mathcal{B}}\lambda_t(\beta) < c\right] - P\left[\sup_{\beta\in\mathcal{B}}\Lambda_t(\beta) > C\right] \\
&= 1 - P\left[\sup_{\beta\in\mathcal{B}}\lambda_t(\beta)^{-1} > c^{-1}\right] - P\left[\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)\| > C\right] \\
&\geq 1 - cE\left[\sup_{\beta\in\mathcal{B}}\lambda_t(\beta)^{-1}\right] - C^{-1}E\left[\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)\|\right],
\end{aligned}$$

by the Markov inequality. The two expectations in the last line are finite by conditions (1) (ii) and (ix). Then, for any $\eta > 0$, there exist $c > 0$ and $C < \infty$ such that $P\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_{c,C}\right] \geq 1 - \eta$.

B.8 Proof of Lemma B.4

Let us define $W_{n,t}(\beta) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} [a(Y_{i,t}, f_t(\beta), \beta) - \mu_t(\beta)]$. Then:

$$\begin{aligned} P\left[\Omega_{3,n,T}(\delta)^c\right] &\leq P\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \sup_{1 \leq t \leq T} \frac{\|W_{n,t}(\beta)\|}{\lambda_t(\beta)} \geq \delta\right] \leq \sum_{t=1}^{T} P\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{\|W_{n,t}(\beta)\|}{\lambda_t(\beta)} \geq \delta\right] \\ &= TP\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{\|W_{n,t}(\beta)\|}{\lambda_t(\beta)} \geq \delta\right]. \end{aligned}$$

Denote by $W_{j,l,n,t}(\beta)$ the elements of matrix $W_{n,t}(\beta)$. Since $\|W_{n,t}(\beta)\|^2 = \sum_{j,l=1}^{r} |W_{j,l,n,t}(\beta)|^2$, we have:

$$P\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{\|W_{n,t}(\beta)\|}{\lambda_t(\beta)} \geq \delta\right] \leq \sum_{j,l=1}^{r} P\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{r}\right],$$

Thus, we have to show that:

$$TP\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{r}\right] \to 0, \tag{B.7}$$

for any j, l.

To control the sup over $\beta \in \mathcal{B}$, let us introduce a finite covering of the compact set $\mathcal{B} \subset \mathbb{R}^q$ by M open balls $B(\beta_m, \varepsilon)$ around β_m and with radius ε, $m = 1, ..., M$. We let $M = M_T$ and $\varepsilon = \varepsilon_T$ depend on sample size T, such that $\varepsilon_T \to 0$, $M_T \to \infty$ and $M_T = O\left(\varepsilon_T^{-q}\right)$. We have:

$$\begin{aligned} \sup_{\beta \in \mathcal{B}} \frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} &\leq \max_{m=1,..,M_T} \sup_{\beta \in B(\beta_m, \varepsilon_T)} \frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \\ &\leq \max_{m=1,..,M_T} \frac{|W_{j,l,n,t}(\beta_m)|}{\lambda_t(\beta_m)} + \sup_{\beta, \beta' \in \mathcal{B}: \left\|\beta' - \beta\right\| \leq \varepsilon_T} \left| \frac{W_{j,l,n,t}(\beta')}{\lambda_t(\beta')} - \frac{W_{j,l,n,t}(\beta)}{\lambda_t(\beta)} \right|. \end{aligned}$$

Thus we get:

$$P\left[\frac{1}{\sqrt{n}}\sup_{\beta\in\mathcal{B}}\frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)}\geq\frac{\delta}{r}\right] \leq P\left[\frac{1}{\sqrt{n}}\sup_{\beta,\beta':\|\beta'-\beta\|\leq\varepsilon_T}\left|\frac{W_{j,l,n,t}(\beta')}{\lambda_t(\beta')}-\frac{W_{j,l,n,t}(\beta)}{\lambda_t(\beta)}\right|\geq\frac{\delta}{2r}\right] + M_T\sup_{\beta\in\mathcal{B}}P\left[\frac{1}{\sqrt{n}}\frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)}\geq\frac{\delta}{2r}\right] =: A_1+A_2. \tag{B.8}$$

B.8.1 Bound of A_1

By the Markov inequality we have:

$$A_1\leq\frac{2r}{\delta\sqrt{n}}E\left[\sup_{\beta,\beta':\|\beta'-\beta\|\leq\varepsilon_T}\left|\frac{W_{j,l,n,t}(\beta')}{\lambda_t(\beta')}-\frac{W_{j,l,n,t}(\beta)}{\lambda_t(\beta)}\right|\right].$$

