Supplementary Material on
"Efficiency in Large Dynamic Panel Models with Common Factors",
Patrick Gagliardini and Christian Gouriéroux

This supplementary material provides the Limit Theorems for uniform stochastic convergence (Appendix B) and the technical Lemmas (Appendix C) used in the proofs of Propositions 1, 2, 3, 5 and 6.

APPENDIX B
LIMIT THEOREMS

In Section B.1 we consider the uniform consistency of the cross-sectional factor approximations (Theorem 1). We provide in Section B.2 the uniform convergence of time series averages of factor approximations (Theorem 2). In Section B.3 we consider the uniform convergence of nonlinear aggregates of cross-sectional and time series averages (Theorem 3). The secondary Lemmas B.1-B.5 used in the proofs of Theorems 1-3 are provided in Section B.4.

B.1 Uniform consistency of the factor approximations

In Limit Theorem 1 we give the convergence rate of the factor approximation $\hat{f}_{n,t}(\beta)$ defined in equation (3.3), uniformly across dates $1 \leq t \leq T$ and micro-parameter values $\beta \in \mathcal{B}$.

THEOREM 1 *Under Assumptions A.1-A.5, Assumptions H.1, H.2, H.5, H.6, H.7 (i)-(ii), H.8-H.10 in Appendix A.1, and if $n, T \to \infty$ such that $T^\nu / n = O(1)$ for a value $\nu > 1$:*

$$\sup_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| = O_p\left(\frac{(\log n)^{\delta_2}}{\sqrt{n}} \right),$$

where $f_t(\beta)$ is defined in equation (4.3), $\delta_2 = \gamma_2 + \gamma_3/2 + 2/d_3 + 1/2$ and constants $\gamma_2, \gamma_3 \geq 0$, $d_3 > 0$ are defined in Assumptions H.8-H.10.

Proof of Theorem 1: Let

$$\varepsilon_n = r \frac{(\log n)^{\delta_2}}{\sqrt{n}}, \tag{b.1}$$

where $r > 0$ is a constant. We have to show that, for any $\eta > 0$, there exists a value of r such that $\mathbb{P}\left[\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n\right] \leq \eta$, for $n, T \to \infty$ such that $T^\nu/n = O(1)$, $\nu > 1$. We have:

$$\begin{aligned}\mathbb{P}\left[\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n\right] &\leq T\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n\right] \\ &= TE\left[\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]\right]. \qquad \text{(b.2)}\end{aligned}$$

Conditional on factor path $\underline{f_t}$, the estimator $\widehat{f}_{n,t}(\beta)$ is the concentrated ML estimator of "parameter" f_t given the "nuisance" parameter β, computed on the sample $(y_{i,t}, y_{i,t-1})$, $i = 1, ..., n$. This sample is i.i.d. conditional on $\underline{f_t}$. Thus, the strategy of the proof is to first use a large deviation result for i.i.d. data to get an upper bound for $\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]$, for given sample size n and date t, as a function of $\underline{f_t}$. Then, we compute the expectation of this bound w.r.t. $\underline{f_t}$, and establish the asymptotic behaviour of the RHS of inequality (b.2).

i) Bound of $\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]$

By equation (3.3) we have $\hat{f}_{n,t}(\beta) = \arg\max_{f\in\mathcal{F}_n}\frac{1}{n}\sum_{i=1}^{n} l_{i,t}(\alpha)$, where $l_{i,t}(\alpha) = \log h(y_{i,t}|y_{i,t-1}, f; \beta)$ and $\alpha = (f', \beta')'$. To bound the probability $\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]$ for a given sample size n and date t, we use the large deviation result of Lemma B.1 in Appendix B.4.1. We replace density $l_i(\alpha)$ in Lemma B.1 by $l_{i,t}(\alpha)$, parameter set $\mathcal{F}$ by $\mathcal{F}_n$, and work with the conditional distribution of the data $(y_{i,t}, y_{i,t-1})$ given the factor path $\underline{f_t}$.

Lemma B.1 differs from large deviation results for ML estimators derived in the literature [1] since it makes fully explicit how the upper bound on the probability of large deviation of the ML estimate depends on the distribution of the data and on the parameter set, for given sample size. In available results, this dependence is partly hidden in some generic constants in the bound. In our framework, the upper bound for $\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left\|\widehat{f}_{n,t}(\beta) - f_t(\beta)\right\| \geq \varepsilon_n \mid \underline{f_t}\right]$ is stochastic and depends on the factor path $\underline{f_t}$. Knowing the pattern of this dependence

[1] See the classical results in Bahadur (1960, 1967) on the asymptotic behavior of the probability of large deviation of ML estimates for a scalar parameter with i.i.d. data, the work along similar lines in e.g. Fu (1982), Lemmas 2 and 3 for Sieve estimators in Shen and Wong (1994), the result used in the proof of Theorem 1 in Chen, Shen (1998), p. 309, with weakly dependent data.

explicitly is necessary when the factor path is integrated out in the second step of the proof. Moreover, Lemma B.1 allows to make explicit how the upper bound depends on the parameter set $\mathcal{F}_n$. This is necessary for the asymptotic analysis, since the parameter set $\mathcal{F}_n$ is expanding w.r.t. n.

Let us check the conditions of Lemma B.1, and consider first the realizations of the factor path $\underline{f_t}$ such that $f_t(\beta) \in \mathcal{F}_n$ for any $\beta \in \mathcal{B}$. Condition i) of Lemma B.1 is implied by Assumptions H.1 and H.7 (i). Condition ii) of Lemma B.1 is satisfied from Assumption H.2. Condition iii) of Lemma B.1 with $\gamma_{11} = 4$ is implied by Assumption H.9 and:

$$E_0 \left[\sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_n} \left\| \frac{\partial \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial (\beta', f')'} \right\|^4 |\underline{f_t} \right] \leq [\log(n)]^{\gamma_3} \mathcal{R}_t.$$

Let us now check Condition iv) of Lemma B.1. By the first-order condition defining the pseudo-true factor value $E_0 \left[\frac{\partial \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f_t} |\underline{f_t} \right] = 0$, and the implicit function theorem, we deduce that function $f_t(\beta)$ is differentiable w.r.t. β, $\mathbb{P}$-a.s., and:

$$\frac{\partial f_t(\beta)}{\partial \beta'} = -I_{t,ff}(\beta)^{-1} I_{t,f\beta}(\beta),$$

where the matrices $I_{t,ff}(\beta)$ and $I_{t,f\beta}(\beta)$ are the (f, f) and (f, β) blocks of the Hessian matrix $I_t(\beta) = E_0 \left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial (\beta', f_t')' \partial (\beta', f_t')} |\underline{f_t} \right]$. Moreover, we have $\sup_{\beta \in \mathcal{B}} \|I_{t,ff}(\beta)^{-1}\| \leq \tilde{c} \xi_{t,1}^*$, for a contant $\tilde{c} > 0$, and $\sup_{\beta \in \mathcal{B}} \|I_{t,f\beta}(\beta)\| \leq (\xi_{t,1}^{**})^{1/2}$, where processes $\xi_{t,1}^*$ and $\xi_{t,1}^{**}$ are defined in Assumption H.5. Therefore, we get:

$$\mathcal{M}_t \equiv \sup_{\beta \in \mathcal{B}} \left\| \frac{\partial f_t(\beta)}{\partial \beta'} \right\| \leq \tilde{c} \xi_{t,1}^* (\xi_{t,1}^{**})^{1/2}, \tag{b.3}$$

and $\mathcal{M}_t < \infty$, $\mathbb{P}$-a.s., from Assumption H.5. Finally, the bounds in equations (b.26) and (b.27) are satisfied, since:

$$\inf_{\beta \in \mathcal{B}} \inf_{f \in \mathcal{F}_n : f \neq f_t(\beta)} \frac{2KL_t(f, f_t(\beta); \beta)}{\|f - f_t(\beta)\|^2} \geq [\log(n)]^{-\gamma_2} \mathcal{K}_t,$$

and:

$$\sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_n} E_0 \left[\left\| \frac{\partial \log h(y_{i,t}|y_{i,t-1}, f; \beta)}{\partial f} \right\|^2 |\underline{f_t} \right] \leq [\log(n)]^{\gamma_3} \Gamma_t,$$

where processes $\mathcal{K}_t$ and Γ_t are defined in Assumptions H.8 and H.10.

From Lemma B.1 and the definition of ε_n in equation (b.1), we get:

$$
\begin{aligned}
& \mathbb{P}\left[\sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| \geq \varepsilon_n \mid \underline{f_t}\right] \\
\leq \; & C_1 Vol(\mathcal{B})(1+\mathcal{M}_t)^{m+q} \frac{n^{m+q}}{\varepsilon_n^q} \exp\left(-C_2 n \varepsilon_n^2 \frac{[\log(n)]^{-\gamma_2}\mathcal{K}_t}{1+[\log(n)]^{\gamma_2+\gamma_3}\Gamma_t/\mathcal{K}_t}\right) + C_3 \varepsilon_n^2 [\log(n)]^{\gamma_2+\gamma_3} \frac{\mathcal{R}_t}{\mathcal{K}_t} \\
\leq \; & \frac{C_1}{r^q} Vol(\mathcal{B})(1+\mathcal{M}_t)^{m+q} n^{m+3q/2} \exp\left(-C_2 r^2 [\log(n)]^{1+4/d_3} \frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right) \\
& + C_3 \frac{r^2}{n} [\log(n)]^{2\delta_2+\gamma_2+\gamma_3} \frac{\mathcal{R}_t}{\mathcal{K}_t},
\end{aligned}
$$

for any factor path such that $f_t(\beta) \in \mathcal{F}_n$ for any $\beta \in \mathcal{B}$, where $Vol(\mathcal{B}) = \int_{\mathcal{B}} d\lambda$ is the Lebesgue measure of set $\mathcal{B}$, and C_1, C_2, C_3 are constants independent of $\underline{f_t}$ and n, T. Thus, we get:

$$
\begin{aligned}
& \mathbb{P}\left[\sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| \geq \varepsilon_n \mid \underline{f_t}\right] \\
& \quad \leq \frac{C_1}{r^q} Vol(\mathcal{B})(1+\mathcal{M}_t)^{m+q} n^{m+3q/2} \exp\left(-C_2 r^2 [\log(n)]^{1+4/d_3} \frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right) \\
& \quad + C_3 \frac{r^2}{n} [\log(n)]^{2\delta_2+\gamma_2+\gamma_3} \frac{\mathcal{R}_t}{\mathcal{K}_t} + 1\left\{\bigcup_{\beta \in \mathcal{B}} [f_t(\beta) \in \mathcal{F}_n^c]\right\},
\end{aligned} \tag{b.4}
$$

for any factor path $\underline{f_t}$, $\mathbb{P}$-a.s.

ii) Integrating out the factor path

By integrating out the factor path $\underline{f_t}$, we get from inequalities (b.2) and (b.4):

$$
\begin{aligned}
& \mathbb{P}\left[\sup_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}} \left\| \widehat{f}_{n,t}(\beta) - f_t(\beta) \right\| \geq \varepsilon_n\right] \\
\leq \; & \frac{C_1}{r^q} Vol(\mathcal{B}) T n^{m+3q/2} E\left[(1+\mathcal{M}_t)^{m+q} \exp\left(-C_2 r^2 [\log(n)]^{1+4/d_3} \frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right)\right] \\
& + C_3 T \frac{r^2}{n} [\log(n)]^{2\delta_2+\gamma_2+\gamma_3} E\left[\frac{\mathcal{R}_t}{\mathcal{K}_t}\right] + T\mathbb{P}\left[\bigcup_{\beta \in \mathcal{B}} [f_t(\beta) \in \mathcal{F}_n^c]\right] \\
\equiv \; & I_{1,n,T} + I_{2,n,T} + I_{3,n,T}.
\end{aligned} \tag{b.5}
$$

Let us now bound these three terms and prove that they are $o(1)$.

(a) From the Cauchy-Schwarz inequality, term $I_{1,n,T}$ is such that:

$$I_{1,n,T} \leq \frac{C_1}{r^q} Vol(\mathcal{B}) T n^{m+3q/2} E\left[(1+\mathcal{M}_t)^{2m+2q}\right]^{1/2} E\left[\exp\left(-2C_2 r^2 [\log(n)]^{1+4/d_3} \frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right)\right]^{1/2}. \tag{b.6}$$

The first expectation in the RHS is finite. Indeed, from inequality (b.3) and Assumption H.5, we have $\mathbb{P}[\mathcal{M}_t \geq u] \leq \tilde{b}_1 \exp(-\tilde{c}_1 u^{\tilde{d}_1})$, as $u \to \infty$, for some constants $\tilde{b}_1, \tilde{c}_1, \tilde{d}_1 > 0$. Thus, the stationary distribution of process $\mathcal{M}_t$ admits finite moments of any order, and $E\left[(1+\mathcal{M}_t)^{2m+2q}\right] < \infty$. To bound the second expectation in the RHS of (b.6) we use Lemma B.2 in Appendix B.4.2, which provides a bound of the expectation $E[\exp(-uW^{-1})]$ from the tail behavior of the positive random variable W. Let us verify that the variable $W \equiv W_t = (1+\Gamma_t/\mathcal{K}_t)/\mathcal{K}_t$ satisfies the condition of Lemma B.2. From Assumption H.10 we have:

$$\begin{aligned}
\mathbb{P}[W \geq u] &\leq P[\mathcal{K}_t^{-1} \geq u/2] + \mathbb{P}[\Gamma_t \mathcal{K}_t^{-2} \geq u/2] \\
&\leq \mathbb{P}[\mathcal{K}_t^{-1} \geq u/2] + \mathbb{P}[\Gamma_t \geq (u/2)^{1/2}] + \mathbb{P}[\mathcal{K}_t^{-1} \geq (u/2)^{1/4}] \leq 3b_3 \exp[-c_3 (u/2)^{d_3/4}].
\end{aligned}$$

By applying Lemma B.2 with $\varrho = d_3/4$, we get:

$$\begin{aligned}
E\left[\exp\left(-2C_2 r^2 [\log(n)]^{1+4/d_3} \frac{\mathcal{K}_t}{1+\Gamma_t/\mathcal{K}_t}\right)\right] &\leq \tilde{C}_1 \exp\left[-\tilde{C}_2 (2C_2 r^2)^{d_3/(d_3+4)} \log(n)\right] \\
&= \tilde{C}_1 n^{-\tilde{C}_2 (2C_2 r^2)^{d_3/(d_3+4)}},
\end{aligned} \tag{b.7}$$

for some constants $\tilde{C}_1, \tilde{C}_2 > 0$. Thus, from inequalities (b.6) and (b.7), we get:

$$I_{1,n,T} \leq \frac{C_1}{r^q} Vol(\mathcal{B}) \tilde{C}_1^{1/2} E\left[(1+\mathcal{M}_t)^{2m+2q}\right]^{1/2} T n^{m+3q/2-(\tilde{C}_2/2)(2C_2 r^2)^{d_3/(d_3+4)}} = O(T/n) = o(1),$$

if $m + 3q/2 - (\tilde{C}_2/2)(2C_2 r^2)^{d_3/(d_3+4)} \leq -1$, i.e., if $r \geq \frac{1}{\sqrt{2C_2}} \left(\frac{m+3q/2+1}{\tilde{C}_2/2}\right)^{1/2+2/d_3}$.

(b) Let us now consider the second term in the RHS of inequality (b.5). From Assumptions H.9 and H.10, $E\left[\frac{\mathcal{R}_t}{\mathcal{K}_t}\right] \leq E\left[\mathcal{R}_t^2\right]^{1/2} E\left[\mathcal{K}_t^{-2}\right]^{1/2} < \infty$. Then, from the condition $T^\nu/n = O(1)$ for $\nu > 1$, we get $I_{2,n,T} = o(1)$.

(c) Finally, from Assumptions H.6 and H.7 (ii), we have:

$$\mathbb{P}\left[\bigcup_{\beta\in\mathcal{B}}[f_t(\beta)\in\mathcal{F}_n^c]\right]\leq\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\|f_t(\beta)\|\geq r_n\right]\leq b_2\exp\left(-c_2r_n^{d_2}\right)=b_2n^{-2}.$$

Since $T/n^2=o(1)$, we get $I_{3,n,T}=o(1)$.

B.2 Uniform consistency of time series averages of factor approximations

Limit Theorem 2 provides a uniform convergence result for time series averages of nonlinear transformations of current and lagged factor approximations $\hat{f}_{n,t}(\beta)$. These nonlinear transformations can involve the macro-parameter θ. The uniformity property concerns both parameters $\beta\in\mathcal{B}$ and $\theta\in\Theta$.

THEOREM 2 *Let Assumptions A.1-A.5, H.1, H.2, H.4 (i), H.5, H.6, H.7 (i)-(ii), H.8-H.10 hold, and assume that function $G(f_t,f_{t-1};\theta)$ satisfies the Regularity Condition RC.1 below. Then, if $n,T\to\infty$ such that $T^\nu/n=O(1)$ for a value $\nu>1$:*

$$\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left\|\frac{1}{T}\sum_{t=1}^{T}G(\hat{f}_{n,t}(\beta),\hat{f}_{n,t-1}(\beta);\theta)-E_0\left[G(f_t(\beta),f_{t-1}(\beta);\theta)\right]\right\|=o_p(1).$$

Regularity Condition RC.1: *The function $G(F_t;\theta)$, where $F_t=(f_t',f_{t-1}')'$, is such that:* (i) *$G(F;\theta)$ is continuous w.r.t. $F\in\mathbb{R}^{2m}$, for any $\theta\in\Theta$.* (ii) *For any $\beta\in\mathcal{B}$ and $\theta\in\Theta$, we have $E_0\left[\|G(F_t(\beta);\theta)\|\right]<\infty$, where $F_t(\beta)=(f_t(\beta)',f_{t-1}(\beta)')'$ and $f_t(\beta)$ is defined in (4.3).* (iii) $E\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial vec[G(F_t(\beta);\theta)]}{\partial(\beta',\theta')}\right\|\right]<\infty$. (iv) $\mathbb{P}\left[\xi_{t,6}\geq u\right]\leq b_6\exp\left(-c_6u^{d_6}\right)$, *as $u\to\infty$, for some constants $b_6,c_6,d_6>0$, where $\xi_{t,6}=\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\sup_{F\in\mathbb{R}^{2m}:\|F-F_t(\beta)\|\leq\eta^*}\left\|\frac{\partial vec[G(F;\theta)]}{\partial F}\right\|$, for some $\eta^*>0$.*

Proof of Theorem 2: Let us denote $\hat{F}_{n,t}(\beta) = \left(\hat{f}_{n,t}(\beta)', \hat{f}_{n,t-1}(\beta)'\right)'$. We have:

$$
\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T} G\left(\hat{F}_{n,t}(\beta);\theta\right) - E_0\left[G(F_t(\beta);\theta)\right] \\
= \;& \frac{1}{T}\sum_{t=1}^{T}\left(G\left(F_t(\beta);\theta\right) - E_0\left[G(F_t(\beta);\theta)\right]\right) + \frac{1}{T}\sum_{t=1}^{T}\left(G\left(\hat{F}_{n,t}(\beta);\theta\right) - G\left(F_t(\beta);\theta\right)\right) \\
\equiv \;& J_{1,T}(\beta,\theta) + J_{2,nT}(\beta,\theta).
\end{aligned}
$$

Let us now prove that the two terms in the RHS are $o_p(1)$ uniformly in $\beta \in \mathcal{B}, \theta \in \Theta$.

i) Proof that $\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}} J_{1,T}(\beta,\theta) = o_p(1)$

We use the Uniform Law of Large Numbers (ULLN) in Newey (1991), Corollary 2.1. Then, we get $\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}} J_{1,T}(\beta,\theta) = o_p(1)$, if the two following conditions hold:
(a) *Pointwise convergence:* $J_{1,T}(\beta,\theta) = o_p(1)$, for all parameter values (β,θ) in set $\mathcal{B}\times\Theta$;
(b) *Stochastic Lipschitz property:*

$$
|G(F_t(\tilde{\beta});\tilde{\theta}) - G\left(F_t(\beta);\theta\right)| \leq B_t\left(\|\tilde{\beta}-\beta\| + \|\tilde{\theta}-\theta\|\right), \tag{b.8}
$$

for all $(\beta,\theta), (\tilde{\beta},\tilde{\theta}) \in \mathcal{B}\times\Theta$ and some process B_t such that $\frac{1}{T}\sum_{t=1}^{T} E[B_t] = O(1)$.

Let us now prove conditions (a) and (b).

(a) Pointwise convergence: Since process (f_t) is strictly stationary and mixing (Assumption A.3), by Proposition 3.44 in White (2001) it follows that process (f_t) is also ergodic. Morever, for given $\beta \in \mathcal{B}$, the pseudo-true factor value $f_t(\beta)$ is a measurable function of the factor path $\underline{f_t}$ [Assumption H.4 (i) in Appendix A.1]. Now, we use that the strict stationarity and ergodicity properties are maintained under measurable transformations, involving possibly an infinite number of coordinates [Breiman (1992), Proposition 6.31]. Thus, process $f_t(\beta)$ is strictly stationary and ergodic, for given $\beta \in \mathcal{B}$. Since, for given $\theta \in \Theta$, the function $F \to G(F;\theta)$ is continuous by Regularity Condition RC.1 (i), by the same argument it follows that process $G(F_t(\beta);\theta)$ is strictly stationary and ergodic, for any given $(\beta,\theta) \in \mathcal{B}\times\Theta$. Then, Regularity Condition RC.1 (ii) and the ergodic theorem [Breiman (1992), Corollary 6.23] imply that the sample average $\frac{1}{T}\sum_{t=1}^{T} G\left(F_t(\beta);\theta\right)$ converges to the population expectation

$E_0\left[G\left(F_t(\beta);\theta\right)\right]$ almost surely, for any given $(\beta,\theta)\in\mathcal{B}\times\Theta$. This implies $J_{1,T}(\beta,\theta)=o_p(1)$, for any given $(\beta,\theta)\in\mathcal{B}\times\Theta$.

(b) Stochastic Lipschitz property: Inequality (b.8) holds for all $(\beta,\theta),(\tilde{\beta},\tilde{\theta})\in\mathcal{B}\times\Theta$ with the strictly stationary process B_t given by:

$$B_t \;=\; \sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial vec[G(F_t(\beta);\theta)]}{\partial(\beta',\theta')'}\right\|.$$

Moreover, from Regularity Condition RC.1 (iii), we have $E[B_t]<\infty$, and Condition (b) follows.

ii) Proof that $\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}} J_{2,nT}(\beta,\theta)=o_p(1)$

Let $\varepsilon>0$ be given. We have:

$$\mathbb{P}\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}} J_{2,nT}(\beta,\theta)\geq\varepsilon\right]\leq\mathbb{P}\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|G(\hat{F}_{n,t}(\beta);\theta)-G(F_t(\beta);\theta)\right\|\geq\varepsilon\right].$$

Now, we use that $\left\|\hat{F}_{n,t}(\beta)-F_t(\beta)\right\|\leq\eta$ implies:

$$\left\|G(\hat{F}_{n,t}(\beta);\theta)-G(F_t(\beta);\theta)\right\|\leq\eta\sup_{F:\|F-F_t(\beta)\|\leq\eta}\left\|\frac{\partial vec[G(F;\theta)]}{\partial F'}\right\|,$$

for any $\eta>0$. Thus, for $\eta_n=\varepsilon[c_6/\log n]^{1/d_6}$, where constants $c_6,d_6>0$ are defined in Regularity Condition RC.1 (iv), we get:

$$\begin{aligned}
\mathbb{P}\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}} J_{2,nT}(\beta,\theta)\geq\varepsilon\right] &\leq \mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|\hat{F}_{n,t}(\beta)-F_t(\beta)\right\|>\eta_n\right]\\
&\quad+\mathbb{P}\left[\sup_{1\leq t\leq T}\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\sup_{F:\|F-F_t(\beta)\|\leq\eta_n}\left\|\frac{\partial vec[G(F,\theta)]}{\partial F'}\right\|\geq\frac{\varepsilon}{\eta_n}\right]\\
&\leq \mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|\hat{F}_{n,t}(\beta)-F_t(\beta)\right\|>\eta_n\right]\\
&\quad+T\mathbb{P}\left[\sup_{\theta\in\Theta}\sup_{\beta\in\mathcal{B}}\sup_{F:\|F-F_t(\beta)\|\leq\eta^*}\left\|\frac{\partial vec[G(F,\theta)]}{\partial F'}\right\|\geq\frac{\varepsilon}{\eta_n}\right]\\
&\equiv P_{1,nT}+P_{2,nT},
\end{aligned}$$

for large n and $\eta^* > 0$ as in Regularity Condition RC.1 (iv). Now, $P_{1,nT} = o(1)$ from Limit Theorem 1 in Appendix B.1. Moreover, from Regularity Condition RC.1 (iv), we get:

$$P_{2,nT} \leq b_6 T \exp\left(-c_6[\varepsilon/\eta_n]^{d_6}\right) = b_6 T/n = o(1).$$

The conclusion follows.

B.3 Uniform consistency of nonlinear aggregates

THEOREM 3 *Let Assumptions A.1-A.5, H.1, H.2, H.4 (i), H.5, H.6, H.7 (i)-(ii), H.8-H.10 hold, and assume that functions a and φ satisfy either the Regularity Condition RC.2, or the Regularity Condition RC.3, below. Then, if $n, T \to \infty$ such that $T^\nu/n = O(1)$ for a value $\nu > 1$:*

$$\sup_{\beta \in \mathcal{B}} \left\| \frac{1}{T} \sum_{t=1}^{T} \varphi\left(\frac{1}{n} \sum_{i=1}^{n} a(y_{i,t}, y_{i,t-1}, \hat{f}_{n,t}(\beta), \beta) \right) - E_0\left[\varphi\left(\mu_t(\beta)\right)\right] \right\| = o_p(1), \tag{b.9}$$

where $\mu_t(\beta) = E_0\left[a(y_{i,t}, y_{i,t-1}, f_t(\beta), \beta) | \underline{f_t}\right]$.

Limit Theorem 3 provides a Uniform Law of Large Numbers (ULLN) for nonlinear aggregates of panel data. These nonlinear aggregates involve a combination of linear and nonlinear time-series and cross-sectional transformations, which explains the novelty of Limit Theorem 3 compared to other ULLN in the literature. More precisely, the nonlinear aggregates correspond to the time series average of the nonlinear transformation by mapping φ of the cross-sectional average of random matrices $a(a_{i,t}, y_{i,t-1}, \hat{f}_{n,t}(\beta), \beta)$ depending on data $y_{i,t}, y_{i,t-1}$, factor approximation $\hat{f}_{n,t}(\beta)$ and micro-parameter β. The large sample limit of such an aggregate is the time-series expectation of the transformation by mapping φ of the cross-sectional expectation $\mu_t(\beta)$.

We distinguish two sets of regularity conditions. Regularity Condition RC.2 requires that mapping φ is Lipschitz continuous. Regularity Condition RC.3 relaxes this condition and allows to apply Limit Theorem 3 for instance when mapping φ corresponds to matrix inversion, or the log-determinant function, on the set of positive definite matrices (see the proofs of Lemmas 1 and 6 in Appendices C.1 and C.6). Regularity Condition RC.3 also introduces tail conditions on the stationary distribution of the reciprocal of the smallest eigenvalue of the positive definite matrix $\mu_t(\beta)$ uniformly w.r.t. $\beta \in \mathcal{B}$.

Regularity Condition RC.2: *The functions a and φ are such that:*

(1) (i) $E_0\left[\sup_{\beta\in\mathcal{B}}\|a(Y_{i,t},f_t(\beta),\beta)\|^4\right]<\infty$, *where* $Y_{i,t}=(y_{i,t},y_{i,t-1})'$.

(ii) $E_0\left[\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial\ vec\ a[Y_{i,t},f_t(\beta),\beta)]}{\partial\beta'}\right\|^4\right]<\infty.$

(iii) *For any* $\beta\in\mathcal{B}$*:* $\mu_t(\beta)=E_0[a(Y_{i,t},f_t(\beta),\beta)|\underline{f_t}]$ *is a measurable function of the factor path* $\underline{f_t}$.

