SUPPLEMENTARY MATERIAL FOR

Bootstrapping two-stage quasi-maximum likelihood estimators of time series models

Sílvia Gonçalves, Ulrich Hounyo, Andrew J. Patton, and Kevin Sheppard

This Version: December 12, 2020

This supplementary appendix is organized as follows. First, we provide a set of definitions useful to understand our assumptions. Next, we provide asymptotic theory and bootstrap theory for general two-stage M estimators under a set of high level conditions (which include uniform laws of large numbers, central limit theorems and an asymptotic linear representation for $\hat{\alpha}_n$ and $\hat{\alpha}_n^*$). Next, we provide proof of Theorem 4.1 appearing in Section 4.1 of the main paper. Then, we provide asymptotic theory and bootstrap theory for general two-stage GMM estimators under a set of high level conditions. Finally, we provide two auxiliary lemmas used in the proof of Theorem 4.4, followed by the proofs of Theorems 4.3 and 4.4.

S1 Definitions

In the following and throughout the appendix, K denotes a constant, which may change from line to line and from (in)equality to (in)equality.

Definition 1. *We define $\{X_t\}$ to be L_q-NED on a mixing process $\{V_t\}$ if $E\left(X_t^q\right) < \infty$ and $v_k \equiv \sup_t \left\|X_t - E_{t-k}^{t+k}\left(X_t\right)\right\|_q \to 0$ as $k \to \infty$. Here, $\left\|X_t\right\|_p \equiv \left(E\left|X_t\right|^p\right)^{1/p}$ is the L_p norm and $E_{t-k}^{t+k}\left(\cdot\right) \equiv E\left(\cdot|\mathcal{F}_{t-k}^{t+k}\right)$, where $\mathcal{F}_{t-k}^{t+k} \equiv \sigma\left(V_{t-k}, \ldots, V_{t+k}\right)$ is the σ-field generated by $V_{t-k}, \ldots, V_{t+k}$. If $v_k = O\left(k^{-a-\delta}\right)$ for some $\delta > 0$, we say $\{X_t\}$ is L_q-NED of size $-a$.*

Definition 2. *$\{V_t\}$ is strong mixing if*

$$\lambda_k \equiv \sup_m \sup_{\{A\in\mathcal{F}_{-\infty}^m, B\in\mathcal{F}_{m+k}^\infty\}} |P(A\cap B) - P(A)P(B)| \to 0$$

as $k\to\infty$ suitably fast.

Definition 3. *A random function $f : X\times\Theta \to R$ is Lipschitz continuous on Θ a.s.-P if for all θ_1 and $\theta_2 \in \Theta$, $|f_t(x,\theta_1) - f_t(x,\theta_2)| \le L_t(x)|\theta_1 - \theta_2|$ for all x in a set with probability one, for some function $L_t(x)$ such that $\sup_n \{n^{-1}\sum_{t=1}^n E(L_t(x))\} = O(1)$.*

Definition 4. *A sequence of random functions $\{f_t : \mathcal{X}\times\Theta \to \mathbb{R}\}$ is $r-$dominated on Θ uniformly in t if there exists $D_t : \mathcal{X}\to\mathbb{R}$ such that $|f_t(x,\theta)| \le D_t(x)$ for all $\theta\in\Theta$ and D_t is measurable such that $\|D_t\|_r \le \Delta < \infty$ for all t.*

Definition 5. *A sequence of random functions $\{f_t : \mathcal{X}\times\Theta \to \mathbb{R}\}$ is L_q-NED on $\{V_t\}$ of size $-a$ on (Θ,ρ) if for each $\theta_0\in\Theta$ there exists $\delta_0 > 0$ such that the random sequences $\left\{\bar{f}_t(\delta) = \sup_{\eta^0(\delta)} f_t(x,\theta)\right\}$ and $\left\{\underline{f}_t(\delta) = \inf_{\eta^0(\delta)} f_t(x,\theta)\right\}$ are L_q-NED on $\{V_t\}$ of size $-a$ for all $0<\delta\le\delta_0$, where $\eta^0(\delta) = \{\theta\in\Theta : \rho(\theta,\theta_0) < \delta\}$.*

S2 General results for two-step M-estimators

In this section, we provide results for a general two-step M estimator $\hat{\beta}_n$ based on a first step estimator $\hat{\alpha}_n$ which has an asymptotic linear representation. Specifically, in the first step, we estimate $\alpha_0 \in \mathcal{A}\subset\mathbb{R}^k$ with some asymptotically linear estimator $\hat{\alpha}_n$ (which does not need to be an M estimator; e.g. it could be a GMM estimator). In the second step, we estimate β_0 with

$$\hat{\beta}_n = \arg\min_{\beta\in\mathcal{B}} Q_{2n}(\hat{\alpha}_n,\beta),$$

where

$$Q_{2n}(\hat{\alpha}_n, \beta) \equiv n^{-1} \sum_{t=1}^{n} q_{2t}\left(X^t, \hat{\alpha}_n, \beta\right),$$

and $q_{2t} : \mathbb{R}^{lt} \times \mathcal{A} \times \mathcal{B} \to \mathbb{R}$ is an objective function that depends on β and α and $X^t \equiv (X_1, \ldots, X_{t-1}, X_t)$. The two-step QMLE of Section 3 is a special case of $\hat{\beta}_n$ when $q_{2t}(X^t, \hat{\alpha}_n, \beta) = -\log f_{2t}(X^t, \hat{\alpha}_n, \beta)$, where f_{2t} denotes the conditional likelihood function of X_t given X^{t-1}, and $\hat{\alpha}_n$ is also a QMLE.

We follow White (1994) and Wooldridge (1994) and provide a set of high level conditions that allow us to derive general results.

Assumption $\mathcal{A}$.

$\mathcal{A}.1$ Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. The observed data are a realization of a stochastic process $\left\{X_t : \Omega \to \mathbb{R}^l,\, t \in \mathbb{N}\right\}$.

$\mathcal{A}.2$ The functions $\{q_{2t}(X^t, \alpha, \beta)\}$ are such that $q_{2t}(\cdot, \alpha, \beta)$ is measurable for each $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are compact subsets of $\mathbb{R}^k$ and $\mathbb{R}^p$, respectively, and $q_{2t}(x^t, \cdot, \cdot)$ is continuous on $\mathcal{A} \times \mathcal{B}$ for all x^t in some set F_t with $P(F_t) = 1$.

$\mathcal{A}.3$ (i) $\hat{\alpha}_n \to^P \alpha_0 \in int(\mathcal{A})$.

(ii) $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = n^{-1/2} \sum_{t=1}^{n} \psi_t(X^t, \alpha_0) + o_P(1)$, for some function $\{\psi_t(X^t, \alpha_0)\}$ such that $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_P(1)$.

$\mathcal{A}.4$ (i) $\bar{Q}_2(\alpha, \beta) \equiv \lim_{n\to\infty} E(Q_{2n}(\alpha, \beta))$ exists and is continuous on $\mathcal{A} \times \mathcal{B}$.

(ii) β_0 is the unique minimizer of $\bar{Q}_2(\alpha_0, \beta) \equiv \lim_{n\to\infty} E(Q_{2n}(\alpha_0, \beta))$ on $\mathcal{B}$.

(iii) $\beta_0 \in int(\mathcal{B})$.

$\mathcal{A}.5$ $\{q_{2t}(X^t, \alpha, \beta)\}$ satisfies a weak ULLN on $\mathcal{A} \times \mathcal{B}$ (i.e. $\sup_{\alpha,\beta} \left|Q_{2n}(\alpha, \beta) - \bar{Q}_2(\alpha, \beta)\right| = o_P(1)$).

$\mathcal{A}$.6 (i) $\{q_{2t}(X^t,\alpha,\beta)\}$ is twice continuously differentiable on $int(\mathcal{A})\times int(\mathcal{B})$.

(ii) The functions $\left\{\frac{\partial}{\partial\alpha'}\varphi_{2t}(X^t,\alpha,\beta)\right\}$ and $\left\{\frac{\partial}{\partial\beta'}\varphi_{2t}(X^t,\alpha,\beta)\right\}$ satisfy a weak ULLN on $\mathcal{A}\times\mathcal{B}$, where $\varphi_{2t}(X^t,\alpha,\beta)\equiv\frac{\partial}{\partial\beta}q_{2t}(X^t,\alpha,\beta)$.

$\mathcal{A}$.7 (i) $H_0\equiv\lim_{n\to\infty}E\left(n^{-1}\sum_{t=1}^n\frac{\partial}{\partial\beta'}\varphi_{2t}(X^t,\alpha_0,\beta_0)\right)>0$.

(ii) $F_0\equiv\lim_{n\to\infty}E\left(n^{-1}\sum_{t=1}^n\frac{\partial}{\partial\alpha'}\varphi_{2t}(X^t,\alpha_0,\beta_0)\right)<\infty$.

$\mathcal{A}$.8 The function $\{\varphi_{2t}(X^t,\alpha_0,\beta_0)+F_0\psi_t(X^t,\alpha_0)\}$ satisfies the CLT, i.e.

$$n^{-1/2}\sum_{t=1}^n\left(\varphi_{2t}\left(X^t,\alpha_0,\beta_0\right)+F_0\psi_t\left(X^t,\alpha_0\right)\right)\to^d N(0,J_0),$$

where

$$J_0\equiv\lim_{n\to\infty}Var\left(n^{-1/2}\sum_{t=1}^n\left(\varphi_{2t}\left(X^t,\alpha_0,\beta_0\right)+F_0\psi_t\left(X^t,\alpha_0\right)\right)\right)>0.$$

Assumption $\mathcal{A}$.3(ii) assumes that $\hat{\alpha}_n$ admits an asymptotic linear representation, which includes not only M-estimators but also other estimators such as GMM estimators.

Theorem S2.1. *Under Assumptions $\mathcal{A}$.1, $\mathcal{A}$.2, $\mathcal{A}$.3(i), $\mathcal{A}$.4(i)-(ii) and $\mathcal{A}$.5, $\hat{\beta}_n\to^P\beta_0$.*

Theorem S2.2. *Under Assumptions $\mathcal{A}.1-\mathcal{A}.8$, $\sqrt{n}\left(\hat{\beta}_n-\beta_0\right)\to^d N\left(0,H_0^{-1}J_0H_0^{-1}\right)$.*

Theorems S2.1 and S2.2 are well known in the literature (see e.g. White (1994), Newey and McFadden (1994) and Wooldridge (1994)) and are only given here for completeness, but their proof is omitted for brevity.

Next, we provide a set of general conditions for bootstrap validity. Suppose that the bootstrap two-step M-estimator is defined as

$$\hat{\beta}_n^*=\arg\min_{\beta\in\mathcal{B}}Q_{2n}^*\left(\hat{\alpha}_n^*,\beta\right),$$

where $\hat{\alpha}_n^*$ is the first-step bootstrap analogue of $\hat{\alpha}_n$, and

$$Q_{2n}^*\left(\hat{\alpha}_n^*, \beta\right) \equiv n^{-1} \sum_{t=1}^{n} q_{2t}^*\left(X^{*t}, \hat{\alpha}_n^*, \beta\right),$$

and where for each $\beta \in \mathcal{B}$, we let $q_{2t}^*\left(X^{*t}, \hat{\alpha}_n^*, \beta\right) = q_{2,\tau_t}\left(X^{\tau_t}, \hat{\alpha}_n^*, \beta\right)$ with τ_t denoting a set of indices chosen by the bootstrap. The first step bootstrap estimator $\hat{\alpha}_n^*$ is not necessarily an M-estimator. All we require in Assumption $\mathcal{B}^*$ below is that it has an asymptotic linear representation of the same type as $\hat{\alpha}_n$ but with $\psi_t\left(X^t, \alpha_0\right)$ replaced with $\psi_t^*\left(X^{*t}, \hat{\alpha}_n\right) = \psi_{\tau_t}\left(X^{\tau_t}, \hat{\alpha}_n\right)$. Thus, both $\hat{\alpha}_n^*$ and $\hat{\beta}_n^*$ depend on the same set of bootstrap indices $\{\tau_t\}$.

Assumption $\mathcal{B}^*$

$\mathcal{B}^*$.1 (i) $\hat{\alpha}_n^* - \hat{\alpha}_n \rightarrow^{P^*} 0$, in prob-$P$.

(ii) $\sqrt{n}\left(\hat{\alpha}_n^* - \hat{\alpha}_n\right) = n^{-1/2} \sum_{t=1}^{n} \psi_t^*\left(X^{*t}, \hat{\alpha}_n\right) + o_{P^*}(1)$, in prob-$P$.

$\mathcal{B}^*$.2 The functions $\{q_{2t}^*\left(X^{*t}, \alpha, \beta\right)\}$ satisfy a bootstrap ULLN on $\mathcal{A} \times \mathcal{B}$, i.e.

$$\sup_{\alpha,\beta} \left|Q_{2n}^*\left(\alpha, \beta\right) - Q_{2n}\left(\alpha, \beta\right)\right| \rightarrow^{P^*} 0,$$

in prob-P.

$\mathcal{B}^*$.3 The functions $\left\{\frac{\partial}{\partial \alpha'} \varphi_{2t}^*\left(X^{*t}, \alpha, \beta\right)\right\}$ and $\left\{\frac{\partial}{\partial \beta'} \varphi_{2t}^*\left(X^{*t}, \alpha, \beta\right)\right\}$ satisfy a bootstrap ULLN on $\mathcal{A} \times \mathcal{B}$, where $\varphi_{2t}^*\left(X^{*t}, \alpha, \beta\right) \equiv \frac{\partial}{\partial \beta} q_{2t}^*\left(X^{*t}, \alpha, \beta\right)$.

$\mathcal{B}^*$.4 $n^{-1/2} \sum_{t=1}^{n} \left(\varphi_{2t}^*\left(X^{*t}, \hat{\alpha}_n, \hat{\beta}_n\right) + F_0 \psi_t^*\left(X^{*t}, \hat{\alpha}_n\right)\right) \rightarrow^{d^*} N\left(0, J_0\right)$, in prob-$P$, where

$$J_0 \equiv \lim_{n \to \infty} Var\left(n^{-1/2} \sum_{t=1}^{n} \left(\varphi_{2t}\left(X^t, \alpha_0, \beta_0\right) + F_0 \psi_t\left(X^t, \alpha_0\right)\right)\right) > 0.$$

Assumption $\mathcal{B}^*$ imposes high level conditions on the bootstrap first step estimator and on the bootstrap second step objective function and its derivatives. These conditions can be verified for any particular bootstrap method used to obtain $\hat{\alpha}_n^*$ and $\hat{\beta}_n^*$, where $\hat{\beta}_n^*$ is a QMLE estimator and $\hat{\alpha}_n^*$ is any estimator admitting an asymptotic linear representation (as specified by Assumption $\mathcal{B}^*.2$). We verify these conditions for the two-step QMLE studied in Section 3.

Theorem S2.3. *Suppose Assumptions $\mathcal{A}.1$, $\mathcal{A}.2$, $\mathcal{A}.3$(i), $\mathcal{A}.4$(i)-(ii) and $\mathcal{A}.5$ hold. If in addition Assumptions $\mathcal{B}^*.1$(i) and $\mathcal{B}^*.2$ are satisfied, then $\hat{\beta}_n^* - \hat{\beta}_n \to^{P^*} 0$, in prob-P.*

Theorem S2.4. *Suppose Assumptions $\mathcal{A}.1 - \mathcal{A}.8$ hold. If in addition Assumptions $\mathcal{B}^*.1 - \mathcal{B}^*.4$ are satisfied, then $\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right) \to^{d^*} N\left(0, H_0^{-1} J_0 H_0^{-1}\right)$, in prob-P.*

Theorems S2.2 and S2.4 imply that

$$\sup_{x \in \mathbb{R}^p} \left| P^*\left(\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right) \le x\right) - P\left(\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \le x\right)\right| \to^P 0,$$

as $n \to \infty$, thus justifying the use of the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right)$ as a consistent estimator of the distribution of $\sqrt{n}\left(\hat{\beta}_n - \beta_0\right)$.

S3 Proofs of Theorems S2.3, S2.4 and 4.1

Proof of Theorem S2.3. Let $\tilde{Q}_n(\beta) = Q_{2n}(\hat{\alpha}_n, \beta) = n^{-1}\sum_{t=1}^n q_{2t}\left(X^t, \hat{\alpha}_n, \beta\right)$. We apply Lemma A.2 of GW (2004) with $Q_n(\cdot, \theta) = \tilde{Q}_n(\beta)$. We can easily verify that $\tilde{Q}_n(\beta)$ satisfies the first part of this lemma, implying that $\hat{\beta}_n \to^P \beta_0$. Next, we verify that the function

$$\tilde{Q}_n^*(\beta) = Q_{2n}^*(\hat{\alpha}_n^*, \beta)$$

satisfies the second part of Lemma A.2. First, note that $\hat{\beta}_n^* = \arg\min_\beta \tilde{Q}_n^*(\beta)$, where $\tilde{Q}_n^*(\beta)$ satisfies the measurability and continuity assumptions given in particular Assumptions $\mathcal{A}.2$.