To bound the expectation we use:

$$\begin{aligned}\sup_{\|\beta'-\beta\|\leq\varepsilon_T}\left|\frac{W_{j,l,n,t}(\beta')}{\lambda_t(\beta')}-\frac{W_{j,l,n,t}(\beta)}{\lambda_t(\beta)}\right| &\leq \sup_{\beta\in\mathcal{B}}\lambda_t(\beta)^{-1}\sup_{\|\beta'-\beta\|\leq\varepsilon_T}|W_{j,l,n,t}(\beta')-W_{j,l,n,t}(\beta)|\\ &+ \sup_{\beta\in\mathcal{B}}|W_{j,l,n,t}(\beta)|\sup_{\|\beta'-\beta\|\leq\varepsilon_T}|\lambda_t(\beta')^{-1}-\lambda_t(\beta)^{-1}|.\end{aligned}$$

We have:

$$\sup_{\beta\in\mathcal{B}}|W_{j,l,n,t}(\beta)|\leq\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\sup_{\beta\in\mathcal{B}}\|a(Y_{i,t},f_t(\beta),\beta)\|+E\left[\sup_{\beta\in\mathcal{B}}\|a(Y_{i,t},f_t(\beta),\beta)\|\mid\underline{f_t}\right]\right\},$$

and:

$$\begin{aligned}&\sup_{\|\beta'-\beta\|\leq\varepsilon_T}|W_{j,l,n,t}(\beta')-W_{j,l,n,t}(\beta)|\\ \leq\;& \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial a(Y_{i,t},f_t(\beta),\beta)}{\partial\beta'}\right\|+E\left[\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial a(Y_{i,t},f_t(\beta),\beta)}{\partial\beta'}\right\|\mid\underline{f_t}\right]\right\}\varepsilon_T.\end{aligned}$$

Moreover, for any $\beta, \beta' \in \mathcal{B}$ such that $\|\beta' - \beta\| \leq \varepsilon_T$:

$$
\begin{aligned}
& |\lambda_t(\beta')^{-1} - \lambda_t(\beta)^{-1}| \\
= & \left| \sup_{x:\|x\|=1} x'\mu_t(\beta')^{-1}x - \sup_{x:\|x\|=1} x'\mu_t(\beta)^{-1}x \right| \\
\leq & \sup_{x:\|x\|=1} \left| x' \left(\mu_t(\beta')^{-1} - \mu_t(\beta)^{-1} \right) x \right| \\
\leq & \|\mu_t(\beta')^{-1} - \mu_t(\beta)^{-1}\| \leq \sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)^{-1}\|^2 E\left[\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial a\left(Y_{i,t}, f_t(\beta), \beta\right)}{\partial \beta'} \right\| |\underline{f_t} \right] \varepsilon_T.
\end{aligned}
$$

Thus, using $\sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)^{-1}\| \leq C \sup_{\beta \in \mathcal{B}} \lambda_t(\beta)^{-1}$, we get:

$$
\begin{aligned}
A_1 \leq & \frac{4Cr\varepsilon_T}{\delta} E\left[\sup_{\beta \in \mathcal{B}} \lambda_t(\beta)^{-1} E\left[\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial a\left(Y_{i,t}, f_t(\beta), \beta\right)}{\partial \beta} \right\| \mid \underline{f_t} \right] \right] \\
+ & \frac{4Cr\varepsilon_T}{\delta} E\left[E\left[\sup_{\beta \in \mathcal{B}} \|a\left(Y_{i,t}, f_t(\beta), \beta\right)\| \mid \underline{f_t} \right] \sup_{\beta \in \mathcal{B}} \lambda_t(\beta)^{-2} E\left[\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial a\left(Y_{i,t}, f_t(\beta), \beta\right)}{\partial \beta'} \right\| |\underline{f_t} \right] \right] \\
\leq & \frac{4Crc_1\varepsilon_T}{\delta},
\end{aligned}
$$

where:

$$
\begin{aligned}
c_1 = & E\left[\sup_{\beta \in \mathcal{B}} \lambda_t(\beta)^{-2} \right]^{1/2} E\left[\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial a\left(Y_{i,t}, f_t(\beta), \beta\right)}{\partial \beta} \right\|^2 \right]^{1/2} \\
+ & E\left[\sup_{\beta \in \mathcal{B}} \lambda_t(\beta)^{-4} \right]^{1/2} E\left[\sup_{\beta \in \mathcal{B}} \|a\left(Y_{i,t}, f_t(\beta), \beta\right)\|^4 \right]^{1/4} E\left[\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial a\left(Y_{i,t}, f_t(\beta), \beta\right)}{\partial \beta} \right\|^4 \right]^{1/4} < \infty,
\end{aligned}
$$

by Conditions (1) (ii), (iii), (ix).