(iv) $\mathbb{P}\left[\xi_{t,7}\geq u\right]\leq b_7\exp\left(-c_7u^{d_7}\right)$, *as* $u\to\infty$, *for some constants* $b_7,c_7,d_7>0$, *where* $\xi_{t,7}=\sup_{\beta\in\mathcal{B}}E_0\left[\|a(Y_{i,t},f_t(\beta),\beta)\|^2|\underline{f_t}\right]$.

(v) $\mathbb{P}\left[\xi_{t,8}\geq u\right]\leq b_8\exp\left(-c_8u^{d_8}\right)$, *as* $u\to\infty$, *for some constants* $b_8,c_8,d_8>0$, *where* $\xi_{t,8}=\sup_{\beta\in\mathcal{B}}E_0\left[\sup_{f\in\mathbb{R}^m:\|f-f_t(\beta)\|\leq\eta^*}\left\|\frac{\partial\ vec[a(Y_{i,t},f,\beta)]}{\partial f'}\right\|^2|\underline{f_t}\right]$, *with* $\eta^*>0$;

(2) *The function* φ *is Lipschitz continuous and such that* $E_0\left[\|\varphi(\mu_t(\beta))\|\right]<\infty$, *for any* $\beta\in\mathcal{B}$.

Regularity Condition RC.3: *The functions a and φ are such that:*

(1) *Regularity Condition* RC.2 (1) *holds. Function* $a(Y,f,\beta)$ *admits values in the set of* (r,r) *symmetric matrices, for some* $r\in\mathbb{N}$. *Moreover:*

(i) $\mu_t(\beta)=E_0\left[a(Y_{i,t},f_t(\beta),\beta)|\underline{f_t}\right]\in\mathcal{U}$, *for any* t *and* $\beta\in\mathcal{B}$, $\mathbb{P}$*-a.s., where* $\mathcal{U}$ *is the open subset of positive definite* (r,r) *matrices.*

(ii) $\mathbb{P}\left[\xi_{t,9}\geq u\right]\leq b_9\exp\left(-c_9u^{d_9}\right)$, *as* $u\to\infty$, *for some constants* $b_9,c_9,d_9>0$, *where* $\xi_{t,9}=\left(\inf_{\beta\in\mathcal{B}}\lambda_t(\beta)\right)^{-1}$ *and* $\lambda_t(\beta)>0$ *is the smallest eigenvalue of matrix* $\mu_t(\beta)$;

(2) *The function* $\varphi:\mathcal{U}\to\mathbb{R}$ *is such that:*

(i) φ *is Lipschitz continuous on any compact subset of* $\mathcal{U}$.

(ii) $|\varphi(w)|\leq C_{10}\|z\|^{\gamma_{10}}\psi(z)$, *for any* $w,z\in\mathcal{U}$ *such that* $w=(Id+\Delta)z$, $\|\Delta\|\leq 1/2$, *where constants* C_{10},γ_{10} *satisfy* $C_{10}>0$, $\gamma_{10}\leq 2$, *and function* $\psi:\mathcal{U}\to\mathbb{R}$ *is such that* $E_0[\sup_{\beta\in\mathcal{B}}|\psi(\mu_t(\beta))|^4]<\infty$.

We first prove Theorem 3 under Regularity Condition RC.2. Then, we give the proof under Regularity Condition RC.3.

B.3.1 Proof of Theorem 3 under Regularity Condition RC.2

Let us write:

$$\begin{aligned}
& \frac{1}{T}\sum_{t=1}^{T}\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right)-E_0\left[\varphi\left(\mu_t(\beta)\right)\right] \\
= & \frac{1}{T}\sum_{t=1}^{T}\varphi\left(\mu_t\left(\beta\right)\right)-E_0\left[\varphi\left(\mu_t\left(\beta\right)\right)\right] \\
& +\frac{1}{T}\sum_{t=1}^{T}\left\{\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},f_t(\beta),\beta)\right)-\varphi\left(\mu_t\left(\beta\right)\right)\right\} \\
& +\frac{1}{T}\sum_{t=1}^{T}\left\{\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right)-\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},f_t(\beta),\beta)\right)\right\} \\
\equiv & J_{3,T}(\beta)+J_{4,n,T}(\beta)+J_{5,n,T}(\beta), \qquad \text{(b.10)}
\end{aligned}$$

where $Y_{i,t}=(y_{i,t},y_{i,t-1})'$. The component $J_{3,T}(\beta)$ is the time series average of a nonlinear transformation of process $\mu_t\left(\beta\right)$. The component $J_{4,n,T}(\beta)$ accounts for the discrepancy between the cross-sectional average $\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},f_t(\beta),\beta)$ and the conditional expectation $\mu_t\left(\beta\right)=E_0[a(Y_{i,t},f_t(\beta),\beta)|\underline{f_t}]$. The component $J_{5,n,T}(\beta)$ is induced by the approximation of the pseudo-true factor value $f_t(\beta)$ with the estimator $\hat{f}_{n,t}(\beta)$. Let us prove that these three components are $o_p(1)$, uniformly in $\beta\in\mathcal{B}$.

i) Proof that $\sup_{\beta\in\mathcal{B}}|J_{3,T}(\beta)|=o_p(1)$

The proof of this uniform convergence is similar to part i) in the proof of Limit Theorem 2 in Section B.2. We replace $\mu_t(\beta)$ for $F_t(\beta)$, and mapping φ for mapping $G(\cdot;\theta)$, and use Regularity Conditions RC.2 (1i)-(1iii) and (2). Since the mapping φ is independent of parameter θ, there is no sup over $\theta\in\Theta$ here.

ii) Proof that $\sup_{\beta\in\mathcal{B}}|J_{4,n,T}(\beta)|=o_p(1)$

Let us now consider term $J_{4,n,T}(\beta)$ in the RHS of equation (b.10). Let $\varepsilon>0$. The condition $\|x-y\|\leq L/\varepsilon$ implies $|\varphi(x)-\varphi(y)|\leq\varepsilon$, since function φ is Lipschitz continuous, with

Lipschitz constant L, say [Regularity Condition RC.2 (2)]. Thus, we get:

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

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

$$\Omega_{1,n,T}(\delta)=\left\{\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\left[a(Y_{i,t},f_t(\beta),\beta)-\mu_t(\beta)\right]\right\|\leq\delta\right\}. \tag{b.11}$$

In Lemma B.3 (i) in Appendix B.4.3 we show that $\mathbb{P}[\Omega_{1,n,T}(\delta)]\to 1$, as $n,T\to\infty$ such that $T/n\to 0$, for any $\delta>0$. Since $P_{1,\varepsilon}=1-\mathbb{P}\left[\Omega_{1,n,T}(L/\varepsilon)\right]$, we get that $P_{1,\varepsilon}\to 0$ as $n,T\to\infty$, $T/n\to 0$, for any $\varepsilon>0$. It follows that $\sup_{\beta\in\mathcal{B}}|J_{4,n,T}(\beta)|=o_p(1)$.

iii) Proof that $\sup_{\beta\in\mathcal{B}}|J_{5,n,T}(\beta)|=o_p(1)$

Let us finally consider term $J_{5,n,T}(\beta)$ in the RHS of equation (b.10). Let $\varepsilon>0$ be given. Then:

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

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

$$\Omega_{2,n,T}(\delta)=\left\{\sup_{\beta\in\mathcal{B}}\sup_{1\leq t\leq T}\left\|\frac{1}{n}\sum_{i=1}^{n}\left[a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)-a(Y_{i,t},f_t(\beta),\beta)\right]\right\|\leq\delta\right\}. \tag{b.12}$$

In Lemma B.4 (i) in Appendix B.4.4 we show that $\mathbb{P}[\Omega_{2,n,T}(\delta)]\to 1$, as $n,T\to\infty$ such that $T/n\to 0$, for any $\delta>0$. Since $P_{2,\varepsilon}=1-\mathbb{P}\left[\Omega_{2,n,T}(L/\varepsilon)\right]$, we get that $P_{2,\varepsilon}\to 0$ as $n,T\to\infty$, $T/n\to 0$, for any $\varepsilon>0$. It follows $\sup_{\beta\in\mathcal{B}}|J_{5,n,T}(\beta)|=o_p(1)$.

B.3.2 Proof of Theorem 3 under Regularity Condition RC.3

Under Regularity Condition RC.3, matrix function φ is defined on the subset $\mathcal{U} \subset \mathbb{R}^{r\times r}$ of positive definite (r, r) matrices. Therefore, the LHS of equation (b.9) is well-defined only when $\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{U}$ for any $1 \leq t \leq T$ and $\beta \in \mathcal{B}$.

i) Let us first prove that this event occurs with probability approaching (w.p.a.) 1. Let $\eta > 0$ be given. In Lemma B.5 in Appendix B.4.5 we prove that there exists a compact set $\mathcal{K} \subset \mathcal{U}$ such that $\mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}\right] \geq 1 - \eta$. Let further $\delta > 0$ be such that $\left\{x \in \mathbb{SR}^{r\times r} : dist(x, \mathcal{K}) \leq \delta\right\} \subset \mathcal{U}$, where $\mathbb{SR}^{r\times r}$ is the set of (r, r) symmetric matrices and $dist(x, \mathcal{K}) \equiv \inf_{y\in\mathcal{K}} \|x - y\|$ is the distance of matrix x from set $\mathcal{K}$. Then:

$$
\begin{aligned}
P_{n,T} &\equiv \mathbb{P}\left[\left\{\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta), 1 \leq t \leq T, \beta \in \mathcal{B}\right\} \subset \mathcal{U}\right] \\
&\geq \mathbb{P}\left[\left(\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}\right) \bigcap \Omega_{1,n,T}(\delta/2) \bigcap \Omega_{2,n,T}(\delta/2)\right] \\
&\geq \mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}\right] + \mathbb{P}\left[\Omega_{1,n,T}(\delta/2)\right] + \mathbb{P}\left[\Omega_{2,n,T}(\delta/2)\right] - 2 \\
&\geq \mathbb{P}\left[\Omega_{1,n,T}(\delta/2)\right] + \mathbb{P}\left[\Omega_{2,n,T}(\delta/2)\right] - 1 - \eta,
\end{aligned}
$$

where events $\Omega_{1,n,T}(\delta/2)$ and $\Omega_{2,n,T}(\delta/2)$ are defined in equations (b.11) and (b.12). From Lemmas B.3 (i) and B.4 (i) in Appendices B.4.3 and B.4.4, respectively, it follows that $\limsup_{n,T\to\infty} P_{n,T} \geq 1 - \eta$. Since constant $\eta > 0$ can be chosen arbitrarily small, we get that $\lim_{n,T\to\infty} P_{n,T} = 1$. Therefore, the event $\left\{\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta), 1 \leq t \leq T, \beta \in \mathcal{B}\right\} \subset \mathcal{U}$ occurs w.p.a. 1.

ii) We can focus on this event in the rest of the proof. Let $\varepsilon, \bar{\eta} > 0$ be given. We have to prove that:

$$
\limsup_{n,T\to\infty} \mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T} \varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta)\right) - E_0\left[\varphi\left(\mu_t(\beta)\right)\right]\right| \geq \varepsilon\right] \leq \bar{\eta}. \qquad \text{(b.13)}
$$

Let us introduce a globally Lipschitz approximation of function φ. More precisely, let $\mathcal{K}_1 \subset \mathcal{U}$ be a compact set and let $\tilde{\varphi}$ be a Lipschitz continuous function on $\mathcal{U}$ such that

$$
\tilde{\varphi} = \varphi \text{ on } \mathcal{K}_1 \text{ and } |\tilde{\varphi}| \leq |\varphi| \text{ on } \mathcal{U}. \qquad \text{(b.14)}
$$

Such a function exists by Regularity Condition RC.3 (2i). Then inequality (b.13) follows if function $\tilde{\varphi}$ can be chosen such that:

$$A_{1,\varepsilon} \equiv \limsup_{n,T\to\infty} \mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right) - E_0\left[\tilde{\varphi}\left(\mu_t(\beta)\right)\right]\right| \geq \varepsilon/3\right] \leq \bar{\eta}/2, \tag{b.15}$$

$$\begin{aligned} A_{2,\varepsilon} \;\equiv\; & \limsup_{n,T\to\infty} \mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\left|\frac{1}{T}\sum_{t=1}^{T}\left[\varphi\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right) - \tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n}a(Y_{i,t},\hat{f}_{n,t}(\beta),\beta)\right)\right]\right| \geq \varepsilon/3\right] \\ & \leq \bar{\eta}/2, \end{aligned} \tag{b.16}$$

and:

$$A_3 \equiv \sup_{\beta\in\mathcal{B}}\left|E_0\left[\tilde{\varphi}\left(\mu_t(\beta)\right)\right] - E_0\left[\varphi\left(\mu_t(\beta)\right)\right]\right| \leq \varepsilon/3. \tag{b.17}$$

The proof proceeds as follows. We first show that $A_{1,\varepsilon} = 0$, which implies inequality (b.15). Then, we derive upper bounds for $A_{2,\varepsilon}$, and A_3. From those bounds we prove that inequalities (b.16) and (b.17) hold.

i) Proof that $A_{1,\varepsilon} = 0$

From the definition of the globally Lipschitz approximation in (b.14), and Regularity Conditions RC.3 (1), (2ii), function $\tilde{\varphi}$ is Lipschitz continuous and such that $E_0\left[|\tilde{\varphi}(\mu_t(\beta))|\right] < \infty$. Indeed, we have:

$$E_0\left[|\tilde{\varphi}(\mu_t(\beta))|\right] \leq E_0\left[|\varphi(\mu_t(\beta))|\right] \leq C_{10}E_0\left[\|\mu_t(\beta)\|^{\gamma_{10}}|\psi(\mu_t(\beta))|\right],$$

where function ψ is defined in Regularity Condition RC.3 (2ii). Then, from the Cauchy-Schwarz inequality, we get:

$$E_0\left[|\tilde{\varphi}(\mu_t(\beta))|\right] \leq C_{10}E_0\left[\|\mu_t(\beta)\|^{2\gamma_{10}}\right]^{1/2} E_0\left[|\psi(\mu_t(\beta))|^2\right]^{1/2} < \infty,$$

for any $\beta \in \mathcal{B}$. Hence, functions $(a, \tilde{\varphi})$ satisfy Regularity Condition RC.2. Thus, we get $A_{1,\varepsilon} = 0$ by applying Limit Theorem 3 under Regularity Condition RC.2.

ii) Upper bound for $A_{2,\varepsilon}$

Let us now consider term $A_{2,\varepsilon}$ in inequality (b.16). Since $\tilde{\varphi} = \varphi$ on set $\mathcal{K}_1$ [see (b.14)], in the event that defines $A_{2,\varepsilon}$ only the dates t with $\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{K}_1^c$ contribute to the sum. Moreover, we have $|\varphi - \tilde{\varphi}| \leq 2|\varphi|$ on set $\mathcal{U}$ [see (b.14)]. Therefore, we have:

$$\begin{aligned}
&\sup_{\beta\in\mathcal{B}} \left| \frac{1}{T}\sum_{t=1}^{T} \left[\varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) - \tilde{\varphi}\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) \right] \right| \\
&\leq 2 \sup_{\beta\in\mathcal{B}} \frac{1}{T}\sum_{t=1}^{T} 1\left\{ \frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{K}_1^c \right\} \left| \varphi\left(\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) \right|. \quad \text{(b.18)}
\end{aligned}$$

Let us now bound the RHS of inequality (b.18) in two steps.

a) Let $\mathcal{K}_2 \subset \mathcal{K}_1$ be a compact set, and $\delta > 0$ a scalar, such that:

$$dist\,(\mathcal{K}_2, \mathcal{K}_1^c) > 2\delta, \quad \text{(b.19)}$$

where $dist\,(\mathcal{K}_2, \mathcal{K}_1^c) \equiv \inf_{x\in\mathcal{K}_2, y\in\mathcal{K}_1^c} \|x - y\|$ denotes the distance between sets $\mathcal{K}_2$ and $\mathcal{K}_1^c$. When the event $\Omega_{1,n,T}(\delta) \cap \Omega_{2,n,T}(\delta)$ occurs, where $\Omega_{j,n,T}(\delta)$, $j = 1, 2$, are defined in equation (b.11) and (b.12), respectively, we have $\left\| \frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) - \mu_t(\beta) \right\| \leq 2\delta$, $\mathbb{P}$-a.s., for any $t = 1, ..., T$ and $\beta \in \mathcal{B}$. By condition (b.19), we get:

$$\frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{K}_1^c \Rightarrow \mu_t(\beta) \in \mathcal{K}_2^c,$$

$\mathbb{P}$-a.s., for any $t = 1, ..., T$ and $\beta \in \mathcal{B}$. It follows:

$$1\left\{ \frac{1}{n}\sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \in \mathcal{K}_1^c \right\} \leq 1\left\{\mu_t(\beta) \in \mathcal{K}_2^c\right\} \leq 1 - 1\left\{(\mu_t(\beta), \beta \in \mathcal{B}) \subset \mathcal{K}_2\right\}, \quad \text{(b.20)}$$

for any $\beta \in \mathcal{B}$, since $1\left\{\mu_t(\beta) \in \mathcal{K}_2^c\right\} = 1$ for some $\beta \in \mathcal{B}$ holds if, and only if, $1\left\{(\mu_t(\beta), \beta \in \mathcal{B}) \subset \mathcal{K}_2\right\} = 0$ holds.

b) Define for $\delta > 0$ as above the events:

$$\Omega_{3,n,T}(\delta) = \left\{ \sup_{\beta\in\mathcal{B}} \sup_{1\leq t\leq T} \frac{1}{\lambda_t(\beta)} \left\| \frac{1}{n}\sum_{i=1}^{n} \left[a(Y_{i,t}, f_t(\beta), \beta) - \mu_t(\beta)\right] \right\| \leq \delta \right\}, \quad \text{(b.21)}$$

and:

$$\Omega_{4,n,T}(\delta) = \left\{ \sup_{\beta \in \mathcal{B}} \sup_{1 \leq t \leq T} \frac{1}{\lambda_t(\beta)} \left\| \frac{1}{n} \sum_{i=1}^{n} \left[a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) - a(Y_{i,t}, f_t(\beta), \beta) \right] \right\| \leq \delta \right\}, \quad \text{(b.22)}$$

where $\lambda_t(\beta)$ is as in Regularity Condition RC.3 (1ii). In Lemmas B.3 (ii) and B.4 (ii) in Appendices B.4.3 and B.4.4, respectively, we prove that $\mathbb{P}[\Omega_{3,n,T}(\delta)] \to 1$ and $\mathbb{P}[\Omega_{4,n,T}(\delta)] \to 1$, as $n, T \to \infty$, $T/n \to 0$. When the event $\Omega_{3,n,T}(\delta) \cap \Omega_{4,n,T}(\delta)$ occurs, with $\delta \leq 1/4$, we have have $\|\Delta_t(\beta)\| \leq 2\delta \leq 1/2$, $\mathbb{P}$-a.s., for any $t = 1, ..., T$ and $\beta \in \mathcal{B}$, where $\Delta_t(\beta) \equiv \left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) - \mu_t(\beta) \right) (\mu_t(\beta))^{-1}$. Thus, from Regularity Condition RC.3 (2ii) we get:

$$\left| \varphi \left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) \right| \leq C_{10} \|\mu_t(\beta)\|^{\gamma_{10}} \psi(\mu_t(\beta)). \quad \text{(b.23)}$$

From inequalities (b.18), (b.20) and (b.23) we get that, when event $\cap_{j=1}^{4} \Omega_{j,n,T}(\delta)$ occurs, we have:

$$\begin{aligned} &\sup_{\beta \in \mathcal{B}} \left| \frac{1}{T} \sum_{t=1}^{T} \left[\varphi \left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) - \tilde{\varphi} \left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) \right] \right| \\ \leq\;& \frac{2C_{10}}{T} \sum_{t=1}^{T} (1 - 1\{(\mu_t(\beta), \beta \in \mathcal{B}) \subset \mathcal{K}_2\}) \sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^{\gamma_{10}} \psi(\mu_t(\beta)). \end{aligned}$$

It follows that:

$$\begin{aligned} &\mathbb{P} \left[\sup_{\beta \in \mathcal{B}} \left| \frac{1}{T} \sum_{t=1}^{T} \left[\varphi \left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) - \tilde{\varphi} \left(\frac{1}{n} \sum_{i=1}^{n} a(Y_{i,t}, \hat{f}_{n,t}(\beta), \beta) \right) \right] \right| \geq \varepsilon/3 \right] \\ \leq\;& \sum_{j=1}^{4} \mathbb{P}[\Omega_{j,n,T}(\delta)^c] \\ &+ \mathbb{P} \left[\frac{2C_{10}}{T} \sum_{t=1}^{T} (1 - 1\{(\mu_t(\beta), \beta \in \mathcal{B}) \subset \mathcal{K}_2\}) \sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^{\gamma_{10}} \psi(\mu_t(\beta)) \geq \varepsilon/3 \right] \\ \leq\;& \sum_{j=1}^{4} \mathbb{P}[\Omega_{j,n,T}(\delta)^c] + \frac{6C_{10}}{\varepsilon} E \left[(1 - 1\{(\mu_t(\beta), \beta \in \mathcal{B}) \subset \mathcal{K}_2\}) \sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^{\gamma_{10}} \psi(\mu_t(\beta)) \right], \end{aligned}$$

by the Markov inequality. By taking the limit for $n, T \to \infty$ such that $T/n \to 0$, from

Lemmas B.3 and B.4 we get:

$$A_{2,\varepsilon} \leq \frac{6C_{10}}{\varepsilon} E\left[\left(1 - 1\left\{(\mu_t(\beta), \beta \in \mathcal{B}) \subset \mathcal{K}_2\right\}\right) \sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^{\gamma_{10}} \psi\left(\mu_t(\beta)\right)\right].$$

By the Minkowsky and Cauchy-Schwarz inequalities, we have:

$$\begin{aligned}
& E\left[\left(1 - 1\left\{(\mu_t(\beta), \beta \in \mathcal{B}) \subset \mathcal{K}_2\right\}\right) \sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^{\gamma_{10}} \psi\left(\mu_t(\beta)\right)\right] \\
\leq \; & \left(1 - \mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_2\right]\right)^{1/p} E\left[\sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^{\gamma_{10} q} \psi\left(\mu_t(\beta)\right)^q\right]^{1/q} \\
\leq \; & \left(1 - \mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_2\right]\right)^{1/p} E\left[\sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^{\gamma_{10} q p'}\right]^{1/(p'q)} E\left[\sup_{\beta \in \mathcal{B}} \psi\left(\mu_t(\beta)\right)^{qq'}\right]^{1/(qq')},
\end{aligned}$$

with $p, q, p^{'}, q^{'} > 1$ such that $1/p + 1/q = 1$ and $1/p^{'} + 1/q^{'} = 1$. Fix $q \in (1, \frac{4}{1+\gamma_{10}})$ and $p^{'} = 4/(\gamma_{10} q)$. We get:

$$A_{2,\varepsilon} \leq \frac{6C_{10}}{\varepsilon} \left(1 - \mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_2\right]\right)^{1/p} C_{11}, \tag{b.24}$$

where $C_{11} = E\left[\sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|^4\right]^{\gamma_{10}/4} E\left[\sup_{\beta \in \mathcal{B}} \psi\left(\mu_t(\beta)\right)^4\right]^{1/q - \gamma_{10}/4} < \infty$ by Regularity Conditions RC.3 (1) and (2ii).

iii) Bound of A_3

Let us now bound A_3 defined in the LHS of inequality (b.17). By similar arguments as above:

$$A_3 \leq 2C_{10} \left(1 - \mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_2\right]\right)^{1/p} C_{11}. \tag{b.25}$$

iv) Proof of inequalities (b.16) and (b.17)

From Lemma B.5 in Appendix B.4.5, we can fix $\mathcal{K}_1$, $\mathcal{K}_2$ and δ such that $\mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_2\right] \geq 1 - \min\left\{\left(\frac{\varepsilon \bar{\eta}}{12 C_{10} C_{11}}\right)^p, \left(\frac{\varepsilon}{6 C_{10} C_{11}}\right)^p\right\}$ and condition (b.19) hold. Then, from inequalities (b.24) and (b.25), inequalities (b.16) and (b.17) follow, and the proof is concluded.

B.4 Secondary Lemmas

B.4.1 Lemma B.1

Lemma B.1 provides a large deviation inequality for $\sup_{\beta \in \mathcal{B}} \|\hat{f}_n(\beta) - f(\beta)\|$ in finite sample, where $\hat{f}_n(\beta)$ denotes the ML estimator of parameter f with sample size n, and $f(\beta)$ denotes the pseudo-true value of parameter f, for given value of the nuisance parameter $\beta \in \mathcal{B}$.

Lemma B.1: *Let n be given and let data y_i, for $i = 1, ..., n$, be i.i.d. with density $h(y_i, \alpha)$ parametrized by $\alpha = (f', \beta')'$, where the parameter of interest is $f \in \mathcal{F} \subset \mathbb{R}^m$, and the nuisance parameter is $\beta \in \mathcal{B} \subset \mathbb{R}^q$. We denote by $\alpha_0 = (f_0', \beta_0')'$ the true parameter value. Let us consider the concentrated ML estimator of parameter f defined by:*

$$\widehat{f}_n(\beta) = \arg\max_{f \in \mathcal{F}} L_n(f, \beta),$$

for any $\beta \in \mathcal{B}$, where $L_n(f, \beta) = \dfrac{1}{n}\sum_{i=1}^{n} l_i(\alpha)$ and $l_i(\alpha) = \log h(y_i, \alpha)$. Denote $L(\alpha) = E_0[l_i(\alpha)]$, and $\mathcal{A} = \mathcal{F} \times \mathcal{B}$. Let us assume:
i) The set $\mathcal{F}$ is compact and convex, and the set $\mathcal{B}$ is compact.
ii) For any given $\beta \in \mathcal{B}$, the function $L(f, \beta)$ is uniquely maximized w.r.t. $f \in \mathcal{F}$ at $f(\beta) = \arg\max_{f \in \mathcal{F}} L(f, \beta)$. The true values of parameters $f_0 \in \mathcal{F}$ and $\beta_0 \in \mathcal{B}$ satisfy $f_0 = f(\beta_0)$, and the matrix $E_0\left[-\dfrac{\partial^2 l_i(f(\beta), \beta)}{\partial f \partial f'}\right]$ is non-singular, for any $\beta \in \mathcal{B}$.
iii) There exists a constant $\gamma_{11} > 2$ such that $\mathcal{R} \equiv E_0\left[\sup_{\alpha \in \mathcal{A}} \left\|\dfrac{\partial \log h(y_i, \alpha)}{\partial \alpha}\right\|^{\gamma_{11}}\right] < \infty$.
iv) The function $f(\beta)$ is differentiable and such that $\mathcal{M} \equiv \sup_{\beta \in \mathcal{B}} \left\|\dfrac{\partial f(\beta)}{\partial \beta'}\right\| < \infty$.
Then, there exist constants $C_1, C_2, C_3 > 0$ (depending on parameter dimensions m and q, but independent of parameter sets $\mathcal{F}$, $\mathcal{B}$ and of the parametric model) such that for any constant $\varepsilon > 0$:

$$\mathbb{P}\left[\sup_{\beta \in \mathcal{B}} \left\|\widehat{f}_n(\beta) - f(\beta)\right\| \geq \varepsilon\right] \leq C_1 Vol(\mathcal{B})(1+\mathcal{M})^{m+q} \frac{n^{m+q}}{\varepsilon^q} \exp\left(-C_2 n \varepsilon^2 \frac{\mathcal{K}}{1 + \Gamma/\mathcal{K}}\right) + C_3 \varepsilon^{\gamma_{11}-2} \frac{\mathcal{R}}{\mathcal{K}},$$

where:

$$\mathcal{K} \equiv \inf_{\beta \in \mathcal{B}} \inf_{f \in \mathcal{F}: f \neq f(\beta)} \frac{2KL(f, f(\beta); \beta)}{\|f - f(\beta)\|^2} > 0, \tag{b.26}$$

$KL(f, f(\beta); \beta) \equiv L(f(\beta), \beta) - L(f, \beta)$ is the Kullback-Leibler discrepancy between f and $f(\beta)$ for given $\beta \in \mathcal{B}$, the scalar Γ is given by:

$$\Gamma \equiv \sup_{\alpha \in \mathcal{A}} E_0 \left[\left\| \frac{\partial \log h(y_i, \alpha)}{\partial f} \right\|^2 \right] < \infty, \tag{b.27}$$

with $Vol(\mathcal{B}) = \int_{\mathcal{B}} d\lambda$ is the Lebesgue measure of set $\mathcal{B}$.