Therefore, the result follows if we show that

$$\sup_{\beta\in\mathcal{B}}\left|\tilde{Q}_n^*\left(\beta\right)-\tilde{Q}_n\left(\beta\right)\right|\to^{P^*}0,\ \text{prob-}P.$$

To see that this is the case, note that

$$\begin{aligned}\sup_{\beta\in\mathcal{B}}\left|\tilde{Q}_n^*\left(\beta\right)-\tilde{Q}_n\left(\beta\right)\right| &= \sup_{\beta\in\mathcal{B}}\left|Q_{2n}^*\left(\hat{\alpha}_n^*,\beta\right)-Q_{2n}\left(\hat{\alpha}_n,\beta\right)\right| \\ &\leq \sup_{\beta\in\mathcal{B}}\left|Q_{2n}^*\left(\hat{\alpha}_n^*,\beta\right)-Q_{2n}\left(\hat{\alpha}_n^*,\beta\right)\right|+\sup_{\beta\in\mathcal{B}}\left|Q_{2n}\left(\hat{\alpha}_n^*,\beta\right)-\bar{Q}_2\left(\hat{\alpha}_n^*,\beta\right)\right| \\ &\quad+\sup_{\beta\in\mathcal{B}}\left|Q_{2n}\left(\hat{\alpha}_n,\beta\right)-\bar{Q}_2\left(\hat{\alpha}_n,\beta\right)\right|+\sup_{\beta\in\mathcal{B}}\left|\bar{Q}_2\left(\hat{\alpha}_n^*,\beta\right)-\bar{Q}_2\left(\hat{\alpha}_n,\beta\right)\right| \\ &\leq \sup_{\alpha\in\mathcal{A},\beta\in\mathcal{B}}\left|Q_{2n}^*\left(\alpha,\beta\right)-Q_{2n}\left(\alpha,\beta\right)\right|+2\sup_{\alpha\in\mathcal{A},\beta\in\mathcal{B}}\left|Q_{2n}\left(\alpha,\beta\right)-\bar{Q}_2\left(\alpha,\beta\right)\right| \\ &\quad+\sup_{\beta\in\mathcal{B}}\left|\bar{Q}_2\left(\hat{\alpha}_n^*,\beta\right)-\bar{Q}_2\left(\hat{\alpha}_n,\beta\right)\right|. \qquad \text{(S3.1)}\end{aligned}$$

The first two terms are $o_{P^*}(1)$ and $o_P(1)$, respectively, given $\mathcal{B}^*.2$ and $\mathcal{A}.5$. The third term is $o_{P^*}(1)$ in prob-P, given the fact that $\bar{Q}_2(\alpha,\beta)$ is continuous on $\mathcal{A}\times\mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are compact subsets of finite dimensional Euclidean spaces, and the fact that $\hat{\alpha}_n^*-\hat{\alpha}_n\to^{P^*}0$, in prob-$P$ by Assumption $\mathcal{B}^*.1$.

Proof of Theorem S2.4. By a mean value expansion of $n^{-1/2}\sum_{t=1}^n\varphi_{2t}^*\left(X^{*t},\hat{\alpha}_n^*,\hat{\beta}_n^*\right)$ around $\hat{\beta}_n$,

$$0=n^{-1/2}\sum_{t=1}^n\varphi_{2t}^*\left(X^{*t},\hat{\alpha}_n^*,\hat{\beta}_n\right)+\left[n^{-1}\sum_{t=1}^n\frac{\partial}{\partial\beta'}\varphi_{2t}^*\left(X^{*t},\hat{\alpha}_n^*,\ddot{\beta}_n^*\right)\right]\sqrt{n}\left(\hat{\beta}_n^*-\hat{\beta}_n\right),$$

where $\ddot{\beta}_n^*$ lies between $\hat{\beta}_n^*$ and $\hat{\beta}_n$. A second mean value expansion of $n^{-1/2}\sum_{t=1}^n\varphi_{2t}^*\left(X^{*t},\hat{\alpha}_n^*,\hat{\beta}_n\right)$

around $\hat{\alpha}_n$ yields

$$\begin{aligned} 0 &= n^{-1/2}\sum_{t=1}^{n}\varphi_{2t}^{*}\left(X^{*t},\hat{\alpha}_n,\hat{\beta}_n\right)+\left[n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}\varphi_{2t}^{*}\left(X^{*t},\ddot{\alpha}_n^{*},\hat{\beta}_n\right)\right]\sqrt{n}\left(\hat{\alpha}_n^{*}-\hat{\alpha}_n\right) \\ &\quad+\left[n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\beta'}\varphi_{2t}^{*}\left(X^{*t},\hat{\alpha}_n^{*},\ddot{\beta}_n^{*}\right)\right]\sqrt{n}\left(\hat{\beta}_n^{*}-\hat{\beta}_n\right), \end{aligned}$$

where $\ddot{\alpha}_n^{*}$ lies between $\hat{\alpha}_n^{*}$ and $\hat{\alpha}_n$. By a ULLN applied to $\frac{\partial}{\partial\alpha'}\varphi_{2t}^{*}\left(X^{*t},\alpha,\beta\right)$ and $\frac{\partial}{\partial\beta'}\varphi_{2t}^{*}\left(X^{*t},\alpha,\beta\right)$ (Assumption $\mathcal{B}^{*}$.3), we have that

$$n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}\varphi_{2t}^{*}\left(X^{*t},\ddot{\alpha}_n^{*},\hat{\beta}_n\right)-n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}\varphi_{2t}\left(X^{t},\alpha_0,\beta_0\right)\to^{P^{*}}0,\text{ in prob-}P,$$

which implies that

$$n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}\varphi_{2t}^{*}\left(X^{*t},\ddot{\alpha}_n^{*},\hat{\beta}_n\right)\to^{P^{*}}F_0,\text{ in prob-}P,$$

since $\hat{\alpha}_n^{*}\to^{P^{*}}\alpha_0$, $\hat{\beta}_n\to^{P}\beta_0$, and $n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}\varphi_{2t}\left(X^{t},\alpha_0,\beta_0\right)\to^{P}F_0$. Similarly,

$$n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\beta'}\varphi_{2t}^{*}\left(X^{*t},\hat{\alpha}_n^{*},\ddot{\beta}_n^{*}\right)\to^{P^{*}}H_0,\ \text{in prob-}P,$$

since $\hat{\alpha}_n^{*}\to^{P^{*}}\alpha_0$ and $\hat{\beta}_n^{*}\to^{P^{*}}\beta_0$. It follows that

$$0=n^{-1/2}\sum_{t=1}^{n}\varphi_{2t}^{*}\left(X^{*t},\hat{\alpha}_n,\hat{\beta}_n\right)+F_0\sqrt{n}\left(\hat{\alpha}_n^{*}-\hat{\alpha}_n\right)+H_0\sqrt{n}\left(\hat{\beta}_n^{*}-\hat{\beta}_n\right)+o_{P^{*}}(1).$$

By Assumption $\mathcal{B}^{*}$.1(ii),

$$\sqrt{n}\left(\hat{\alpha}_n^{*}-\hat{\alpha}_n\right)=n^{-1/2}\sum_{t=1}^{n}\psi_t^{*}\left(X^{*t},\hat{\alpha}_n\right)+o_{P^{*}}(1),$$

which implies that

$$0 = n^{-1/2} \sum_{t=1}^{n} \varphi_{2t}^{*} \left(X^{*t}, \hat{\alpha}_n, \hat{\beta}_n \right) + F_0 \left(n^{-1/2} \sum_{t=1}^{n} \psi_t^{*} \left(X^{*t}, \hat{\alpha}_n \right) \right) + H_0 \sqrt{n} \left(\hat{\beta}_n^{*} - \hat{\beta}_n \right) + o_{P^*} (1) .$$

Hence,

$$\sqrt{n} \left(\hat{\beta}_n^{*} - \hat{\beta}_n \right) = -H_0^{-1} n^{-1/2} \sum_{t=1}^{n} \left(\varphi_{2t}^{*} \left(X^{*t}, \hat{\alpha}_n, \hat{\beta}_n \right) + F_0 \psi_t^{*} \left(X^{*t}, \hat{\alpha}_n \right) \right) + o_{P^*} (1) .$$

The result now follows from Assumption $\mathcal{B}^*$.4.

Proof of Theorem 4.1. We verify that the high level conditions of Theorem S2.4 are satisfied for the two-step QMLE under Assumption A as strengthened by Assumption B. In particular, we can show that Assumption $\mathcal{B}^*$.1(i) is satisfied for $\hat{\alpha}_n^{*} = \arg\max_{\alpha} Q_{1n}^{*} (\alpha) \equiv n^{-1} \sum_{t=1}^{n} \log f_{1t}^{*} (X^{*t}, \alpha)$ by relying on GW (2004)'s Theorem 2.1 under Assumption A.1., A.6 and part (i) of Assumptions A.2-A.5 and A.7, A.8. Similarly, we can apply Theorem 2.2 of GW (2004) to conclude that $\mathcal{B}^*$.1(ii) is verified with $\psi_t^{*} (X^{*t}, \hat{\alpha}_n) = -A_0^{-1} s_{1t}^{*} (X^{*t}, \hat{\alpha}_n)$. To verify Assumption $\mathcal{B}^*$.2, we let $q_{2t}^{*} (X^{*t}, \alpha, \beta) = -\log f_{2t}^{*} (X^{*t}, \alpha, \beta)$ and apply Lemmas A.4 and A.5 of GW (2004). Assumptions A.4(ii) and A.5(ii) together with the requirement that $\ell_n = o(n)$ suffice to prove that $\mathcal{B}^*$.2 holds. $\mathcal{B}^*$.3 can be verified similarly by showing that a bootstrap ULLN applies to the derivatives of $s_{2t}^{*} (X^{*t}, \alpha, \beta)$ with respect to α and β under A.4(ii) and A.5(ii) and the rate condition on the block size ℓ_n. Finally, to check that the bootstrap CLT (cf. Assumption $\mathcal{B}^*$.4) holds for $s_t^{*} \left(\hat{\alpha}_n, \hat{\beta}_n \right) \equiv \varphi_{2t}^{*} \left(X^{*t}, \hat{\alpha}_n, \hat{\beta}_n \right) + F_0 \psi_t^{*} (X^{*t}, \hat{\alpha}_n) = -s_{2t}^{*} \left(X^{*t}, \hat{\alpha}_n, \hat{\beta}_n \right) + F_0 A_0^{-1} s_{1t}^{*} (X^{*t}, \hat{\alpha}_n)$ we proceed as in the proof of Theorem 2.2 of GW (2004). Specifically, we write

$$-n^{-1/2} \sum_{t=1}^{n} s_t^{*} \left(\hat{\alpha}_n, \hat{\beta}_n \right) = n^{-1/2} \sum_{t=1}^{n} \left(s_{2t}^{*} \left(\hat{\alpha}_n, \hat{\beta}_n \right) - F_0 A_0^{-1} s_{1t}^{*} (\hat{\alpha}_n) \right) \equiv \xi_{1n} + \xi_{2n} + \xi_{3n} + \xi_{4n},$$

with

$$
\begin{aligned}
\xi_{1n} &= n^{-1/2}\sum_{t=1}^{n}\left(\left(s_{2t}^{*}\left(\alpha_0,\beta_0\right)-F_0A_0^{-1}s_{1t}^{*}\left(\alpha_0\right)\right)-\left(s_{2t}\left(\alpha_0,\beta_0\right)-F_0A_0^{-1}s_{1t}\left(\alpha_0\right)\right)\right);\\
\xi_{2n} &= n^{-1/2}\sum_{t=1}^{n}\left(s_{2t}\left(\hat{\alpha}_n,\hat{\beta}_n\right)-s_{2t}\left(\alpha_0,\beta_0\right)\right)-F_0A_0^{-1}n^{-1/2}\sum_{t=1}^{n}\left(s_{1t}\left(\hat{\alpha}_n\right)-s_{1t}\left(\alpha_0\right)\right);\\
\xi_{3n} &= n^{-1/2}\sum_{t=1}^{n}\left(s_{2t}^{*}\left(\hat{\alpha}_n,\hat{\beta}_n\right)-s_{2t}^{*}\left(\alpha_0,\beta_0\right)\right)-F_0A_0^{-1}n^{-1/2}\sum_{t=1}^{n}\left(s_{1t}^{*}\left(\hat{\alpha}_n\right)-s_{1t}^{*}\left(\alpha_0\right)\right);\\
\xi_{4n} &= n^{-1/2}\sum_{t=1}^{n}s_{2t}\left(\hat{\alpha}_n,\hat{\beta}_n\right)-F_0A_0^{-1}n^{-1/2}\sum_{t=1}^{n}s_{1t}\left(\hat{\alpha}_n\right).
\end{aligned}
$$

By arguing exactly as in GW (2004), we can show that under Assumption A strengthened by Assumption B, $\xi_{1n}\to^{d^*}N\left(0,J_0\right)$, in prob-$P$, and $\xi_{2n}+\xi_{3n}=o_{P^*}\left(1\right)$ in prob-P, whereas $\xi_{4n}=o_P\left(1\right)$ by the first order conditions that define $\hat{\alpha}_n$ and $\hat{\beta}_n$.

S4 General results for two-step GMM-estimators

In this section, we provide results for a general two-step GMM estimator $\hat{\beta}_n$ based on a first step estimator $\hat{\alpha}_n$ which has an asymptotic linear representation. As in Section S2, in the first step, we estimate $\alpha_0\in\mathcal{A}\subset\mathbb{R}^k$ with some asymptotically linear estimator $\hat{\alpha}_n$ (e.g. it could be an M estimator or a GMM estimator). In the second step, we now estimate β_0 with a GMM estimator define as:

$$
\hat{\beta}_n=\arg\min_{\beta\in\mathcal{B}}Q_{2n}\left(\hat{\alpha}_n,\beta\right),
$$

where

$$
Q_{2n}\left(\hat{\alpha}_n,\beta\right)\equiv\bar{m}_n'\left(\hat{\alpha}_n,\beta\right)W_n\bar{m}_n\left(\hat{\alpha}_n,\beta\right),
$$

such that $\bar{m}_n\left(\hat{\alpha}_n,\beta\right)\equiv n^{-1}\sum_{t=1}^{n}m_{2t}\left(X^t,\hat{\alpha}_n,\beta\right)$, $m_{2t}:\mathbb{R}^{lt}\times\mathcal{A}\times\mathcal{B}\to\mathbb{R}^r$ is an objective function that depends on β and α and $X^t\equiv\left(X_1,\ldots,X_{t-1},X_t\right)$. The weighting matrix

W_n is a $r \times r$ symmetric and positive definite (random) matrix. We make the following assumptions.

Assumption $\mathcal{AG}$.

$\mathcal{AG}.1$ Let $(\Omega, \mathcal{F}, P)$ be a complete probability space. The observed data are a realization of a stochastic process $\left\{X_t : \Omega \to \mathbb{R}^l,\ t \in \mathbb{N}\right\}$.

$\mathcal{AG}.2$ The functions $\{m_{2t}(X^t, \alpha, \beta)\}$ are such that $m_{2t}(\cdot, \alpha, \beta)$ is measurable for each $(\alpha, \beta) \in \mathcal{A} \times \mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are compact subsets of $\mathbb{R}^k$ and $\mathbb{R}^p$, respectively, and $m_{2t}(x^t, \cdot, \cdot)$ is continuous on $\mathcal{A} \times \mathcal{B}$ for all x^t in some set F_t with $P(F_t) = 1$.

$\mathcal{AG}.3$ (i) $\hat{\alpha}_n \to^P \alpha_0 \in int(\mathcal{A})$.

(ii) $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = n^{-1/2} \sum_{t=1}^{n} \psi_t(X^t, \alpha_0) + o_P(1)$, for some function $\{\psi_t(X^t, \alpha_0)\}$ such that $\sqrt{n}(\hat{\alpha}_n - \alpha_0) = O_P(1)$.

$\mathcal{AG}.4$ (i) $\bar{Q}_2(\alpha, \beta) \equiv \lim_{n\to\infty} E(Q_{2n}(\alpha, \beta))$ exists and is continuous on $\mathcal{A} \times \mathcal{B}$.

(ii) β_0 is the unique solution in $\mathcal{B}$ to the equation $E(m_{2t}(X^t, \alpha_0, \beta)) = 0$, and the weighting matrix W_n is such that $W_n \to^P W$, where W a non-random symmetric and positive definite matrix.

(iii) $\beta_0 \in int(\mathcal{B})$.