B.8.2 Bound of A_2

To bound A_2, write for $\beta \in \mathcal{B}$:

$$
\begin{aligned}
P\left[\frac{1}{\sqrt{n}} \frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{2r} \right] = & E\left[P\left[\frac{1}{\sqrt{n}} |W_{j,l,n,t}(\beta)| \geq \frac{\delta}{2r} \lambda_t(\beta) \mid (f_t) \right] \right] \\
= & E\left[P\left[\left| \sum_{i=1}^{n} [a_{j,l}(Y_{i,t}, f_t(\beta), \beta) - \mu_{j,l,t}(\beta)] \right| \geq \frac{\delta \lambda_t(\beta)}{2r} n \mid (f_t) \right] \right],
\end{aligned}
$$

where $a_{j,l}$ and $\mu_{j,l,t}$ denote the elements of matrices a and μ_t, respectively. To bound the inner conditional probability, we use the independence property of the $Y_{i,t}$ conditional on (f_t), and the Bernstein's inequality [e.g., Bosq (1998), Theorem 1.2]. Let us first check the Cramer's conditions. From conditions (v) and (x), we have: [3]

$$E\left[\|a_{j,l}(Y_{i,t}, f_t(\beta), \beta) - \mu_{j,l,t}(\beta)\|^k \,|\underline{f_t}\right] \leq E\left[\|a(Y_{i,t}, f_t(\beta), \beta) - \mu_t(\beta)\|^k \,|\underline{f_t}\right] \leq \gamma_t(\beta)^k k!,$$

for $k = 3, 4, \cdots$. Then, from the Bernstein's inequality applied conditional on (f_t), we get:

$$\begin{aligned} & P\left[\left|\sum_{i=1}^{n} [a_{j,l}(Y_{i,t}, f_t(\beta), \beta) - \mu_{j,l,t}(\beta)]\right| \geq n\frac{\delta}{2r}\lambda_t(\beta) \mid (f_t)\right] \\ \leq & \; 2\exp\left(-\frac{n\frac{\delta^2}{4r^2}\lambda_t(\beta)^2}{4\gamma_t(\beta)^2 + \frac{\delta}{r}\lambda_t(\beta)\gamma_t(\beta)}\right) \qquad \text{(B.9)} \\ \leq & \; 2\exp\left(-c_4 n \frac{\lambda_t(\beta)^2}{\gamma_t(\beta)^2 + \lambda_t(\beta)\gamma_t(\beta)}\right), \end{aligned}$$

P-a.s., where $c_4 = \dfrac{\delta^2}{4r^2}\dfrac{1}{\max\{4, \delta/r\}}$. By using:

$$\frac{\lambda_t(\beta)^2}{\gamma_t(\beta)^2 + \lambda_t(\beta)\gamma_t(\beta)} = \left[\frac{\gamma_t(\beta)}{\lambda_t(\beta)}\left(1 + \frac{\gamma_t(\beta)}{\lambda_t(\beta)}\right)\right]^{-1} \geq \xi_t,$$

for any $\beta \in \mathcal{B}$, we get:

$$\sup_{\beta \in \mathcal{B}} P\left[\frac{1}{\sqrt{n}}\frac{|W_{j,l,n,t}(\beta)|}{\nu_t(\beta)} \geq \frac{\delta}{2r}\right] \leq 2E\left[\exp\left(-c_4 n \xi_t\right)\right].$$

Using conditions (vi) and (x), and condition (4) on the rate of divergence of n and T, we get $E\left[\exp\left(-c_4 n \xi_t\right)\right] \leq C_1 \exp\left(-c_5 n^{\delta}\right) \leq C_1 \exp\left(-c_6 T^{\delta d}\right)$. Thus:

$$A_2 \leq 2C_1 M_T \exp\left(-c_6 T^{\delta d}\right).$$

[3] For the purpose of simplifying the regularity assumptions, the Cramer's conditions are stated in a slightly different form compared to Theorem 1.2 in Bosq (1998). The inequality (B.9) can be proved by the same arguments as in the proof of Theorem 1.2 in Bosq (1998).