Proof of Lemma B.1: Let us first relate probability $\mathbb{P}\left[\sup_{\beta \in \mathcal{B}} \|\hat{f}_n(\beta) - f(\beta)\| > \varepsilon\right]$ to the probability of large deviations of the empirical process associated with the log-likelihood function.

i) Probability of large deviation of the likelihood process

Define the set:

$$\mathcal{F}_k(\beta) = \left\{ f \in \mathcal{F} : 2^k \varepsilon \geq \|f - f(\beta)\| \geq 2^{k-1} \varepsilon \right\}, \tag{b.28}$$

for any $k = 1, 2, \cdots$, and $\beta \in \mathcal{B}$. Then, we have:

$$\begin{aligned} \mathbb{P}\left[\sup_{\beta \in \mathcal{B}} \|\hat{f}_n(\beta) - f(\beta)\| > \varepsilon\right] &\leq \mathbb{P}\left[\bigcup_{k=1}^{\infty} \bigcup_{\beta \in \mathcal{B}} \left\{ \hat{f}_n(\beta) \in \mathcal{F}_k(\beta) \right\}\right] \\ &\leq \sum_{k=1}^{\infty} \mathbb{P}\left[\bigcup_{\beta \in \mathcal{B}} \left\{ \hat{f}_n(\beta) \in \mathcal{F}_k(\beta) \right\}\right]. \end{aligned}$$

Moreover, for any integer k:

$$\begin{aligned} \mathbb{P}\left[\bigcup_{\beta \in \mathcal{B}} \left\{ \hat{f}_n(\beta) \in \mathcal{F}_k(\beta) \right\}\right] &\leq \mathbb{P}\left[\bigcup_{\beta \in \mathcal{B}} \left\{ \sup_{f \in \mathcal{F}_k(\beta)} L_n(f, \beta) \geq L_n(\hat{f}_n(\beta), \beta) \right\}\right] \\ &\leq \mathbb{P}\left[\bigcup_{\beta \in \mathcal{B}} \left\{ \sup_{f \in \mathcal{F}_k(\beta)} L_n(f, \beta) \geq L_n(f(\beta), \beta) \right\}\right] \\ &= \mathbb{P}\left[\sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_k(\beta)} \left(L_n(f, \beta) - L_n(f(\beta), \beta)\right) \geq 0\right]. \end{aligned}$$

Now, let us introduce the sets:

$$\mathcal{A}_k = \{(f,\beta) : f \in \mathcal{F}_k(\beta), \beta \in \mathcal{B}\} \subset \mathcal{A}, \quad k = 1,2,..., \tag{b.29}$$

and the mapping π that maps $\alpha = (f', \beta')'$ into $\pi(\alpha) = (f(\beta)', \beta')$. [2] Thus, we have:

$$\mathbb{P}\left[\sup_{\beta\in\mathcal{B}} \|\hat{f}_n(\beta) - f(\beta)\| > \varepsilon\right] \leq \sum_{k=1}^{\infty} \mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k} [L_n(\alpha) - L_n(\pi(\alpha))] \geq 0\right].$$

Define:

$$\Psi_n(\alpha) = L_n(\alpha) - L_n(\pi(\alpha)) - [L(\alpha) - L(\pi(\alpha))] = \frac{1}{n}\sum_{i=1}^{n} \psi_i(\alpha), \tag{b.30}$$

where $\psi_i(\alpha) = l_i(\alpha) - l_i(\pi(\alpha)) - E_0[l_i(\alpha) - l_i(\pi(\alpha))]$. Then, we have:

$$\begin{aligned}
&\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k} [L_n(\alpha) - L_n(\pi(\alpha))] \geq 0\right] \\
\leq\ &\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k} (L_n(\alpha) - L_n(\pi(\alpha)) - [L(\alpha) - L(\pi(\alpha))]) \geq \inf_{\alpha\in\mathcal{A}_k} (L(\pi(\alpha)) - L(\alpha))\right] \\
=\ &\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k} \Psi_n(\alpha) \geq \inf_{\alpha\in\mathcal{A}_k} KL(\alpha, \pi(\alpha))\right],
\end{aligned}$$

where $KL(\alpha,\pi(\alpha)) = L(\pi(\alpha)) - L(\alpha) = KL(f, f(\beta);\beta)$. Now, from the definitions of sets $\mathcal{F}_k(\beta)$ and $\mathcal{A}_k$ in (b.28) and (b.29), respectively, we get:

$$\begin{aligned}
\inf_{\alpha\in\mathcal{A}_k} KL(\alpha,\pi(\alpha)) &= \inf_{\beta\in\mathcal{B}} \inf_{f\in\mathcal{F}_k(\beta)} K(f, f(\beta);\beta) \\
&\geq \inf_{\beta\in\mathcal{B}} \inf_{f\in\mathcal{F}:\|f-f(\beta)\|\geq 2^{k-1}\varepsilon} KL(f, f(\beta);\beta) \geq \frac{1}{2}\mathcal{K}\left(2^{k-1}\varepsilon\right)^2,
\end{aligned}$$

where constant $\mathcal{K}$ is defined in (b.26). Thus, we get:

$$\mathbb{P}\left[\sup_{\beta\in\mathcal{B}} \|\hat{f}_n(\beta) - f(\beta)\| > \varepsilon\right] \leq \sum_{k=1}^{\infty} \mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k} \Psi_n(\alpha) \geq \lambda_k\right], \tag{b.31}$$

where:

$$\lambda_k \equiv \frac{1}{2}\mathcal{K}\left(2^{k-1}\varepsilon\right)^2. \tag{b.32}$$

[2] Geometrically, the set $\mathcal{A}_k$ consists of two strips of width $2^{k-1}\varepsilon$ in the (f,β) plane, which are parallel to the curve $f(\beta)$, $\beta \in \mathcal{B}$, with a distance $2^{k-1}\varepsilon$ from the latter. The mapping π is the projection onto the curve $f(\beta)$, $\beta \in \mathcal{B}$, along the f-axis.

To bound the series in the RHS of inequality (b.31), let us decompose the likelihood empirical process $\Psi_n(\alpha)$ as:

$$\Psi_n(\alpha) = \tilde{\Psi}_n(\alpha) + R_n(\alpha),$$

where:

$$\tilde{\Psi}_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n}[l_i(\alpha) - l_i(\pi(\alpha))]\,1\{U_i \leq B\} - E[[l_i(\alpha) - l_i(\pi(\alpha))]\,1\{U_i \leq B\}] \equiv \frac{1}{n}\sum_{i=1}^{n}\tilde{\psi}_i(\alpha), \tag{b.33}$$

with:

$$U_i = \sup_{\alpha\in\mathcal{A}}\left\|\frac{\partial \log h(y_i,\alpha)}{\partial\alpha}\right\|, \quad B = \varepsilon^{-1}, \tag{b.34}$$

and:

$$R_n(\alpha) = \frac{1}{n}\sum_{i=1}^{n}[l_i(\alpha) - l_i(\pi(\alpha))]\,1\{U_i > B\} - E[[l_i(\alpha) - l_i(\pi(\alpha))]\,1\{U_i > B\}]. \tag{b.35}$$

Thus, we have:

$$\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\|\hat{f}_n(\beta) - f(\beta)\| > \varepsilon\right] \leq \sum_{k=1}^{\infty}\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k}\left|\tilde{\Psi}_n(\alpha)\right| \geq \frac{1}{2}\lambda_k\right] + \sum_{k=1}^{\infty}\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k}|R_n(\alpha)| \geq \frac{1}{2}\lambda_k\right]. \tag{b.36}$$

ii) Bound of the second series in the RHS of inequality (b.36)

Let us first bound the second series in the RHS of inequality (b.36). By using that $\|\alpha - \pi(\alpha)\| \leq 2^k\varepsilon$ for any $\alpha \in \mathcal{A}_k$, from (b.28) and (b.29) we get:

$$|R_n(\alpha)| \leq 2^k\varepsilon\left(\frac{1}{n}\sum_{i=1}^{n}U_i 1\{U_i > B\} + E[U_i 1\{U_i > B\}]\right),$$

by the mean value Theorem. Thus, from equations (b.32) and (b.35), we have:

$$\begin{aligned}
\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k}|R_n(\alpha)| \geq \frac{1}{2}\lambda_k\right] &\leq \mathbb{P}\left[2^k\varepsilon\left(\frac{1}{n}\sum_{i=1}^{n}U_i 1\{U_i > B\} + E[U_i 1\{U_i > B\}]\right) \geq \frac{1}{2}\lambda_k\right] \\
&= \mathbb{P}\left[\frac{1}{n}\sum_{i=1}^{n}(U_i 1\{U_i > B\} + E[U_i 1\{U_i > B\}]) \geq \frac{1}{16}\mathcal{K}2^k\varepsilon\right].
\end{aligned}$$

By using:

$$E\left[U_i 1\left\{U_i > B\right\}\right] \leq B^{-(\gamma_{11}-1)} E\left[U_i^{\gamma_{11}} 1\left\{U_i > B\right\}\right] \leq \mathcal{R}\varepsilon^{\gamma_{11}-1},$$

from condition iii) and $B = \varepsilon^{-1}$, and by using the Markov inequality, we get:

$$\mathbb{P}\left[\sup_{\alpha \in \mathcal{A}_k} \left|R_n\left(\alpha\right)\right| \geq \frac{1}{2}\lambda_k\right] \leq \left(\frac{16}{\mathcal{K}2^k\varepsilon}\right) 2E\left[U_i 1\left\{U_i > B\right\}\right] \leq \frac{32\mathcal{R}\varepsilon^{\gamma_{11}-2}}{2^k\mathcal{K}}.$$

Thus, we get:

$$\sum_{k=1}^{\infty} \mathbb{P}\left[\sup_{\alpha \in \mathcal{A}_k} \left|R_n\left(\alpha\right)\right| \geq \frac{1}{2}\lambda_k\right] \leq \sum_{k=1}^{\infty} \frac{32\mathcal{R}\varepsilon^{\gamma_{11}-2}}{2^k\mathcal{K}} = \frac{32\mathcal{R}\varepsilon^{\gamma_{11}-2}}{\mathcal{K}}. \tag{b.37}$$

iii) Bound of the first series in the RHS of inequality (b.36)

Now let us consider the first series in the RHS of inequality (b.36). Let us introduce a covering of set $\mathcal{A}_k$ defined in (b.29) by means of $N \equiv N_k$ balls $B\left(\alpha_j, \eta\right)$, $j = 1, 2, \cdots, N$, with center $\alpha_j \equiv \alpha_{j,k}$ and radius:

$$\eta \equiv \eta_k = \frac{1}{64}\frac{\mathcal{K}}{1+\mathcal{M}}2^{2k}\varepsilon^3. \tag{b.38}$$

The number, centers and radii of the balls may depend on index k, but we suppress this dependence to simplify notation. By Fubini's Theorem, the Lebesgue measure of set $\mathcal{A}_k$ is such that:

$$Vol(\mathcal{A}_k) = \int_{\mathcal{A}_k} d\lambda = \int_{\mathcal{B}}\int_{\mathcal{F}_k(\beta)} \lambda(df)\lambda(d\beta) \leq \tilde{C}_m\left(2^k\varepsilon\right)^m \int_{\mathcal{B}} \lambda(d\beta) = \tilde{C}_m\left(2^k\varepsilon\right)^m Vol(\mathcal{B}),$$

where set $\mathcal{F}_k(\beta)$ is defined in equation (b.28), and $\tilde{C}_m$ is a constant depending on dimension m only. Thus, we can chose the number $N \in \mathbb{N}$ of balls covering set $\mathcal{A}_k$ such that:

$$N \leq C^*_{m+q} Vol(\mathcal{A}_k)\eta^{-(m+q)} \leq 64^{m+q} Vol(\mathcal{B})\tilde{C}_{m,q}\varepsilon^{-2m-3q}\left(\frac{1+\mathcal{M}}{\mathcal{K}}\right)^{m+q}, \tag{b.39}$$

where C^*_{m+q} is a constant depending on $m+q$ only, and $\tilde{C}_{m,q} = \tilde{C}_m C^*_{m+q}$. Then, for $\alpha \in \mathcal{A}_k$:

$$\begin{aligned}\left|\tilde{\Psi}_n(\alpha)\right| &\leq \max_{j=1,\ldots,N}\left|\tilde{\Psi}_n(\alpha_j)\right| + \sup_{\alpha,\alpha' \in \mathcal{A}_k : \left\|\alpha-\alpha'\right\| \leq \eta}\left|\tilde{\Psi}_n\left(\alpha'\right) - \tilde{\Psi}_n(\alpha)\right| \\ &\leq \max_{j=1,\ldots,N}\left|\tilde{\Psi}_n(\alpha_j)\right| + 2\eta B(1+\mathcal{M}),\end{aligned}$$

since $B = \varepsilon^{-1}$ bounds the U_i in the definition of $\tilde{\Psi}_n$ [see equations (b.33) and (b.34)], and $\|\pi(\alpha') - \pi(\alpha)\| \leq (1+\mathcal{M})\|\beta' - \beta\|$ [see Condition iv)]. Using $2\eta B(1+\mathcal{M}) = \frac{1}{4}\lambda_k$ from equations (b.32) and (b.38), we get:

$$\mathbb{P}\left[\sup_{\alpha\in\mathcal{A}_k}\left|\tilde{\Psi}_n(\alpha)\right| \geq \frac{\lambda_k}{2}\right] \leq \mathbb{P}\left[\max_{j=1,\ldots,N}\left|\tilde{\Psi}_n(\alpha_j)\right| \geq \frac{\lambda_k}{4}\right] \leq N \sup_{\alpha\in\mathcal{A}_k}\mathbb{P}\left[\left|\tilde{\Psi}_n(\alpha)\right| \geq \frac{\lambda_k}{4}\right]. \tag{b.40}$$

Let us now bound $\mathbb{P}\left[\left|\tilde{\Psi}_n(\alpha)\right| \geq \frac{1}{4}\lambda_k\right]$ for $\alpha \in \mathcal{A}_k$. Since $\tilde{\Psi}_n(\alpha)$ in (b.33) is an average of zero-mean independent random variables, we can use Bernstein's inequality [see Bosq (1998), Theorem 1.2]. Let us first check the conditions of this theorem. We use that $\|\alpha - \pi(\alpha)\| \leq 2^k\varepsilon$ for any $\alpha \in \mathcal{A}_k$. Then, from equation (b.34), for any $\alpha \in \mathcal{A}_k$ we have:

$$\left|\tilde{\psi}_i(\alpha)\right| = \left|[l_i(\alpha) - l_i(\pi(\alpha))]\,1\{U_i \leq B\} - E\left[[l_i(\alpha) - l_i(\pi(\alpha))]\,1\{U_i \leq B\}\right]\right| \leq 2B2^k\varepsilon = 2^{k+1}, \tag{b.41}$$

and:

$$\begin{aligned}E\left[\tilde{\psi}_i(\alpha)^2\right] &= V\left[[l_i(\alpha) - l_i(\pi(\alpha))]\,1\{U_i \leq B\}\right] \leq E_0\left[|l_i(\alpha) - l_i(\pi(\alpha))|^2\right] \\ &\leq \sup_{\alpha\in\mathcal{A}_k} E_0\left[\left(\frac{|l_i(\alpha) - l_i(\pi(\alpha))|}{\|\alpha - \pi(\alpha)\|}\right)^2\right]\left(2^k\varepsilon\right)^2.\end{aligned} \tag{b.42}$$

To bound $\sup_{\alpha\in\mathcal{A}_k} E_0\left[\left(\frac{|l_i(\alpha) - l_i(\pi(\alpha))|}{\|\alpha - \pi(\alpha)\|}\right)^2\right]$ we use:

$$l_i(\alpha) - l_i(\pi(\alpha)) = l_i(f,\beta) - l_i(f(\beta),\beta) = \int_0^1 \frac{\partial l_i(f(\beta) + \tau(f - f(\beta)),\beta)}{\partial f'}(f - f(\beta))d\tau,$$

by the convexity of set $\mathcal{F}$ in condition i) of Lemma B.1. Then, we get:

$$\begin{aligned}\frac{|l_i(\alpha) - l_i(\pi(\alpha))|}{\|\alpha - \pi(\alpha)\|} &\leq \int_0^1 \left\| \frac{\partial l_i(f(\beta) + \tau(f - f(\beta)), \beta)}{\partial f} \right\| d\tau \\ &\leq \left(\int_0^1 \left\| \frac{\partial l_i(f(\beta) + \tau(f - f(\beta)), \beta)}{\partial f} \right\|^2 d\tau \right)^{1/2}.\end{aligned}$$

Then, by the Cauchy-Schwarz inequality, for any $\alpha \in \mathcal{A}_k$ we have:

$$E_0 \left[\left(\frac{|l_i(\alpha) - l_i(\pi(\alpha))|}{\|\alpha - \pi(\alpha)\|} \right)^2 \right] \leq \int_0^1 E_0 \left[\left\| \frac{\partial l_i(f(\beta) + \tau(f - f(\beta)), \beta)}{\partial f} \right\|^2 \right] d\tau \leq \Gamma,$$

where constant Γ is defined in equation (b.27). Thus, from inequality (b.42), for any $\alpha \in \mathcal{A}_k$ we have:

$$E\left[\tilde{\psi}_i(\alpha)^2 \right] \leq \Gamma \left(2^k \varepsilon \right)^2. \tag{b.43}$$

By applying the Bernstein's inequality [see Bosq (1998), Theorem 1.2], and using the definition of λ_k in equation (b.32), we get:

$$\begin{aligned}P\left[\left| \tilde{\Psi}_n(\alpha) \right| \geq \frac{\lambda_k}{4} \right] &\leq 2 \exp\left(-\frac{n(\lambda_k/4)^2}{4\Gamma \left(2^k \varepsilon\right)^2 + 2\left(\frac{\lambda_k}{4}\right) 2^{k+1}} \right) \\ &\leq 2\exp\left(-n\varepsilon^2 2^{k-12} \frac{\mathcal{K}^2}{\Gamma + \mathcal{K}} \right),\end{aligned} \tag{b.44}$$

for any $\alpha \in \mathcal{A}_k$. Thus, from inequalities (b.39), (b.40) and (b.44) we get:

$$\begin{aligned}\sum_{k=1}^{\infty} P\left[\sup_{\alpha \in \mathcal{A}_k} \left| \tilde{\Psi}_n(\alpha) \right| \geq \frac{1}{2}\lambda_k \right] &\leq 64^{m+q} 2Vol(\mathcal{B}) \tilde{C}_{m,q} \varepsilon^{-2m-3q} \left(\frac{1+\mathcal{M}}{\mathcal{K}} \right)^{m+q} \\ &\quad \cdot \sum_{k=1}^{\infty} \exp\left(-n\varepsilon^2 2^{k-12} \frac{\mathcal{K}^2}{\Gamma + \mathcal{K}} \right).\end{aligned}$$

Now, by using that:

$$\begin{aligned}\sum_{k=1}^{\infty} \exp\left(-n\varepsilon^2 2^{k-12} \frac{\mathcal{K}^2}{\Gamma+\mathcal{K}} \right) &\leq \sum_{k=1}^{\infty} \exp\left(-kn\varepsilon^2 2^{-12} \frac{\mathcal{K}^2}{\Gamma+\mathcal{K}} \right) \\ &\leq \frac{1}{1-e^{-1}} \exp\left(-n\varepsilon^2 2^{-12} \frac{\mathcal{K}^2}{\Gamma+\mathcal{K}} \right),\end{aligned}$$

and:

$$\mathcal{K} \geq 2^{12} \frac{1}{n\varepsilon^2},$$

if $n\varepsilon^2 2^{-12} \dfrac{\mathcal{K}^2}{\Gamma + \mathcal{K}} \geq 1$, we get:

$$\sum_{k=1}^{\infty} \mathbb{P}\left[\sup_{\alpha \in \mathcal{A}_k} \left|\tilde{\Psi}_n(\alpha)\right| \geq \frac{1}{2}\lambda_k\right] \leq Vol(\mathcal{B}) \frac{2e}{e-1} \left(\frac{1+\mathcal{M}}{64}\right)^{m+q} \tilde{C}_{m,q} \frac{n^{m+q}}{\varepsilon^q} \exp\left(-n\varepsilon^2 2^{-12} \frac{\mathcal{K}^2}{\Gamma + \mathcal{K}}\right), \tag{b.45}$$

if $n\varepsilon^2 2^{-12} \dfrac{\mathcal{K}^2}{\Gamma + \mathcal{K}} \geq 1$.

iv) Conclusion

Thus, from inequalities (b.36), (b.37) and (b.45) we get:

$$\begin{aligned} P\left[\sup_{\beta \in \mathcal{B}} \|\hat{f}_n(\beta) - f(\beta)\| > \varepsilon\right] &\leq C_1^* Vol(\mathcal{B})(1+\mathcal{M})^{m+q} \frac{n^{m+q}}{\varepsilon^q} \exp\left(-C_2 n\varepsilon^2 \frac{\mathcal{K}^2}{\Gamma + \mathcal{K}}\right) + C_3 \varepsilon^{\gamma_{11}-2} \frac{\mathcal{R}}{\mathcal{K}} \\ &\quad + 1\left\{C_2 n\varepsilon^2 \frac{\mathcal{K}^2}{\Gamma + \mathcal{K}} \leq 1\right\}, \end{aligned}$$

where $C_1^* = \dfrac{2e}{e-1}\left(\dfrac{1}{64}\right)^{m+q} \tilde{C}_{m,q}$, $C_2 = 2^{-12}$ and $C_3 = 32$. Finally, by using the Bernstein's trick $1\{x \leq 1\} \leq e^{1-x}$, we get:

$$P\left[\sup_{\beta \in \mathcal{B}} \|\hat{f}_n(\beta) - f(\beta)\| > \varepsilon\right] \leq C_1 Vol(\mathcal{B})(1+\mathcal{M})^{m+q} \frac{n^{m+q}}{\varepsilon^q} \exp\left(-C_2 n\varepsilon^2 \frac{\mathcal{K}^2}{\Gamma + \mathcal{K}}\right) + C_3 \varepsilon^{\gamma_{11}-2} \frac{\mathcal{R}}{\mathcal{K}},$$

where $C_1 = C_1^* + e$, for n large and ε small enough.

B.4.2 Lemma B.2

Lemma B.2: *Let W be a positive random variable such that $\mathbb{P}[W \geq u] \leq C_1 \exp(-C_2 u^{\varrho})$, for any $u \in \mathbb{R}$ sufficiently large and some constants $C_1, C_2, \varrho > 0$. Then $E[\exp(-uW^{-1})] \leq \tilde{C}_1 \exp\left(-\tilde{C}_2 u^{\varrho/(1+\varrho)}\right)$, for any $u \in \mathbb{R}$ sufficiently large and some constants $\tilde{C}_1, \tilde{C}_2 > 0$.*

Proof of Lemma B.2: Let $Z = W^{-1}$ and $\varepsilon > 0$. We have:

$$\begin{aligned} E[\exp(-uW^{-1})] &= E[\exp(-uZ)] = \int_0^\varepsilon e^{-uz} f(z)dz + \int_\varepsilon^\infty e^{-uz} f(z)dz \\ &= e^{-u\varepsilon}F(\varepsilon) + u\int_0^\varepsilon e^{-uz}F(z)dz + \int_\varepsilon^\infty e^{-uz} f(z)dz, \end{aligned} \tag{b.46}$$

where f and F denote the pdf and cdf of Z, respectively, and we apply integration by part. The second integral in the RHS of equation (b.46) is such that:

$$\int_\varepsilon^\infty e^{-uz} f(z)dz \leq e^{-u\varepsilon} \int_\varepsilon^\infty f(z)dz \leq e^{-u\varepsilon}.$$

Thus, the conclusion follows if we show that:

$$I(u) \equiv u \int_0^\varepsilon e^{-uz} F(z)dz \leq C_3 \exp\left(-C_4 u^{\varrho/(1+\varrho)}\right), \tag{b.47}$$

for some constants $C_3, C_4 > 0$. Now, for $\varepsilon > 0$ small enough, we have $F(z) = \mathbb{P}[W \geq 1/z] \leq C_1 \exp\left[-C_2(1/z)^\varrho\right]$, for $z \leq \varepsilon$. Thus:

$$I(u) \leq C_1 u \int_0^\varepsilon \exp\left[-uz - C_2(1/z)^\varrho\right] dz = C_1 \int_0^{u\varepsilon} \exp\left[-y - C_2(u/y)^\varrho\right] dy.$$

For large u and any $a \in (0,1)$ we get:

$$\begin{aligned} I(u) &\leq C_1 \int_0^{u^a} \exp\left[-y - C_2(u/y)^\varrho\right] dy + C_1 \int_{u^a}^{u\varepsilon} \exp\left[-y - C_2(u/y)^\varrho\right] dy \\ &\leq C_1 e^{-C_2 u^{(1-a)\varrho}} \int_0^{u^a} \exp(-y)dy + C_1 \int_{u^a}^{u\varepsilon} \exp(-y)dy \\ &\leq C_1 e^{-C_2 u^{(1-a)\varrho}} + C_1 e^{-u^a} - C_1 e^{-\varepsilon u}. \end{aligned}$$

Then, for $a = \varrho/(1+\varrho)$, the bound in (b.47) follows, and Lemma B.2 is proved.

B.4.3 Lemma B.3

Lemma B.3: *Suppose Assumptions A.1-A.5, and Assumption H.1 in Appendix A.1 hold. Then:*
(i) *Under Regularity Condition RC.2 (1) in Section B.3, we have $\mathbb{P}\left[\Omega_{1,n,T}(\delta)\right] \to 1$ as $n, T \to \infty$, such that $T/n \to 0$, for any $\delta > 0$, where the event $\Omega_{1,n,T}(\delta)$ is defined in equation (b.11).*

(ii) *Under Regularity Condition RC.3 (1) in Section B.3, we have* $\mathbb{P}\left[\Omega_{3,n,T}(\delta)\right] \to 1$ *as* $n, T \to \infty$, $T/n \to 0$, *for any* $\delta > 0$, *where the event* $\Omega_{3,n,T}(\delta)$ *is defined in equation (b.21).*

Proof of Lemma B.3: We provide the proof of Lemma B.3 (ii) only, since the proof of Lemma B.3 (i) is similar after replacing $\lambda_t(\beta)$ in event $\Omega_{3,n,T}(\delta)$ with 1.

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

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

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

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

Thus, we have to show that:

$$T\mathbb{P}\left[\frac{1}{\sqrt{n}} \sup_{\beta\in\mathcal{B}} \frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{r}\right] \to 0, \tag{b.48}$$

for any $j, l = 1, ..., r$.