$\mathcal{AG}.5$ $\sup_{\alpha,\beta} \left|Q_{2n}(\alpha, \beta) - \bar{Q}_2(\alpha, \beta)\right| = o_P(1)$.

$\mathcal{AG}.6$ (i) $\{m_{2t}(X^t, \alpha, \beta)\}$ is continuously differentiable on $int(\mathcal{A}) \times int(\mathcal{B})$.

(ii) The functions $\left\{\frac{\partial}{\partial \alpha'} m_{2t}(X^t, \alpha, \beta)\right\}$ and $\left\{\frac{\partial}{\partial \beta'} m_{2t}(X^t, \alpha, \beta)\right\}$ satisfy a weak ULLN on $\mathcal{A} \times \mathcal{B}$.

$\mathcal{AG}.7$ (i) $\Gamma_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^{n} \frac{\partial}{\partial \beta'} m_{2t}(X^t, \alpha_0, \beta_0)\right)$ is of full rank.

(ii) $\Phi_0 \equiv \lim_{n\to\infty} E\left(n^{-1} \sum_{t=1}^{n} \frac{\partial}{\partial \alpha'} m_{2t}(X^t, \alpha_0, \beta_0)\right) < \infty$.

$\mathcal{AG}.8$ The function $\{m_{2t}(X^t, \alpha_0, \beta_0) + \Phi_0 \psi_t(X^t, \alpha_0)\}$ satisfies the CLT, i.e.

$$n^{-1/2} \sum_{t=1}^{n} \left(m_{2t}\left(X^t, \alpha_0, \beta_0\right) + \Phi_0 \psi_t\left(X^t, \alpha_0\right)\right) \to^d N\left(0, \Upsilon_0\right),$$

where

$$\Upsilon_0 \equiv \lim_{n\to\infty} Var\left(n^{-1/2} \sum_{t=1}^{n} \left(m_{2t}\left(X^t, \alpha_0, \beta_0\right) + \Phi_0 \psi_t\left(X^t, \alpha_0\right)\right)\right) > 0.$$

Theorem S4.1. *Under Assumptions* $\mathcal{AG}.1$, $\mathcal{AG}.2$, $\mathcal{AG}.3$*(i)*, $\mathcal{AG}.4$*(i)-(ii) and* $\mathcal{AG}.5$, $\hat{\beta}_n \to^P \beta_0$.

Theorem S4.2. *Under Assumptions* $\mathcal{AG}.1 - \mathcal{AG}.8$,

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) \to^d N\left(0, \left(\Gamma_0' W \Gamma_0\right)^{-1} \Gamma_0' W \Upsilon_0 W \Gamma_0 \left(\Gamma_0' W \Gamma_0\right)^{-1}\right).$$

Next, we provide a set of general conditions for bootstrap validity. Suppose that the bootstrap two-step GMM-estimator is defined as

$$\hat{\beta}_n^* = \arg\min_{\beta\in\mathcal{B}} Q_{2n}^*\left(\hat{\alpha}_n^*, \beta\right),$$

where $\hat{\alpha}_n^*$ is the first-step bootstrap analogue of $\hat{\alpha}_n$,

$$Q_{2n}^*\left(\hat{\alpha}_n^*, \beta\right) \equiv \bar{m}_n^{*\prime}\left(\hat{\alpha}_n^*, \beta\right) W_n \bar{m}_n^*\left(\hat{\alpha}_n^*, \beta\right),$$

such that $\bar{m}_n^*\left(\hat{\alpha}_n^*, \beta\right) \equiv n^{-1} \sum_{t=1}^{n} m_{2t}^*\left(X^{*t}, \hat{\alpha}_n^*, \beta\right)$, and for each $\beta \in \mathcal{B}$, we let $m_{2t}^*\left(X^{*t}, \hat{\alpha}_n^*, \beta\right) = m_{2,\tau_t}\left(X^{\tau_t}, \hat{\alpha}_n^*, \beta\right) - E^*\left(m_{2,\tau_t}\left(X^{\tau_t}, \hat{\alpha}_n^*, \hat{\beta}_n\right)\right)$ with τ_t denoting a set of indices chosen by the bootstrap. Note that recentering ensures that the bootstrap moment conditions $E^*\left(\bar{m}_n^*\left(\hat{\alpha}_n^*, \hat{\beta}_n\right)\right) = 0$ hold (even when the model is overidentified). For the first step bootstrap estimator $\hat{\alpha}_n^*$, all we require in Assumption $\mathcal{BG}^*$ below is that it has an asymptotic linear representation of the same type as $\hat{\alpha}_n$ but with $\psi_t(X^t, \alpha_0)$ replaced with $\psi_t^*(X^{*t}, \hat{\alpha}_n) = \psi_{\tau_t}(X^{\tau_t}, \hat{\alpha}_n)$. Thus, both $\hat{\alpha}_n^*$ and $\hat{\beta}_n^*$ depend on the same set of bootstrap indices $\{\tau_t\}$.

Assumption $\mathcal{BG}^*$

$\mathcal{BG}^*.1$ (i) $\hat{\alpha}_n^* - \hat{\alpha}_n \to^{P^*} 0$, in prob-$P$.

(ii) $\sqrt{n}(\hat{\alpha}_n^* - \hat{\alpha}_n) = n^{-1/2}\sum_{t=1}^n \psi_t^*(X^{*t}, \hat{\alpha}_n) + o_{P^*}(1)$, in prob-$P$.

$\mathcal{BG}^*.2$ $\sup_{\alpha,\beta} |Q_{2n}^*(\alpha, \beta) - Q_{2n}(\alpha, \beta)| \to^{P^*} 0$, in prob-$P$.

$\mathcal{BG}^*.3$ The functions $\left\{\frac{\partial}{\partial \alpha'} m_{2t}^*(X^{*t}, \alpha, \beta)\right\}$ and $\left\{\frac{\partial}{\partial \beta'} m_{2t}^*(X^{*t}, \alpha, \beta)\right\}$ satisfy a bootstrap ULLN on $\mathcal{A} \times \mathcal{B}$.

$\mathcal{BG}^*.4$ $n^{-1/2}\sum_{t=1}^n \left(m_{2t}^*\left(X^{*t}, \hat{\alpha}_n, \hat{\beta}_n\right) + \Phi_0 \psi_t^*(X^{*t}, \hat{\alpha}_n)\right) \to^{d^*} N(0, \Upsilon_0)$, in prob-$P$, where

$$\Upsilon_0 \equiv \lim_{n\to\infty} Var\left(n^{-1/2}\sum_{t=1}^n \left(m_{2t}\left(X^t, \alpha_0, \beta_0\right) + \Phi_0 \psi_t\left(X^t, \alpha_0\right)\right)\right) > 0.$$

Assumption $\mathcal{BG}^*$ imposes high level conditions on the bootstrap first step estimator and on the bootstrap second step objective function and its derivatives. These conditions can be verified for any particular bootstrap method used to obtain $\hat{\alpha}_n^*$ and $\hat{\beta}_n^*$, where $\hat{\beta}_n^*$ is a GMM estimator and $\hat{\alpha}_n^*$ is any estimator admitting an asymptotic linear representation (as specified by Assumption $\mathcal{BG}^*.2$).

Theorem S4.3. *Suppose Assumptions $\mathcal{AG}.1$, $\mathcal{AG}.2$, $\mathcal{AG}.3$(i), $\mathcal{AG}.4$(i)-(ii) and $\mathcal{AG}.5$ hold. If in addition Assumptions $\mathcal{BG}^*.1$(i) and $\mathcal{B}^*.2$ are satisfied, then $\hat{\beta}_n^* - \hat{\beta}_n \to^{P^*} 0$, in prob-P.*

Theorem S4.4. *Suppose Assumptions $\mathcal{AG}.1 - \mathcal{AG}.8$ hold. If in addition Assumptions $\mathcal{BG}^*.1$–$\mathcal{BG}^*.4$ are satisfied, then $\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right) \to^{d^*} N\left(0, (\Gamma_0' W \Gamma_0)^{-1} \Gamma_0' W \Upsilon_0 W \Gamma_0 (\Gamma_0' W \Gamma_0)^{-1}\right)$, in prob-P.*

Theorems S4.2 and S4.4 imply that

$$\sup_{x\in\mathbb{R}^p}\left|P^*\left(\sqrt{n}\left(\hat{\beta}_n^*-\hat{\beta}_n\right)\le x\right)-P\left(\sqrt{n}\left(\hat{\beta}_n-\beta_0\right)\le x\right)\right|\to^P 0,$$

as $n\to\infty$, thus justifying the use of the bootstrap distribution of $\sqrt{n}\left(\hat{\beta}_n^*-\hat{\beta}_n\right)$ as a consistent estimator of the distribution of $\sqrt{n}\left(\hat{\beta}_n-\beta_0\right)$.

S5 Proofs of Theorems S4.1, S4.2, S4.3 and S4.4

Proof of Theorem S4.1. Under $\mathcal{AG}$.2, $\{Q_{2n}\}$ is a sequence of measurable continuous functions on $\mathcal{A}\times\mathcal{B}$, a.s.-$P$. By $\mathcal{A}$.3(i), $\hat{\alpha}_n\to^P\alpha_0\in\mathcal{A}$ and by $\mathcal{AG}$.5, $\sup_{\alpha,\beta}\left|Q_{2n}(\alpha,\beta)-\bar{Q}_2(\alpha,\beta)\right|=o_P(1)$, where $\bar{Q}_2(\alpha,\beta)$ is continuous by $\mathcal{AG}$.4(i). Hence, we can apply Theorem 3.7 of White (1993) to conclude that $Q_{2n}(\hat{\alpha}_n,\beta)-\bar{Q}_2(\alpha_0,\beta)=o_P(1)$ uniformly on $\mathcal{B}$. Next, let $\tilde{Q}_n(\beta)\equiv Q_{2n}(\hat{\alpha}_n,\beta)$ and note that $\tilde{Q}_n(\beta)$ satisfies the conditions of Theorem 3.4 of White (1993). In particular, $\tilde{Q}_n(\beta)$ converges to $\bar{Q}_2(\alpha_0,\beta)$ uniformly on $\mathcal{B}$, as just showed above. Next, note that Assumption $\mathcal{AG}$.4(ii) ensures that β_0 is the unique minimizer of $\bar{Q}_2(\alpha_0,\beta)$ on $\mathcal{B}$. Thus, it follows that $\hat{\beta}_n-\beta_0=o_P(1)$.

Proof of Theorem S4.2. Given the first order condition of the GMM estimator $\hat{\beta}_n$, we have

$$\left(\frac{\partial}{\partial\beta'}\bar{m}_n\left(\hat{\alpha}_n,\hat{\beta}_n\right)\right)'W_n\sqrt{n}\bar{m}_n\left(\hat{\alpha}_n,\hat{\beta}_n\right)=0,$$

where $\bar{m}_n\left(\hat{\alpha}_n,\hat{\beta}_n\right)=n^{-1}\sum_{t=1}^n m_{2t}\left(X^t,\hat{\alpha}_n,\hat{\beta}_n\right)$. Next, we consider a mean value expansion of $\sqrt{n}\bar{m}_n\left(\hat{\alpha}_n,\hat{\beta}_n\right)$ around β_0. Thus, we have

$$0=\left(\frac{\partial}{\partial\beta'}\bar{m}_n\left(\hat{\alpha}_n,\hat{\beta}_n\right)\right)'W_n\left(\sqrt{n}\bar{m}_n(\hat{\alpha}_n,\beta_0)+\frac{\partial}{\partial\beta'}\bar{m}_n\left(\hat{\alpha}_n,\ddot{\beta}_n\right)\sqrt{n}\left(\hat{\beta}_n-\beta_0\right)\right),$$

where $\ddot{\beta}_n$ lies between $\hat{\beta}_n$ and β_0 and the equality to zero holds with probability approaching one (w.p.a.1.) as $n \to \infty$, because $\hat{\beta}_n$ is interior to $\mathcal{B}$ w.p.a.1. by Assumption $\mathcal{AG}$.4(iii). Another mean value expansion of $\sqrt{n}\bar{m}_n(\hat{\alpha}_n, \beta_0)$ around α_0 yields

$$\sqrt{n}\bar{m}_n(\hat{\alpha}_n, \beta_0) = \sqrt{n}\bar{m}_n(\alpha_0, \beta_0) + \frac{\partial}{\partial \alpha'}\bar{m}_n(\ddot{\alpha}_n, \beta_0)\sqrt{n}(\hat{\alpha}_n - \alpha_0),$$

where $\ddot{\alpha}_n$ lies between $\hat{\alpha}_n$ and α_0, implying that

$$\begin{aligned} 0 \;\; = \;\; & \left(\frac{\partial}{\partial \beta'}\bar{m}_n\left(\hat{\alpha}_n, \hat{\beta}_n\right)\right)' W_n \left(\sqrt{n}\bar{m}_n(\alpha_0, \beta_0) + \frac{\partial}{\partial \alpha'}\bar{m}_n(\ddot{\alpha}_n, \beta_0)\sqrt{n}(\hat{\alpha}_n - \alpha_0)\right) \\ & + \left(\frac{\partial}{\partial \beta'}\bar{m}_n\left(\hat{\alpha}_n, \hat{\beta}_n\right)\right)' W_n \frac{\partial}{\partial \beta'}\bar{m}_n\left(\hat{\alpha}_n, \ddot{\beta}_n\right)\sqrt{n}\left(\hat{\beta}_n - \beta_0\right). \end{aligned}$$

By a ULLN applied to $\frac{\partial}{\partial \alpha'} m_{2t}(X^t, \alpha, \beta)$ and $\frac{\partial}{\partial \beta'} m_{2t}(X^t, \alpha, \beta)$ (Assumption $\mathcal{AG}$.6), we have that

$$\frac{\partial}{\partial \alpha'}\bar{m}_n(\ddot{\alpha}_n, \beta_0) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial \alpha'}m_{2t}\left(X^t, \ddot{\alpha}_n, \beta_0\right) \to^P \lim_{n\to\infty} E\left(n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial \alpha'}m_{2t}\left(X^t, \alpha_0, \beta_0\right)\right) = \Phi_0,$$

$$\frac{\partial}{\partial \beta'}\bar{m}_n\left(\hat{\alpha}_n, \ddot{\beta}_n\right) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial \beta'}m_{2t}\left(X^t, \hat{\alpha}_n, \ddot{\beta}_n\right) \to^P \lim_{n\to\infty} E\left(n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial \beta'}m_{2t}\left(X^t, \alpha_0, \beta_0\right)\right) = \Gamma_0,$$

$$\frac{\partial}{\partial \beta'}\bar{m}_n\left(\hat{\alpha}_n, \hat{\beta}_n\right) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial \beta'}m_{2t}\left(X^t, \hat{\alpha}_n, \hat{\beta}_n\right) \to^P \lim_{n\to\infty} E\left(n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial \beta'}m_{2t}\left(X^t, \alpha_0, \beta_0\right)\right) = \Gamma_0.$$

since $\hat{\alpha}_n \to^P \alpha_0$ and $\hat{\beta}_n \to^P \beta_0$. Furthermore, note that by Assumptions $\mathcal{AG}$.3(ii) and $\mathcal{AG}$.4(ii), we have

$$\sqrt{n}(\hat{\alpha}_n - \alpha_0) = n^{-1/2}\sum_{t=1}^{n}\psi_t\left(X^t, \alpha_0\right) + o_P(1),$$

and $W_n \to^P W$, where W a non-random symmetric and positive definite matrix, respectively. It follows that

$$\Gamma_0' W \Gamma_0 \sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = -\Gamma_0' W n^{-1/2} \sum_{t=1}^{n} \left(m_{2t}\left(X^t, \alpha_0, \beta_0\right) + \Phi_0 \psi_t\left(X^t, \alpha_0\right)\right) + o_P(1).$$

Implying that,

$$\sqrt{n}\left(\hat{\beta}_n - \beta_0\right) = -\left(\Gamma_0' W \Gamma_0\right)^{-1} \Gamma_0' W n^{-1/2} \sum_{t=1}^{n} \left(m_{2t}\left(X^t, \alpha_0, \beta_0\right) + \Phi_0 \psi_t\left(X^t, \alpha_0\right)\right) + o_P(1).$$

Note that because W is positive definite (and hence non singular) and Γ_0 is of full rank (Assumption $\mathcal{AG}$.7(i)), ensure that $\Gamma_0' W \Gamma_0$ is non singular. The result now follows from Assumption $\mathcal{AG}$.8.