B.8.3 Proof of (B.7)

From (B.8) we get:

$$TP\left[\frac{1}{\sqrt{n}}\sup_{\beta\in\mathcal{B}}\frac{|W_{j,l,n,t}(\beta)|}{\nu_t(\beta)}\geq\frac{\delta}{r}\right]\leq\frac{4Crc_1}{\delta}T\varepsilon_T+2C_1TM_T\exp\left(-c_6T^{\delta d}\right).$$

Now chose $\varepsilon_T=T^{-b}$ for $b>1$. Using $M_T=O\left(\varepsilon_T^{-q}\right)=O\left(T^{qb}\right)$, (B.7) follows.

B.9 Proof of Lemma B.5

For any $\eta>0$, if $\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|\hat{f}_{n,t}(\beta)-f_t(\beta)\right\|\leq\eta$ then:

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\left[a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)-a(y_{i,t},f_t(\beta),\beta)\right]\right\|\leq\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\frac{1}{n}\sum_{i=1}^{n}\sup_{f:\|f-f_t(\beta)\|\leq\eta}\left\|\frac{\partial a\left(Y_{i,t},f,\beta\right)}{\partial f'}\right\|\eta.$$

Thus, for any sequence $\eta_T\to 0$ and constant $\eta^*>0$ we get:

$$\begin{aligned}P\left[\Omega_{4,n,T}(\delta)^c\right]\quad\leq\quad & P\left[\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|\hat{f}_{n,t}(\beta)-f_t(\beta)\right\|>\eta_T\right]\\ & +P\left[\eta_T\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\frac{1}{\lambda_t(\beta)}\frac{1}{n}\sum_{i=1}^{n}\sup_{f:\|f-f_t(\beta)\|\leq\eta^*}\left\|\frac{\partial a\left(Y_{i,t},f,\beta\right)}{\partial f'}\right\|>\delta\right].\end{aligned}$$

By denoting $b(Y_{i,t},f_t(\beta),\beta)=\sup_{f:\|f-f_t(\beta)\|\leq\eta^*}\left\|\frac{\partial a\left(Y_{i,t},f,\beta\right)}{\partial f'}\right\|$, $\nu_t(\beta)=E_0\left[b(Y_{i,t},f_t(\beta),\beta)|\underline{f_t}\right]$ and $\varsigma_t=\sup_{\beta\in\mathcal{B}}\frac{1}{\lambda_t(\beta)}\nu_t(\beta)$, we get:

$$\begin{aligned}P\left[\Omega_{4,n,T}(\delta)^c\right]\quad\leq\quad & P\left[\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|\hat{f}_{n,t}(\beta)-f_t(\beta)\right\|>\eta_T\right]\\ +\quad & P\left[\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\frac{1}{\lambda_t(\beta)}\frac{1}{n}\sum_{i=1}^{n}|b(Y_{i,t},f_t(\beta),\beta)-\nu_t(\beta)|\geq\frac{\delta}{2\eta_T}\right]\\ +\quad & P\left[\sup_{1\leq t\leq T}\varsigma_t\geq\frac{\delta}{2\eta_T}\right]=:P_{1,n,T}+P_{2,n,T}+P_{3,T}.\end{aligned}$$

Let $\eta_T=R(logT)^{-1/\bar{\delta}}$, where $R>0$ is a constant and $\bar{\delta}$ is defined in condition (1) (x). Then $P_{1,n,T}=o(1)$ as $n,T\to\infty$ from condition (3). We prove that $P_{2,n,T}=o(1)$ and

$P_{3,T} = o(1)$ in Sections B.9.1 and B.9.2, respectively. Then the conclusion follows.

B.9.1 Proof that $P_{2,n,T} = o(1)$

Since $\dfrac{\delta}{2\eta_T} \to \infty$, we have:

$$P_{2,n,T} \leq P\left[\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\frac{1}{\lambda_t(\beta)}\frac{1}{n}\sum_{i=1}^{n}|b(Y_{i,t}, f_t(\beta), \beta) - \nu_t(\beta)| \geq \delta^*\right],$$

for any $\delta^* > 0$ and large T. The RHS probability converges to zero by the same argument as in the proof of Lemma B.4 in Section B.8 and using conditions (1) (i)-(iii), (vii), (ix)-(x).

B.9.2 Proof that $P_{3,T} = o(1)$

We have from condition (1) (x):

$$P_{3,T} \leq TP\left[\varsigma_t \geq \frac{\delta}{2\eta_T}\right] \leq C_3 T \exp\left(-c_4 R^{-\bar{\delta}} logT\right) = C_3 T^{1-c_4 R^{-\bar{\delta}}},$$

for $c_4 = C_4(\delta/2)^{\bar{\delta}}$. Then, for $R < (1/c_4)^{-1/\bar{\delta}}$ we get $P_{3,T} = o(1)$.