Let us write $W_{j,l,n,t}(\beta)$ as:

$$\begin{aligned}
W_{j,l,n,t}(\beta) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(a_{j,l}(Y_{i,t}, f_t(\beta), \beta) 1\{U_{i,t} \leq B_n\} - E\left[a_{j,l}(Y_{i,t}, f_t(\beta), \beta) 1\{U_{i,t} \leq B_n\} | \underline{f_t}\right]\right) \\
&\quad + \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(a_{j,l}(Y_{i,t}, f_t(\beta), \beta) 1\{U_{i,t} > B_n\} - E\left[a_{j,l}(Y_{i,t}, f_t(\beta), \beta) 1\{U_{i,t} > B_n\} | \underline{f_t}\right]\right) \\
&\equiv \tilde{W}_{j,l,n,t}(\beta) + R_{j,l,n,t}(\beta), \qquad \text{(b.49)}
\end{aligned}$$

where $a_{j,l}$ denotes the element (j,l) of matrix function a,

$$U_{i,t} = \sup_{\beta \in \mathcal{B}} \|a(Y_{i,t}, f_t(\beta), \beta)\|, \text{ and } B_n = \frac{4r}{\delta}\sqrt{n}. \qquad \text{(b.50)}$$

Then:

$$\begin{aligned} T\mathbb{P}\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{r}\right] &\leq T\mathbb{P}\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{\left|\tilde{W}_{j,l,n,t}(\beta)\right|}{\lambda_t(\beta)} \geq \frac{\delta}{2r}\right] \\ &\quad + T\mathbb{P}\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{|R_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{2r}\right]. \qquad \text{(b.51)} \end{aligned}$$

Let us now bound the two terms in the RHS of inequality (b.51).

i) Bound of the second term in the RHS of inequality (b.51)

Let us first bound the second term in the RHS of inequality (b.51). By using that $|R_{j,l,n,t}(\beta)| \leq \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left(U_{i,t} 1\{U_{i,t} > B_n\} + E\left[U_{i,t} 1\{U_{i,t} > B_n\} | \underline{f_t}\right]\right)$ uniformly in $\beta \in \mathcal{B}$, and the Markov inequality conditional on $\underline{f_t}$, we get:

$$\begin{aligned} T\mathbb{P}\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} \frac{|R_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{2r}\right] &\leq TE\left[\mathbb{P}\left[\frac{1}{\sqrt{n}} \sup_{\beta \in \mathcal{B}} |R_{j,l,n,t}(\beta)| \geq \frac{\delta}{2r} \inf_{\beta \in \mathcal{B}} \lambda_t(\beta) \middle| \underline{f_t}\right]\right] \\ &\leq \frac{4rT}{\delta} E\left[\frac{E\left[U_{i,t} 1\{U_{i,t} > B_n\} | \underline{f_t}\right]}{\inf_{\beta \in \mathcal{B}} \lambda_t(\beta)}\right]. \end{aligned}$$

Moreover, by the Minkowsky inequality, Regularity Conditions RC.2 (1i) [which is implied by Regularity Conditon RC.3 (1)] and Regularity Condition RC.3 (1ii), we get:

$$\begin{aligned} E\left[\frac{E\left[U_{i,t} 1\{U_{i,t} > B_n\} | \underline{f_t}\right]}{\inf_{\beta \in \mathcal{B}} \lambda_t(\beta)}\right] &\leq B_n^{-2} E\left[\frac{E\left[U_{i,t}^3 | \underline{f_t}\right]}{\inf_{\beta \in \mathcal{B}} \lambda_t(\beta)}\right] \\ &\leq B_n^{-2} E\left[\sup_{\beta \in \mathcal{B}} \lambda_t(\beta)^{-4}\right]^{1/4} E\left[\sup_{\beta \in \mathcal{B}} \|a(Y_{i,t}, f_t(\beta), \beta)\|^4\right]^{3/4} \\ &= O(1/n). \end{aligned}$$

Thus, since $T/n = o(1)$, we get:

$$T\mathbb{P}\left[\frac{1}{\sqrt{n}}\sup_{\beta\in\mathcal{B}}\frac{|R_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{2r}\right] = O(T/n) = o(1). \tag{b.52}$$

ii) Bound of the first term in the RHS of inequality (b.51)

Let us now bound the first term in the RHS of inequality (b.51). To control the supremum over $\mathcal{B}$, let us introduce a finite covering of the compact set $\mathcal{B} \subset \mathbb{R}^q$ by means of M open balls $B(\beta_m, \varepsilon)$ with center β_m and radius ε, $m = 1, ..., M$. We let $M = M_T$ and $\varepsilon = \varepsilon_T$ depend on sample size T, such that $\varepsilon_T \to 0$, $M_T \to \infty$ and $M_T = O\left(\varepsilon_T^{-q}\right)$. We have:

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

Thus, we get:

$$\begin{aligned}
\mathbb{P}\left[\frac{1}{\sqrt{n}}\sup_{\beta\in\mathcal{B}}\frac{\left|\tilde{W}_{j,l,n,t}(\beta)\right|}{\lambda_t(\beta)} \geq \frac{\delta}{2r}\right] &\leq \mathbb{P}\left[\frac{1}{\sqrt{n}}\sup_{\beta,\beta':\left\|\beta'-\beta\right\|\leq\varepsilon_T}\left|\frac{\tilde{W}_{j,l,n,t}(\beta')}{\lambda_t(\beta')} - \frac{\tilde{W}_{j,l,n,t}(\beta)}{\lambda_t(\beta)}\right| \geq \frac{\delta}{4r}\right] \\
&\quad + M_T\sup_{\beta\in\mathcal{B}}\mathbb{P}\left[\frac{1}{\sqrt{n}}\frac{\left|\tilde{W}_{j,l,n,t}(\beta)\right|}{\lambda_t(\beta)} \geq \frac{\delta}{4r}\right] \equiv A_1 + A_2, \text{ say.}
\end{aligned} \tag{b.53}$$

i) Bound of term A_1 in inequality (b.53)

By the Markov inequality we have:

$$A_1 \leq \frac{4r}{\delta\sqrt{n}}E\left[\sup_{\beta,\beta':\left\|\beta'-\beta\right\|\leq\varepsilon_T}\left|\frac{\tilde{W}_{j,l,n,t}(\beta')}{\lambda_t(\beta')} - \frac{\tilde{W}_{j,l,n,t}(\beta)}{\lambda_t(\beta)}\right|\right]. \tag{b.54}$$

To bound the expectation we use:

$$\sup_{\|\beta'-\beta\|\leq\varepsilon_T}\left|\frac{\tilde{W}_{j,l,n,t}(\beta')}{\lambda_t(\beta')}-\frac{\tilde{W}_{j,l,n,t}(\beta)}{\lambda_t(\beta)}\right| \leq \sup_{\beta\in\mathcal{B}}[\lambda_t(\beta)^{-1}]\sup_{\|\beta'-\beta\|\leq\varepsilon_T}\left|\tilde{W}_{j,l,n,t}(\beta')-\tilde{W}_{j,l,n,t}(\beta)\right| + \sup_{\beta\in\mathcal{B}}|\tilde{W}_{j,l,n,t}(\beta)|\sup_{\|\beta'-\beta\|\leq\varepsilon_T}|\lambda_t(\beta')^{-1}-\lambda_t(\beta)^{-1}|. \quad \text{(b.55)}$$

From the definition of $\tilde{W}_{j,l,n,t}(\beta)$ in equation (b.49), we have:

$$\sup_{\beta\in\mathcal{B}}|\tilde{W}_{j,l,n,t}(\beta)| \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\sup_{\beta\in\mathcal{B}}\|a(Y_{i,t},f_t(\beta),\beta)\| + E\left[\sup_{\beta\in\mathcal{B}}\|a(Y_{i,t},f_t(\beta),\beta)\| \mid \underline{f_t}\right]\right\}, \quad \text{(b.56)}$$

and:

$$\sup_{\|\beta'-\beta\|\leq\varepsilon_T}\left|\tilde{W}_{j,l,n,t}(\beta')-\tilde{W}_{j,l,n,t}(\beta)\right| \leq \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left\{\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial vec[a(Y_{i,t},f_t(\beta),\beta)]}{\partial\beta'}\right\| + E\left[\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial vec[a(Y_{i,t},f_t(\beta),\beta)]}{\partial\beta'}\right\| \mid \underline{f_t}\right]\right\}\varepsilon_T. \quad \text{(b.57)}$$

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

$$\begin{aligned}|\lambda_t(\beta')^{-1}-\lambda_t(\beta)^{-1}| &= \left|\sup_{x\in\mathbb{R}^r:\|x\|=1}x'\mu_t(\beta')^{-1}x - \sup_{x\in\mathbb{R}^r:\|x\|=1}x'\mu_t(\beta)^{-1}x\right| \\ &\leq \sup_{x\in\mathbb{R}^r:\|x\|=1}\left|x'\left(\mu_t(\beta')^{-1}-\mu_t(\beta)^{-1}\right)x\right| = \|\mu_t(\beta')^{-1}-\mu_t(\beta)^{-1}\|_{op},\end{aligned}$$

where $\|.\|_{op}$ denotes the matrix operator norm. Since matrix norms are equivalent, we have:

$$\begin{aligned}\|\mu_t(\beta')^{-1}-\mu_t(\beta)^{-1}\|_{op} &\leq c^*\|\mu_t(\beta')^{-1}-\mu_t(\beta)^{-1}\| \\ &\leq c^*\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)^{-1}\|^2 E\left[\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial vec[a(Y_{i,t},f_t(\beta),\beta)]}{\partial\beta'}\right\| |\underline{f_t}\right]\varepsilon_T,\end{aligned}$$

and $\sup_{\beta\in\mathcal{B}}\|\mu_t(\beta)^{-1}\| \leq c^{**}\sup_{\beta\in\mathcal{B}}[\lambda_t(\beta)^{-1}]$, for some constants $c^*, c^{**}>0$. Thus, we get:

$$|\lambda_t(\beta')^{-1}-\lambda_t(\beta)^{-1}| \leq C_{12}\sup_{\beta\in\mathcal{B}}[\lambda_t(\beta)^{-2}]E\left[\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial vec[a(Y_{i,t},f_t(\beta),\beta)]}{\partial\beta'}\right\| |\underline{f_t}\right]\varepsilon_T, \quad \text{(b.58)}$$

where $C_{12} = c^*(c^{**})^2$. From bounds (b.54)-(b.58) and the Cauchy-Schwarz inequality, we get:

$$\begin{aligned} A_1 &\leq \frac{8r\varepsilon_T}{\delta} E\left[\sup_{\beta\in\mathcal{B}} [\lambda_t(\beta)^{-1}] E\left[\sup_{\beta\in\mathcal{B}} \left\| \frac{\partial a\left(Y_{i,t}, f_t(\beta), \beta\right)}{\partial \beta} \right\| \mid \underline{f_t}\right]\right] \\ &\quad + \frac{8C_{12}r\varepsilon_T}{\delta} E\left[E\left[\sup_{\beta\in\mathcal{B}} \left\| a\left(Y_{i,t}, f_t(\beta), \beta\right)\right\| \mid \underline{f_t}\right] \sup_{\beta\in\mathcal{B}} \lambda_t(\beta)^{-2} E\left[\sup_{\beta\in\mathcal{B}} \left\| \frac{\partial vec[a\left(Y_{i,t}, f_t(\beta), \beta\right)]}{\partial \beta'} \right\| |\underline{f_t}\right]\right] \\ &\leq \frac{8C_{13}r\varepsilon_T}{\delta}, \end{aligned} \tag{b.59}$$

where:

$$\begin{aligned} C_{13} &= E\left[\sup_{\beta\in\mathcal{B}} \lambda_t(\beta)^{-2}\right]^{1/2} E\left[\sup_{\beta\in\mathcal{B}} \left\| \frac{\partial vec[a\left(Y_{i,t}, f_t(\beta), \beta\right)]}{\partial \beta'} \right\|^2\right]^{1/2} \\ &\quad + C_{12} E\left[\sup_{\beta\in\mathcal{B}} \lambda_t(\beta)^{-4}\right]^{1/2} E\left[\sup_{\beta\in\mathcal{B}} \left\| a\left(Y_{i,t}, f_t(\beta), \beta\right)\right\|^4\right]^{1/4} \\ &\quad \cdot E\left[\sup_{\beta\in\mathcal{B}} \left\| \frac{\partial vec[a\left(Y_{i,t}, f_t(\beta), \beta\right)]}{\partial \beta'} \right\|^4\right]^{1/4} < \infty, \end{aligned}$$

by Regularity Conditions RC.2 (1i-ii) and RC.3 (1ii).

ii) Bound of term A_2 in inequality (b.53)

To bound A_2, by using the definition of $\tilde{W}_{j,l,n,t}(\beta)$ in equation (b.49) we can write:

$$\begin{aligned} \mathbb{P}\left[\frac{1}{\sqrt{n}} \frac{\left|\tilde{W}_{j,l,n,t}(\beta)\right|}{\lambda_t(\beta)} \geq \frac{\delta}{4r}\right] &= E\left[\mathbb{P}\left[\frac{1}{\sqrt{n}} \left|\tilde{W}_{j,l,n,t}(\beta)\right| \geq \frac{\delta}{4r} \lambda_t(\beta) |\underline{f_t}\right]\right] \\ &= E\left[\mathbb{P}\left[\left|\sum_{i=1}^n \psi_{i,t}(\beta)\right| \geq \frac{\delta \lambda_t(\beta)}{4r} n |\underline{f_t}\right]\right], \end{aligned} \tag{b.60}$$

for $\beta \in \mathcal{B}$, where $\psi_{i,t}(\beta) \equiv a_{j,l}(Y_{i,t}, f_t(\beta), \beta) 1\{U_{i,t} \leq B_n\} - E\left[a_{j,l}(Y_{i,t}, f_t(\beta), \beta) 1\{U_{i,t} \leq B_n\} |\underline{f_t}\right]$. To bound the inner conditional probability in the RHS of equation (b.60), we use the independence property of the $Y_{i,t}$, for i varying, conditional on $\underline{f_t}$, and the Bernstein's inequality [e.g., Bosq (1998), Theorem 1.2]. For any $\beta \in \mathcal{B}$, we have:

$$|\psi_{i,t}(\beta)| \leq 2B_n,$$

from the definitions of $U_{i,t}$ and B_n in (b.50), and:

$$V\left[\psi_{i,t}(\beta)|\underline{f_t}\right] = V\left[a_{j,l}(Y_{i,t}, f_t(\beta), \beta)1\left\{U_{i,t} \le B\right\}|\underline{f_t}\right] \le E\left[\|a(Y_{i,t}, f_t(\beta), \beta)\|^2 |\underline{f_t}\right] \equiv \sigma_t^2(\beta).$$

Then, from Bernstein's inequality applied conditional on $\underline{f_t}$, we get:

$$\begin{aligned}
\mathbb{P}\left[\left|\sum_{i=1}^{n}\psi_{i,t}(\beta)\right| \ge n\frac{\delta}{4r}\lambda_t(\beta)|\underline{f_t}\right] &\le 2\exp\left(-\frac{n\frac{\delta^2}{16r^2}\lambda_t(\beta)^2}{4\sigma_t^2(\beta)+\frac{\delta}{r}\lambda_t(\beta)B_n}\right) \\
&\le 2\exp\left(-C_{14}\sqrt{n}\frac{\lambda_t(\beta)^2}{\sigma_t^2(\beta)+\lambda_t(\beta)}\right), \qquad \text{(b.61)}
\end{aligned}$$

$\mathbb{P}$-a.s., where $C_{14} = \dfrac{\delta^2}{64r^2}$, since $B_n = \dfrac{4r}{\delta}\sqrt{n}$. From inequalities (b.60) and (b.61), we get:

$$\sup_{\beta\in\mathcal{B}}\mathbb{P}\left[\frac{1}{\sqrt{n}}\frac{\left|\tilde{W}_{j,l,n,t}(\beta)\right|}{\lambda_t(\beta)} \ge \frac{\delta}{4r}\right] \le 2E\left[\exp\left(-C_{14}\sqrt{n}\zeta_t^{-1}\right)\right], \qquad \text{(b.62)}$$

where:

$$\begin{aligned}
\zeta_t &\equiv \left(\inf_{\beta\in\mathcal{B}}\lambda_t(\beta)\right)^{-1} + \sup_{\beta\in\mathcal{B}}\frac{\sigma_t^2(\beta)}{\lambda_t(\beta)^2} \le \left(\inf_{\beta\in\mathcal{B}}\lambda_t(\beta)\right)^{-1} + \left(\inf_{\beta\in\mathcal{B}}\lambda_t(\beta)\right)^{-2}\sup_{\beta\in\mathcal{B}}\sigma_t^2(\beta) \\
&= \xi_{t,9} + \xi_{t,9}^2\xi_{t,7},
\end{aligned}$$

where processes $\xi_{t,7}$ and $\xi_{t,9}$ are defined in Regularity Conditions RC.2 (1iv) and RC.3 (1ii). To bound the expectation in the RHS of inequality (b.62), we use Lemma B.2 in Section B.4.2. Let us check the condition of Lemma B.2. From Regularity Conditions RC.2 (1iv) and RC.3 (1ii) in Appendix B.3, we have:

$$\begin{aligned}
\mathbb{P}[\zeta_t \ge u] &\le \mathbb{P}\left[\xi_{t,9} \ge u/2\right] + \mathbb{P}\left[\xi_{t,9}^2\xi_{t,7} \ge u/2\right] \\
&\le \mathbb{P}\left[\xi_{t,9} \ge u/2\right] + \mathbb{P}\left[\xi_{t,9} \ge (u/2)^{1/4}\right] + \mathbb{P}\left[\xi_{t,7} \ge (u/2)^{1/2}\right] \\
&\le b_9\exp[-c_9(u/2)^{d_9}] + b_9\exp[-c_9(u/2)^{d_9/4}] + b_7\exp[-c_7(u/2)^{d_7/2}].
\end{aligned}$$

Thus, the condition of Lemma B.2 is satisfied with $\varrho = \min\{d_7/2, d_9/4\}$. Then, by using Lemma B.2 and the condition on the rate of divergence of n and T, we get:

$$E\left[\exp\left(-C_{14}\sqrt{n}\zeta_t^{-1}\right)\right] \le \tilde{C}_1\exp\left(-\tilde{C}_2(C_{14}\sqrt{n})^{\varrho/(1+\varrho)}\right) \le \tilde{C}_1\exp\left(-C_{15}T^{\varrho/(2+2\varrho)}\right), \qquad \text{(b.63)}$$

for some constants $\tilde{C}_1$, $\tilde{C}_2$ and $C_{15} > 0$. Thus, from inequalities (b.62) and (b.63), we get:

$$A_2 \leq 2\tilde{C}_1 M_T \exp\left(-C_{15} T^{\varrho/(2+2\varrho)}\right). \tag{b.64}$$

iii) Proof of convergence (b.48)

From inequality (b.51), convergence (b.52), and inequalities (b.53), (b.59) and (b.64), we get:

$$T\mathbb{P}\left[\frac{1}{\sqrt{n}}\sup_{\beta\in\mathcal{B}}\frac{|W_{j,l,n,t}(\beta)|}{\lambda_t(\beta)} \geq \frac{\delta}{r}\right] \leq \frac{8C_{13}r}{\delta}T\varepsilon_T + 2\tilde{C}_1 T M_T \exp\left(-C_{15}T^{\varrho/(2+2\varrho)}\right) + o(1).$$

Now choose $\varepsilon_T = T^{-C_{16}}$ for $C_{16} > 1$. Since $M_T = O\left(\varepsilon_T^{-q}\right) = O\left(T^{qC_{16}}\right)$, the convergence (b.48) follows.

B.4.4 Lemma B.4

Lemma B.4: *Suppose Assumptions A.1-A.5, and Assumptions H.1, H.2, H.5, H.6, H.7 (i-ii), H.8-H.10 in Appendix A.1 hold. Then:*
(i) *Under Regularity Condition RC.2 (1) in Section B.3, we have* $\mathbb{P}[\Omega_{2,n,T}(\delta)] \to 1$ *as* $n, T \to \infty$, $T/n \to 0$, *for any* $\delta > 0$, *where the event* $\Omega_{2,n,T}(\delta)$ *is defined in equation (b.12).*
(ii) *Under Regularity Condition RC.3 (1) in Section B.3, we have* $\mathbb{P}[\Omega_{4,n,T}(\delta)] \to 1$ *as* $n, T \to \infty$, $T/n \to 0$, *for any* $\delta > 0$, *where the event* $\Omega_{4,n,T}(\delta)$ *is defined in equation (b.22).*

Proof of Lemma B.4: We give the proof of Lemma B.4 (ii) only, since the proof of Lemma B.4 (i) is similar after replacing $\lambda_t(\beta)$ in event $\Omega_{4,n,T}(\delta)$ with 1.

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

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

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

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

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

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

Now, let sequence η_T be such that:

$$
\eta_T = (C_{17} \log T)^{-2/C_{18}}, \quad C_{17} > 0, \quad 0 < C_{18} \leq \min\{2d_8, d_9\}, \tag{b.65}
$$

where constants $d_8 > 0$ and $d_9 > 0$ are defined in Regularity Condition RC.2 (1v) and RC.3 (1ii) in Section B.3.

i) Proof that $P_{1,n,T} = o(1)$

We have $\frac{(\log n)^{\delta_2}}{\sqrt{n}} = o(\eta_T)$, as $n, T \to \infty$ such that $T/n \to 0$, for any constant δ_2. Thus, we get $P_{1,n,T} = o(1)$ as $n, T \to \infty$ from Limit Theorem 1 in Appendix B.1.

ii) Proof that $P_{2,n,T} = o(1)$

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

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

for any constant $\delta^* > 0$ and large T. The RHS probability converges to zero by the same argument as in the proof of Lemma B.3 (ii) in Section B.4.3 and using Regularity Conditions RC.2 (1i-ii), (1v) and RC.3 (1i-ii).

iii) Proof that $P_{3,T} = o(1)$

We have $P_{3,T} \leq T\mathbb{P}\left[\varsigma_t \geq \dfrac{\delta}{2\eta_T}\right]$. By using $\varsigma_t \leq \left(\inf\limits_{\beta\in\mathcal{B}} \lambda_t(\beta)\right)^{-1} \sup\limits_{\beta\in\mathcal{B}} E_0[b(Y_{i,t}, f_t(\beta), \beta)^2|\underline{f_t}]^{1/2} \leq \xi_{t,9}\xi_{t,8}^{1/2}$, where processes $\xi_{t,8}$ and $\xi_{t,9}$ are defined in Regularity Conditions RC.2 (1v) and RC.3 (1ii). We get:

$$
\begin{aligned}
\mathbb{P}\left[\varsigma_t \geq \frac{\delta}{2\eta_T}\right] &\leq \mathbb{P}\left[\xi_{t,9} \geq \left(\frac{\delta}{2\eta_T}\right)^{1/2}\right] + \mathbb{P}\left[\xi_{t,8} \geq \frac{\delta}{2\eta_T}\right] \\
&\leq b_9 \exp\left(-c_9(\delta/(2\eta_T))^{d_9/2}\right) + b_8 \exp\left(-c_8(\delta/(2\eta_T))^{d_8}\right).
\end{aligned}
$$

Then, by the definition of η_T in (b.65), we deduce:

$$
\begin{aligned}
P_{3,T} &\leq Tb_9 \exp\left(-c_9(\delta/2)^{d_9/2} C_{17} \log T\right) + Tb_8 \exp\left(-c_8(\delta/2)^{d_8} C_{17} \log T\right) \\
&= b_9 T^{1-c_9 C_{17}(\delta/2)^{d_9/2}} + b_8 T^{1-c_8 C_{17}(\delta/2)^{d_8}}.
\end{aligned}
$$

Then, for $C_{17} > \max\{c_9^{-1}(\delta/2)^{-d_9/2}, c_8^{-1}(\delta/2)^{-d_8}\}$, we get $P_{3,T} = o(1)$.

B.4.5 Lemma B.5

Lemma B.5: *Let mapping a admit values in the set of (r,r) symmetric matrices and satisfy Regularity Condition RC.3 (1) in Section B.3, and let $\mu_t(\beta) = E_0[a(Y_{i,t}, f_t(\beta), \beta)|\underline{f_t}]$. Then, for any $\eta > 0$, there exists a compact subset $\mathcal{K} \subset \mathcal{U}$ of the set $\mathcal{U}$ of positive definite (r,r) matrices, such that $\mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}\right] \geq 1 - \eta$.*

Proof of Lemma B.5: The matrix $\mu_t(\beta)$ is positive definite, for any t and $\beta \in \mathcal{B}$, $\mathbb{P}$-a.s. Let $eig_{min}(x)$ and $eig_{max}(x)$ denote the smallest and the largest eigenvalues of the symmetric matrix $x \in \mathbb{SR}^{r\times r}$, respectively, and let $\lambda_t(\beta) = eig_{min}(\mu_t(\beta))$ and $\Lambda_t(\beta) = eig_{max}(\mu_t(\beta))$. For any constants C_1, C_2 such that $0 < C_1 \leq C_2 < \infty$, let us define the set $\mathcal{K}_{C_1,C_2} = \left\{x \in \mathbb{SR}^{r\times r} : C_1 \leq eig_{min}(x) \leq eig_{max}(x) \leq C_2\right\} \subset \mathcal{U}$. This is a compact subset of the set of (r,r) positive definite matrices. Then:

$$
\begin{aligned}
\mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_{C_1,C_2}\right] &= \mathbb{P}\left[\inf_{\beta\in\mathcal{B}} \lambda_t(\beta) \geq C_1,\ \sup_{\beta\in\mathcal{B}} \Lambda_t(\beta) \leq C_2\right] \\
&\geq 1 - \mathbb{P}\left[\inf_{\beta\in\mathcal{B}} \lambda_t(\beta) < C_1\right] - \mathbb{P}\left[\sup_{\beta\in\mathcal{B}} \Lambda_t(\beta) > C_2\right].
\end{aligned}
$$

Now, we use $\Lambda_t(\beta) \leq c^* \|\mu_t(\beta)\|$, for any t and $\beta \in \mathcal{B}$, $\mathbb{P}$-a.s., and a positive constant c^* that depends on dimension r only. Indeed, the largest eigenvalue $eig_{max}(A)$ of a symmetric matrix $A \in \mathbb{SR}^{r \times r}$ coincides with the operator norm $\|A\|_{op} \equiv \sup_{\xi \in \mathbb{R}^r : \|\xi\|=1} \xi' A \xi$ of the matrix, i.e. $eig_{max}(A) = \|A\|_{op}$, and all norms in an Euclidean space are equivalent. Then, we get:

$$\begin{aligned}
\mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_{C_1,C_2}\right] &\geq 1 - \mathbb{P}\left[\sup_{\beta \in \mathcal{B}} [\lambda_t(\beta)^{-1}] > C_1^{-1}\right] - \mathbb{P}\left[\sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\| > C_2/c^*\right] \\
&\geq 1 - C_1 E\left[\sup_{\beta \in \mathcal{B}} [\lambda_t(\beta)^{-1}]\right] - (c^*/C_2) E\left[\sup_{\beta \in \mathcal{B}} \|\mu_t(\beta)\|\right],
\end{aligned}$$

by the Markov inequality. The two expectations in the last line are finite by Regularity Condition RC.3 (1ii), and Regularity Condition RC.2 (1i), which is implied by Regularity Condition RC.3 (1). Then, for any $\eta > 0$, there exist $C_1 > 0$ and $C_2 < \infty$ such that $\mathbb{P}\left[\{\mu_t(\beta), \beta \in \mathcal{B}\} \subset \mathcal{K}_{C_1,C_2}\right] \geq 1 - \eta$.

APPENDIX C
TECHNICAL LEMMAS

We provide Lemmas 1-8 in Sections C.1-C.8. The secondary Lemmas C.1-C.4 used in the proofs of Lemmas 1-8 are given in Section C.9.

C.1 Lemma 1

LEMMA 1 *Under Assumptions A.1-A.5 and H.1-H.6, H.7 (i)-(ii), H.8-H.10, H.13, and if $n, T \to \infty$ such that $T^{\nu}/n = O(1)$, for $\nu > 1$, we have: (i) $\sup_{\beta \in \mathcal{B}} |\mathcal{L}^*_{nT}(\beta) - \mathcal{L}^*(\beta)| = o_p(1)$, where functions $\mathcal{L}^*_{nT}(\beta)$ and $\mathcal{L}^*(\beta)$ are defined in equations (3.7) and (4.4), respectively; (ii) $\sup_{\beta \in \mathcal{B},\ \theta \in \Theta} |\mathcal{L}_{1,nT}(\beta, \theta) - \mathcal{L}_1(\beta, \theta)| = o_p(1)$, where functions $\mathcal{L}_{1,nT}(\beta, \theta)$ and $\mathcal{L}_1(\beta, \theta)$ are defined in equations (3.8) and (a.10), respectively.*

Proof of Lemma 1 (i): We apply Limit Theorem 3 in Appendix B.3 with $a(y_{i,t}, y_{i,t-1}, f_t, \beta) = \log h(y_{i,t}|y_{i,t-1}, f_t; \beta)$ and φ being the identity mapping. Let us check Regularity Condition RC.2 in Appendix B.3. Regularity Condition RC.2 (1i) is implied by Assumption H.3 (ii) in Appendix A.1. To check Regularity Condition RC.2 (1ii), we use $\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial \beta} \right\|$ $\leq \sup_{\beta \in \mathcal{B}} \left\| \frac{\partial \log h}{\partial \beta}(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta) \right\| + \sup_{\beta \in \mathcal{B}} \left\| \frac{\partial \log h}{\partial f}(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta) \right\| \sup_{\beta \in \mathcal{B}} \left\| \frac{\partial f_t(\beta)}{\partial \beta'} \right\|$ and $\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial f_t(\beta)}{\partial \beta'} \right\| \leq \tilde{c}\xi^*_{t,1}(\xi^{**}_{t,1})^{1/2}$, from equation (b.3) in Section B.1, where processes $\xi^*_{t,1}$ and $\xi^{**}_{t,1}$ are defined in Assumption H.5 in Appendix A.1, and $\tilde{c} > 0$ is a constant. Then, Regularity Condition RC.2 (1ii) is implied by Assumptions H.3 (ii) and H.5 in Appendix A.1. Regularity Conditions RC.2 (1iii, iv, v) are implied by Assumptions H.4 (ii) and H.5 in Appendix A.1. Finally, Regularity Condition RC.2 (2) in Appendix B.3 is satisfied, since the identity mapping is Lipschitz continuous and $E_0[|\varphi(\mu_t(\beta))|] \leq E_0[|\log h(y_{i,t}|y_{i,t-1}, f_t; \beta)|] < \infty$ from Assumption H.3 (ii). Thus, the smoothness regularity conditions to apply Limit Theorem 3 are satisfied.