Proof of Theorem S4.3. Under our assume conditions, $\hat{\beta}_n - \beta_0 = o_P(1)$, by Theorem S4.1. Given that $\hat{\beta}_n^* - \hat{\beta}_n = \left(\hat{\beta}_n^* - \beta_0\right) - \left(\hat{\beta}_n - \beta_0\right)$, the desired result follows by showing that $\hat{\beta}_n^* - \beta_0 = o_{P^*}(1)$, in prob-$P$. Under Assumption $\mathcal{AG}$.4(ii), observed that β_0 is the unique minimizer of $\bar{Q}_2(\alpha_0, \beta)$ on $\mathcal{B}$. Thus, for any $\varepsilon > 0$ such that $|\beta - \beta_0| > \varepsilon$, there is $\delta > 0$ such that $\bar{Q}_2(\alpha_0, \beta) - \bar{Q}_2(\alpha_0, \beta_0) \geq \delta > 0$. It follows that

$$\begin{aligned} P\left(P^*\left(\left|\hat{\beta}_n^* - \beta_0\right|\right) > \varepsilon\right) \leq & \frac{1}{\varepsilon/2} P\left(2 \sup_{\beta} \left|Q_{2n}(\hat{\alpha}_n, \beta) - \bar{Q}_2(\alpha_0, \beta)\right| > \delta/2\right) \\ & + P\left(P^*\left(2 \sup_{\beta} \left|Q_{2n}^*(\hat{\alpha}_n^*, \beta) - Q_{2n}(\hat{\alpha}_n, \beta)\right| > \delta/2\right) > \varepsilon/2\right), \end{aligned}$$

where we use inequality (B.6) in Dovonon and Goncalves (2017), with $Q_T(\theta) = Q_{2n}(\hat{\alpha}_n, \beta)$, $Q_T^*(\theta) = Q_{2n}^*(\hat{\alpha}_n^*, \beta)$, and $Q(\theta) = \bar{Q}_2(\alpha_0, \beta)$. Therefore, in order to have $\hat{\beta}_n^* - \beta_0 = o_{P^*}(1)$, in prob-$P$, it suffices to show that the following hold: (a) $\sup_\beta \left|Q_{2n}(\hat{\alpha}_n, \beta) - \bar{Q}_2(\alpha_0, \beta)\right| = o_P(1)$ and (b) $\sup_\beta \left|Q_{2n}^*(\hat{\alpha}_n^*, \beta) - Q_{2n}(\hat{\alpha}_n, \beta)\right| = o_{P^*}(1)$, in prob-$P$. (a) by triangular inequality, we have

$$\sup_{\beta \in \mathcal{B}} \left|Q_{2n}(\hat{\alpha}_n, \beta) - \bar{Q}_2(\alpha_0, \beta)\right| \leq \sup_{\alpha \in \mathcal{A}, \beta \in \mathcal{B}} \left|Q_{2n}(\hat{\alpha}_n, \beta) - Q_{2n}(\alpha_0, \beta)\right| + \sup_{\alpha \in \mathcal{A}, \beta \in \mathcal{B}} \left|Q_{2n}(\alpha_0, \beta) - \bar{Q}_2(\alpha_0, \beta)\right|.$$

The first term is $o_P(1)$, given the fact that $Q_{2n}(\alpha,\beta)$ is continuous on $\mathcal{A}\times\mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are compact subsets of finite dimensional Euclidean spaces (by Assumption $\mathcal{AG}$.2), and the fact that $\hat{\alpha}_n - \alpha_0 \to^P 0$, by Assumption $\mathcal{AG}$.3. The second term is $o_P(1)$ given $\mathcal{AG}$.5. To obtain (b), we use similar arguments as in (S3.1). Specifically, note that we can write

$$\begin{aligned}\sup_{\beta\in\mathcal{B}}\left|Q_{2n}^*(\hat{\alpha}_n^*,\beta)-Q_{2n}(\hat{\alpha}_n,\beta)\right| \leq& \sup_{\alpha\in\mathcal{A},\beta\in\mathcal{B}}\left|Q_{2n}^*(\alpha,\beta)-Q_{2n}(\alpha,\beta)\right|+2\sup_{\alpha\in\mathcal{A},\beta\in\mathcal{B}}\left|Q_{2n}(\alpha,\beta)-\bar{Q}_2(\alpha,\beta)\right|\\ &+\sup_{\beta\in\mathcal{B}}\left|\bar{Q}_2(\hat{\alpha}_n^*,\beta)-\bar{Q}_2(\hat{\alpha}_n,\beta)\right|,\end{aligned}$$

where the first two terms are $o_{P^*}(1)$ and $o_P(1)$, respectively, given $\mathcal{BG}^*$.2 and $\mathcal{AG}$.5. The third term is $o_{P^*}(1)$ in prob -P, given the fact that $\bar{Q}_2(\alpha,\beta)$ is continuous on $\mathcal{A}\times\mathcal{B}$, where $\mathcal{A}$ and $\mathcal{B}$ are compact subsets of finite dimensional Euclidean spaces, and the fact that $\hat{\alpha}_n^* - \hat{\alpha}_n \to^{P^*} 0$, in prob-$P$ by Assumption $\mathcal{BG}^*$.1.

Proof of Theorem S4.4. The proof follows closely that of Theorem S4.2. Given the first order condition of the GMM estimator $\hat{\beta}_n^*$, we have

$$\left(\frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*,\hat{\beta}_n^*\right)\right)' W_n\sqrt{n}\bar{m}_n^*\left(\hat{\alpha}_n^*,\hat{\beta}_n^*\right)=0,$$

where $\bar{m}_n^*\left(\hat{\alpha}_n^*,\hat{\beta}_n^*\right)=n^{-1}\sum_{t=1}^n m_{2t}^*\left(X^{*t},\hat{\alpha}_n^*,\hat{\beta}_n^*\right)$. Next, we consider a mean value expansion of $\sqrt{n}\bar{m}_n^*\left(\hat{\alpha}_n^*,\hat{\beta}_n^*\right)$ around $\hat{\beta}_n$. Thus, we have

$$0=\left(\frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*,\hat{\beta}_n^*\right)\right)' W_n\left(\sqrt{n}\bar{m}_n^*\left(\hat{\alpha}_n^*,\hat{\beta}_n\right)+\frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*,\ddot{\beta}_n^*\right)\sqrt{n}\left(\hat{\beta}_n^*-\hat{\beta}_n\right)\right),$$

where $\ddot{\beta}_n^*$ lies between $\hat{\beta}_n^*$ and $\hat{\beta}_n$. A second mean value expansion of $\sqrt{n}\bar{m}_n\left(\hat{\alpha}_n^*,\hat{\beta}_n\right)$ around $\hat{\alpha}_n$ yields

$$\sqrt{n}\bar{m}_n^*\left(\hat{\alpha}_n^*,\hat{\beta}_n\right)=\sqrt{n}\bar{m}_n^*\left(\hat{\alpha}_n,\hat{\beta}_n\right)+\frac{\partial}{\partial\alpha'}\bar{m}_n^*\left(\ddot{\alpha}_n^*,\hat{\beta}_n\right)\sqrt{n}\left(\hat{\alpha}_n^*-\hat{\alpha}_n\right),$$

where $\ddot{\alpha}_n^*$ lies between $\hat{\alpha}_n^*$ and $\hat{\alpha}_n$, implying that

$$\begin{aligned} 0 &= \left(\frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right)\right)' W_n \left(\sqrt{n}\bar{m}_n^*\left(\hat{\alpha}_n, \hat{\beta}_n\right) + \frac{\partial}{\partial\alpha'}\bar{m}_n^*\left(\ddot{\alpha}_n^*, \hat{\beta}_n\right)\sqrt{n}\left(\hat{\alpha}_n - \alpha_0\right)\right) \\ &\quad + \left(\frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right)\right)' W_n \frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*, \ddot{\beta}_n^*\right)\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right). \end{aligned}$$

By a ULLN applied to $\frac{\partial}{\partial\alpha'}m_{2t}^*\left(X^{*t}, \alpha, \beta\right)$ and $\frac{\partial}{\partial\beta'}m_{2t}^*\left(X^{*t}, \alpha, \beta\right)$ (Assumption $\mathcal{BG}^*.3$) , we have that

$$n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}m_{2t}^*\left(X^{*t}, \ddot{\alpha}_n^*, \hat{\beta}_n\right) - n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}m_{2t}\left(X^t, \alpha_0, \beta_0\right) \to^{P^*} 0, \text{ in prob-}P,$$

which implies that

$$\frac{\partial}{\partial\alpha'}\bar{m}_n^*\left(\ddot{\alpha}_n^*, \hat{\beta}_n\right) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}m_{2t}^*\left(X^{*t}, \ddot{\alpha}_n^*, \hat{\beta}_n\right) \to^{P^*} \Phi_0, \text{ in prob-}P,$$

since $\hat{\alpha}_n^* \to^{P^*} \alpha_0$, $\hat{\beta}_n \to^P \beta_0$, and $n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\alpha'}m_{2t}\left(X^t, \alpha_0, \beta_0\right) \to^P \Phi_0$. Similarly,

$$\frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*, \ddot{\beta}_n^*\right) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\beta'}m_{2t}^*\left(X^{*t}, \hat{\alpha}_n^*, \ddot{\beta}_n^*\right) \to^{P^*} \Gamma_0, \text{ in prob-}P,$$

$$\frac{\partial}{\partial\beta'}\bar{m}_n^*\left(\hat{\alpha}_n^*, \hat{\beta}_n^*\right) = n^{-1}\sum_{t=1}^{n}\frac{\partial}{\partial\beta'}m_{2t}^*\left(X^{*t}, \hat{\alpha}_n^*, \hat{\beta}_n^*\right) \to^{P^*} \Gamma_0, \text{ in prob-}P,$$

since $\hat{\alpha}_n^* \to^{P^*} \alpha_0$ and $\hat{\beta}_n^* \to^{P^*} \beta_0$. Furthermore, note that by Assumptions $\mathcal{BG}^*.1$(ii) , $\mathcal{AG}.4$(ii), and $\mathcal{AG}.7$(i), we have

$$\sqrt{n}\left(\hat{\alpha}_n^* - \hat{\alpha}_n\right) = n^{-1/2}\sum_{t=1}^{n}\psi_t^*\left(X^{*t}, \hat{\alpha}_n\right) + o_{P^*}(1),$$

$W_n \to^P W$, where W a non-random symmetric and positive definite matrix, and Γ_0 is of full rank, respectively. Hence

$$\sqrt{n}\left(\hat{\beta}_n^* - \hat{\beta}_n\right) = -\left(\Gamma_0' W \Gamma_0\right)^{-1} \Gamma_0' W n^{-1/2} \sum_{t=1}^{n} \left(m_{2t}^*\left(X^{*t}, \hat{\alpha}_n, \hat{\beta}_n\right) + \Phi_0 \psi_t^*\left(X^{*t}, \hat{\alpha}_n\right)\right) + o_{P^*}(1).$$

The result now follows from Assumption $\mathcal{BG}^*$.4.

S6 Auxiliary lemmas used in the proof of Theorem 4.4

The main goal of this section is to show that a bootstrap version of the L_p maximal inequality stated in Assumption B6(iii) holds under our assumptions. In particular, we show that for some $p > 2+\delta$, $\left(\mathbb{E}\left|\mathbb{G}_n^*\right|_{\mathcal{N}_\eta}^p\right)^{1/p} \leq \eta$ holds when $\mathcal{N}_\eta$ is as defined in Assumption B6(iii) and $\mathbb{G}_n^*$ is defined as

$$\begin{aligned} & \mathbb{G}_n^*\left(q_2\left(\alpha_0, \beta\right) - q_2\left(\alpha_0, \beta_0\right)\right) \\ = \;& n^{-1} \sum_{t=1}^{n} \left(q_{2t}^*\left(\alpha_0, \beta\right) - q_{2t}^*\left(\alpha_0, \beta_0\right)\right) - E^*\left(n^{-1} \sum_{t=1}^{n} \left(q_{2t}^*\left(\alpha_0, \beta\right) - q_{2t}^*\left(\alpha_0, \beta_0\right)\right)\right), \end{aligned}$$

where $q_{2t}^*\left(\alpha, \beta\right)$ is a MBB version of $q_{2t}\left(\alpha, \beta\right) = \log f_{2t}\left(\alpha, \beta\right)$. This result is as follows.

Lemma S6.1. *Suppose that Assumption B6(iii) holds, and assume that* $\{\log f_{2t}\left(\alpha, \beta\right)\}$ *satisfies a Lipschitz continuity condition on* $\mathcal{A} \times \mathcal{B}$*, a.s.-P, with Lipschitz functions* $\{L_t\}$ *such that* $E\left|L_t\right|^p < \infty$ *for* $p > 2 + \delta$*, for some* $\delta > 0$*. Then,* $\left(\mathbb{E}\left|\mathbb{G}_n^*\right|_{\mathcal{N}_\eta}^p\right)^{1/p} \leq \eta$ *for any* $\eta > 0$.

To prove Lemma S6.1, we rely on the following L_p multiplier inequality, which extends Lemma 4.1 of Praestgaard and Wellner (1993) by allowing for $p \geq 1$ rather than just $p = 1$.

To state this result, we need to introduce some notation. Recall that for a generic time series $\{X_t : t = 1, \ldots, n\}$, letting $k = \frac{n}{\ell}$ denote the number of blocks of size ℓ needed to define a MBB sample of size n and letting $\{I_j : j = 1, \ldots, k\}$ be an i.i.d. uniform sequence of indices distributed on $\{1, \ldots, n-\ell+1\}$ allows us to write the MBB average as

$$\bar{X}_n^* = n^{-1} \sum_{t=1}^{n} X_t^* = k^{-1} \sum_{j=1}^{k} \left(\ell^{-1} \sum_{t=1}^{\ell} X_{t+(j-1)\ell}^* \right) = k^{-1} \sum_{j=1}^{k} \left(\sum_{t=1}^{\ell} X_{t+I_j-1} \right) = n^{-1} \sum_{j=1}^{k} Z_{I_j}.$$

Another way to write this average is as follows. Let $N = n - \ell + 1$, and let $\mathbf{W}_N = (W_1, \ldots, W_N)'$ denote a triangular array of weights whose distribution is the Multinomial $(k, (N^{-1}, \ldots, N^{-1}))$ distribution[1]. Note that these are non-negative exchangeable random variables. We can then think of $\bar{X}_n^*$ as a weighted average of the block sums $Z_j = \sum_{t=1}^{\ell} X_{t+j-1}$, weighted by W_j :

$$\bar{X}_n^* = n^{-1} \sum_{j=1}^{N} W_j Z_j,$$

where W_j denotes the number of times the j^{th} block sum Z_j is drawn in the bootstrap sample. Note that if $\ell = 1$, then $N = k = n$, and this way of writing the bootstrap average is exactly the same as when studying the nonparametric i.i.d. bootstrap using the Multinomial distribution $(n, (n^{-1}, \ldots, n^{-1}))$. Thus, our framework is an extension of the usual framework to the MBB. Our goal in Lemma S6.1 is to bound the L_p moment of the bootstrap empirical process

$$\mathbb{G}_n^* f = n^{-1/2} \sum_{t=1}^{n} \left(f_t^* - E^* \left(f_t^* \right) \right).$$

With this new notation, we can write

$$\mathbb{G}_n^* f = n^{-1/2} \sum_{j=1}^{N} \left(W_j - E_W \left(W_j \right) \right) \left(\sum_{t=1}^{\ell} f_{t+j-1} \right),$$

where $E_W(\cdot)$ (and $P_W(\cdot)$) denotes expectation (and probability) with respect to the random

[1]For simplicity, we will drop the array notation and will write W_j rather than $W_{N,j}$. Similarly, we will omit the index n in the definition N_n.

vector $\mathbf{W}_N$ defined above. The L_p-multiplier we are about to state gives a bound on the L_p moments of averages defined as $n^{-1/2}\sum_{j=1}^N W_j Z_j$, where Z_j will play the role of the block sum $\sum_{t=1}^{\ell} f_{t+j-1}$ in our application.

To state this result, define the joint probability $\mathbb{P} = P \times P_W$, which we wrote before as $P \times P^*$, and let $\|W_1\|_{2,1} = \int_0^\infty \sqrt{P_W(W_1 \geq u)}du$. Some expressions below may be non-measurable; probability and expectation of these expressions are understood in terms of outer probability and outer expectation (see, e.g. van der Vaart and Wellner, 1996, p. 6). Application of Fubini's theorem to such expectations requires additional care. We assume that a measurability condition that restores the Fubini theorem is satisfied in all our applications below.