Proof of Lemma 1 (ii): Let us write $\mathcal{L}_{1,nT}(\beta, \theta) = \mathcal{L}_{11,nT}(\beta) + \mathcal{L}_{12,nT}(\beta, \theta)$, where $\mathcal{L}_{11,nT}(\beta) = -\frac{1}{2}\sum_{t=1}^{T} \log \det I_{n,t}(\beta)$ and $\mathcal{L}_{12,nT}(\beta, \theta) = \frac{1}{T}\sum_{t=1}^{T} \log g\left(\hat{f}_{n,t}(\beta)|\hat{f}_{n,t-1}(\beta); \theta\right)$. To show the uniform convergence of $\mathcal{L}_{11,nT}(\beta)$, we apply Limit Theorem 3 with $a(y_{i,t}, y_{i,t-1}, f_t, \beta) =$

$-\dfrac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t; \beta)}{\partial f_t \partial f_t'}$, $\mu_t(\beta) = E_0 \left[-\dfrac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f_t \partial f_t'} | \underline{f_t} \right] = I_{t,ff}(\beta)$, and $\varphi(x) = \log \det(x)$, for x a symmetric positive definite (m, m) matrix. Regularity Condition RC.3 (1) in Appendix B.3 is implied by Assumptions H.3, H.4 (iii) and H.5. In Lemma C.1 in Appendix C.9.1 we show that mapping φ satisfies Regularity Condition RC.3 (2). Then, from Limit Theorem 3 it follows that $\mathcal{L}_{11,nT}(\beta)$ converges to $-\dfrac{1}{2} E_0[\log \det I_{t,ff}(\beta)]$ in probability, uniformly w.r.t. $\beta \in \mathcal{B}$.

To show the uniform convergence of $\mathcal{L}_{12,nT}(\beta, \theta)$, we apply Limit Theorem 2 with $G(f_t, f_{t-1}; \theta) = \log g(f_t|f_{t-1}; \theta)$. Regularity Condition RC.1 in Appendix B.2 is implied by Assumptions H.5 and H.13 in Appendix A.1. Then, $\mathcal{L}_{12,nT}(\beta, \theta)$ converges to $E_0[\log g(f_t(\beta)|f_{t-1}(\beta); \theta)]$ in probability, uniformly w.r.t. $\beta \in \mathcal{B}, \theta \in \Theta$.

C.2 Lemma 2

LEMMA 2 *Under Assumptions A.1-A.5 and H.1-H.3, H.5-H.11, and if* $T^\nu/n = O(1)$, $\nu > 1$, *we have* $\inf\limits_{1 \le t \le T} \inf\limits_{\beta \in \mathcal{B}} \inf\limits_{f_t \in \mathcal{F}_n} \dfrac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta); \beta) - \mathcal{L}_{n,t}(f_t; \beta)}{\|\hat{f}_{n,t}(\beta) - f_t\|^2} \ge \dfrac{C_2}{[\log(n)]^{C_3}}$, *w.p.a. 1, for some constants* $C_2, C_3 > 0$, *where* $\mathcal{L}_{n,t}(f; \beta) = \dfrac{1}{n} \sum\limits_{i=1}^{n} \log h(y_{i,t}|y_{i,t-1}, f; \beta)$.

Proof of Lemma 2: To simplify the notation, we assume that f_t is scalar, i.e., $m = 1$. Let $\eta > 0$. We have:

$$
\begin{aligned}
&\mathbb{P} \left[\inf_{1 \le t \le T} \inf_{\beta \in \mathcal{B}} \inf_{f_t \in \mathcal{F}_n} \frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta); \beta) - \mathcal{L}_{n,t}(f_t; \beta)}{[\hat{f}_{n,t}(\beta) - f_t]^2} \le \frac{C_2}{[\log(n)]^{C_3}} \right] \\
&\le \mathbb{P} \left[\inf_{1 \le t \le T} \inf_{\beta \in \mathcal{B}} \inf_{f_t \in \mathcal{F}_n : |f_t - \hat{f}_{n,t}(\beta)| \le \eta} \frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta); \beta) - \mathcal{L}_{n,t}(f_t; \beta)}{[\hat{f}_{n,t}(\beta) - f_t]^2} \le \frac{C_2}{[\log(n)]^{C_3}} \right] \\
&+ \mathbb{P} \left[\inf_{1 \le t \le T} \inf_{\beta \in \mathcal{B}} \inf_{f_t \in \mathcal{F}_n : |f_t - \hat{f}_{n,t}(\beta)| \ge \eta} \frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta); \beta) - \mathcal{L}_{n,t}(f_t; \beta)}{[\hat{f}_{n,t}(\beta) - f_t]^2} \le \frac{C_2}{[\log(n)]^{C_3}} \right] \equiv P_{1,nT} + P_{2,nT}.
\end{aligned}
\tag{c.1}
$$

Let us now show that probabilities $P_{1,nT}$ and $P_{2,nT}$ are $o(1)$, for suitable constants $C_2, C_3 > 0$.

i) Proof that $P_{1,nT} = o(1)$

By a Taylor expansion of function $\mathcal{L}_{n,t}(f_t;\beta)$ around $f_t = \hat{f}_{n,t}(\beta)$, and by using that $\dfrac{\partial \mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)}{\partial f_t} = 0$, w.p.a. 1, we get:

$$P_{1,nT} \leq \mathbb{P}\left[\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-\hat{f}_{n,t}(\beta)|\leq\eta} -\frac{\partial^2\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^2} \leq \frac{2C_2}{[\log(n)]^{C_3}}\right] + o(1).$$

Since $\hat{f}_{n,t}(\beta)$ converges uniformly to $f_t(\beta)$ (Limit Theorem 1 in Appendix B.1), we have w.p.a. 1:

$$\begin{aligned}
&\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-\hat{f}_{n,t}(\beta)|\leq\eta} -\frac{\partial^2\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^2} \geq \inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-f_t(\beta)|\leq 2\eta} -\frac{\partial^2\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^2}\\
&\geq \inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f\in\mathcal{F}_n:|f-f_t(\beta)|\leq 2\eta} E_0\left[-\frac{\partial^2\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f^2}|\underline{f_t}\right]\\
&- \sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}\left|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f^2} - E_0\left[\frac{\partial^2\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f^2}|\underline{f_t}\right]\right|. \qquad \text{(c.2)}
\end{aligned}$$

If constant $\eta > 0$ is such that $2\eta \leq \eta^*$, where η^* is defined in Assumption H.5 in Appendix A.1, then the first term in the RHS of inequality (c.2) is such that:

$$\inf_{\beta\in\mathcal{B}}\inf_{f\in\mathcal{F}_n:|f-f_t(\beta)|\leq 2\eta} E_0\left[-\frac{\partial^2\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f^2}|\underline{f_t}\right] \geq (\xi_{t,1})^{-1},$$

where process $\xi_{t,1}$ is defined in Assumption H.5 in Appendix A.1. Moreover, in Lemma C.2 in Appendix C.9.2 we show that the second term in the RHS of inequality (c.2) is $O_p\left(\dfrac{[\log(n)]^{\delta_3}}{\sqrt{n}}\right)$, for a constant $\delta_3 > 0$. Then, from inequality (c.2) we get w.p.a. 1:

$$\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-\hat{f}_{n,t}(\beta)|\leq\eta} -\frac{\partial^2\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^2} \geq \inf_{1\leq t\leq T}(\xi_{t,1})^{-1} - \frac{C_2}{[\log(n)]^{C_3}}.$$

Then, it follows:

$$\begin{aligned}
P_{1,nT} &\leq \mathbb{P}\left[\inf_{1\leq t\leq T}(\xi_{t,1})^{-1} \leq \frac{3C_2}{[\log(n)]^{C_3}}\right] + o(1) = \mathbb{P}\left[\sup_{1\leq t\leq T}\xi_{t,1} \geq \frac{[\log(n)]^{C_3}}{3C_2}\right] + o(1)\\
&\leq T\mathbb{P}\left[\xi_{t,1} \geq \frac{[\log(n)]^{C_3}}{3C_2}\right] + o(1).
\end{aligned}$$

Thus, from Assumption H.5 we get:

$$P_{1,nT} \leq b_1 T \exp\left(-c_1\left(\frac{[\log(n)]^{C_3}}{3C_2}\right)^{d_1}\right) + o(1) = O(T/n) + o(1) = o(1),$$

if C_2 and C_3 are such that $C_3 \geq 1/d_1$ and $c_1(1/(3C_2))^{d_1} \geq 1$, i.e., $C_2 \leq \frac{1}{3}c_1^{1/d_1}$.

ii) Proof that $P_{2,nT} = o(1)$

Let us first derive a lower bound for $\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-\hat{f}_{n,t}(\beta)|\geq\eta}\frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t;\beta)}{[\hat{f}_{n,t}(\beta)-f_t]^2}$.
From Assumption H.7 (iii), the uniform convergence of $\hat{f}_{n,t}(\beta)$ to $f_t(\beta)$ (Limit Theorem 1 in Appendix B.1) and by using that $\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta) - \mathcal{L}_{n,t}(f_t;\beta) \geq 0$ for $f_t \in \mathcal{F}_n$ and $\beta \in \mathcal{B}$, we have w.p.a. 1:

$$\begin{aligned}
&\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-\hat{f}_{n,t}(\beta)|\geq\eta}\frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t;\beta)}{[\hat{f}_{n,t}(\beta)-f_t]^2}\\
&\quad\geq \frac{1}{4R_n^2}\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-\hat{f}_{n,t}(\beta)|\geq\eta}[\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t;\beta)]\\
&\quad\geq \frac{1}{4R_n^2}\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-f_t(\beta)|\geq\eta/2}[\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t;\beta)], \qquad \text{(c.3)}
\end{aligned}$$

where R_n is defined in Assumption H.7 (iii). Moreover, we have:

$$\begin{aligned}
&\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-f_t(\beta)|\geq\eta/2}[\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t;\beta)]\\
&\quad\geq \inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-f_t(\beta)|\geq\eta/2}[\mathcal{L}_t(f_t(\beta);\beta)-\mathcal{L}_t(f_t;\beta)]\\
&\quad-\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\left|\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t(\beta);\beta)\right| - 2\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\sup_{f_t\in\mathcal{F}_n}\left|\mathcal{L}_{n,t}(f_t;\beta)-\mathcal{L}_t(f_t;\beta)\right|, \qquad \text{(c.4)}
\end{aligned}$$

where:

$$\mathcal{L}_t(f;\beta) = E_0\left[\log h(y_{i,t}|y_{i,t-1},f;\beta)|\underline{f_t}\right]. \qquad \text{(c.5)}$$

From Assumption H.8, the first term in the RHS of inequality (c.4) is such that:

$$\begin{aligned}
&\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-f_t(\beta)|\geq\eta/2}[\mathcal{L}_t(f_t(\beta);\beta)-\mathcal{L}_t(f_t;\beta)]\\
&\quad\geq\frac{\eta^2}{4}\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-f_t(\beta)|\geq\eta/2}\frac{\mathcal{L}_t(f_t(\beta);\beta)-\mathcal{L}_t(f_t;\beta)}{[f_t-f_t(\beta)]^2}\geq\frac{\eta^2}{8[\log(n)]^{\gamma_2}}\inf_{1\leq t\leq T}\mathcal{K}_t. \qquad \text{(c.6)}
\end{aligned}$$

Moreover, in Lemma C.3 in Appendix C.9.3 we prove that the second and third terms in the RHS of inequality (c.4) are $O_p\left(\frac{[\log(n)]^{\delta_4}}{n}\right)$ and $O_p\left(\frac{[\log(n)]^{\delta_5}}{\sqrt{n}}\right)$, respectively, for some constants $\delta_4, \delta_5 > 0$. Then, from inequalities (c.3)-(c.6) and by using $R_n \leq C_4[\log(n)]^{\gamma_1}$, with $C_4, \gamma_1 > 0$ [see Assumption H.7 (iii)], we get w.p.a. 1:

$$
\begin{aligned}
&\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}\inf_{f_t\in\mathcal{F}_n:|f_t-\hat{f}_{n,t}(\beta)|\geq\eta}\frac{\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t;\beta)}{[\hat{f}_{n,t}(\beta)-f_t]^2}\\
&\quad\geq\frac{\eta^2}{32R_n^2[\log(n)]^{\gamma_2}}\inf_{1\leq t\leq T}\mathcal{K}_t+O_p\left(\frac{[\log(n)]^{\delta_4}}{nR_n^2}\right)+O_p\left(\frac{[\log(n)]^{\delta_5}}{\sqrt{n}R_n^2}\right),\\
&\quad\geq\frac{\eta^2}{32C_4^2[\log(n)]^{\gamma_2+2\gamma_1}}\inf_{1\leq t\leq T}\mathcal{K}_t-\frac{C_2}{[\log(n)]^{C_3}}.
\end{aligned}
$$

From the definition of probability $P_{2,nT}$ in equation (c.1), we get:

$$
\begin{aligned}
&P_{2,nT}\leq\mathbb{P}\left[\frac{\eta^2}{32C_4^2[\log(n)]^{\gamma_2+2\gamma_1}}\inf_{1\leq t\leq T}\mathcal{K}_t\leq\frac{2C_2}{[\log(n)]^{C_3}}\right]+o(1)\\
&\leq\mathbb{P}\left[\inf_{1\leq t\leq T}\mathcal{K}_t\leq\frac{64C_2C_4^2}{\eta^2[\log(n)]^{C_3-\gamma_2-2\gamma_1}}\right]+o(1)=\mathbb{P}\left[\sup_{1\leq t\leq T}\mathcal{K}_t^{-1}\geq\frac{\eta^2[\log(n)]^{C_3-\gamma_2-2\gamma_1}}{64C_2C_4^2}\right]+o(1)\\
&\leq T\mathbb{P}\left[\mathcal{K}_t^{-1}\geq\frac{\eta^2[\log(n)]^{C_3-\gamma_2-2\gamma_1}}{64C_2C_4^2}\right]+o(1).
\end{aligned}
$$

From Assumption H.10 we get:

$$
P_{2,nT}\leq b_3T\exp\left(-c_3\left[\frac{\eta^2[\log(n)]^{C_3-\gamma_2-2\gamma_1}}{64C_2C_4^2}\right]^{d_3}\right)+o(1)=O(T/n)+o(1)=o(1),
$$

if C_2, C_3 are such that $(C_3-\gamma_2-2\gamma_1)d_3\geq 1$ and $c_3[\eta^2/(64C_2C_4^2)]^{d_3}\geq 1$, i.e., $C_3\geq\gamma_2+2\gamma_1+1/d_3$ and $C_2\leq\frac{\eta^2c_3^{1/d_3}}{64C_4^2}$.

C.3 Lemma 3

LEMMA 3 *Let us define the sequence $\kappa_n = 2[\log(n)/C_6]^{C_7}$, for $n\in\mathbb{N}$, where constants $C_6, C_7 > 0$ are such that $C_6\leq\min\{c_1, c_5\}$ and $C_7\geq\max\{3/d_1, 2/d_5\}$, for $c_1, d_1 > 0$ and $c_5, d_5 > 0$ defined in Assumptions H.5 and H.13 (iii), respectively. Then, under Assumptions A.1-A.5 and H.1, H.2, H.5-H.11, H.13 (iii) and if $T^\nu/n = O(1)$, $\nu > 1$, w.p.a. 1, we have: (i) $\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}}I_{n,t}(\beta)\geq\kappa_n^{-1}$, (ii) $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}I_{n,t}(\beta)\leq\kappa_n$, (iii) $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\tilde{J}_{3,n,t}(\beta)\leq\kappa_n$, and (iv)*

$\sup_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}, \theta \in \Theta} \tilde{D}_{pq,n,t}(\beta, \theta) \leq \kappa_n$, *for* $p + q = 1$, *where* $I_{n,t}(\beta)$ *is defined in equation (3.4), and* $\tilde{J}_{3,n,t}(\beta)$ *and* $\tilde{D}_{pq,n,t}(\beta, \theta)$ *are as in equation (a.14).*

Proof of Lemma 3 (i): By using Limit Theorem 1 in Appendix B.1 and the mean-value theorem, we have w.p.a. 1:

$$\begin{aligned}
\inf_{1 \leq t \leq T} \inf_{\beta \in \mathcal{B}} I_{n,t}(\beta) \geq & \inf_{1 \leq t \leq T} \inf_{\beta \in \mathcal{B}} E_0 \left[-\frac{\partial^2 \log h(y_{i,t} | y_{i,t-1}, f_t(\beta); \beta)}{\partial f^2} | \underline{f_t} \right] \\
& - \sup_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}} \sup_{f \in \mathcal{F}_n} \left| \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log h(y_{i,t} | y_{i,t-1}, f; \beta)}{\partial f^2} - E_0 \left[\frac{\partial^2 \log h(y_{i,t} | y_{i,t-1}, f; \beta)}{\partial f^2} | \underline{f_t} \right] \right| \\
& - \sup_{1 \leq t \leq T} \sup_{\beta \in \mathcal{B}} E_0 \left[\sup_{f : |f - f_t(\beta)| \leq \eta^*} \left| \frac{\partial^3 \log h(y_{i,t} | y_{i,t-1}, f; \beta)}{\partial f^3} \right| | \underline{f_t} \right] \left| \hat{f}_{n,t}(\beta) - f_t(\beta) \right|,
\end{aligned}$$

for $\eta^* > 0$. The first term in the RHS is such that $\inf_{\beta \in \mathcal{B}} E_0 \left[-\frac{\partial^2 \log h(y_{i,t} | y_{i,t-1}, f_t(\beta); \beta)}{\partial f^2} | \underline{f_t} \right] \geq (\xi^*_{t,1})^{-1} \geq (\xi_{t,1})^{-1}$, where processes $\xi_{t,1}$ are $\xi^*_{t,1}$ are defined in Assumption H.5 in Appendix A.1. Moreover, from Lemma C.2 in Appendix C.9.2, Limit Theorem 1 in Appendix B.1 and Assumption H.5, the second and third terms in the RHS are $O_p \left(\frac{(\log n)^{\max\{\delta_2, \delta_3\}}}{\sqrt{n}} \right)$, where constants $\delta_2 > 0$ and $\delta_3 > 0$ are defined in Limit Theorem 1 and Lemma C.2, respectively. Therefore, we get w.p.a. 1:

$$\inf_{1 \leq t \leq T} \inf_{\beta \in \mathcal{B}} I_{n,t}(\beta) \geq \inf_{1 \leq t \leq T} (\xi_{t,1})^{-1} + O_p \left(\frac{(\log n)^{\max\{\delta_2, \delta_3\}}}{\sqrt{n}} \right) \geq \inf_{1 \leq t \leq T} (\xi_{t,1})^{-1} - \kappa_n^{-1}. \qquad \text{(c.7)}$$

Thus:

$$\begin{aligned}
\mathbb{P} \left[\inf_{1 \leq t \leq T} \inf_{\beta \in \mathcal{B}} I_{n,t}(\beta) \geq \kappa_n^{-1} \right] & \geq \mathbb{P} \left[\inf_{1 \leq t \leq T} (\xi_{t,1})^{-1} \geq 2\kappa_n^{-1} \right] + o(1) \\
& = 1 - \mathbb{P} \left[\sup_{1 \leq t \leq T} \xi_{t,1} \geq \kappa_n / 2 \right] + o(1) \\
& \geq 1 - T\mathbb{P} \left[\xi_{t,1} \geq \kappa_n / 2 \right] + o(1).
\end{aligned}$$

From Assumption H.5 and the definition of κ_n, we have $\mathbb{P} [\xi_{t,1} \geq \kappa_n / 2] \leq b_1 \exp \left(-c_1 (\kappa_n / 2)^{d_1} \right) \leq b_1 / n$, since $c_1 (\kappa_n / 2)^{d_1} \geq \log(n)$. Then, we get $\mathbb{P} \left[\inf_{1 \leq t \leq T} \inf_{\beta \in \mathcal{B}} I_{n,t}(\beta) \geq \kappa_n^{-1} \right] \geq 1 - O(T/n) + o(1) = 1 - o(1)$, since $T/n \to 0$.

Proof of Lemma 3 (ii): Similarly, we have w.p.a. 1:

$$\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}} I_{n,t}(\beta) \leq \sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}} E_0\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1},f_t(\beta);\beta)}{\partial f^2}|\underline{f_t}\right]+\kappa_n/2.$$

Moreover, $\sup_{\beta\in\mathcal{B}} E_0\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1},f_t(\beta);\beta)}{\partial f^2}|\underline{f_t}\right] \leq (\xi^{**}_{t,1})^{1/2} \leq (\xi_{t,1})^{1/2}$, where processes ξ_t and $\xi^{**}_{t,1}$ are defined in Assumption H.5. Then, we get:

$$\begin{aligned}
\mathbb{P}\left[\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}} I_{n,t}(\beta)\leq\kappa_n\right] &\geq \mathbb{P}\left[\sup_{1\leq t\leq T}(\xi_{t,1})^{1/2}\leq\kappa_n/2\right]+o(1)\\
&\geq 1-T\mathbb{P}\left[\xi_{t,1}\geq(\kappa_n/2)^2\right]+o(1)\\
&\geq 1-Tb_1\exp\left(-c_1(\kappa_n/2)^{2d_1}\right)=1-O(T/n)-o(1)=1-o(1),
\end{aligned}$$

from Assumption H.5, the definition of κ_n and the condition $T/n\to 0$.

Proof of Lemma 3 (iii): From the uniform convergence of $\hat{f}_{n,t}(\beta)$ to $f_t(\beta)$ (Limit Theorem 1 in Appendix B.1), and since sequence ε_n involved in the definition of $\tilde{J}_{3,nt}(\beta)$ is such that $\varepsilon_n=o(1)$ (see Appendix A.2.1), we have for any $\eta>0$, w.p.a. 1:

$$\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\tilde{J}_{3,nt}(\beta)\leq\left(\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}} I_{n,t}(\beta)\right)^{-3/2}\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\sup_{f_t:|f_t-f_t(\beta)|\leq\eta}\left|\frac{\partial^3\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^3}\right|.$$

Moreover, we have $\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}} I_{n,t}(\beta)\geq\inf_{1\leq t\leq T}(\xi_{t,1})^{-1}+O_p\left(\frac{(\log n)^{\max\{\delta_2,\delta_3\}}}{\sqrt{n}}\right)$ from inequality (c.7), and:

$$\begin{aligned}
\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\sup_{f_t:|f_t-f_t(\beta)|\leq\eta}\left|\frac{\partial^3\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^3}\right| &\leq \sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\sup_{f_t:|f_t-f_t(\beta)|\leq\eta}\left|\frac{\partial^3\mathcal{L}_{t}(f_t;\beta)}{\partial f_t^3}\right|\\
&+\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\sup_{f_t:|f_t-f_t(\beta)|\leq\eta}\left|\frac{\partial^3\mathcal{L}_{n,t}(f_t;\beta)}{\partial f_t^3}-\frac{\partial^3\mathcal{L}_{t}(f_t;\beta)}{\partial f_t^3}\right|\\
&\leq\sup_{1\leq t\leq T}(\xi_{t,1})^{1/2}+O_p\left(\frac{[\log(n)]^{\delta_6}}{\sqrt{n}}\right),
\end{aligned}$$

for some constant $\delta_6>0$, by similar arguments as in Lemma C.2 in Appendix C.9.2. Thus, we have w.p.a. 1:

$$\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\tilde{J}_{3,nt}(\beta) \leq \sup_{1\leq t\leq T}(\xi_{t,1})^2+\kappa_n/2.$$

Then, from Assumption H.5 and the condition $T/n = o(1)$, we get:

$$\begin{aligned}\mathbb{P}\left[\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\tilde{J}_{3,nt}(\beta)\le\kappa_n\right] &\ge \mathbb{P}\left[\sup_{1\le t\le T}(\xi_{t,1})^2\le\kappa_n/2\right]+o(1)\ge 1-T\mathbb{P}\left[\xi_{t,1}\ge\sqrt{\kappa_n/2}\right]+o(1)\\ &\ge 1-Tb_1\exp\left(-c_1(\kappa_n/2)^{d_1/2}\right)+o(1)=1-o(1),\end{aligned}$$

since $c_1(\kappa_n/2)^{d_1/2}\ge\log(n)$ and $T/n=o(1)$. Lemma 3 (iii) follows.

Proof of Lemma 3 (iv): By similar arguments as in the proofs of Lemmas 3 (i-iii), we have w.p.a. 1:

$$\begin{aligned}\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B},\theta\in\Theta}\tilde{D}_{pq,nt}(\beta,\theta) &\le \sup_{1\le t\le T}(\xi_{t,1})^{1/2}\sup_{\beta\in\mathcal{B},\theta\in\Theta}\sup_{F_t:\|F_t-F_t(\beta)\|\le\eta^*}\left|\frac{\partial^{p+q}\log g(f_t|f_{t-1};\theta)}{\partial f_t^p\partial f_{t-1}^q}\right|+\kappa_n/2\\ &\le \sup_{1\le t\le T}(\xi_{t,1})^{1/2}\xi_{t,5}+\kappa_n/2,\end{aligned}$$

for $p+q=1$, where $F_t=(f_t',f_{t-1}')'$, $\eta^*>0$ and process $\xi_{t,5}$ is defined in Assumption H.13 (iii). Then:

$$\begin{aligned}\mathbb{P}\left[\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B},\theta\in\Theta}\tilde{D}_{pq,nt}(\beta,\theta)\le\kappa_n\right] &\ge \mathbb{P}\left[\sup_{1\le t\le T}(\xi_{t,1})^{1/2}\xi_{t,5}\le\kappa_n/2\right]\\ &\ge 1-T\mathbb{P}\left[(\xi_{t,1})^{1/2}\ge\sqrt{\kappa_n/2}\right]-T\mathbb{P}\left[\xi_{t,5}\ge\sqrt{\kappa_n/2}\right]+o(1).\end{aligned}$$

Thus, from Assumptions H.5 and H.13 (iii) we get:

$$\begin{aligned}\mathbb{P}\left[\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B},\theta\in\Theta}\tilde{D}_{pq,nt}(\beta,\theta)\le\kappa_n\right] &\ge 1-Tb_1\exp\left(-c_1(\kappa_n/2)^{d_1}\right)-Tb_5\exp\left(-c_5(\kappa_n/2)^{d_5/2}\right)\\ &\quad +o(1)=1-o(1),\end{aligned}$$

since $c_1(\kappa_n/2)^{d_1}\ge\log(n)$, $c_5(\kappa_n/2)^{d_5/2}\ge\log(n)$ and $T/n=o(1)$.