Lemma S6.2. *Let $\mathbf{W}_N = (W_1, \ldots, W_N)'$ be an array of non-negative exchangeable random variables such that, for every N, $\|W_1\|_{2,1} = \int_0^\infty \sqrt{P_W(W_1 \geq u)}du < \infty$, and let R denote a random permutation uniformly distributed on Π_N, the set of permutations of $1, 2, \ldots, N$. Let $Z_1, \ldots, Z_N$ be a sequence of random elements such that $(\mathbf{W}_N, R)$ and $(Z_1, \ldots, Z_N)$ are independent, and write $\|Z_j\| = \sup_{h \in \mathcal{F}} |Z_j(h)|$. Then for any N_0 such that $1 \leq N_0 < \infty$ and any $N > N_0$, the following inequality holds for any $p \geq 1$:*

$$\left(\mathbb{E}\left\|n^{-1/2}\sum_{j=1}^N W_j Z_j\right\|^p\right)^{1/p} \leq \frac{N_0}{\sqrt{n}}\left(E_W\left|\max_{1\leq j\leq N} W_j\right|^p\right)^{1/p}\left(\frac{1}{N}\sum_{j=1}^N E\|Z_j\|^p\right)^{1/p}$$
$$+\|W_1\|_{2,1}^{1/p}\cdot\left(E_{Z,R}\left(\max_{N_0<k\leq N}\left\|\frac{1}{\sqrt{k}}\sum_{i=N_0+1}^{k} Z_{R(i)}\right\|\right)^p\right)^{1/p},$$

where we let $E_{Z,R}(\cdot)$ denote the expectation with respect to $Z_1, \ldots, Z_N$ and R jointly.

This result extends Lemma 4.1 of Praestgaard and Wellner (1993) from $p = 1$ to $p \geq 1$. As in Praestgaard and Wellner (1993), we do not assume any particular dependence structure on the vector $(Z_1, \ldots, Z_N)$, the only assumption being that it is independent of the pair $(\mathbf{W}_N, R)$. This is in contrast with the L_p multiplier provided by Cheng (2015, p.

17), which assumes $Z_1, \ldots, Z_N$ to be i.i.d., while also allowing for any $p \geq 1$. The i.i.d. assumption on the random functions Z_j is too strong for our applications, where Z_j will be given by block sums of contributions to the log likelihood function. These are typically serially dependent in the time series context and this is the reason for given Lemma S6.2, a result that might be of independent interest.

Next, we prove Lemma S6.1 and then we prove Lemma S6.2.

Proof of Lemma S6.1 In the following '$\lesssim$' denote smaller than, up to an universal constant $K > 0$. Recalling the definition of $\mathbb{G}_n^* f$, where f is in the function class $\mathcal{N}_\eta$, and the property of the MBB weights, in particular, $\sum_{j=1}^N W_j = k$, implying that $E_W(W_j) = \frac{k}{N}$, we can rewrite $\mathbb{G}_n^* f$ as follows:

$$
\begin{aligned}
\mathbb{G}_n^* f &= n^{-1/2} \sum_{j=1}^N \left(W_j - E_W(W_j)\right) \left(\sum_{t=1}^{\ell} f_{t+j-1}\right) \\
&= n^{-1/2} \sum_{j=1}^N \left(W_j - \frac{k}{N}\right) \left(\sum_{t=1}^{\ell} f_{t+j-1}\right), \text{ since } E_W(W_j) = \frac{k}{N} \\
&= n^{-1/2} \sum_{j=1}^N \left(W_j - \frac{k}{N}\right) \left[\left(\sum_{t=1}^{\ell} f_{t+j-1}\right) - E\left(\sum_{t=1}^{\ell} f_{t+j-1}\right)\right],
\end{aligned}
$$

since $\sum_{t=1}^N \left(W_j - \frac{k}{N}\right) = 0$, and the expectation of $E\left(\sum_{t=1}^{\ell} f_{t+j-1}\right)$ is time invariant under Assumption B6(i). For $j = 1, 2, \ldots, N$, let

$$
Y_j(f) = \sum_{t=1}^{\ell} f_{t+j-1} - E\left(\sum_{t=1}^{\ell} f_{t+j-1}\right) = \sum_{t=1}^{\ell} \left(f_{t+j-1} - E\left(f_{t+j-1}\right)\right). \qquad \text{(S6.1)}
$$

With this notation, $\mathbb{G}_n^* f$ can be rewritten as

$$
\mathbb{G}_n^* f = n^{-1/2} \sum_{j=1}^N \left(W_j - \frac{k}{N}\right) Y_j(f). \qquad \text{(S6.2)}
$$

Our goal is to bound the L_p moment of the supremum of this empirical process. To do so, we follow the same arguments as in Cheng (2015, p. 19) to show that

$$\begin{aligned}\left(\mathbb{E}\left\|\mathbb{G}_n^*\right\|_{\mathcal{N}_\eta}^p\right)^{1/p} &= \left(\mathbb{E}\left(\sup_{f\in\mathcal{N}_\eta}\left|n^{-1/2}\sum_{t=1}^{N}\left(W_j-\frac{k}{N}\right)Y_j\left(f\right)\right|\right)^p\right)^{1/p}\\ &\lesssim 2\left(\mathbb{E}\left(\sup_{f\in\mathcal{N}_\eta}\left|n^{-1/2}\sum_{t=1}^{N}W_jY_j\left(f\right)\right|\right)^p\right)^{1/p}. \qquad \text{(S6.3)}\end{aligned}$$

Next, we apply the L_p multiplier inequality in Lemma S6.2 (using (S6.3)) with $Z_j = Y_j\left(f\right)$ and $\mathcal{F}=\mathcal{N}_\eta$. This yields

$$\begin{aligned}\left(\mathbb{E}\left\|\mathbb{G}_n^*\right\|_{\mathcal{N}_\eta}^p\right)^{1/p} &\lesssim \frac{N_0}{\sqrt{n}}\left(E_W\left|\max_{1\le j\le N}W_j\right|^p\right)^{1/p}\left(\frac{1}{N}\sum_{i=1}^{N}E\left\|Z_i\right\|_{\mathcal{N}_\eta}^p\right)^{1/p}\\ &\quad+\left(\ell\left\|W_{N,1}\right\|_{2,1}\right)^{1/p}\left(\ell^{-1}E_{Z,R}\left(\max_{N_0<k\le N}\left\|\frac{1}{\sqrt{k}}\sum_{i=N_0+1}^{k}Z_{R(i)}\right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p}\\ &\lesssim N_0\frac{\ell}{\sqrt{n}}\left(E_W\left|\max_{1\le j\le N}W_j\right|^p\right)^{1/p}\left(\frac{1}{N\ell^p}\sum_{i=1}^{N}E\left\|Z_i\right\|_{\mathcal{N}_\eta}^p\right)^{1/p}\\ &\quad+\left(\ell^{-1}E_{Z,R}\left(\max_{N_0<k\le N}\left\|\frac{1}{\sqrt{k}}\sum_{i=N_0+1}^{k}Z_{R(i)}\right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p}\\ &\lesssim \mathrm{I}+\mathrm{II}. \qquad \text{(S6.4)}\end{aligned}$$

for any $1\le N_0<\infty$ and $N>N_0$, (the second inequality follows because the MBB weight verifies the condition $\limsup_{N\to\infty}\ell\left\|W_{N,1}\right\|_{2,1}<\infty$, where $W_1=W_{N,1}$). We first bound the first term in the preceding equation, then we bound the second term.

For the first term, note that

$$\begin{aligned}&\frac{1}{N\ell^p}\sum_{i=1}^{N}E\left\|Z_i\right\|_{\mathcal{N}_\eta}^p\\ \le\ &\frac{1}{N\ell^p}\sum_{j=1}^{N}\ell^{p-1}\sum_{t=1}^{\ell}E\left\|\left(f_{t+j-1}-E\left(f_{t+j-1}\right)\right)\right\|_{\mathcal{N}_\eta}^p = n^{-1/2}\frac{\sqrt{n}}{N\ell}\sum_{j=1}^{N}\sum_{t=1}^{\ell}E\left\|\left(f_{t+j-1}-E\left(f_{t+j-1}\right)\right)\right\|_{\mathcal{N}_\eta}^p\end{aligned}$$

from Minkowski's inequality. Using the same arguments as in the proof of Theorem 4.4 (see equation (12))), it follows that (and given Assumption B6(iii)),

$$\begin{aligned}\left(\frac{1}{N\ell^p}\sum_{i=1}^{N}E\left\|Z_i\right\|_{\mathcal{N}_\eta}^p\right)^{1/p} &\lesssim \left(\frac{1}{N\ell}\sum_{j=1}^{N}\sum_{t=1}^{\ell}E\left\|\left(f_{t+j-1}-E\left(f_{t+j-1}\right)\right)\right\|_{\mathcal{N}_\eta}^p\right)^{1/p} \\ &= \left(EE^*\left(\frac{1}{n}\sum_{t=1}^{n}\left\|\left(f_{t+j-1}^*-E\left(f_{t+j-1}\right)\right)\right\|_{\mathcal{N}_\eta}^p\right)\right)^{1/p} \\ &\lesssim \left(\left(n^{-1}\sum_{t=1}^{n}E\left\|\left(f_t-E\left(f_t\right)\right)\right\|_{\mathcal{N}_\eta}^p\right)^{1/p}+\eta O\left(\frac{\ell}{\sqrt{n}}\right)\right) \quad \text{(S6.5)}\end{aligned}$$

where the last term is asymptotically negligible given the condition $\ell = o\left(\sqrt{n}\right)$. Next, we can show that

$$\left(n^{-1}\sum_{t=1}^{n}E\left\|\left(f_t-E\left(f_t\right)\right)\right\|_{\mathcal{N}_\eta}^p\right)^{1/p} \lesssim \left(E\left\|N_\eta\right\|^p\right)^{1/p}, \quad \text{(S6.6)}$$

where N_η is the envelope of the function class $\mathcal{N}_\eta$. Given the Lipschitz continuity assumption (cf. Assumption B6(iv)), we can show that $\left(E\left\|N_\eta\right\|^p\right)^{1/p} \leq \eta$. This implies

$$\begin{aligned}& N_0\frac{\ell}{\sqrt{n}}\left(E_W\left|\max_{1\leq j\leq N}W_j\right|^p\right)^{1/p}\left(\frac{1}{N\ell^p}\sum_{i=1}^{N}E\left\|Z_i\right\|_{\mathcal{N}_\eta}^p\right)^{1/p} \\ \lesssim\ & \underbrace{\left[\frac{\ell}{\sqrt{n}}\left(E_W\left|\max_{1\leq j\leq N}W_j\right|^p\right)^{1/p}\right]}_{o(1)}\underbrace{\left(E\left\|N_\eta\right\|^p\right)^{1/p}}_{\lesssim\eta} = o\left(\eta\right),\end{aligned}$$

provided the second factor is $o\left(1\right)$. Given that $\max_{1\leq j\leq N}W_j^p \geq 1$, $\left(E_W\left|\max_{1\leq j\leq N}W_j\right|^p\right)^{1/p} \leq E_W\left(\max_{1\leq j\leq N}W_j^p\right)$. Therefore, we have

$$\frac{\ell}{\sqrt{n}}\left(E_W\left|\max_{1\leq j\leq N}W_j\right|^p\right)^{1/p} \lesssim \underbrace{\sqrt{\frac{N}{n}}}_{\to 1}\frac{\ell}{\sqrt{N}}E_W\left(\max_{1\leq j\leq N}W_j^p\right).$$

Next, we appeal to Lemma 4.7 of Praestgaard and Wellner (1993) to show that

$$\frac{\ell}{\sqrt{N}} E_W \left(\max_{1 \leq j \leq N} W_j^p \right) = o\left(1\right).$$

To do so, we verify that ℓW_1^p satisfies the necessary conditions of Lemma 4.7 of Praestgaard and Wellner (1993), i.e., the following two conditions

$$\limsup_{N \to \infty} \|\ell W_1^p\|_{2,1} < \infty, \tag{S6.7}$$

and

$$\lim_{\lambda \to \infty} \limsup_{N \to \infty} \sup_{u \geq \lambda} u^2 P_W \left(\ell W_1^p > u\right) = 0, \tag{S6.8}$$

where we recall that W_1 is a an element of a triangular array, i.e. $W_1 = W_{N,1}$. As argued by Cheng (2015), cf. his equation (29), a sufficient condition to obtain both conditions (S6.7) and (S6.8) is that

$$\limsup_{N \to \infty} E_W \left(\ell W_1^{(2+\varepsilon)p} \right) < \infty, \tag{S6.9}$$

for some $\varepsilon > 0$, which in turn is implied by

$$\limsup_{N \to \infty} E_W \left(\ell W_1^5 \right) < \infty,$$

because for a small enough $\epsilon > 0$, we can always choose $p = 5/\left(2 + \epsilon\right) > 2$. Using the property of multinomial distribution, we have

$$\begin{aligned}
E_W\left(W_1^5\right) &= \frac{k}{N_n}+15\frac{k\left(k-1\right)}{N_n^2}+25\frac{k\left(k-1\right)\left(k-2\right)}{N_n^3}+10\frac{k\left(k-1\right)\left(k-2\right)\left(k-3\right)}{N_n^4} \\
&\quad +\frac{k\left(k-1\right)\left(k-2\right)\left(k-3\right)\left(k-4\right)}{N_n^5} \\
&= \frac{\frac{n}{\ell_n}N_n^4+15\frac{n}{\ell_n}\left(\frac{n}{\ell_n}-1\right)N_n^3+25\frac{n}{\ell_n}\left(\frac{n}{\ell_n}-1\right)\left(\frac{n}{\ell_n}-2\right)N_n^2}{N_n^5} \\
&\quad +\frac{10\frac{n}{\ell_n}\left(\frac{n}{\ell_n}-1\right)\left(\frac{n}{\ell_n}-2\right)\left(\frac{n}{\ell_n}-3\right)N_n}{N_n^5} \\
&\quad +\frac{\frac{n}{\ell_n}\left(\frac{n}{\ell_n}-1\right)\left(\frac{n}{\ell_n}-2\right)\left(\frac{n}{\ell_n}-3\right)\left(\frac{n}{\ell_n}-4\right)}{N_n^5}.
\end{aligned}$$

Given the condition $\ell = o\left(\sqrt{n}\right)$, it follows that

$$\limsup_{N\to\infty} E_W\left(\ell W_1^5\right) = 1 < \infty.$$

We follow the same arguments as in Cheng (2015, p. 19) and write

$$\begin{aligned}
&\left(E_{Z,R}\left(\max_{N_0<k\le N}\left\|\frac{1}{\sqrt{k}}\sum_{i=N_0+1}^{k} Z_{R(i)}\right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p} \\
\lesssim\ &\left(E_{Z,R}\left(\max_{N_0<k\le N}\left\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k} Z_{R(i)}\right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p} + \left(E_{Z,R}\left(\left\|\frac{1}{\sqrt{N_0}}\sum_{i=N_0+1}^{N} Z_{R(i)}\right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p} \\
\le\ & 2\left(E_{Z,R}\left(\max_{N_0\le k\le N}\left\|\frac{1}{\sqrt{k}}\sum_{i=1}^{k} Z_{R(i)}\right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p}
\end{aligned}$$

where the last inequality follows by the triangular inequality. Thus, the proof of Lemma S6.1 is completed when

$$\mathrm{II} \lesssim \left(\ell^{-1}E_{Z,R}\left(\max_{N_0<k\le N}\left\|\frac{1}{\sqrt{k}}\sum_{i=N_0+1}^{k} Z_{R(i)}\right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p},$$

holds for $p > 2 + \delta$. Let

$$\widetilde{\mathbb{G}}_k = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} Z_{R(i)},$$

for $N_0 \leq k \leq N$. It follows that when $k = N$, we have

$$\widetilde{\mathbb{G}}_N = \frac{1}{\sqrt{N}} \sum_{i=1}^{N} Z_{R(i)},$$

Recall that for any positive random variable Y, the following holds

$$EY^q = \int_0^\infty q u^{q-1} P\left(Y > u\right) du,$$

for any $q > 0$. The Levy inequality (see e.g., proposition A.1.2 of van der Vaart and Wellner (1996)) implies that,

$$P\left(\max_{k \leq N} \left\| \widetilde{\mathbb{G}}_k \right\|_{\mathcal{N}_\eta} > \lambda\right) \leq K P\left(\left\| \widetilde{\mathbb{G}}_N \right\|_{\mathcal{N}_\eta} > \lambda\right), \tag{S6.10}$$

for every $\lambda > 0$. Hence, we can deduce that

$$\mathrm{II} \lesssim \left(\ell^{-1} E_{Z,R} \left(\max_{N_0 < k \leq N} \left\| \frac{1}{\sqrt{k}} \sum_{i=N_0+1}^{k} Z_{R(i)} \right\|_{\mathcal{N}_\eta}\right)^p\right)^{1/p} \lesssim K^{1/p} \left(\ell^{-1} E_{Z,R} \left\| \widetilde{\mathbb{G}}_N \right\|_{\mathcal{N}_\eta}^p\right)^{1/p}.$$

Proof of Lemma S6.2 The proof follows closely that of Lemma 4.1 in Praestgaard and Wellner (1993). Define a random permutation S of $\{1, \ldots, N\}$ such that $W_{S(1)} \geq \ldots \geq W_{S(N)}$, and if $W_{S(t)} = W_{S(t+1)}$ then $S\left(t\right) < S\left(t+1\right)$. Then, let R be a random permutation uniformly distributed on Π_N (i.e., the set of permutations of $1, 2, \ldots, N$) and independent of $\left(\mathbf{W}, S\right)$. Using the same arguments as in Praestgaard and Wellner (1993), and given the exchangeability of $\mathbf{W}_N$, we have that

$$\begin{aligned}\left(\mathbb{E}\left\|n^{-1/2}\sum_{j=1}^{N}W_jZ_j\right\|^p\right)^{1/p} &= \left(\mathbb{E}\left\|n^{-1/2}\sum_{j=1}^{N}W_{(j)}Z_{R(j)}\right\|^p\right)^{1/p}\\ &\leq \left(\mathbb{E}\left\|n^{-1/2}\sum_{j=1}^{N_0}W_{(j)}Z_{R(j)}\right\|^p\right)^{1/p}+\left(\mathbb{E}\left\|n^{-1/2}\sum_{j=N_0+1}^{N}W_{(j)}Z_{R(j)}\right\|^p\right)^{1/p}\\ &\equiv \mathrm{I}\left(N_0,N\right)+\mathrm{II}\left(N_0,N\right).\end{aligned}$$

where $W_{(j)}=W_{S(j)}$. We can bound the term $\mathrm{I}(N_0,N)$ by

$$\begin{aligned}\mathrm{I}\left(N_0,N\right) &\leq n^{-1/2}\sum_{j=1}^{N_0}\left(\mathbb{E}\left\|W_{(j)}Z_{R(j)}\right\|^p\right)^{1/p} \text{ by Minkowski's inequality}\\ &\leq n^{-1/2}\sum_{j=1}^{N_0}\left(E_W\left|W_{(j)}\right|^p\right)^{1/p}\left(\mathbb{E}\left\|Z_{R(j)}\right\|^p\right)^{1/p} \text{ by independence between } W \text{ and } R\\ &\leq \left(E_W\left(n^{-1/2}\max_{1\leq j\leq N}W_j\right)^p\right)^{1/p}\sum_{j=1}^{N_0}\left(\mathbb{E}\left\|Z_{R(j)}\right\|^p\right)^{1/p}\\ &= \left(E_W\left(n^{-1/2}\max_{1\leq j\leq N}W_j\right)^p\right)^{1/p}\sum_{j=1}^{N_0}\left(\frac{1}{N}\sum_{i=1}^{N}E\left\|Z_i\right\|^p\right)^{1/p} \text{ by the properties of } R\\ &\leq \frac{N_0}{\sqrt{n}}\left(E_W\left(\max_{1\leq j\leq N}W_j\right)^p\right)^{1/p}\left(\frac{1}{N}\sum_{i=1}^{N}E\left\|Z_i\right\|^p\right)^{1/p}.\end{aligned}$$

Note in particular that

$$\begin{aligned}\mathbb{E}\left\|Z_{R(j)}\right\|^p &= E_Z E_{R|Z}\left(\left\|Z_{R(j)}\right\|^p\right) \text{ by the LIE}\\ &= E_Z\left(\frac{1}{N}\sum_{i=1}^{N}\left\|Z_i\right\|^p\right)=\frac{1}{N}\sum_{i=1}^{N}E\left\|Z_i\right\|^p.\end{aligned}$$

If $E\left\|Z_i\right\|^p$ does not depend on i, then this is equal to $E\left\|Z_1\right\|^p$ and we get that

$$\mathrm{I}\left(N_0,N\right)\leq\frac{N_0}{\sqrt{n}}\left(E_W\left(\max_{1\leq j\leq N}W_j\right)^p\right)^{1/p}\left(E\left\|Z_1\right\|^p\right)^{1/p},$$

which is what Cheng (2015) get.

Next, in order to bound the second term i.e., $\mathrm{II}(N_0, N)$, we follow Cheng (2015) and write

$$\sum_{j=N_0+1}^{N} W_{(j)} Z_{R(j)} = \sum_{j=N_0+1}^{N} \sqrt{j}\left(W_{(j)} - W_{(j+1)}\right) T_j,$$

where $T_j = j^{-1/2} \sum_{i=N_0+1}^{j} Z_{R(i)}$ and $W_{(N+1)} = 0$. Hence, following Cheng (2015),

$$\begin{aligned}
& \mathrm{II}\left(N_0, N\right) \\
= & \left(\mathbb{E}\left\| n^{-1/2} \sum_{j=N_0+1}^{N} W_{(j)} Z_{R(j)} \right\|^p\right)^{1/p} \\
= & \left(\mathbb{E}\left\| n^{-1/2} \sum_{j=N_0+1}^{N} \sqrt{j}\left(W_{(j)} - W_{(j+1)}\right) T_j \right\|^p\right)^{1/p} \\
\leq & \left(E_{Z,R}\left\| \max_{N_0<k\leq N} \|T_k\| \right\|^p\right)^{1/p} \cdot \left(E_W \left\| n^{-1/2} \sum_{j=N_0+1}^{N} \sqrt{j}\left(W_{(j)} - W_{(j+1)}\right) \right\|^p\right)^{1/p} \\
= & \left(E_{Z,R}\left\| \max_{N_0<j\leq N} \left| \frac{1}{\sqrt{k}} \sum_{j=N_0+1}^{k} Z_{R(j)} \right| \right\|^p\right)^{1/p} \cdot \left(E_W \left\| n^{-1/2} \sum_{j=N_0+1}^{N} \sqrt{j}\left(W_{(j)} - W_{(j+1)}\right) \right\|^p\right)^{1/p}.
\end{aligned}$$

Thus, the proof is completed if we can show that

$$\begin{aligned}
\left(E_W \left\| n^{-1/2} \sum_{j=N_0+1}^{N} \sqrt{j}\left(W_{(j)} - W_{(j+1)}\right) \right\|^p\right)^{1/p} & \leq n^{-1/2} \left(E_W \left\| \sum_{j=N_0+1}^{N} \sqrt{j}\left(W_{(j)} - W_{(j+1)}\right) \right\|^p\right)^{1/p} \\
& \leq n^{-1/2}\left(n^{1/2} \|W_1\|_{2,1}^{1/p}\right) \leq \|W_1\|_{2,1}^{1/p}
\end{aligned}$$

i.e., if we can show that

$$E_W \left\| \sum_{j=N_0+1}^{N} \sqrt{j}\left(W_{(j)} - W_{(j+1)}\right) \right\|^p \leq n^{p/2} \|W_1\|_{2,1}. \tag{S6.11}$$

It is easy to see that the proof is completed by using exactly the same arguments as in Cheng (2015) (cf. the proof of their equation (43)).

S7 Proofs of Theorems 4.3 and 4.4

Proof of Theorem 4.3. For some small $\delta > 0$,

$$E^* \left| \sqrt{n} \left(\hat{\beta}_{1,n}^* - \hat{\beta}_n \right) \right|^{2+\delta} \leq \left\| \hat{H}_n^{-1} \right\|_1^{2+\delta} E^* \left| n^{-1/2} \sum_{t=1}^n s_t^* \left(\hat{\alpha}_n, \hat{\beta}_n \right) \right|^{2+\delta},$$

where $\|A\|_1$ is the spectral norm of a matrix A, i.e. $\|A\|_1^2 = \max_{x \neq 0} \frac{x'A'Ax}{x'x}$. Since $\hat{H}_n$ is a symmetric matrix, $\left\| \hat{H}_n^{-1} \right\|_1^{2+\delta} = \left(\lambda_{\min}^{-1} \left(\hat{H}_n \right) \right)^{2+\delta} = O_P(1)$ since $\lambda_{\min} \left(\hat{H}_n \right) \to^P \lambda_{\min}(H_0) \neq 0$ by the assumption that H_0 is nonsingular. Thus, it suffices to show that $E^* \left| n^{-1/2} \sum_{t=1}^n s_t^* \left(\hat{\alpha}_n, \hat{\beta}_n \right) \right|^{2+\delta} = O_P(1)$. Using the definition of s_t^*, we can decompose this expectation as

$$\begin{aligned} & E^* \left| n^{-1/2} \sum_{t=1}^n s_t^* \left(\hat{\alpha}_n, \hat{\beta}_n \right) \right|^{2+\delta} \\ \leq \; & E^* \left| n^{-1/2} \sum_{t=1}^n s_{2t}^* \left(\hat{\alpha}_n, \hat{\beta}_n \right) \right|^{2+\delta} + \left\| \hat{F}_n \right\|^{2+\delta} \left\| \hat{A}_n^{-1} \right\|_1^{2+\delta} E^* \left| n^{-1/2} \sum_{t=1}^n s_{1t}^* \left(\hat{\alpha}_n \right) \right|^{2+\delta}. \end{aligned}$$

Each of the bootstrap expectations on the RHS of the display can be shown to be $O_P(1)$ under our assumptions. The arguments are similar to those used by GW (2005). Take e.g. the second of these expectations. Adding and subtracting appropriately, we can bound it by $I_1 + I_2$, where $I_1 = 2^{1+\delta} E^* \left| n^{-1/2} \sum_{t=1}^n s_{1t}^* (\alpha_0) \right|^{2+\delta}$ and $I_2 = 2^{1+\delta} E^* \left| n^{-1/2} \sum_{t=1}^n \left(s_{1t}^* (\hat{\alpha}_n) - s_{1t}^* (\alpha_0) \right) \right|^{2+\delta}$. Under Assumption B4$'$, $E(s_{1t}(\alpha_0)) = 0$ and we can show that $\{s_{1t}(\alpha_0)\}$ is $L_{2+\delta}$-mixingale with bounded mixingale constants and absolutely summable coefficients given in particular the $L_{2+\delta}$-NED assumption on the score function $s_{1t}(\alpha)$. Hence, by Lemma A.1 of GW (2005), we have that $E(I_1) = O(1) + O\left(\left(\frac{\ell_n^2}{n} \right)^{(2+\delta)/2} \right) = O(1)$ since $\ell_n^2 / n \to 0$ by assumption. To show that $I_2 = O_P(1)$, we rely on Assumption B5, the Lipschitz continuity assumption on $s_{1t}(\alpha)$. This assumption implies that

$$E^* \left| n^{-1/2} \sum_{t=1}^{n} \left(s_{1t}^* \left(\hat{\alpha}_n \right) - s_{1t}^* \left(\alpha_0 \right) \right) \right|^{2+\delta} \leq \left(n^{-1} \sum_{t=1}^{n} E^* \left| L_{1t}^* \right|^{2+\delta} \right) \left| \sqrt{n} \left(\hat{\alpha}_n - \alpha_0 \right) \right|^{2+\delta},$$

where $\left| \sqrt{n} \left(\hat{\alpha}_n - \alpha_0 \right) \right|^{2+\delta} = O_P(1)$ and

$$n^{-1} \sum_{t=1}^{n} E^* \left| L_t^* \right|^{2+\delta} = n^{-1} \sum_{t=1}^{n} \left| L_{1t} \right|^{2+\delta} + O_P \left(\frac{\ell_n}{n} \right).$$

where $n^{-1} \sum_{t=1}^{n} E \left| L_{1t} \right|^{2+\delta} = O(1)$ under Assumption B5. The proof that

$$E^* \left| n^{-1/2} \sum_{t=1}^{n} s_{2t}^* \left(\hat{\alpha}_n, \hat{\beta}_n \right) \right|^{2+\delta} = O_P(1)$$

follows under similar arguments.

Proof of Theorem 4.4. The result follows from the triangle inequality if

$$\sup_n \mathbb{E} \left| \sqrt{n} \left(\hat{\beta}_n^* - \beta_0 \right) \right|^{2+\delta} < \infty \text{ and } \sup_n E \left| \sqrt{n} \left(\hat{\beta}_n - \beta_0 \right) \right|^{2+\delta} < \infty.$$

The moment condition on $\sqrt{n} \left(\hat{\beta}_n - \beta_0 \right)$ holds by assumption. Then, the moment condition on $\sqrt{n} \left(\hat{\beta}_n^* - \beta_0 \right)$ follows by an argument similar to that used in Kato (2011). In particular, note that for any positive random variable Z and any $q \geq 1$, we can write $E|Z|^q = q \int_0^\infty t^{q-1} P(Z > t)\, dt$. Hence,

$$\mathbb{E} \left| \sqrt{n} \left(\hat{\beta}_n^* - \beta_0 \right) \right|^{2+\delta} = (2+\delta) \int_0^\infty t^{2+\delta-1} \mathbb{P} \left(\left| \sqrt{n} \left(\hat{\beta}_n^* - \beta_0 \right) \right| > t \right) dt.$$

We will show that $P \left(\left| \sqrt{n} \left(\hat{\beta}_n^* - \beta_0 \right) \right| > t \right) \leq K t^{-p}$ for $p > 2 + \delta$ and some constant K.

This will imply the result since

$$\mathbb{E}\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right|^{2+\delta} \leq K \int_0^\infty t^{2+\delta-p-1} dt < \infty \text{ if } p > 2+\delta.$$

Let $\tilde{Q}_n(\beta) = Q_{2n}(\hat{\alpha}_n, \beta) = n^{-1}\sum_{t=1}^n q_{2t}(X^t, \hat{\alpha}_n, \beta)$, such that $q_{2t}(X^t, \hat{\alpha}_n, \beta) = \log f_{2t}(X^t, \hat{\alpha}_n, \beta)$. Note that $\hat{\beta}_n^* = \arg\max_\beta \tilde{Q}_n^*(\beta)$, where

$$\tilde{Q}_n^*(\beta) = Q_{2n}^*(\hat{\alpha}_n^*, \beta).$$

Partition the parameter space $\mathcal{B}$ into "shells" $S_{j,n} = \{\beta \in \mathcal{B} : 2^{j-1} < |\sqrt{n}(\beta - \beta_0)| \leq 2^j\}$ for any integer $j \geq 1$. If $\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right|$ is larger than 2^{j_0} for a given integer j_0, then $\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right|$ is in one of the shells $S_{j,n}$ with $j \geq j_0$. In that case, the supremum of the map $\beta \longmapsto \tilde{Q}_n^*(\beta) - \tilde{Q}_n^*(\beta_0)$ must be nonnegative by the definition of $\hat{\beta}_n^*$. This implies

$$\mathbb{P}\left(\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right| > 2^{j_0}\right) \leq \sum_{j=j_0}^\infty \mathbb{P}\left(\sup_{\beta \in S_{j,n}} \left\{\tilde{Q}_n^*(\beta) - \tilde{Q}_n^*(\beta_0)\right\} \geq 0\right). \tag{S7.1}$$

Next decompose $\tilde{Q}_n^*(\beta) - \tilde{Q}_n^*(\beta_0)$ as follows:

$$\begin{aligned}
\tilde{Q}_n^*(\beta) - \tilde{Q}_n^*(\beta_0) &= \left[Q_{2n}^*(\hat{\alpha}_n^*, \beta) - Q_{2n}^*(\hat{\alpha}_n^*, \beta_0)\right] - \left[Q_{2n}^*(\alpha_0, \beta) - Q_{2n}^*(\alpha_0, \beta_0)\right] \\
&\quad + Q_{2n}^*(\alpha_0, \beta) - Q_{2n}^*(\alpha_0, \beta_0) - E^*\left[Q_{2n}^*(\alpha_0, \beta) - Q_{2n}^*(\alpha_0, \beta_0)\right] \\
&\quad + E^*\left[Q_{2n}^*(\alpha_0, \beta) - Q_{2n}^*(\alpha_0, \beta_0)\right] - E\left(E^*\left[Q_{2n}^*(\alpha_0, \beta) - Q_{2n}^*(\alpha_0, \beta_0)\right]\right) \\
&\quad + E\left[E^*\left(Q_{2n}^*(\alpha_0, \beta) - Q_{2n}^*(\alpha_0, \beta_0)\right)\right] \\
&\equiv I_{\text{2-step},n}(\beta) + I_{1,n}(\beta) + I_{2,n}(\beta) + I_{3,n}(\beta).
\end{aligned}$$