C.4 Lemma 4

LEMMA 4 *Let $\kappa_n = 2[\log(n)/C_6]^{C_7}$, for $n\in\mathbb{N}$, be the sequence in Lemma 3, where the constants $C_6, C_7>0$ are such that $C_6\le\min\{c_1,c_5\}$ and $C_7\ge\max\{5/d_1, 2/d_5\}$, for $c_1, d_1>0$ and $c_5, d_5>0$ defined in Assumptions H.5 and H.13 (iii), respectively. Then, under Assumptions A.1-A.5, H.1, H.2, H.5-H.11, H.13 (iii), and if $T^\nu/n=O(1)$,*

$\nu > 1$, w.p.a. 1 we have: (i) $\sup_{1 \le t \le T} \sup_{\beta \in \mathcal{B}} |J_{4,n,t}(\beta)| \le \kappa_n$, (ii) $\sup_{1 \le t \le T} \sup_{\beta \in \mathcal{B}} \tilde{J}_{5,n,t}(\beta) \le \kappa_n$, (iii) $\sup_{1 \le t \le T} \sup_{\beta \in \mathcal{B}, \theta \in \Theta} |D_{pq,n,t}(\beta, \theta)| \le \kappa_n$, for $p + q \le 2$ and (iv) $\sup_{1 \le t \le T} \sup_{\beta \in \mathcal{B}, \theta \in \Theta} \tilde{D}_{pq,n,t}(\beta, \theta) \le \kappa_n$, for $p + q = 3$, where $J_{4,n,t}(\beta)$ and $D_{pq,n,t}(\beta, \theta)$ are defined in Proposition 1, and $\tilde{J}_{5,n,t}(\beta)$ and $\tilde{D}_{pq,n,t}(\beta, \theta)$ are defined as in equation (a.21).

Proof of Lemma 4: The proof of Lemma 4 is similar to the proof of Lemma 3 in Section C.3.

C.5 Lemma 5

LEMMA 5 *Under Assumptions A.1-A.5, H.1, H.2, H.5-H.11, H.13 (iii), and if $T^{\nu}/n = O(1)$, $\nu > 1$, we have for any integer $j \ge 3$:*

$$\Lambda_{j,nT}(\beta, \theta) \le C_j^* \left(\frac{T^2 \kappa_n^j}{n^2} \right), \tag{c.8}$$

and:

$$\Lambda_{j,nT}(\beta, \theta) \le C_8 \kappa_n^{2j} j! \left(\frac{T}{n} + \sqrt{T} \varepsilon_n^2 \right)^j, \tag{c.9}$$

uniformly in $\beta \in \mathcal{B}, \theta \in \Theta$, w.p.a. 1, for some constants $C_j^* > 0$, $j = 3, 4, ...$, and $C_8 > 0$, where functions $\Lambda_{j,nT}(\beta, \theta)$, for $j \in \mathbb{N}$, are defined in equation (a.22), sequence $\varepsilon_n \downarrow 0$ involved in the definition of $\Lambda_{j,nT}(\beta, \theta)$ is such that $\frac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, $\mu_1 > 0$, and constants κ_n, $n \in \mathbb{N}$, are defined in Lemma 3.

Proof of Lemma 5: The bound in (c.8) is derived by similar arguments as in parts a)-c) of the proof of Proposition A.4 in Appendix A.2.1 iii). Let us derive the bound given in (c.9) for $m = 1$. Lemma 3 (ii) implies that, if $z \in \mathcal{Z}_{nT}(\beta)$, then $\|z\|^2 \le n\varepsilon_n^2 \kappa_n$, and hence $z \in [-\sqrt{n\varepsilon_n^2 \kappa_n}, \sqrt{n\varepsilon_n^2 \kappa_n}]^T$, uniformly in $\beta \in \mathcal{B}$, w.p.a. 1, where $\mathcal{Z}_{nT}(\beta)$ is defined in Proposition A.1 in Appendix A.2.1. The mass of the hypercube $[-\sqrt{n\varepsilon_n^2 \kappa_n}, \sqrt{n\varepsilon_n^2 \kappa_n}]^T$ in $\mathbb{R}^T$ under a standard multivariate Gaussian distribution is V_n^T, where $V_n \equiv \int_{-\sqrt{n\varepsilon_n^2 \kappa_n}}^{\sqrt{n\varepsilon_n^2 \kappa_n}} \phi(s) ds$ and ϕ denotes the pdf of the standard Gaussian distribution. We have $V_n^T = 1 - o(1)$ under condition $\dfrac{T}{n\varepsilon_n^2} = O(n^{-\mu_1})$, $\mu_1 > 0$. Then, we have:

$$\frac{\Lambda_{j,nT}(\beta,\theta)}{V_n^T} \leq \frac{1}{V_n^T}\frac{1}{(2\pi)^{T/2}}\int_{\left[-\sqrt{n\varepsilon_n^2\kappa_n},\sqrt{n\varepsilon_n^2\kappa_n}\right]^T} \exp\left(-\frac{1}{2}\|z\|^2\right)$$
$$\cdot\left[\sum_{t=1}^T \psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}}z_t, \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}}z_{t-1};\beta,\theta\right)\right]^j dz$$
$$= E_{nT}\left[\left(\sum_{t=1}^T \psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}}z_t, \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}}z_{t-1};\beta,\theta\right)\right)^j\right],$$

w.p.a. 1, where $E_{nT}[.]$ denotes expectation w.r.t. a random vector z in $\mathbb{R}^T$ with truncated standard Gaussian density on $[-\sqrt{n\varepsilon_n^2\kappa_n}, \sqrt{n\varepsilon_n^2\kappa_n}]^T$. Let us now use the bound for $\psi_{n,t}$ in equation (a.20). By applying the triangular inequality, we get:

$$\left[\frac{\Lambda_{j,nT}(\beta,\theta)}{V_n^T}\right]^{1/j} \leq E_{nT}\left[\left(\frac{1}{3!\sqrt{n}}\sum_{t=1}^T J_{3,nt}(\beta)z_t^3\right)^j\right]^{1/j} + E_{nT}\left[\left(\frac{1}{4!n}\sum_{t=1}^T J_{4,nt}(\beta)z_t^4\right)^j\right]^{1/j}$$
$$+E_{nT}\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^T D_{10,nt}(\beta,\theta)z_t\right)^j\right]^{1/j} + E_{nT}\left[\left(\frac{1}{\sqrt{n}}\sum_{t=1}^T D_{01,nt}(\beta,\theta)z_{t-1}\right)^j\right]^{1/j}$$
$$+E_{nT}\left[\left(\frac{1}{2n}\sum_{t=1}^T D_{20,nt}(\beta,\theta)z_t^2\right)^j\right]^{1/j} + E_{nT}\left[\left(\frac{1}{2n}\sum_{t=1}^T D_{02,nt}(\beta,\theta)z_{t-1}^2\right)^j\right]^{1/j}$$
$$+E_{nT}\left[\left(\frac{1}{n}\sum_{t=1}^T D_{11,nt}(\beta,\theta)z_t z_{t-1}\right)^j\right]^{1/j} + E_{nT}\left[\left(\sum_{t=1}^T R_{n,t}(z_t,z_{t-1};\beta,\theta)\right)^j\right]^{1/j}$$
$$\equiv \sum_{k=1}^8 A_{k,j,nT}(\beta,\theta)^{1/j}, \tag{c.10}$$

w.p.a. 1. Let us now show that:

$$A_{k,j,nT}(\beta,\theta) \leq \tilde{C}_8 j!\kappa_n^{2j}\left(\frac{T}{n} + \sqrt{T}\varepsilon_n^2\right)^j, \tag{c.11}$$

uniformly in $\beta \in \mathcal{B}$, $\theta \in \Theta$, for all $j = 3, 4, ...$ and $k = 1, ..., 8$, and for some constant $\tilde{C}_8 > 0$. Then, by using that $V_n^T = 1 - o(1)$, inequality (c.9) follows.

We prove the upper bound for term $A_{k,j,nT}(\beta,\theta)$ with $k=2$ and j even; the proof for the other indices k, and for j odd, is similar. We have from Lemma 4 (i):

$$E_{nT}\left[\left(\frac{1}{4!n}\sum_{t=1}^{T}J_{4,nt}(\beta)z_t^4\right)^j\right] \leq \frac{\kappa_n^j}{(4!n)^j}\sum_{l=1}^{\min\{j,T\}}\sum_{t_1,...,t_l}\sum_{m_1+...+m_l=j}E_n\left[z_{t_1}^{4m_1}\right]\cdots E_n\left[z_{t_l}^{4m_l}\right]$$
$$\leq \frac{\kappa_n^j}{(4!n)^j}\sum_{l=1}^{\min\{j,T\}}T^l\sum_{m_1+...+m_l=j}E_n\left[z_t^{4m_1}\right]\cdots E_n\left[z_t^{4m_l}\right], \tag{c.12}$$

uniformly in $\beta\in\mathcal{B}$, where $E_n[.]$ denotes expectation w.r.t. a random variable z_t with truncated standard Gaussian density on the interval $[-\sqrt{n\varepsilon_n^2\kappa_n},\sqrt{n\varepsilon_n^2\kappa_n}]$, $\sum_{t_1,...,t_l}$ denotes summation over all l-tuples $(t_1,...,t_l)$ of different indices from $1,2,...,T$, and $\sum_{m_1+...+m_l=j}$ denotes summation over all l-tuples $(m_1,...,m_l)$ of integers from $\mathbb{N}^*$ such that $m_1+...+m_l=j$. The number of such l-tuples $(t_1,...,t_l)$ and $(m_1,...,m_l)$ is $T(T-1)\cdots(T-l+1)\leq T^l$, and $\binom{j-1}{l-1}$, respectively. Let us now show that the product $E_n\left[z_t^{4m_1}\right]\cdots E_n\left[z_t^{4m_l}\right]$, for $l\leq j$ and $m_1+\cdots+m_l=j$, satisfies the following two bounds:

$$E_n\left[z_t^{4m_1}\right]\cdots E_n\left[z_t^{4m_l}\right] \leq 2^{j+1}j!(n\varepsilon_n^2\kappa_n)^j, \tag{c.13}$$
$$E_n\left[z_t^{4m_1}\right]\cdots E_n\left[z_t^{4m_l}\right] \leq 4^l(n\varepsilon_n^2\kappa_n)^{2(j-l)}. \tag{c.14}$$

a) Proof of inequality (c.13). To prove the bound in (c.13), we distinguish two cases.

(*) The first case is when $2m_k\geq j$ for an index $k\in\{1,...,l\}$. Without loss of generality, let $k=1$ be that index. Then, we deduce that:

$$E_n\left[z_t^{4m_1}\right]\cdots E_n\left[z_t^{4m_l}\right]\leq E_n\left[z_t^{2j}\right](n\varepsilon_n^2\kappa_n)^{2m_1-j}E_n\left[z_t^4\right](n\varepsilon_n^2\kappa_n)^{2m_2-2}\cdots E_n\left[z_t^4\right](n\varepsilon_n^2\kappa_n)^{2m_l-2}$$
$$\leq V_n^{-l}2^jj!3^{l-1}(n\varepsilon_n^2\kappa_n)^{j-2(l-1)}\leq 2^jj!(n\varepsilon_n^2\kappa_n)^j,$$

for large n, since $E_n[z_t^4]\leq 3V_n^{-1}$, $E_n[z_t^{2j}]\leq 2^jj!V_n^{-1}$, and $V_n=1-o(1)$.

(**) The second case is when $2m_k<j$ for all $k=1,...,l$. Let $a_1,...,a_l\geq 1$ be such that $a_k\leq 2m_k$, for all $k=1,...,l$, and $a_1+a_2+...+a_l=j$. Then, by the Holder inequality:

$$E_n\left[z_t^{4m_1}\right]\cdots E_n\left[z_t^{4m_l}\right]\leq E_n\left[z_t^{2a_1}\right](n\varepsilon_n^2\kappa_n)^{2m_1-a_1}\cdots E_n\left[z_t^{2a_l}\right](n\varepsilon_n^2\kappa_n)^{2m_l-a_l}$$
$$\leq E_n\left[z_t^{2j}\right]^{a_1/j}\cdots E_n\left[z_t^{2j}\right]^{a_l/j}(n\varepsilon_n^2\kappa_n)^j=E_n\left[z_t^{2j}\right](n\varepsilon_n^2\kappa_n)^j\leq V_n^{-1}2^jj!(n\varepsilon_n^2\kappa_n)^j,$$

which yields inequality (c.13).

b) Proof of inequality (c.14). The upper bound in (c.14) follows from:

$$\begin{aligned} E_n\left[z_t^{4m_1}\right]\cdots E_n\left[z_t^{4m_l}\right] &\leq E_n\left[z_t^4\right](n\varepsilon_n^2\kappa_n)^{2(m_1-1)}\cdots E_n\left[z_t^4\right](n\varepsilon_n^2\kappa_n)^{2(m_l-1)} \leq V_n^{-l}3^l(n\varepsilon_n^2\kappa_n)^{2(j-l)} \\ &\leq 4^l(n\varepsilon_n^2\kappa_n)^{2(j-l)}. \end{aligned}$$

Now, let us upper bound the RHS of inequality (c.12) by using the bound in (c.13) for the terms with $l \leq j/2$, and the bound in (c.14) for the terms with $l > j/2$. We get:

$$\begin{aligned} & E_{nT}\left[\left(\frac{1}{4!n}\sum_{t=1}^T J_{4,nt}(\beta)z_t^4\right)^j\right] \leq \frac{\kappa_n^j}{(4!n)^j}\sum_{l=1}^{\min\{j/2,T\}} T^l\binom{j-1}{l-1}2^{j+1}j!(n\varepsilon_n^2\kappa_n)^j \\ & +\frac{\kappa_n^j}{(4!n)^j}\sum_{l=j/2}^{j} T^l\binom{j-1}{l-1}4^l(n\varepsilon_n^2\kappa_n)^{2(j-l)} \\ & \leq 2\left(\frac{\sqrt{T}\varepsilon_n^2\kappa_n^2}{12}\right)^j j!\sum_{l=1}^{j/2}\binom{j-1}{l-1} + \frac{\kappa_n^j}{(4!n)^j}\sum_{l=0}^{j/2}(4T)^{j-l}\binom{j-1}{j-l-1}(n\varepsilon_n^2\kappa_n)^{2l} \\ & \leq 2\left(\frac{\sqrt{T}\varepsilon_n^2\kappa_n^2}{6}\right)^j j! + \left(\frac{\kappa_n^2}{6}\right)^j j!\sum_{l=0}^{j/2}\left(\frac{T}{n}\right)^{j-2l}\left(\sqrt{T}\varepsilon_n^2\right)^{2l} \leq 3j!\left(\frac{\kappa_n^2}{6}\right)^j\left(\frac{T}{n}+\sqrt{T}\varepsilon_n^2\right)^j. \end{aligned}$$

Then, the bound in (c.11) for $k = 2$ follows.

C.6 Lemma 6

LEMMA 6 *Under Assumptions A.1-A.5 and H.1-H.10, H.13 (iii), H.14, and if* $n, T \to \infty$ *such that* $T^\nu/n = O(1)$, $\nu > 1$, *we have:*

(1) (i) $\sup_{\beta\in\mathcal{B}}\left\|\frac{\partial^2\mathcal{L}_{nT}^*(\beta)}{\partial\beta\partial\beta'} - \frac{\partial^2\mathcal{L}^*(\beta)}{\partial\beta\partial\beta'}\right\| = o_p(1)$, *(ii)* $\sup_{\beta\in\mathcal{B},\ \theta\in\Theta}\left\|\frac{\partial^2\mathcal{L}_{1,nT}(\beta,\theta)}{\partial\theta\partial\theta'} - \frac{\partial^2\mathcal{L}_1(\beta,\theta)}{\partial\theta\partial\theta'}\right\| = o_p(1)$, *where functions* $\mathcal{L}_{nT}^*(\beta)$, $\mathcal{L}^*(\beta)$, $\mathcal{L}_{1,nT}(\beta,\theta)$ *and* $\mathcal{L}_1(\beta,\theta)$ *are as in Lemma 1;*

(2) (i) $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\mathcal{L}_{1,nT}(\beta,\theta)}{\partial\beta}\right\| = O_p(1)$, *(ii)* $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\mathcal{L}_{2,nT}(\beta,\theta)}{\partial(\beta',\theta')'}\right\| = O_p(1)$, *(iii)* $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial^2\mathcal{L}_{1,nT}(\beta,\theta)}{\partial\beta\partial\beta'}\right\| = O_p(1)$, *(iv)* $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial^2\mathcal{L}_{1,nT}(\beta,\theta)}{\partial\beta\partial\theta'}\right\| = O_p(1)$, *(v)* $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial^2\mathcal{L}_{2,nT}(\beta,\theta)}{\partial(\beta',\theta')'\partial(\beta',\theta')}\right\| = O_p(1)$, *where function* $\mathcal{L}_{2,nT}(\beta,\theta)$ *is defined in equation (3.10);*

(3) (i) $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\Psi_{nT}(\beta,\theta)}{\partial\beta}\right\| = o_p(1/n)$, *(ii)* $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\Psi_{nT}(\beta,\theta)}{\partial\theta}\right\| = O_p\left(\frac{[\log(n)]^{C_9}}{n^{3/2}}\right)$, *for a constant* $C_9 > 0$, *where* $\Psi_{nT}(\beta,\theta)$ *is the remainder term in the log-likelihood expansion (3.6).*

Moreover, if $n, T \to \infty$ *such that* $T^\nu/n = O(1)$, $\nu > 3/2$, *we have:*

(4) $\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial\tilde{\Psi}_{nT}(\beta,\theta)}{\partial(\beta',\theta')'}\right\| = o_p(1/n^2)$, *where* $\tilde{\Psi}_{nT}(\beta,\theta)$ *is the remainder term in the log-likelihood expansion (3.9).*

Proof of Lemma 6 (1i): From the definition of function $\mathcal{L}^*_{nT}(\beta)$ given in equation (3.7), we get by differentiation:

$$
\begin{aligned}
\frac{\partial\mathcal{L}^*_{nT}(\beta)}{\partial\beta} &= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial\log h}{\partial\beta}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right) \\
&\quad+\frac{1}{nT}\sum_{t=1}^{T}\frac{\partial\hat{f}_{n,t}(\beta)'}{\partial\beta}\underbrace{\sum_{i=1}^{n}\frac{\partial\log h}{\partial f_t}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right)}_{=0} \\
&= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial\log h}{\partial\beta}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right),
\end{aligned}
$$

and:

$$
\begin{aligned}
\frac{\partial^2\mathcal{L}^*_{nT}(\beta)}{\partial\beta\partial\beta'} &= \frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial^2\log h}{\partial\beta\partial\beta'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right) \\
&\quad+\frac{1}{nT}\sum_{t=1}^{T}\sum_{i=1}^{n}\frac{\partial^2\log h}{\partial\beta\partial f_t'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right)\frac{\partial\hat{f}_{n,t}(\beta)}{\partial\beta'}.
\end{aligned}
$$

By differentiating the f.o.c. $\sum_{i=1}^{n}\frac{\partial\log h}{\partial f_t}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right) = 0$ w.r.t. β, we get:

$$
\sum_{i=1}^{n}\frac{\partial^2\log h}{\partial f_t\partial\beta'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right)+\sum_{i=1}^{n}\frac{\partial^2\log h}{\partial f_t\partial f_t'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{n,t}(\beta);\beta\right)\frac{\partial\hat{f}_{n,t}(\beta)}{\partial\beta'} = 0.
$$

Let us introduce the notation:

$$\hat{I}_{t,\beta\beta}(\beta) \equiv -\frac{1}{n}\sum_{i=1}^{n}\frac{\partial^2 \log h}{\partial\beta\partial\beta'}\left(y_{i,t}|y_{i,t-1},\hat{f}_{nt}\left(\beta\right);\beta\right),$$

and similarly $\hat{I}_{t,\beta f}(\beta)$, $\hat{I}_{t,ff}(\beta)$. Then, we get:

$$\frac{\partial\hat{f}_{n,t}\left(\beta\right)}{\partial\beta'} = -\hat{I}_{t,ff}(\beta)^{-1}\hat{I}_{t,f\beta}(\beta), \tag{c.15}$$

and

$$-\frac{\partial^2\mathcal{L}^*_{nT}\left(\beta\right)}{\partial\beta\partial\beta'} = \frac{1}{T}\sum_{t=1}^{T}\left[\hat{I}_{t,\beta\beta}(\beta) - \hat{I}_{t,\beta f}(\beta)\hat{I}_{t,ff}(\beta)^{-1}\hat{I}_{t,f\beta}(\beta)\right].$$

Then, Lemma 6 (1i) follows by applying Limit Theorem 3 in Appendix B.3 with $a(y_{i,t}, y_{i,t-1}, f_t, \beta) = -\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t; \beta)}{\partial(\beta', f_t')'\partial(\beta', f_t')}$ and function $\varphi(x) = (x^{11})^{-1}$, where x is a symmetric positive definite matrix in $\mathbb{R}^{q+m,q+m}$ and x^{11} denotes the upper-left (q, q) block of the inverse x^{-1}. Indeed, Regularity Condition RC.3 (1) in Appendix B.3 is satisfied by Assumptions H.3, H.4 (iii), H.5 in Appendix A.1. Moreover, we prove in Lemma C.4 in Appendix C.9.4 that Regularity Condition RC.3 (2) in Appendix B.3 is satisfied.

Proof of Lemma 6 (1ii): From the definition of function $\mathcal{L}_{1,nT}(\beta, \theta)$ given in equation (3.8) we have:

$$\frac{\partial^2\mathcal{L}_{1,nT}\left(\beta,\theta\right)}{\partial\theta\partial\theta'} = \frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\log g}{\partial\theta\partial\theta'}\left(\hat{f}_{n,t}\left(\beta\right)|\hat{f}_{n,t-1}\left(\beta\right);\theta\right).$$

Then, Lemma 6 (1ii) follows by applying Limit Theorem 2 in Appendix B.2 with function $G(F_t; \theta) = \frac{\partial^2 \log g(f_t|f_{t-1}; \theta)}{\partial\theta\partial\theta'}$. Regularity Condition RC.1 in Appendix B.2 is implied by Assumptions H.5 and H.14 in Appendix A.1.

Proof of Lemma 6 (3ii): From the proof of the CSA expansion of the log-likelihood function [see Appendix A.2.1 ii)], we have $\Psi_{nT}(\beta, \theta) = \frac{1}{nT}\log[\Lambda_{nT}(\beta, \theta) + \Delta_{nT}(\beta, \theta)] \simeq \frac{1}{nT}\log[\Lambda_{nT}(\beta, \theta)]$. We get:

$$\frac{\partial\Psi_{nT}(\beta,\theta)}{\partial\theta} \simeq \frac{1}{nT}\frac{1}{\Lambda_{nT}(\beta,\theta)}\frac{\partial\Lambda_{nT}(\beta,\theta)}{\partial\theta}. \tag{c.16}$$

From the definition of $\Lambda_{nT}(\beta,\theta)$ in equation (a.2) we have (for $m=1$):

$$
\begin{aligned}
\frac{\partial \Lambda_{nT}(\beta,\theta)}{\partial\theta} &= \frac{1}{(2\pi)^{T/2}} \int_{\mathcal{Z}_{nT}(\beta)} \exp\left(-\frac{1}{2}\sum_{t=1}^{T} z_t^2\right) \\
&\cdot \exp\left[\sum_{t=1}^{T} \psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}} z_t, \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}} z_{t-1}; \beta,\theta\right)\right] \\
&\cdot \left(\sum_{t=1}^{T}\left[\frac{\partial \log g}{\partial\theta}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}} z_t | \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}} z_{t-1};\theta\right)\right.\right. \\
&\qquad \left.\left. - \frac{\partial \log g}{\partial\theta}\left(\hat{f}_{n,t}(\beta)|\hat{f}_{n,t-1}(\beta);\theta\right)\right]\right) dz.
\end{aligned}
$$

Thus, from (c.16) we get:

$$
\begin{aligned}
&\frac{\partial \Psi_{nT}(\beta,\theta)}{\partial\theta} \\
&\simeq \frac{1}{nT}\sum_{t=1}^{T} E_{nT,\beta,\theta}\left[\frac{\partial \log g}{\partial\theta}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}} z_t | \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}} z_{t-1};\theta\right)\right. \\
&\qquad \left. - \frac{\partial \log g}{\partial\theta}\left(\hat{f}_{n,t}(\beta)|\hat{f}_{n,t-1}(\beta);\theta\right)\right],
\end{aligned}
$$

where $E_{nT,\beta,\theta}[\cdot]$ denotes the expectation w.r.t. the random vector z in $\mathbb{R}^T$ with density proportional to

$$
\exp\left[-\frac{1}{2}\sum_{t=1}^{T} z_t^2 + \sum_{t=1}^{T} \psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}} z_t, \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}} z_{t-1}; \beta,\theta\right)\right]
$$

on the support $\mathcal{Z}_{nT}(\beta)$. By the mean value Theorem, we get:

$$
\sup_{\beta\in\mathcal{B},\theta\in\Theta}\left\|\frac{\partial \Psi_{nT}(\beta,\theta)}{\partial\theta}\right\| \lesssim \frac{1}{n^{3/2}}\left(\inf_{1\le t\le T}\inf_{\beta\in\mathcal{B}} I_{n,t}(\beta)\right)^{-1/2} C_{nT} \sup_{1\le t\le T}\sup_{\beta\in\mathcal{B},\theta\in\Theta} E_{nT,\beta,\theta}[|z_t|], \quad \text{(c.17)}
$$

where:

$$
\begin{aligned}
C_{nT} \equiv \sup_{\beta\in\mathcal{B},\theta\in\Theta}\sup_{1\le t\le T}\sup_{f_t,f_{t-1}} &\left\{\left\|\frac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_t}\right\| + \right. \\
&\left.\left\|\frac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_{t-1}}\right\| : \|f_t - \hat{f}_{n,t}(\beta)\| + \|f_{t-1} - \hat{f}_{n,t-1}(\beta)\| \le \varepsilon_n\right\},
\end{aligned}
$$

and sequence $\varepsilon_n \downarrow 0$ is involved in the definition of set $\mathcal{Z}_{nT}(\beta)$ (see Appendix A.2.1). From Lemma 3 (i) we have $\left(\inf_{1\leq t\leq T}\inf_{\beta\in\mathcal{B}} I_{n,t}(\beta)\right)^{-1/2} = O_p([\log(n)]^{C_7/2})$, for a constant $C_7 > 0$. Then, Lemma 6 (3ii) follows from inequality (c.17) and the next statements:
(a) $C_{nT} = O_p([\log(n)]^{\delta_7})$, for a constant $\delta_7 > 0$, and
(b) $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B},\theta\in\Theta} E_{nT,\beta,\theta}[|z_t|] = O_p([\log(n)]^{\delta_8})$, for a constant $\delta_8 > 0$.

Proof of statement (a): We use Limit Theorem 1 in Appendix B.1 and the convergence $\varepsilon_n = o(1)$. We have w.p.a. 1:

$$
\begin{aligned}
C_{nT} \leq \sup_{\beta\in\mathcal{B},\theta\in\Theta}\sup_{1\leq t\leq T}\sup_{f_t,f_{t-1}} \Bigg\{ &\left\| \frac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_t} \right\| + \\
&\left\| \frac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_{t-1}} \right\| : \|f_t - f_t(\beta)\| + \|f_{t-1} - f_{t-1}(\beta)\| \leq \eta^* \Bigg\} \\
&\leq \xi^*_{t,5} + \xi^{**}_{t,5},
\end{aligned}
$$

where $\eta^* > 0$ is defined in Assumption H.13 (iii), and processes $\xi^*_{t,5}$ and $\xi^{**}_{t,5}$ are defined as process $\xi_{t,5}$ in Assumption H.13 (iii) with $G(F_t;\theta) = \dfrac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_t}$, and $G(F_t;\theta) = \dfrac{\partial^2 \log g(f_t|f_{t-1};\theta)}{\partial\theta\partial f_{t-1}}$, respectively. Then, statement (a) follows from Assumptions H.13 (iii) and H.14.

Proof of statement (b): We use inequality (a.14) in Appendix A.2, $|z_t| \leq \sqrt{n}\varepsilon_n\kappa_n^{1/2}$ for $z \in \mathcal{Z}_{nT}(\beta)$, and Lemma 3 to get:

$$
\begin{aligned}
&\left|\psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{1}{\sqrt{n}}\left[I_{n,t}(\beta)\right]^{-1/2} z_t, \hat{f}_{n,t-1}(\beta) + \frac{1}{\sqrt{n}}\left[I_{n,t-1}(\beta)\right]^{-1/2} z_{t-1};\beta,\theta\right)\right| \\
&\quad \leq \frac{\kappa_n^{3/2}}{3!}\varepsilon_n|z_t|^2 + \frac{\kappa_n}{\sqrt{n}}|z_t| + \frac{\kappa_n}{\sqrt{n}}|z_{t-1}| \leq o(1)[z_t^2 + z_{t-1}^2],
\end{aligned}
$$

for $z \in \mathcal{Z}_{nT}(\beta)$ such that $|z_t| \geq 1$ for all $t = 1,...,T$, where term $o(1)$ tends to zero. We deduce that the distribution with density proportional to

$$
\exp\left[-\frac{1}{2}\sum_{t=1}^T z_t^2 + \sum_{t=1}^T \psi_{n,t}\left(\hat{f}_{n,t}(\beta) + \frac{[I_{n,t}(\beta)]^{-1/2}}{\sqrt{n}}z_t, \hat{f}_{n,t-1}(\beta) + \frac{[I_{n,t-1}(\beta)]^{-1/2}}{\sqrt{n}}z_{t-1};\beta,\theta\right)\right]
$$

on the support $\mathcal{Z}_{nT}(\beta)$ has Gaussian tails.

C.7 Lemma 7

LEMMA 7 *Let us define the process* $\zeta_{n,t} = \begin{bmatrix} \psi_{n,\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} \psi_{n,f}(t) \\ \dfrac{\partial \log g}{\partial \theta}(f_t | f_{t-1}; \theta_0) \end{bmatrix}$, $t \in \mathbb{N}$, *where* $\psi_{n,\beta}(t) = \dfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \dfrac{\partial \log h}{\partial \beta}(y_{i,t} | y_{i,t-1}, f_t; \beta_0)$, $\psi_{n,f}(t) = \dfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \dfrac{\partial \log h}{\partial f_t}(y_{i,t} | y_{i,t-1}, f_t; \beta_0)$, *and* $I_{ff}(t)$, $I_{\beta f}(t)$ *are the* (f, f) *and* (β, f) *blocks of the information matrix* $I(t)$ *defined in equation (4.6). Then, under Assumptions A.1-A.5 and H.3 (ii), H.4 (iii), H.5, H.13 (iii), H.15, and if* $T, n \to \infty$ *such that* $T^{\nu}/n = O(1)$, $\nu > 1$, *we have:*

(i) $\dfrac{1}{\sqrt{T}} \max_{1 \le t \le T} \|\zeta_{n,t}\| \xrightarrow{p} 0$;

(ii) $\dfrac{1}{T} \sum_{t=1}^{T} \zeta_{n,t} \zeta_{n,t}' \xrightarrow{p} E[\zeta_{n,t} \zeta_{n,t}'] = \Omega$, *where* $\Omega = \begin{pmatrix} I_0^* & 0 \\ 0 & I_{1,\theta\theta} \end{pmatrix}$ *and matrices* I_0^*, $I_{1,\theta\theta}$ *are defined in Proposition 3;*

(iii) $\dfrac{1}{T} E\left(\max_{1 \le t \le T} \|\zeta_{n,t}\|^2 \right) = O(1)$.

C.7.1 Proof of Lemma 7 (i)

Let $\varepsilon > 0$ be given. We have to prove that $\mathbb{P}\left[\max_{1 \le t \le T} \|\zeta_{n,t}\| \ge \varepsilon \sqrt{T} \right] = o(1)$. We use that $\mathbb{P}\left[\max_{1 \le t \le T} \|\zeta_{n,t}\| \ge \varepsilon \sqrt{T} \right] \le T \mathbb{P}\left[\|\zeta_{n,t}\| \ge \varepsilon \sqrt{T} \right]$, and $\|\zeta_{n,t}\| \le \|\zeta_{n,t}^*\| + \|\zeta_t^{**}\|$, where $\zeta_{n,t}^* = \psi_{n,\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} \psi_{n,f}(t)$ and $\zeta_t^{**} = \dfrac{\partial \log g}{\partial \theta}(f_t | f_{t-1}; \theta_0)$. Thus, we get:

$$\mathbb{P}\left[\max_{1 \le t \le T} \|\zeta_{n,t}\| \ge \varepsilon \sqrt{T} \right] \le T \mathbb{P}\left[\|\zeta_{n,t}^*\| \ge \frac{1}{2} \varepsilon \sqrt{T} \right] + T \mathbb{P}\left[\|\zeta_t^{**}\| \ge \frac{1}{2} \varepsilon \sqrt{T} \right]. \tag{c.18}$$

The second term in the RHS of inequality (c.18) is bounded by using Assumption H.13 (iii):

$$T \mathbb{P}\left[\|\zeta_t^{**}\| \ge \frac{1}{2} \varepsilon \sqrt{T} \right] \le T b_5 \exp\left(-c_5 (\varepsilon \sqrt{T}/2)^{d_5} \right) = o(1). \tag{c.19}$$

Let us now focus on the first term in the RHS of inequality (c.18). Let us write:

$$
\begin{aligned}
\zeta_{n,t}^* &= \frac{1}{\sqrt{n}} \sum_{i=1}^n W_{i,t} \\
&= \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{W}_{n,i,t} + \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(W_{i,t} 1\{|W_{i,t}| \geq B_n\} - E[W_{i,t} 1\{|W_{i,t}| \geq B_n\} | \underline{f_t}] \right) \\
&\equiv \tilde{\zeta}_{n,t} + R_{n,t},
\end{aligned}
$$

where:

$$
\begin{aligned}
W_{i,t} &= \frac{\partial \log h}{\partial \beta}(y_{i,t}|y_{i,t-1}, f_t; \beta_0) - I_{\beta f}(t) I_{ff}(t)^{-1} \frac{\partial \log h}{\partial f_t}(y_{i,t}|y_{i,t-1}, f_t; \beta_0), && \text{(c.20)} \\
\tilde{W}_{n,i,t} &= W_{i,t} 1\{|W_{i,t}| \leq B_n\} - E[W_{i,t} 1\{|W_{i,t}| \leq B_n\} | \underline{f_t}], && \text{(c.21)}
\end{aligned}
$$

and:

$$
B_n = \frac{\sqrt{n}}{\varepsilon}. \tag{c.22}
$$

We have:

$$
\mathbb{P}\left[\|\zeta_{n,t}^*\| \geq \frac{1}{2} \varepsilon \sqrt{T} \right] \leq \mathbb{P}\left[\|\tilde{\zeta}_{n,t}\| \geq \frac{1}{4} \varepsilon \sqrt{T} \right] + \mathbb{P}\left[\|R_{n,t}\| \geq \frac{1}{4} \varepsilon \sqrt{T} \right] \equiv P_{1,nT} + P_{2,nT}. \tag{c.23}
$$

Let us now bound the two probabilities in the RHS.

a) Bound of $P_{1,nT}$. We have:

$$
P_{1,nT} = E\left[\mathbb{P}\left[\left\| \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{W}_{n,i,t} \right\| \geq \frac{1}{4} \varepsilon \sqrt{T} \middle| \underline{f_t} \right] \right]. \tag{c.24}
$$

For expository purpose, let us assume that the micro-parameter β is scalar, i.e. $q = 1$, so that the $\tilde{W}_{n,i,t}$ are scalar random variables. To bound the inner conditional probability, we use Bernstein's inequality [Bosq (1998), Theorem 1.2]. From (c.20) and (c.21), the random variables $\tilde{W}_{n,i,t}$, for $i = 1, ..., n$, are i.i.d., conditional on the factor path $\underline{f_t}$, with $E[\tilde{W}_{n,i,t}|\underline{f_t}] = 0$ and $V[\tilde{W}_{n,i,t}|\underline{f_t}] \leq E[W_{i,t}^2|\underline{f_t}] = I_{\beta\beta}(t) - I_{\beta f}(t) I_{ff}(t)^{-1} I_{f\beta}(t) = 1/I^{\beta\beta}(t)$, where $I^{\beta\beta}(t)$ denotes the upper-left element of the inverse matrix $I(t)^{-1}$, and the conditional information matrix $I(t)$ is defined in equation (4.6). Moreover, $|\tilde{W}_{n,i,t}| \leq 2B_n$. Then, by the Bernstein's inequality [Bosq (1998), Theorem 1.2] and (c.22), we get:

$$
\begin{aligned}
\mathbb{P}\left[\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tilde{W}_{n,i,t}\right\| \geq \frac{1}{4}\varepsilon\sqrt{T}\,\middle|\,\underline{f_t}\right] &\leq 2\exp\left(-\frac{nT\varepsilon^2/16}{4n/I^{\beta\beta}(t)+B_n\sqrt{nT}\varepsilon}\right) \\
&\leq 2\exp\left(-\frac{1}{64}\sqrt{T}\varepsilon^2(1/I^{\beta\beta}(t)+1)^{-1}\right). \qquad \text{(c.25)}
\end{aligned}
$$

From (c.24), we get:

$$
P_{1,nT} \leq 2E\left[\exp\left(-\frac{1}{64}\sqrt{T}\varepsilon^2(1/I^{\beta\beta}(t)+1)^{-1}\right)\right]. \qquad \text{(c.26)}
$$

To bound the expectation in the RHS, we use Lemma B.2 in Section B.4.2 applied to the stationary distribution of process $1/I^{\beta\beta}(t)+1$. We use:

$$
1/I^{\beta\beta}(t) \leq \left(eig_{min}([I(t)]^{-1})\right)^{-1} = eig_{max}(I(t)) \leq \tilde{c}(\xi_{t,1}^{**})^{1/2},
$$

for a constant $\tilde{c} > 0$, where $eig_{min}(A)$ and $eig_{max}(A)$ denote the smallest and the largest eigenvalue of the symmetric matrix A, and process $\xi_{t,1}^{**}$ is defined in Assumption H.5. Then, the condition of Lemma B.2 is satisfied with $\varrho = 2d_1$, where constant $d_1 > 0$ is defined in Assumption H.5. From Lemma B.2 we get:

$$
E\left[\exp\left(-\frac{1}{64}\sqrt{T}\varepsilon^2(1/I^{\beta\beta}(t)+1)^{-1}\right)\right] \leq \tilde{C}_1\exp\left(-\tilde{C}_2(\frac{1}{64}\sqrt{T}\varepsilon^2)^{2d_1/(2d_1+1)}\right), \qquad \text{(c.27)}
$$

for some constants $\tilde{C}_1, \tilde{C}_2 > 0$. It follows:

$$
TP_{1,nT} \leq 2T\tilde{C}_1\exp\left(-\tilde{C}_2(\frac{1}{64}\sqrt{T}\varepsilon^2)^{2d_1/(2d_1+1)}\right) = o(1). \qquad \text{(c.28)}
$$

b) Bound of $P_{2,nT}$. From the expression of $P_{2,nT}$ in (c.23), and by using the Markov inequality and equation (c.22), we have:

$$
\begin{aligned}
P_{2,nT} &\leq \frac{4}{\varepsilon\sqrt{T}}E[\|R_{n,t}\|] \leq \frac{8\sqrt{n}}{\varepsilon\sqrt{T}}E[|W_{i,t}|1\{|W_{i,t}| \geq B_n\}] \\
&\leq \frac{8\sqrt{n}}{\varepsilon\sqrt{T}}B_n^{-3}E[|W_{i,t}|^4] = \frac{8\varepsilon^2}{\sqrt{T}n}E[|W_{i,t}|^4].
\end{aligned}
$$

By using $E[|W_{i,t}|^4] < \infty$ from Assumptions H.3 (ii) and H.5, and the condition $T^\nu/n = O(1)$,

$\nu > 1$, we get:

$$TP_{2,nT} = O(\sqrt{T}/n) = o(1). \tag{c.29}$$

From bounds (c.23), (c.28) and (c.29), we get:

$$T\mathbb{P}\left[\|\zeta^*_{n,t}\| \geq \frac{1}{2}\varepsilon\sqrt{T}\right] = o(1). \tag{c.30}$$

Then, from bounds (c.18), (c.19) and (c.30), Lemma 7 (i) follows.

C.7.2 Proof of Lemma 7 (ii)

Let us write:

$$\begin{aligned} \frac{1}{T}\sum_{t=1}^{T}\zeta_{n,t}\zeta'_{n,t} &= E[\zeta_{n,t}\zeta'_{n,t}] + \frac{1}{T}\sum_{t=1}^{T}(E[\zeta_{n,t}\zeta'_{n,t}|\underline{f_t}] - E[\zeta_{n,t}\zeta'_{n,t}]) \\ &\quad + \frac{1}{T}\sum_{t=1}^{T}(\zeta_{n,t}\zeta'_{n,t} - E[\zeta_{n,t}\zeta'_{n,t}|\underline{f_t}]). \end{aligned}$$

We first prove that $E[\zeta_{n,t}\zeta'_{n,t}] = \Omega$, and then show that the other two terms in the RHS are asymptotically negligible.

a) Proof that $E[\zeta_{n,t}\zeta'_{n,t}] = \Omega$

We have:

$$E[\zeta_{n,t}\zeta'_{n,t}|\underline{f_t}] = \begin{pmatrix} I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t) & 0 \\ 0 & \dfrac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}\dfrac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta'} \end{pmatrix}. \tag{c.31}$$

By taking expectation on both sides of the equation, and using the information matrix equality in the lower-right block, we get:

$$E[\zeta_{n,t}\zeta'_{n,t}] = \begin{pmatrix} E[I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t)] & 0 \\ 0 & E\left[-\dfrac{\partial^2 \log g(f_t|f_{t-1};\theta_0)}{\partial\theta\partial\theta'}\right] \end{pmatrix} = \Omega. \tag{c.32}$$

b) Proof that $T^{-1} \sum_{t=1}^{T} (E[\zeta_{n,t}\zeta'_{n,t}|\underline{f_t}] - E[\zeta_{n,t}\zeta'_{n,t}]) = o_p(1)$

From equations (c.31) and (c.32), and Assumptions H.4 (iii) and H.13, process $Z_t \equiv E[\zeta_{n,t}\zeta'_{n,t}|\underline{f_t}] - E[\zeta_{n,t}\zeta'_{n,t}]$ is independent of n and is a measurable transformation of the factor path $\underline{f_t}$. Moreover, process (f_t) is strictly stationary and ergodic by Assumption A.3 and Proposition 3.44 in White (2001). Since strict stationarity and ergodicity are maintained under measurable transformations possibly involving an infinite number of process coordinates [Breiman (1992), Proposition 6.31], it follows that process (Z_t) is strictly stationary and ergodic. Then, the ergodic theorem [Breiman (1992), Corollary 6.23] implies that $\frac{1}{T}\sum_{t=1}^{T} Z_t$ converges to $E[Z_t] = 0$ almost surely, and thus in probability.

c) Proof that $T^{-1} \sum_{t=1}^{T} (\zeta_{n,t}\zeta'_{n,t} - E[\zeta_{n,t}\zeta'_{n,t}|\underline{f_t}]) = o_p(1)$

Let us define $Z_{n,t} = \zeta_{n,t}\zeta'_{n,t} - E[\zeta_{n,t}\zeta'_{n,t}|\underline{f_t}]$. We prove that $\frac{1}{T}\sum_{t=1}^{T} Z_{n,t} = o_p(1)$ by using the WLLN for mixingale arrays in Theorem 2 in Andrews (1988). Let us check the conditions of this theorem. [3]

**) Mixingale property.* First, we prove that $\{Z_{n,t}, \mathcal{G}_{n,t}\}$ is a L^1-mixingale array, where $\mathcal{G}_{n,t} = (\underline{y_{i,t}}, i = 1, ..., n, \underline{f_{t+1}})$, namely:

$$\|E[Z_{n,t}|\mathcal{G}_{n,t-s}]\|_1 \leq b_s, \tag{c.33}$$

for all $n \in \mathbb{N}$ and a positive sequence b_s such that $b_s = o(1)$ as $s \to \infty$, where $\|.\|_1$ denotes the L^1-norm. We have:

$$\begin{aligned} \|E[Z_{n,t}|\mathcal{G}_{n,t-s}]\|_1 &= E\left[\|E[Z_{n,t}|\mathcal{G}_{n,t-s}]\|\right] = E\left[\|E[E[Z_{n,t}|\mathcal{G}_{n,t-s}, \underline{f_t}]|\mathcal{G}_{n,t-s}]\|\right] \\ &\leq E\left[E[\|E[Z_{n,t}|\mathcal{G}_{n,t-s}, \underline{f_t}]\||\mathcal{G}_{n,t-s}]\right] \\ &= E\left[E[\|E[Z_{n,t}|\mathcal{G}_{n,t-s}, \underline{f_t}]\||\underline{f_t}]\right], \end{aligned}$$

by the Law of Iterated Expectation. Now, let us consider $E[Z_{n,t}|\mathcal{G}_{n,t-s}, \underline{f_t}]$. By writing $\zeta_{n,t} = \left(\frac{1}{\sqrt{n}}\sum_{i=1}^{n} W'_{i,t}, \frac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial \theta'}\right)'$, where variables $W_{i,t}$ are defined in equation (c.20),

[3] We replace $Z_{n,t}$ for $X_{n,i}$ in Theorem 2 in Andrews (1988), and T_n for k_n, where T_n denotes the time dimension T of the panel indexed by the cross-sectional dimension n in the double asymptotics. Moreover, we use the mixingale constants $c_{n,i} = 1$ in Theorem 2 in Andrews (1988).

and using the conditional independence and the Markov property of the individual histories given the factor path $\underline{f_t}$ (Assumption A.1), we have:

$$E[Z_{n,t}|\mathcal{G}_{n,t-s},\underline{f_t}] = \begin{pmatrix} E[W_{i,t}W_{i,t}'|y_{i,t-s},\underline{f_t}] - E[W_{i,t}W_{i,t}'|\underline{f_t}] & 0 \\ 0 & 0 \end{pmatrix},$$

where $E[W_{i,t}W_{i,t}'|\underline{f_t}] = I_{\beta\beta}(t) - I_{\beta f}(t)I_{ff}(t)^{-1}I_{f\beta}(t)$. Thus, we get:

$$\|E[Z_{n,t}|\mathcal{G}_{n,t-s}]\|_1 \leq E\left[E[\|E[W_{i,t}W_{i,t}'|y_{i,t-s},\underline{f_t}] - E[W_{i,t}W_{i,t}'|\underline{f_t}]\||\underline{f_t}]\right]. \quad \text{(c.34)}$$

The conditional expectation $E[\|E[W_{i,t}W_{i,t}'|y_{i,t-s},\underline{f_t}] - E[W_{i,t}W_{i,t}'|\underline{f_t}]\||\underline{f_t}]$ can be bounded by using that the individual histories are conditionally beta-mixing given the factor path (Assumption A.4). Indeed, by applying the Ibragimov inequality [see e.g. Davidson (1994), Theorem 14.2] conditionally on $\underline{f_t}$, and the fact that an alpha-mixing coefficient is upper bounded by the corresponding beta-mixing coefficient [see e.g. Davidson (1994), inequality (13.48)], we have $\mathbb{P}$-a.s.:

$$E[\|E[W_{i,t}W_{i,t}'|y_{i,t-s},\underline{f_t}] - E[W_{i,t}W_{i,t}'|\underline{f_t}]\||\underline{f_t}] \leq 6\beta_t(s)^{1/2}E[\|W_{i,t}\|^4|\underline{f_t}]^{1/2}, \quad \text{(c.35)}$$

where $\beta_t(s)$ denotes the conditional beta-mixing coefficient for lag s of the individual process $(y_{i,t})$ given $\underline{f_t}$. From inequalities (c.34) and (c.35), and the Cauchy-Schwarz inequality, we get:

$$\|E[Z_{n,t}|\mathcal{G}_{n,t-s}]\|_1 \leq 6E[\beta_t(s)]^{1/2}E[\|W_{i,t}\|^4]^{1/2},$$

where $E[\|W_{i,t}\|^4] < \infty$ from Assumptions H.3 (ii) and H.5. Hence, we get inequality (c.33) with sequence $b_s = 6E[\beta_t(s)]^{1/2}E[\|W_{i,t}\|^4]^{1/2}$. Since $0 \leq \beta_t(s) \leq 1$, for any t, s and $\mathbb{P}$-a.s., we can apply the Lebesgue Theorem. From Assumption A.4, we get $E[\beta_t(s)] = o(1)$, as $s \to \infty$. Hence, $b_s = o(1)$, as $s \to \infty$.

***) Uniform integrability.* Let us now prove that array $Z_{n,t}$ is uniformly integrable, namely $\lim_{M\to\infty} \sup_{n\in\mathbb{N}} E\left[\|Z_{n,t}\|1\{\|Z_{n,t}\| \geq M\}\right] = 0$. [4] By Theorem 12.11 in Davidson (1994), uniform integrability is implied by uniform L^p-boundedness, namely $\sup_{n\in\mathbb{N}} E\left[\|Z_{n,t}\|^p\right] < \infty$, for a $p > 1$. Let us prove uniform L^2-boundedness of array $Z_{n,t}$. By using $\|Z_{n,t}\| \leq \|\zeta_{n,t}\|^2 + E\left[\|\zeta_{n,t}\|^2|\underline{f_t}\right]$, and the Cauchy-Schwarz and triangular inequalities, we have $E\left[\|Z_{n,t}\|^2\right]^{1/2} \leq$

[4] By strict stationarity, the sup over t is unnecessary.

$2E\left[\|\zeta_{n,t}\|^4\right]^{1/2}$. Moreover, by using $\|\zeta_{n,t}\|^2 = \left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{i,t}\right\|^2 + \left\|\frac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}\right\|^2$ and the triangular inequality, $E\left[\|\zeta_{n,t}\|^4\right]^{1/2} \leq E\left[\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{i,t}\right\|^4\right]^{1/2} + E\left[\left\|\frac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}\right\|^4\right]^{1/2}$. Expectation $E\left[\left\|\frac{\partial \log g(f_t|f_{t-1};\theta_0)}{\partial\theta}\right\|^4\right]$ is finite by Assumption H.15. Hence, uniform L^2-boundedness of array $Z_{n,t}$ follows, if we show that:

$$\sup_{n\in\mathbb{N}} E\left[\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{i,t}\right\|^4\right] < \infty. \tag{c.36}$$

For expository purpose, let us assume a scalar micro-parameter, i.e. $q = 1$, so that the $W_{i,t}$ are scalar random variables. By using the i.i.d. property of the individual histories given the factor path (Assumption A.1), and $E[W_{i,t}|\underline{f_t}] = 0$, we have:

$$\begin{aligned} E\left[\left|\frac{1}{\sqrt{n}}\sum_{i=1}^n W_{i,t}\right|^4 |\underline{f_t}\right] &= \frac{1}{n^2}\sum_{i=1}^n E\left[W_{i,t}^4|\underline{f_t}\right] + \frac{1}{n^2}\sum_{i,j=1,i\neq j}^n E\left[W_{i,t}^2|\underline{f_t}\right] E\left[W_{j,t}^2|\underline{f_t}\right] \\ &= \frac{1}{n}E\left[W_{i,t}^4|\underline{f_t}\right] + \frac{n-1}{n}E\left[W_{i,t}^2|\underline{f_t}\right]^2 \leq E\left[W_{i,t}^4|\underline{f_t}\right]. \end{aligned}$$

By taking expectation on both sides, and using that $E\left[W_{i,t}^4\right] < \infty$ from Assumptions H.3 (ii) and H.5, bound (c.36) follows.

By Theorem 2 in Andrews (1988), it follows that $\frac{1}{T}\sum_{t=1}^T Z_{n,t} = o_p(1)$.

C.7.3 Proof of Lemma 7 (iii)

We have:

$$\frac{1}{T}E\left(\max_{1\leq t\leq T}\|\zeta_{n,t}\|^2\right) \leq \frac{1}{T}E\left[\sum_{t=1}^T \|\zeta_{n,t}\|^2\right] = TrE[\zeta_{n,t}\zeta_{n,t}'] = Tr(\Omega),$$

for all $T \in \mathbb{N}$, from equation (c.32).

C.8 Lemma 8

LEMMA 8 *Under Assumptions A.1-A.5 and H.1, H.3 (ii), H.5, H.6, H.7 (i)-(ii), H.8-H.10, we have* $\sup_{\beta \in \mathcal{B}} \left\| \frac{\partial \hat{f}_{n,t}(\beta)}{\partial \beta'} \right\| = O_p(1)$*, conditionally on* $\underline{f_t}$*, for* $\mathbb{P}$*-almost every (a.e.)* $\underline{f_t}$*.*

Proof of Lemma 8: From equation (c.15) we have:

$$\frac{\partial \hat{f}_{n,t}(\beta)}{\partial \beta'} = -\hat{I}_{t,ff}(\beta)^{-1} \hat{I}_{t,f\beta}(\beta), \tag{c.37}$$

where:

$$\begin{aligned} \hat{I}_{t,ff}(\beta) &\equiv -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f \partial f'} \Big(y_{i,t} | y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta \Big), \\ \hat{I}_{t,f\beta}(\beta) &\equiv -\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log h}{\partial f \partial \beta'} \Big(y_{i,t} | y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta \Big). \end{aligned}$$

Let us write:

$$\begin{aligned} \hat{I}_{t,ff}(\beta) - I_{t,ff}(\beta) &= -\frac{1}{n} \sum_{i}^{n} \left[\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, \hat{f}_{n,t}(\beta); \beta)}{\partial f \partial f'} - \frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f \partial f'} \right] \\ &\quad - \left(\frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f \partial f'} - E\left[\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta); \beta)}{\partial f \partial f'} | \underline{f_t} \right] \right) \\ &\equiv I_{1,n,t}(\beta) + I_{2,n,t}(\beta). \end{aligned} \tag{c.38}$$

We have:

$$\sup_{\beta \in \mathcal{B}} |I_{1,n,t}(\beta)| = o_p(1), \tag{c.39}$$

conditionally on $\underline{f_t}$, for $\mathbb{P}$-a.e. $\underline{f_t}$, by using that $\sup_{\beta \in \mathcal{B}} \|\hat{f}_{n,t}(\beta) - f_t(\beta)\| = o_p(1)$, conditionally on $\underline{f_t}$, for $\mathbb{P}$-a.e. $\underline{f_t}$, and Assumption H.5. We have:

$$\sup_{\beta \in \mathcal{B}} |I_{2,n,t}(\beta)| = o_p(1), \tag{c.40}$$

conditionally on $\underline{f_t}$, for $\mathbb{P}$-a.e. $\underline{f_t}$, by applying the ULLN in Lemma 2.4 in Newey, McFadden (1994) conditionally on $\underline{f_t}$. We can apply Lemma 2.4 in Newey, McFadden (1994) since, for any date t and $\mathbb{P}$-a.e. $\underline{f_t}$, we have:

a) Function $H_t(Y_{i,t};\beta) \equiv \dfrac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta)}{\partial f \partial f'}$ is continuous w.r.t. β, for almost any $Y_{i,t} = (y_{i,t}, y_{i,t-1})' \in \mathbb{R}^2$;

b) Parameter set $\mathcal{B} \subset \mathbb{R}^q$ is compact;

c) Random vectors $Y_{i,t}$, for i varying, are i.i.d. conditionally on $\underline{f_t}$;

d) We have $E\left[\sup_{\beta\in\mathcal{B}} \|H_t(Y_{i,t};\beta)\| |\underline{f_t}\right] < \infty$.

Condition a) is implied by continuity of function $\partial^2 \log h/\partial f\partial f'$ w.r.t. (β, f), and continuity of pseudo-true factor value $f_t(\beta)$ w.r.t. β (see the proof of Limit Theorem 1 in Section B.1). Conditions b), c) and d) are implied by Assumptions H.1, A.1, and H.3 ii), respectively.

From (c.38), (c.39) and (c.40), we get $\hat{I}_{t,ff}(\beta) - I_{t,ff}(\beta) = o_p(1)$, uniformly in $\beta \in \mathcal{B}$ and conditional on $\underline{f_t}$, for $\mathbb{P}$-a.e. $\underline{f_t}$. Similarly, we can prove $\hat{I}_{t,f\beta}(\beta) - I_{t,f\beta}(\beta) = o_p(1)$, uniformly in $\beta \in \mathcal{B}$ and conditional on $\underline{f_t}$, for $\mathbb{P}$-a.e. $\underline{f_t}$. The conclusion follows.