Note that

$$\begin{aligned} E^*\left(Q_{2n}^*\left(\alpha_0,\beta\right)-Q_{2n}^*\left(\alpha_0,\beta_0\right)\right) &= E^*\left(n^{-1}\sum_{t=1}^{n} q_{2t}^*\left(\alpha_0,\beta\right)-q_{2t}^*\left(\alpha_0,\beta_0\right)\right) \\ &= \sum_{t=1}^{n}\gamma_{nt}\left(q_{2t}\left(\alpha_0,\beta\right)-q_{2t}\left(\alpha_0,\beta_0\right)\right), \end{aligned}$$

where the weighting function γ_{nt} is defined as

$$\gamma_{nt}=\begin{cases} \frac{t}{\ell_n(n-\ell_n+1)}, & \text{if} \quad t\in\{1,\ldots,\ell_n\} \\ \frac{1}{n-\ell_n+1}, & \text{if} \quad i\in\{\ell_n+1,\ldots,n-\ell_n\} \\ \frac{n-t+1}{\ell_n(n-\ell_n+1)}, & \text{if} \quad i\in\{n-\ell_n+1,\ldots,n\} \end{cases},$$

such that $\sum_{t=1}^{n}\gamma_{nt}=1$. It follows that

$$I_{3n}\left(\beta\right)=\sum_{t=1}^{n}\gamma_{nt}E\left(q_{2t}\left(\alpha_0,\beta\right)-q_{2t}\left(\alpha_0,\beta_0\right)\right)=\bar{Q}_2\left(\alpha_0,\beta\right)-\bar{Q}_2\left(\alpha_0,\beta_0\right),$$

given the time homogeneity of the moments $E\left(q_{2t}\left(\alpha,\beta\right)\right)$ (which is part of Assumption B6(i)) and the fact that $\sum_{t=1}^{n}\gamma_{nt}=1$. By the quadratic behavior assumption, we can conclude that $-I_{3,n}\left(\beta\right)\geq K\left|\beta-\beta_0\right|^2\geq K\frac{2^{2j-2}}{n}$ on $S_{j,n}$, for some $K>0$. Then, for each j the following inclusion holds

$$\begin{aligned} &\left\{\sup_{\beta\in S_{j,n}}\left\{\tilde{Q}_n^*\left(\beta\right)-\tilde{Q}_n^*\left(\beta_0\right)\right\}\geq 0\right\} \\ \subset\ &\left\{\sup_{\beta\in S_{j,n}}\left|I_{\text{2-step},n}\left(\beta\right)\right|+\sup_{\beta\in S_{j,n}}\left|I_{1,n}\left(\beta\right)\right|+\sup_{\beta\in S_{j,n}}\left|I_{2,n}\left(\beta\right)\right|\geq K\frac{2^{2j-2}}{n}\right\}. \end{aligned}$$

It follows that the right-hand side of (S7.1) i.e., $\sum_{j=j_0}^{\infty}P\left(\sup_{\beta\in S_{j,n}}\left\{\tilde{Q}_n^*\left(\beta\right)-\tilde{Q}_n^*\left(\beta_0\right)\right\}\geq 0\right)$

can be bounded by

$$\sum_{j=j_0}^{\infty}\mathbb{P}\left(\sup_{\beta\in S_{j,n}}\left|I_{2\text{-step},n}\left(\beta\right)\right|\geq K\frac{2^{2(j-1)}}{n}\right)$$
$$+\sum_{j=j_0}^{\infty}\mathbb{P}\left(\sup_{\beta\in S_{j,n}}\left|I_{1,n}\left(\beta\right)\right|\geq K\frac{2^{2(j-1)}}{n}\right)+\sum_{j=j_0}^{\infty}\mathbb{P}\left(\sup_{\beta\in S_{j,n}}\left|I_{2,n}\left(\beta\right)\right|\geq K\frac{2^{2(j-1)}}{n}\right).$$

Thus, by Markov's inequality (with $p>2+\delta$) we have

$$\begin{aligned}
&\sum_{j=j_0}^{\infty}\mathbb{P}\left(\sup_{\beta\in S_{j,n}}\left\{\tilde{Q}_n^*\left(\beta\right)-\tilde{Q}_n^*\left(\beta_0\right)\right\}\geq 0\right)\\
\leq\ & K\left[\begin{array}{c}\sum_{j=j_0}^{\infty}\left(\frac{2^{2(j-1)}}{n}\right)^{-p}\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{2\text{-step},n}\left(\beta\right)\right|^p\right)\\ +\sum_{j=j_0}^{\infty}\left(\frac{2^{2(j-1)}}{n}\right)^{-p}\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{1,n}\left(\beta\right)\right|^p\right)+\sum_{j=j_0}^{\infty}\left(\frac{2^{2(j-1)}}{n}\right)^{-p}\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{2,n}\left(\beta\right)\right|^p\right)\end{array}\right]\\
\leq\ & K\left[\begin{array}{c}\sum_{j=j_0}^{\infty}2^{-2pj}n^p\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{2\text{-step},n}\left(\beta\right)\right|^p\right)\\ +\sum_{j=j_0}^{\infty}2^{-2pj}n^p\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{1,n}\left(\beta\right)\right|^p\right)+\sum_{j=j_0}^{\infty}2^{-2pj}n^p\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{2,n}\left(\beta\right)\right|^p\right)\end{array}\right],
\end{aligned}$$

where the constant K has changed from the first to second inequality. The crucial part of the proof is to bound each expectation by $O\left(n^{-p}2^{pj}\right)$. This will imply that

$$\begin{aligned}
\mathbb{P}\left(\left|\sqrt{n}\left(\hat{\beta}_n^*-\beta_0\right)\right|>2^{j_0}\right)\ &\leq\ \sum_{j=j_0}^{\infty}\mathbb{P}\left(\sup_{\beta\in S_{j,n}}\left\{\tilde{Q}_n^*\left(\beta\right)-\tilde{Q}_n^*\left(\beta_0\right)\right\}\geq 0\right)\leq K\sum_{j\geq j_0}2^{-pj}\\
&=\ \sum_{j\geq j_0}\left(\frac{1}{2}\right)^{pj}=(1/2)^{pj_0}+(1/2)^{p(j_0+1)}+\dots\\
&=\ \left(\frac{1}{2}\right)^{pj_0}\underbrace{\left(1+(1/2)^p+(1/2)^{2p}+\dots\right)}_{=\frac{1}{1-(1/2)^p}<K}\leq K2^{-pj_0}.
\end{aligned}$$

Since

$$\mathbb{E}\left|\sqrt{n}\left(\hat{\beta}_n^*-\beta_0\right)\right|^{2+\delta}=p\int_0^{\infty}t^{2+\delta-1}\mathbb{P}\left(\left|\sqrt{n}\left(\hat{\beta}_n^*-\beta_0\right)\right|>t\right)dt,$$

we can take the above result with $j_0 = \log_2 t$. This implies

$$\mathbb{P}\left(\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right| > t\right) \leq K 2^{-p\log_2 t} = K 2^{\log_2 t^{-p}} = K t^{-p},$$

and since $p > 2 + \delta$, we can conclude

$$\begin{aligned}\mathbb{E}\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right|^{2+\delta} &= p\int_0^\infty t^{2+\delta-1}\mathbb{P}\left(\left|\sqrt{n}\left(\hat{\beta}_n^* - \beta_0\right)\right| > t\right)dt \\ &\leq K\int_0^\infty t^{2+\delta-1}t^{-p}dt = K\int_0^\infty t^{-1-(p-2+\delta)}dt < \infty.\end{aligned}$$

Bounding $E\left(\sup_{\beta\in S_{j,n}}\left|I_{\text{2-step},n}(\beta)\right|^p\right)$: Recall that

$$\begin{aligned}I_{\text{2-step},n}(\beta) &= \left[Q_{2n}^*(\hat{\alpha}_n^*, \beta) - Q_{2n}^*(\hat{\alpha}_n^*, \beta_0)\right] - \left[Q_{2n}^*(\alpha_0, \beta) - Q_{2n}^*(\alpha_0, \beta_0)\right] \\ &= \left[Q_{2n}^*(\hat{\alpha}_n^*, \beta) - Q_{2n}^*(\alpha_0, \beta)\right] - \left[Q_{2n}^*(\hat{\alpha}_n^*, \beta_0) - Q_{2n}^*(\alpha_0, \beta_0)\right] \\ &= n^{-1}\sum_{t=1}^n\left(q_{2t}^*(\hat{\alpha}_n^*, \beta) - q_{2t}^*(\alpha_0, \beta)\right) - n^{-1}\sum_{t=1}^n\left(q_{2t}^*(\hat{\alpha}_n^*, \beta_0) - q_{2t}^*(\alpha_0, \beta_0)\right)\end{aligned}$$

By taking the Taylor series expansion of q_{2t} around $(\alpha, \beta) = (\alpha_0, \beta_0)$, we have

$$q_{2t}(\alpha, \beta) = q_{2t}(\alpha_0, \beta_0) + \frac{\partial}{\partial\alpha'}q_{2t}(\alpha_0, \beta_0)(\alpha - \alpha_0) + \frac{\partial}{\partial\beta'}q_{2t}(\alpha_0, \beta_0)(\beta - \beta_0) + R_2(\alpha, \beta), \tag{S7.2}$$

such that

$$R_2 = \frac{1}{2!}\left[\begin{array}{c}(\alpha - \alpha_0)'\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}\left(\bar{\alpha}, \bar{\beta}\right)(\alpha - \alpha_0) + (\beta - \beta_0)'\frac{\partial}{\partial\beta\partial\beta'}q_{2t}\left(\bar{\alpha}, \bar{\beta}\right)(\beta - \beta_0) \\ +(\alpha - \alpha_0)'\frac{\partial}{\partial\alpha\partial\beta'}q_{2t}\left(\bar{\alpha}, \bar{\beta}\right)(\beta - \beta_0) + (\beta - \beta_0)'\frac{\partial}{\partial\beta\partial\alpha'}q_{2t}\left(\bar{\alpha}, \bar{\beta}\right)(\alpha - \alpha_0)\end{array}\right]$$

where $\bar{\alpha}$ lies between α and α_0 and $\bar{\beta}$ lies between β and β_0.

Then using (S7.2), we can write

$$\begin{aligned} q_{2t}^{*}\left(\hat{\alpha}_{n}^{*},\beta\right)-q_{2t}^{*}\left(\alpha_{0},\beta\right) &= \frac{\partial}{\partial\alpha'}q_{2t}^{*}\left(\alpha_{0},\beta\right)\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right) \\ &+\frac{1}{2!}\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right)'\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}^{*}\left(\bar{\alpha}_{1},\beta\right)\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right)\end{aligned}$$

where $\bar{\alpha}_1$ lies between $\hat{\alpha}_n^*$ and α_0. Similarly, we have

$$\begin{aligned} q_{2t}^{*}\left(\hat{\alpha}_{n}^{*},\beta_{0}\right)-q_{2t}^{*}\left(\alpha_{0},\beta_{0}\right) &= \frac{\partial}{\partial\alpha'}q_{2t}^{*}\left(\alpha_{0},\beta_{0}\right)\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right) \\ &+\frac{1}{2!}\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right)'\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}^{*}\left(\bar{\alpha}_{2},\beta_{0}\right)\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right),\end{aligned}$$

where $\bar{\alpha}_2$ lies between $\hat{\alpha}_n^*$ and α. It follows that

$$\begin{aligned} I_{\text{2-step},n}\left(\beta\right) &= n^{-1}\sum_{t=1}^{n}\left(\frac{\partial}{\partial\alpha'}q_{2t}^{*}\left(\alpha_{0},\beta\right)-\frac{\partial}{\partial\alpha'}q_{2t}^{*}\left(\alpha_{0},\beta_{0}\right)\right)\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right) \\ &+\frac{1}{2!}n^{-1}\sum_{t=1}^{n}\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right)'\left(\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}^{*}\left(\bar{\alpha}_{1},\beta\right)-\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}^{*}\left(\bar{\alpha}_{2},\beta_{0}\right)\right)\left(\hat{\alpha}_{n}^{*}-\alpha_{0}\right).\end{aligned}$$

Suppose that $\left\{\frac{\partial}{\partial\alpha'}q_{2t}\left(\alpha,\beta\right)\right\}$ and $\left\{\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}\left(\alpha,\beta\right)\right\}$ are Lipschitz continuous in $\left(\alpha,\beta\right)$:

$$\left|\frac{\partial}{\partial\alpha'}q_{2t}\left(\alpha,\beta\right)-\frac{\partial}{\partial\alpha'}q_{2t}\left(\alpha_{0},\beta_{0}\right)\right|\leq L_{1t}\left(X^{t}\right)\left(\left|\alpha-\alpha_{0}\right|+\left|\beta-\beta_{0}\right|\right),$$

and

$$\left|\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}\left(\alpha,\beta\right)-\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}\left(\alpha_{0},\beta_{0}\right)\right|\leq L_{2t}\left(X^{t}\right)\left(\left|\alpha-\alpha_{0}\right|+\left|\beta-\beta_{0}\right|\right),$$

where the functions $L_{1t}\left(X^t\right)$ and $L_{2t}\left(X^t\right)$ do not depend on α nor β. Thus, we have

$$\left|\frac{\partial}{\partial\alpha'}q_{2t}^{*}\left(\alpha_{0},\beta\right)-\frac{\partial}{\partial\alpha'}q_{2t}^{*}\left(\alpha_{0},\beta_{0}\right)\right|\leq L_{1t}^{*}\left(\left|\beta-\beta_{0}\right|\right), \tag{S7.3}$$

and similarly,

$$\begin{aligned}\left|\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}^*(\alpha_1,\beta)-\frac{\partial}{\partial\alpha\partial\alpha'}q_{2t}^*(\alpha_2,\beta_0)\right| &\leq L_{2t}^*\left(|\bar{\alpha}_1-\bar{\alpha}_2|+|\beta-\beta_0|\right)\\ &\leq L_{2t}^*\left(|\hat{\alpha}_n^*-\alpha_0|+|\beta-\beta_0|\right), \qquad \text{(S7.4)}\end{aligned}$$

where the last inequality follows because both $\bar{\alpha}_1$ and $\bar{\alpha}_2$ lie between $\hat{\alpha}_n^*$ and α_0. Therefore by the triangular inequality and using (S7.3) and (S7.4), we have

$$\begin{aligned}\left|I_{\text{2-step},n}(\beta)\right| &\leq n^{-1}\left(n^{-1}\sum_{t=1}^{n}L_{1t}^*\right)\left|\sqrt{n}\left(\hat{\alpha}_n^*-\alpha_0\right)\right|\left|\sqrt{n}\left(\beta-\beta_0\right)\right|\\ &\quad +n^{-3/2}\frac{1}{2!}\left(n^{-1}\sum_{t=1}^{n}L_{2t}^*\right)\left|\sqrt{n}\left(\hat{\alpha}_n^*-\alpha_0\right)\right|^2\left(\sqrt{n}\left|\hat{\alpha}_n^*-\alpha_0\right|+\left|\sqrt{n}\left(\beta-\beta_0\right)\right|\right).\end{aligned}$$

Hence, successive applications of the Hölder's inequality yields

$$\begin{aligned}&E\left(\sup_{\beta\in S_{j,n}}\left|I_{\text{2-step},n}(\beta)\right|^p\right)\\ \leq\ & Kn^{-p}\left(\mathbb{E}\left(\left|\sqrt{n}\left(\hat{\alpha}_n^*-\alpha_0\right)\right|^{\varepsilon p}\right)\right)^{\frac{1}{\varepsilon}}\left(\mathbb{E}\left(\left(n^{-1}\sum_{t=1}^{n}L_{1t}^*\right)^{\frac{\varepsilon}{\varepsilon-1}p}\sup_{\beta\in S_{j,n}}\left|\sqrt{n}\left(\beta-\beta_0\right)\right|^{\frac{\varepsilon}{\varepsilon-1}p}\right)\right)^{\frac{\varepsilon-1}{\varepsilon}}\\ &+Kn^{-\frac{3p}{2}}\left(\mathbb{E}\left(\left|\sqrt{n}\left(\hat{\alpha}_n^*-\alpha_0\right)\right|^{3\varepsilon p}\right)\right)^{\frac{1}{\varepsilon}}\left(\mathbb{E}\left(n^{-1}\sum_{t=1}^{n}L_{2t}^*\right)^{\frac{\varepsilon}{\varepsilon-1}p}\right)^{\frac{\varepsilon-1}{\varepsilon}}\\ &+Kn^{-\frac{3p}{2}}\left(\mathbb{E}\left(\left|\sqrt{n}\left(\hat{\alpha}_n^*-\alpha_0\right)\right|^{2\varepsilon p}\right)\right)^{\frac{1}{\varepsilon}}\left(\mathbb{E}\left(\left(n^{-1}\sum_{t=1}^{n}L_{2t}^*\right)^{\frac{\varepsilon}{\varepsilon-1}p}\sup_{\beta\in S_{j,n}}\left|\sqrt{n}\left(\beta-\beta_0\right)\right|^{\frac{\varepsilon}{\varepsilon-1}p}\right)\right)^{\frac{\varepsilon-1}{\varepsilon}}\end{aligned}$$

for some $\varepsilon > 1$. Note that for $\beta \in S_{j,n}$, we have $|\sqrt{n}\,(\beta - \beta_0)| \leq 2^j$. This implies that

$$
\begin{aligned}
E\left(\sup_{\beta \in S_{j,n}} |I_{2\text{-step},n}(\beta)|^p\right) \leq\; & Kn^{-p}\left(\mathbb{E}\left(\left|\sqrt{n}\,(\hat{\alpha}_n^* - \alpha_0)\right|^{\varepsilon p}\right)\right)^{\frac{1}{\varepsilon}} 2^{pj}\left(\mathbb{E}\left(\left(n^{-1}\sum_{t=1}^{n} L_{1t}^*\right)^{\frac{\varepsilon}{\varepsilon-1}p}\right)\right)^{\frac{\varepsilon-1}{\varepsilon}} \\
& + Kn^{-\frac{3p}{2}}\left(\mathbb{E}\left(\left|\sqrt{n}\,(\hat{\alpha}_n^* - \alpha_0)\right|^{3\varepsilon p}\right)\right)^{\frac{1}{\varepsilon}}\left(\mathbb{E}\left(n^{-1}\sum_{t=1}^{n} L_{2t}^*\right)^{\frac{\varepsilon}{\varepsilon-1}p}\right)^{\frac{\varepsilon-1}{\varepsilon}} \\
& + Kn^{-\frac{3p}{2}}\left(\mathbb{E}\left(\left|\sqrt{n}\,(\hat{\alpha}_n^* - \alpha_0)\right|^{2\varepsilon p}\right)\right)^{\frac{1}{\varepsilon}} 2^{pj}\left(\mathbb{E}\left(\left(n^{-1}\sum_{t=1}^{n} L_{2t}^*\right)^{\frac{\varepsilon}{\varepsilon-1}p}\right)\right)^{\frac{\varepsilon-1}{\varepsilon}}
\end{aligned}
$$

Suppose we assume that $E\left(|\sqrt{n}\,(\hat{\alpha}_n^* - \alpha_0)|^{3\varepsilon p}\right) < \infty$. If in addition we assume that $E\left(|L_{1t}|^{\frac{\varepsilon}{\varepsilon-1}p}\right) < \infty$ and $E\left(|L_{2t}|^{\frac{\varepsilon}{\varepsilon-1}p}\right) < \infty$, we can show that the expectations of average of the functions involving L_{1t}^* and L_{2t}^* are bounded. For instance,

$$
\begin{aligned}
\mathbb{E}\left(n^{-1}\sum_{t=1}^{n} L_{1t}^*\right)^{\frac{\varepsilon}{\varepsilon-1}p} &\leq Kn^{-1}\sum_{t=1}^{n}\mathbb{E}\left(|L_{1t}^*|^{\frac{\varepsilon}{\varepsilon-1}p}\right) \\
&= Kn^{-1}\sum_{t=1}^{n}\left(E\left(E^*\left(|L_{1t}^*|^{\frac{\varepsilon}{\varepsilon-1}p}\right)\right)\right) \\
&= KEE^*\left(n^{-1}\sum_{t=1}^{n}|L_{1t}^*|^{\frac{\varepsilon}{\varepsilon-1}p}\right) \\
&= KE\left(\sum_{t=1}^{n}\gamma_{nt}\,|L_{1t}|^{\frac{\varepsilon}{\varepsilon-1}p}\right) < \infty \text{ if } E\left(|L_{1t}|^{\frac{\varepsilon}{\varepsilon-1}p}\right) < \infty.
\end{aligned}
$$

Thus, under these assumptions

$$
\mathbb{E}\left(\sup_{\beta \in S_{j,n}} |I_{2\text{-step},n}(\beta)|^p\right) \leq Kn^{-p}2^{pj},
$$

which implies

$$\begin{aligned}
\sum_{j=j_0}^{\infty} 2^{-2pj} n^p \mathbb{E}\left(\sup_{\beta \in S_{j,n}} \left| I_{\text{2-step},n}(\beta)\right|^p\right) &\leq K \sum_{j=j_0}^{\infty} 2^{-2pj} 2^{pj} \underbrace{n^p n^{-p}}_{=1} \\
&= K \sum_{j=j_0}^{\infty} 2^{-pj} \\
&\leq K 2^{-pj_0}.
\end{aligned}$$

Bounding $\mathbb{E}\left(\sup_{\beta \in S_{j,n}} \left| I_{1,n}(\beta)\right|^p\right)$: Note that by definition of $I_{1,n}(\beta)$, we have that

$$\begin{aligned}
I_{1,n}(\beta) &= Q_{2n}^*(\alpha_0,\beta) - Q_{2n}^*(\alpha_0,\beta_0) - E^*\left[Q_{2n}^*(\alpha_0,\beta) - Q_{2n}^*(\alpha_0,\beta_0)\right] \\
&= n^{-1}\sum_{t=1}^{n}\left(q_{2t}^*(\alpha_0,\beta) - q_{2t}^*(\alpha_0,\beta_0)\right) - E^*\left(n^{-1}\sum_{t=1}^{n}\left(q_{2t}^*(\alpha_0,\beta) - q_{2t}^*(\alpha_0,\beta_0)\right)\right) \\
&\equiv n^{-1/2}\mathbb{G}_n^*\left(q_2(\alpha_0,\beta) - q_2(\alpha_0,\beta_0)\right),
\end{aligned}$$

where for a class of functions $\mathcal{F} = \{f\}$, we define the empirical process $\mathbb{G}_n^* f$ as

$$\mathbb{G}_n^* f = n^{-1/2}\sum_{t=1}^{n}\left(f_t^* - E^* f_t^*\right).$$

Define the L^p norm of $\mathbb{G}_n^* f$ over $\mathcal{F}$ as

$$\left(\mathbb{E}\left|\mathbb{G}_n^*\right|_{\mathcal{F}}^p\right)^{1/p} = \left(\mathbb{E}\left(\sup_{f\in\mathcal{F}}\left|\mathbb{G}_n^* f\right|\right)^p\right)^{1/p}.$$

With this notation,

$$\begin{aligned}
\mathbb{E}\left(\sup_{\beta\in S_{j,n}}|I_{1,n}(\beta)|^p\right) &= \mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|n^{-1/2}\mathbb{G}_n^*\left(q_{2t}(\alpha_0,\beta)-q_{2t}(\alpha_0,\beta_0)\right)\right|^p\right)\\
&= n^{-p/2}\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|\mathbb{G}_n^*\left(q_{2t}(\alpha_0,\beta)-q_{2t}(\alpha_0,\beta_0)\right)\right|^p\right)\\
&= n^{-p/2}\left\{\left(\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|\mathbb{G}_n^*\left(q_{2t}(\alpha_0,\beta)-q_{2t}(\alpha_0,\beta_0)\right)\right|^p\right)\right)^{1/p}\right\}^p\\
&= n^{-p/2}\left(\left(\mathbb{E}\left|\mathbb{G}_n^*\right|^p_{\mathcal{N}_\delta}\right)^{1/p}\right)^p,
\end{aligned}$$

where we let $\mathcal{N}_\eta = \{q_2(\alpha_0,\beta)-q_2(\alpha_0,\beta_0) : |\beta-\beta_0|\leq\eta,\ (\alpha,\beta)\in\mathcal{A}\times\mathcal{B}\}$. Lemma S6.2 shows that for any $\eta>0$, $\left(\mathbb{E}\left|\mathbb{G}_n^*\right|^p_{\mathcal{N}_\eta}\right)^{1/p}\leq\eta$ holds under our assumptions. Thus, letting $\eta=\frac{2^j}{\sqrt{n}}$ yields $\left(\mathbb{E}\left|\mathbb{G}_n^*\right|^p_{\mathcal{N}_\eta}\right)^{1/p}\leq\left(\frac{2^j}{\sqrt{n}}\right)^p$, implying that

$$\mathbb{E}\left(\sup_{\beta\in S_{j,n}}|I_{1,n}(\beta)|^p\right)\leq n^{-p/2}\frac{2^{pj}}{n^{p/2}}=n^{-p}2^{pj}$$

It follows that

$$\sum_{j=j_0}^{\infty}2^{-2pj}n^p\mathbb{E}\left(\sup_{\beta\in S_{j,n}}|I_{1,n}(\beta)|^p\right)\leq\sum_{j=j_0}^{\infty}2^{-2pj}n^pn^{-p}2^{pj}=\sum_{j=j_0}^{\infty}2^{-pj}\leq K2^{-pj_0},$$

as above.

Bounding $E\left(\sup_{\beta\in S_{j,n}}|I_{2,n}(\beta)|^p\right)$: The argument is similar. By definition of $I_{2,n}(\beta)$, we have

$$
\begin{aligned}
I_{2,n}(\beta) &= E^*\left[Q_{2n}^*(\alpha_0,\beta) - Q_{2n}^*(\alpha_0,\beta_0)\right] - E\left(E^*\left[Q_{2n}^*(\alpha_0,\beta) - Q_{2n}^*(\alpha_0,\beta_0)\right]\right) \\
&= n^{-1}\sum_{t=1}^{n} E^*\left(q_{2t}^*(\alpha_0,\beta) - q_{2t}^*(\alpha_0,\beta_0)\right) - n^{-1}\sum_{t=1}^{n} E\left(E^*\left(q_{2t}^*(\alpha_0,\beta) - q_{2t}^*(\alpha_0,\beta_0)\right)\right) \\
&= \sum_{t=1}^{n}\gamma_{nt}\left[\left(q_{2t}(\alpha_0,\beta) - q_{2t}(\alpha_0,\beta_0)\right) - E\left(q_{2t}(\alpha_0,\beta) - q_{2t}(\alpha_0,\beta_0)\right)\right] \\
&= n^{-1/2}\left(\sum_{t=1}^{n}\sqrt{n}\gamma_{nt}\left[\left(q_{2t}(\alpha_0,\beta) - q_{2t}(\alpha_0,\beta_0)\right) - E\left(q_{2t}(\alpha_0,\beta) - q_{2t}(\alpha_0,\beta_0)\right)\right]\right) \\
&= n^{-1/2}\mathbb{G}_{n,\gamma}\left(q_2(\alpha_0,\beta) - q_2(\alpha_0,\beta_0)\right),
\end{aligned}
$$

where we define the empirical process $\mathbb{G}_{n,\gamma}$ as

$$
\mathbb{G}_{n,\gamma}f = \sum_{t=1}^{n}\sqrt{n}\gamma_{nt}\left(f_t - Ef_t\right),
$$

with weights defined as above. Similarly, we define the L^p norm of $\mathbb{G}_{n,\gamma}f$ over $\mathcal{F} = \{f\}$ as

$$
\left(E\left|\mathbb{G}_{n,\gamma}\right|_{\mathcal{F}}^{p}\right)^{1/p} = \left(E\left(\sup_{f\in\mathcal{F}}\left|\mathbb{G}_{n,\gamma}f\right|\right)^{p}\right)^{1/p}.
$$

With this notation

$$
\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{2,n}(\beta)\right|^{p}\right) = n^{-p/2}\left(\left(E\left|\mathbb{G}_{n,\gamma}\right|_{\mathcal{F}}^{p}\right)^{1/p}\right)^{p}.
$$

It suffices to bound $\left(E\left|\mathbb{G}_{n,\gamma}\right|_{\mathcal{F}}^{p}\right)^{1/p}$. Assumption B6(ii) provides a bound on the L_p-norm of the empirical process $\mathbb{G}_n$, which differs from $\mathbb{G}_{n,\gamma}$ due to presence of the weights γ_{nt}. It is well known that these weights are introduced by the fact that the MBB puts less weight on the first and last ℓ observations in the sample. In particular, we can show that for any function f_t, the MBB expectation $E^*\left(\bar{f}_n^*\right) = \sum_{t=1}^{n}\gamma_{nt}f_t = n^{-1}\sum_{t=1}^{n}f_t + O_P\left(\frac{\ell}{n}\right)$. Using this insight, we can show that

$$\mathbb{G}_{n,\gamma}f = \frac{n}{n-\ell+1}\mathbb{G}_n f - \frac{n}{n-\ell+1}\mathbb{R}_{1n}f - \frac{n}{n-\ell+1}\mathbb{R}_{2n}f,$$

where

$$\begin{aligned}
\mathbb{R}_{1n}f &= \frac{1}{\sqrt{n}}\sum_{t=1}^{\ell}\left(1-\frac{t}{\ell}\right)(f_t - Ef_t), \\
\mathbb{R}_{2n}f &= \frac{1}{\sqrt{n}}\sum_{t=1}^{\ell}\left(1-\frac{t}{\ell}\right)(f_{n-t+1} - Ef_{n-t+1}).
\end{aligned}$$

By Minkowski's inequality,

$$\begin{aligned}
\left(E\left|\mathbb{G}_{n,\gamma}\right|_{\mathcal{F}}^{p}\right)^{1/p} &\leq \frac{n}{n-\ell+1}\left\{\left(E\left|\mathbb{G}_n\right|_{\mathcal{F}}^{p}\right)^{1/p} + \left(E\left|\mathbb{R}_{1n}\right|_{\mathcal{F}}^{p}\right)^{1/p} + \left(E\left|\mathbb{R}_{2n}\right|_{\mathcal{F}}^{p}\right)^{1/p}\right\} \\
&\leq K\left(\left(E\left|\mathbb{G}_n\right|_{\mathcal{F}}^{p}\right)^{1/p} + \left(E\left|\mathbb{R}_{1n}\right|_{\mathcal{F}}^{p}\right)^{1/p} + \left(E\left|\mathbb{R}_{2n}\right|_{\mathcal{F}}^{p}\right)^{1/p}\right), \qquad \text{(S7.5)}
\end{aligned}$$

for some constant K since $\ell \to \infty$ such that $\ell = o(\sqrt{n})$ under our assumptions. This implies

$$\begin{aligned}
\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{2,n}(\beta)\right|^{p}\right) &= n^{-p/2}\left(\left(E\left|\mathbb{G}_{n,\gamma}\right|_{\mathcal{F}}^{p}\right)^{1/p}\right)^{p} \\
&\leq Kn^{-p/2}\left(\left(E\left|\mathbb{G}_n\right|_{\mathcal{F}}^{p}\right)^{1/p} + \left(E\left|\mathbb{R}_{1n}\right|_{\mathcal{F}}^{p}\right)^{1/p} + \left(E\left|\mathbb{R}_{2n}\right|_{\mathcal{F}}^{p}\right)^{1/p}\right)^{p} \\
&\leq Kn^{-p/2}\left(\frac{2^{jp}}{n^{p/2}} + E\left|\mathbb{R}_{1n}\right|_{\mathcal{F}}^{p} + E\left|\mathbb{R}_{2n}\right|_{\mathcal{F}}^{p}\right),
\end{aligned}$$

where we have used Assumption B6(ii) with $\eta = \frac{2^j}{\sqrt{n}}$ to bound $\left(E\left|\mathbb{G}_n\right|_{\mathcal{F}}^{p}\right)^{1/p}$. The remainder terms can be bounded by $O\left(\left(\frac{\ell}{\sqrt{n}}\right)^{p}\frac{2^{jp}}{n^{p}}\right)$ using the Lipschitz condition given in Assumption B6(iii), where the Lipschitz function for the log likelihood function $\{q_{2t}(\alpha,\beta)\}$ has a finite p^{th} order moment. Since $\ell = o(\sqrt{n})$ by assumption, the contribution of the two remainder terms is smaller than that of the first term. We can then claim that

$$\mathbb{E}\left(\sup_{\beta\in S_{j,n}}\left|I_{2,n}\left(\beta\right)\right|^{p}\right)\leq Kn^{-p/2}\frac{2^{jp}}{n^{p/2}}=Kn^{-p}2^{jp},$$

and the proof follows as above.

References

[1] Cheng, G. (2015). Moment Consistency of the Exchangeably Weighted Bootstrap for Semiparametric M estimation. *Scandinavian Journal of Statistics*, 42, 665-684.

[2] Dovonon, P., and S. Goncalves (2017). Bootstrapping the GMM overidentification test under first-order underidentification. *Journal of Econometrics*, 201, 43-71.

[3] Gonçalves, S. and H. White (2004). Maximum likelihood and the bootstrap for nonlinear dynamic models. *Journal of Econometrics*, 119, 199-219.

[4] Newey, W.K., McFadden, D. (1994). Large sample estimation and hypothesis testing. In: Engle, R. McFadden, D. (Eds.), *Handbook of Econometrics, vol IV*, North Holland.

[5] Praestgaard, J., and J. A. Wellner (1993). Exchangeably weighted bootstraps of the general empirical process. *The Annals of Probability*, 2053-2086.

[6] van der Vaart, A. W., and J. A. Wellner (1996). Weak convergence and empirical processes: with applications to statistics, *Springer,* New York.

[7] White, H. (1994). *Estimation, Inference and Specification Analysis.* Econometric Society Monographs 22, Cambridge University Press: Cambridge.

[8] Wooldridge, J. (1994). Estimation and inference for dependent processes. In: Engle, R. McFadden, D. (Eds.), *Handbook of Econometrics, vol IV*, North Holland.