C.9 Secondary Lemmas

C.9.1 Lemma C.1

Lemma C.1: *Under Assumption H.5 in Appendix A.1, the function φ that maps a symmetric positive definite (m,m) matrix x into $\varphi(x) = \log\det(x)$ satisfies Regularity Condition RC.3 (2) in Appendix B.3 with $\mu_t(\beta) = I_{t,ff}(\beta) = E_0\left[-\dfrac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta)}{\partial f_t \partial f_t'}|\underline{f_t}\right]$.*

Proof: Let us first prove that Regularity Condition RC.3 (2i) in Appendix B.3 is satisfied. Let $\mathcal{K}$ be a compact subset of the set $\mathcal{U}$ of positive definite (m,m) matrices. Let $A, B \in \mathcal{K}$ and define $x(\xi) = (1-\xi)A + \xi B$ and the function $f(\xi) = \log\det x(\xi)$, for $\xi \in [0,1]$. Its derivative is given by $f'(\xi) = Tr\left[x(\xi)^{-1}\dfrac{dx(\xi)}{d\xi}\right] = Tr\left[((1-\xi)A + \xi B)^{-1}(B-A)\right]$, where Tr denotes the trace of a matrix. By the mean value Theorem, we get:

$$|\log\det(B) - \log\det(A)| = |f(1) - f(0)| \leq \sup_{\xi\in[0,1]} |f'(\xi)| \leq \sup_{x\in\bar{\mathcal{K}}} \|x^{-1}\| \|B - A\|,$$

where $\bar{\mathcal{K}}$ is the convex hull of set $\mathcal{K}$ and $\sup_{x\in\bar{\mathcal{K}}} \|x^{-1}\| < \infty$ by the compactness of set $\bar{\mathcal{K}}$.

Let us now prove that Regularity Condition RC.3 (2ii) in Appendix B.3 is satisfied. For $w = (Id + \Delta)z$, with $\|\Delta\| \leq 1/2$, we have $\varphi(w) = \log\det(Id + \Delta) + \log\det(z) \leq C_1 + C_2 \log\|z\|$, where constants $C_1, C_2 > 0$ are independent of z. Thus, we can choose $\gamma_{10} = 0$ and $\psi(z) = 1 + |\log\|z\||$ in Regularity Condition RC.3 (2ii). Now, by using that for $\mu_t(\beta) = I_{t,ff}(\beta)$ we have $\tilde{c}_1(\xi^*_{t,1})^{-1} \leq \|\mu_t(\beta)\| \leq \tilde{c}_2(\xi^{**}_{t,1})^{1/2}$, for any $\beta \in \mathcal{B}$ and some constants $\tilde{c}_1, \tilde{c}_2 > 0$, where processes $\xi^*_{t,1}$ and $\xi^{**}_{t,1}$ are defined in Assumption H.5. Then, we get $E_0[\sup_{\beta\in\mathcal{B}} |\psi(\mu_t(\beta))|^4] < \infty$ from Assumption H.5.

C.9.2 Lemma C.2

Lemma C.2: *Under Assumptions A.1-A.5, H.7 (iii) and H.11, and if $T^\nu/n = O(1)$, $\nu > 1$, we have:*

$$\sup_{1\leq t\leq T} \sup_{\beta\in\mathcal{B}} \sup_{f\in\mathcal{F}_n} \left\| \frac{1}{n}\sum_{i=1}^n \left(\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial f \partial f'} - E_0\left[\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial f \partial f'} |\underline{f_t}\right]\right)\right\|$$
$$= O_p\left(\frac{[\log(n)]^{\delta_3}}{\sqrt{n}}\right),$$

for a constant $\delta_3 > 0$.

Proof: For expository purpose, we consider the case of a scalar factor, i.e. $m = 1$. Define:

$$a(Y_{i,t}, f, \beta) = \frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f;\beta)}{\partial f^2},$$

where $Y_{i,t} = (y_{i,t}, y_{i,t-1})'$, and $W_{n,t}(f,\beta) = \dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n \big(a(Y_{i,t}, f, \beta) - E[a(Y_{i,t}, f, \beta)|\underline{f_t}]\big)$. Let:

$$\delta_3 = \max\{\gamma_4, 1 + 1/d_4\}, \tag{c.41}$$

where constants $\gamma_4 > 0$ and $d_4 > 0$ are defined in Assumptions H.11 (i), (iii). We now show that the probability

$$P_{n,T} \equiv \mathbb{P}\left[\sup_{1\leq t\leq T} \sup_{\beta\in\mathcal{B}} \sup_{f\in\mathcal{F}_n} |W_{n,t}(f,\beta)| \geq C_3[\log(n)]^{\delta_3}\right]$$

can be made arbitrarily small as $n, T \to \infty$, $T^\nu/n = O(1)$, $\nu > 1$, for a suitable constant $C_3 > 0$.

We have $P_{nT} \leq T\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}|W_{n,t}(f,\beta)| \geq C_3[\log(n)]^{\delta_3}\right]$. Moreover, let us write:

$$\begin{aligned} W_{n,t}(f,\beta) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \big(a(Y_{i,t},f,\beta)1\{U_{n,it} \leq B_n\} - E[a(Y_{i,t},f,\beta)1\{U_{n,it} \leq B_n\}|\underline{f_t}]\big) \\ &\quad + \frac{1}{\sqrt{n}}\sum_{i=1}^n \big(a(Y_{i,t},f,\beta)1\{U_{n,it} \geq B_n\} - E[a(Y_{i,t},f,\beta)1\{U_{n,it} \geq B_n\}|\underline{f_t}]\big) \\ &\equiv \tilde{W}_{n,t}(f,\beta) + R_{n,t}(f,\beta), \end{aligned}$$

where:

$$U_{n,it} = \sup_{f\in\mathcal{F}_n}\sup_{\beta\in\mathcal{B}}|a(Y_{i,t},f,\beta)|, \quad B_n = \sqrt{n}. \tag{c.42}$$

Then:

$$P_{nT} \leq T\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}|\tilde{W}_{n,t}(f,\beta)| \geq \frac{1}{2}C_3[\log(n)]^{\delta_3}\right] + T\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}|R_{n,t}(f,\beta)| \geq \frac{1}{2}C_3[\log(n)]^{\delta_3}\right]. \tag{c.43}$$

Let us now bound the two terms in the RHS.

i) Bound of the second term in the RHS of (c.43)

We have $\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}|R_{n,t}(f,\beta)| \leq \frac{1}{\sqrt{n}}\sum_{i=1}^n\big(U_{n,it}1\{U_{n,it} \geq B_n\} + E[U_{n,it}1\{U_{n,it} \geq B_n\}|\underline{f_t}]\big)$. Then, by the Markov inequality we get:

$$\begin{aligned} & T\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}|R_{n,t}(f,\beta)| \geq \frac{1}{2}C_3[\log(n)]^{\delta_3}\right] \\ & \leq \frac{4T\sqrt{n}}{C_3[\log(n)]^{\delta_3}}E[U_{n,it}1\{U_{n,it} \geq B_n\}] \leq \frac{4T\sqrt{n}}{C_3[\log(n)]^{\delta_3}B_n^3}E[U_{n,it}^4] = O\left(\frac{T[\log(n)]^{\gamma_5-\delta_3}}{n}\right) = o(1), \end{aligned}$$

for some constant $\gamma_5 > 0$, by Assumptions H.11 (ii)-(iii), $B_n = \sqrt{n}$ and the condition $T^\nu/n = O(1)$, $\nu > 1$.

ii) Bound of the first term in the RHS of (c.43)

Let us introduce a covering of set $\mathcal{B}$ by means of N_n open balls $B(\beta_j,\eta_n)$, $j = 1,...,N_n$, with center $\beta_j \in \mathbb{R}^q$ and radius $\eta_n = n^{-3/2}$ depending on n. Similarly, let $B(\xi_i,\eta_n)$, $i = 1,...,M_n$, be a covering of set $\mathcal{F}_n$. Since set $\mathcal{B} \subset \mathbb{R}^q$ is independent of n, while the Lebesgue mass of set $\mathcal{F}_n \subset \mathbb{R}$ is $O([\log(n)]^{\gamma_1})$ [see Assumption H.7 (iii)], we have $N_n, M_n \to \infty$ as $n \to \infty$, such that:

$$N_n = O(\eta_n^{-q}) = O(n^{3q/2}), \quad M_n = O([\log(n)]^{\gamma_1}\eta_n^{-1}) = O([\log(n)]^{\gamma_1} n^{3/2}). \tag{c.44}$$

We have:

$$\begin{aligned}
&\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n} |\tilde{W}_{n,t}(f,\beta)| \leq \max_{i=1,\dots,M_n,\ j=1,\dots,N_n} \sup_{\beta\in B(\beta_j,\eta_n), f\in B(\xi_i,\eta_n)} |\tilde{W}_{n,t}(f,\beta)| \\
&\leq \max_{i=1,\dots,M_n,\ j=1,\dots,N_n} |\tilde{W}_{n,t}(\xi_i,\beta_j)| + \sup_{\beta,\beta':\|\beta-\beta'\|\leq\eta_n, f,f':|f-f'|\leq\eta_n} |\tilde{W}_{n,t}(f,\beta) - \tilde{W}_{n,t}(f',\beta')|.
\end{aligned}$$

Thus:

$$\begin{aligned}
&T\mathbb{P}\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n} |\tilde{W}_{n,t}(f,\beta)| \geq \frac{1}{2}C_3[\log(n)]^{\delta_3}\right] \\
\leq\ & T\mathbb{P}\left[\sup_{\beta,\beta':\|\beta-\beta'\|\leq\eta_n, f,f':|f-f'|\leq\eta_n} |\tilde{W}_{n,t}(f,\beta) - \tilde{W}_{n,t}(f',\beta')| \geq \frac{1}{4}C_3[\log(n)]^{\delta_3}\right] \\
&+TN_nM_n \sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n} \mathbb{P}\left[|\tilde{W}_{n,t}(f,\beta)| \geq \frac{1}{4}C_3[\log(n)]^{\delta_3}\right] \equiv A_{1,nT} + A_{2,nT}.
\end{aligned} \tag{c.45}$$

Let us now bound $A_{1,nT}$ and $A_{2,nT}$.

a) Bound of term $A_{1,nT}$ in (c.45)

We use that:

$$\begin{aligned}
&\sup_{\beta,\beta':\|\beta-\beta'\|\leq\eta_n, f,f':|f-f'|\leq\eta_n} |\tilde{W}_{n,t}(f,\beta) - \tilde{W}_{n,t}(f',\beta')| \\
&\leq \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}\left\|\frac{\partial a(Y_{i,t},f,\beta)}{\partial(\beta',f)'}\right\| + \sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n} E\left[\left\|\frac{\partial a(Y_{i,t},f,\beta)}{\partial(\beta',f)'}\right\| |\underline{f_t}\right]\right)2\eta_n.
\end{aligned}$$

Then, by the Markov inequality, we get:

$$A_{1,nT} \leq \frac{16T\sqrt{n}\eta_n}{C_3[\log(n)]^{\delta_3}} E\left[\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}\left\|\frac{\partial a(Y_{i,t},f,\beta)}{\partial(\beta',f)'}\right\|\right] = o(1),$$

from Assumption H.11 (iii), $\eta_n = n^{-3/2}$ and the condition $T^\nu/n = O(1)$, $\nu > 1$.

b) Bound of term $A_{2,nT}$ in (c.45)

For given $\beta \in \mathcal{B}$, $f \in \mathcal{F}_n$, let us write:

$$\begin{aligned}
\mathbb{P}\left[|\tilde{W}_{n,t}(f,\beta)| \geq \frac{1}{4}C_3[\log(n)]^{\delta_3}\right] &= E\left[\mathbb{P}\left[|\tilde{W}_{n,t}(f,\beta)| \geq \frac{1}{4}C_3[\log(n)]^{\delta_3}|\underline{f_t}\right]\right] \\
&= E\left[\mathbb{P}\left[\left|\sum_{i=1}^{n}\psi_{n,it}(f,\beta)\right| \geq \frac{1}{4}\sqrt{n}C_3[\log(n)]^{\delta_3}|\underline{f_t}\right]\right],
\end{aligned}$$

where $\psi_{n,it}(f,\beta) \equiv a(Y_{i,t},f,\beta)1\{U_{n,it} \leq B_n\} - E[a(Y_{i,t},f,\beta)1\{U_{n,it} \leq B_n\}|\underline{f_t}]$. To bound the conditional probability within the expectation, we use that the variables $\psi_{n,it}(f,\beta)$, $i = 1,...,n$, are independent and zero-mean, conditionally on the factor path $\underline{f_t}$, and we apply the Bernstein's inequality [see Bosq (1998), Theorem 1.2]. We have:

$$|\psi_{n,it}(f,\beta)| \leq 2B_n,$$

and:

$$V[\psi_{n,it}(f,\beta)|\underline{f_t}] \leq \sup_{f\in\mathcal{F}_n}\sup_{\beta\in\mathcal{B}} E[|a(Y_{i,t},f,\beta)|^2|\underline{f_t}] \leq \xi^*_{t,4}[\log(n)]^{\gamma_4},$$

where $\xi^*_{t,4} \equiv \sup_{n\geq 1}\sup_{f\in\mathcal{F}_n}\sup_{\beta\in\mathcal{B}}[\log(n)]^{-\gamma_4}E\left[\left|\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f^2}\right|^2|\underline{f_t}\right]$, and constant $\gamma_4 \geq 0$ is defined in Assumptions H.11 (i), (iii). Then, from the Bernstein's inequality:

$$\begin{aligned}
\mathbb{P}\left[\left|\sum_{i=1}^{n}\psi_{n,it}(f,\beta)\right| \geq \frac{1}{4}\sqrt{n}C_3[\log(n)]^{\delta_3}|\underline{f_t}\right] &\leq 2\exp\left(-\frac{\left(\frac{1}{4}\sqrt{n}C_3[\log(n)]^{\delta_3}\right)^2}{4n[\log(n)]^{\gamma_4}\xi^*_{t,4} + 4B_n\left(\frac{1}{4}\sqrt{n}C_3[\log(n)]^{\delta_3}\right)}\right) \\
&\leq 2\exp\left(-\frac{1}{64}C_3[\log(n)]^{\delta_3}(\xi^*_{t,4}+1)^{-1}\right),
\end{aligned}$$

as long as $C_3 \geq 1$, since $B_n = \sqrt{n}$ and $\delta_3 \geq \gamma_4$ from (c.41). Thus, we get:

$$\mathbb{P}\left[|\tilde{W}_{n,t}(f,\beta)| \geq \frac{1}{4}C_3[\log(n)]^{\delta_3}\right] \leq 2E\left[\exp\left(-\frac{1}{64}C_3[\log(n)]^{\delta_3}(\xi^*_{t,4}+1)^{-1}\right)\right].$$

To bound the expectation in the RHS we use Lemma B.2 in Appendix B.4.2 applied to the stationary distribution of process $\xi^*_{t,4}+1$. From Assumption H.11 (iii), the condition of Lemma B.2 is satisfied with $\varrho = d_4$, where constant $d_4 > 0$ is defined in Assumption H.11.

We get:

$$
\begin{aligned}
E\left[\exp\left(-\frac{1}{64}C_3[\log(n)]^{\delta_3}(\xi_{t,4}^*+1)^{-1}\right)\right] &\leq \tilde{C}_1\exp\left(-\tilde{C}_2\left[\frac{1}{64}C_3[\log(n)]^{\delta_3}\right]^{d_4/(1+d_4)}\right)\\
&\leq \tilde{C}_1 n^{-\tilde{C}_2(C_3/64)^{d_4/(1+d_4)}},
\end{aligned}
$$

for some constants $\tilde{C}_1, \tilde{C}_2 > 0$ independent of C_3, since $\delta_3 d_4/(1+d_4) \geq 1$ from (c.41). Thus:

$$
\mathbb{P}\left[|\tilde{W}_{n,t}(f,\beta)| \geq \frac{1}{4}C_3[\log(n)]^{\delta_3}\right] \leq 2\tilde{C}_1 n^{-\tilde{C}_2(C_3/64)^{d_4/(1+d_4)}}.
$$

From the expression of $A_{2,nT}$ in (c.45), and the bounds on N_n and M_n in (c.44), we get:

$$
A_{2,nT} = O\left(Tn^{3(q+1)/2}[\log(n)]^{\gamma_1}n^{-\tilde{C}_2(C_3/64)^{d_4/(1+d_4)}}\right) = O\left(\frac{T}{n}[\log(n)]^{\gamma_1}\right) = o(1),
$$

from the condition $T^\nu/n = O(1)$, $\nu > 1$, if $\tilde{C}_2(C_3/64)^{d_4/(1+d_4)} \geq 3(q+1)/2+1$, i.e., if $C_3 \geq 64\left(\frac{3(q+1)+2}{2\tilde{C}_2}\right)^{1+1/d_4}$.

C.9.3 Lemma C.3

Lemma C.3: *Under Assumptions A.1-A.5 and H.1, H.2, H.5-H.11, and if $T^\nu/n = O(1)$, $\nu > 1$:*

(i) $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\left|\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta) - \mathcal{L}_{n,t}(f_t(\beta);\beta)\right| = O_p\left(\frac{[\log(n)]^{\delta_4}}{n}\right)$,

(ii) $\sup_{1\leq t\leq T}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}|\mathcal{L}_{n,t}(f;\beta) - \mathcal{L}_t(f;\beta)| = O_p\left(\frac{[\log(n)]^{\delta_5}}{\sqrt{n}}\right)$,

for some constants $\delta_4 > 0$ and $\delta_5 > 0$, where $\mathcal{L}_{n,t}(f;\beta)$ is defined as in Lemma 2, and $\mathcal{L}_t(f;\beta)$ is defined in equation (c.5).

Proof of Lemma C.3 (i): By a second-order Taylor expansion around $\hat{f}_{n,t}(\beta)$, we have:

$$
\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta) - \mathcal{L}_{n,t}(f_t(\beta);\beta) = -\frac{1}{2}[\hat{f}_{n,t}(\beta) - f_t(\beta)]'\frac{\partial^2\mathcal{L}_{n,t}(\tilde{f}_{n,t}(\beta);\beta)}{\partial f_t\partial f_t'}[\hat{f}_{n,t}(\beta) - f_t(\beta)],
$$

where $\tilde{f}_{n,t}(\beta)$ is a mean value, since $\dfrac{\partial\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)}{\partial f_t} = 0$, w.p.a. 1. Thus, from the uniform convergence of $\hat{f}_{n,t}(\beta)$ to $f_t(\beta)$ (Limit Theorem 1 in Appendix B.1), for any $\eta > 0$, we get w.p.a. 1:

$$
\begin{aligned}
&\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\left|\mathcal{L}_{n,t}(\hat{f}_{n,t}(\beta);\beta)-\mathcal{L}_{n,t}(f_t(\beta);\beta)\right| \\
&\le \sup_{\beta\in\mathcal{B}}\sup_{1\le t\le T}\|\hat{f}_{n,t}(\beta)-f_t(\beta)\|^2 \sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_{n,t}(f;\beta)}{\partial f_t\partial f_t'}\right\|, \\
&= O_p\left(\frac{[\log(n)]^{2\delta_2}}{n}\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_{n,t}(f;\beta)}{\partial f_t\partial f_t'}\right\|\right).
\end{aligned}
$$

Moreover, from Lemma C.2 we have:

$$
\begin{aligned}
\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_{n,t}(f;\beta)}{\partial f_t\partial f_t'}\right\| &\le \sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_t(f;\beta)}{\partial f_t\partial f_t'}\right\| \\
&\quad + \sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n}\left\|\frac{\partial^2\mathcal{L}_{n,t}(f;\beta)}{\partial f_t\partial f_t'}-\frac{\partial^2\mathcal{L}_t(f;\beta)}{\partial f_t\partial f_t'}\right\| \\
&= \sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f\in\mathcal{F}_n:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_t(f;\beta)}{\partial f_t\partial f_t'}\right\| + O_p\left(\frac{[\log(n)]^{\delta_3}}{\sqrt{n}}\right),
\end{aligned}
$$

for a constant $\delta_3>0$. Then, Lemma C.3 (i) follows from the next bound:

$$
\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_t(f;\beta)}{\partial f_t\partial f_t'}\right\| = O_p\left([\log(n)]^{1/d_1}\right), \tag{c.46}
$$

where $d_1>0$ is defined in Assumption H.5. To prove bound (c.46), we use:

$$
\sup_{\beta\in\mathcal{B}}\sup_{f:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_t(f;\beta)}{\partial f_t\partial f_t'}\right\| \le \sup_{\beta\in\mathcal{B}}E_0\left[\sup_{f:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\log h(y_{i,t}|y_{i,t-1},f;\beta)}{\partial f\partial f'}\right\| |\underline{f_t}\right] \le \xi_{t,1}^{**},
$$

if $\eta\le\eta^*$, where process $\xi_{t,1}^{**}$ and constant η^* are defined in Assumption H.5. Then, we get:

$$
\begin{aligned}
&\mathbb{P}\left[\sup_{1\le t\le T}\sup_{\beta\in\mathcal{B}}\sup_{f:\|f-f_t(\beta)\|<\eta}\left\|\frac{\partial^2\mathcal{L}_t(f;\beta)}{\partial f_t\partial f_t'}\right\|\ge C_1(\log n)^{1/d_1}\right] \le T\mathbb{P}\left[\xi_{t,1}^{**}\ge C_1(\log n)^{1/d_1}\right] \\
&\le Tb_1\exp\left(-c_1{C_1}^{d_1}\log n\right) = b_1Tn^{-c_1{C_1}^{d_1}} = O(T/n) = o(1),
\end{aligned}
$$

if constant C_1 is such that $C_1\ge c_1^{-1/d_1}$. Then, the bound in (c.46) follows.

Proof of Lemma C.3 (ii): The proof of Lemma C.3 (ii) is similar to the proof of Lemma C.2 in Section C.9.2, by using $a(Y_{i,t},f,\beta)=\log h(y_{i,t}|y_{i,t-1},f,\beta)$.

C.9.4 Lemma C.4

Lemma C.4: *Let function φ be either:*
(i) *The matrix inversion $\varphi : \mathcal{U} \rightarrow \mathbb{R}^{r\times r}$, $\varphi(x) = x^{-1}$, where $\mathcal{U}$ denotes the set of positive definite (r,r) matrices, or*
(ii) *The mapping $\varphi : \mathcal{U} \rightarrow \mathbb{R}^{s\times s}$, $\varphi(x) = (x^{11})^{-1}$, where x^{11} is the upper-left s-dimensional block of matrix x^{-1}, for $s < r$.*
Then, under Assumption H.5 in Appendix A.1, Regularity Condition RC.3 (2) in Appendix B.3 is satisfied with $\mu_t(\beta) = I_t(\beta) = E_0\left[-\frac{\partial^2 \log h(y_{i,t}|y_{i,t-1}, f_t(\beta);\beta)}{\partial(\beta', f_t')'\partial(\beta', f_t')}|\underline{f_t}\right]$.

Proof of Lemma C.4 (i): Let us verify Regularity Condition RC.3 (2i) in Appendix B.3. Let $\mathcal{K} \subset \mathcal{U}$ be compact, and let $w, z \in \mathcal{K}$. Since $w^{-1} - z^{-1} = -z^{-1}(w-z)w^{-1}$, we deduce that φ is Lipschitz continuous on $\mathcal{K}$ with Lipschitz constant $L = \sup_{z\in\mathcal{K}} \left\|z^{-1}\right\|^2 < \infty$. Hence, Regularity Condition RC.3 (2i) is satisfied. Let us now consider Regularity Condition RC.3 (2ii) in Appendix B.3. Let $w, z \in \mathcal{U}$, $w = (Id + \Delta)z$, $\|\Delta\| \leq 1/2$. Then $Id + \Delta$ is a nonsingular matrix. From $w^{-1} = z^{-1}(Id+\Delta)^{-1}$ and $\|(Id+\Delta)^{-1}\| \leq (1 - \|\Delta\|)^{-1} = 2$, we see that Regularity Condition RC.3 (2ii) is satisfied with $C_{10} = 2$, $\gamma_{10} = 0$ and $\psi(z) = \|z^{-1}\|$. Indeed, $E\left[\sup_{\beta\in\mathcal{B}} |\psi(\mu_t(\beta))|^4\right] = E\left[\sup_{\beta\in\mathcal{B}} \|\mu_t(\beta)^{-1}\|^4\right] \leq C_1 E\left[(\xi^*_{t,1})^4\right] < \infty$, for some constant $C_1 > 0$, where process $\xi^*_{t,1}$ is defined in Assumption H.5.

Proof of Lemma C.4 (ii): Let us consider the block decomposition:

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

Then $\varphi(x) = x_{11} - x_{12}x_{22}^{-1}x_{21}$. Regularity Condition RC.3 (2i) is satisfied, since φ consists of summation and product of mappings that are Lipschitz continuous on compact sets. To check Regularity Condition RC.3 (2ii), let $w, z \in \mathcal{U}$, $w = (Id+\Delta)z$, $\|\Delta\| \leq 1/2$. Then:

$$\begin{aligned} \|\varphi(w)\| &\leq \|w_{11}\| + \|w_{12}\| \left\|w_{22}^{-1}\right\| \|w_{21}\| \leq \|w\| + \|w\|^2 \left\|w_{22}^{-1}\right\| \\ &\leq \|Id+\Delta\| \|z\| + \|Id+\Delta\|^2 \|z\|^2 \left\|w_{22}^{-1}\right\|. \end{aligned}$$

Denote by $d = r - s$ the dimension of w_{22}. Since matrices w and w_{22} are positive definite, and matrix norms are equivalent, we have:

$$\begin{aligned}
\left\|w_{22}^{-1}\right\| &\leq C_1^* \sup_{u\in\mathbb{R}^d:\|u\|=1} u^{'}w_{22}^{-1}u = C_1^*\left(\inf_{u\in\mathbb{R}^d:\|u\|=1} u^{'}w_{22}u\right)^{-1} \leq C_1^*\left(\inf_{u\in\mathbb{R}^r:\|u\|=1} u^{'}wu\right)^{-1} \\
&= C_1^* \sup_{u\in\mathbb{R}^r:\|u\|=1} u^{'}w^{-1}u \leq C_1^*C_1^{**}\left\|w^{-1}\right\|,
\end{aligned}$$

where $C_1^*, C_1^{**} > 0$ are constants. Moreover, $\|w^{-1}\| \leq \left\|(Id+\Delta)^{-1}\right\| \|z^{-1}\| \leq 2\|z^{-1}\|$. We get that $\|\varphi(w)\| \leq C_2\left(\|z\| + \|z\|^2\|z^{-1}\|\right) \leq 2C_2\|z\|^2\|z^{-1}\|$, for a constant $C_2 > 0$. Thus, Regularity Condition RC.3 (2ii) is satisfied with $\gamma_{10} = 2$ and $\psi(z) = \|z^{-1}\|$.

References

[1] Amemiya, T. (1985): *Advanced Econometrics*, Harvard University Press.

[2] Andrews, D. (1988): "Laws of Large Numbers for Dependent Non-Identically Distributed Random Variables", Econometric Theory, 4, 458-67.

[3] Bahadur, R. (1960): "Asymptotic Efficiency of Tests and Estimates", Sankhya, 22, 229-252.

[4] Bahadur, R. (1967): "Rates of Convergence of Estimates and Test Statistics", Annals of Mathematical Statistics, 38, 303-324.

[5] Bosq, D. (1998): *Nonparametric Statistics for Stochastic Processes*, Springer.

[6] Breiman, L. (1992): *Probability*, Classics in Applied Mathematics, SIAM.

[7] Chen, X., and X., Shen (1998): "Sieve Extremum Estimates for Weakly Dependent Data“, Econometrica, 66, 289-214.

[8] Davidson, J. (1994): *Stochastic Limit Theory*, Advanced Texts in Econometrics, Oxford University Press, Oxford.

[9] Fu, J. (1982): "Large Sample Point Estimation: A Large Deviation Theory Approach", Annals of Statistics, 10, 762-771.

[10] Newey, W. (1991): "Uniform Convergence in Probability and Stochastic Equicontinuity“, Econometrica, 59, 1161-1167.

[11] Newey, W., and D., McFadden (1994): "Large Sample Estimation and Hypothesis Testing“, *Handbook of Econometrics*, Volume IV, R. Engle and D. McFadden (Eds.), North-Holland, Chapter 36.

[12] Shen, X., and W., Wong (1994): "Convergence Rate of Sieve Estimates", Annals of Statistics, 22, 580-615.

[13] White, H. (2001): *Asymptotic Theory for Econometricians*, Academic Press.