ONLINE APPENDIX: The ESP estimator

Benjamin Holcblat and Fallaw Sowell

This appendix mainly consists of the proofs of Theorem 1, consistency and asymptotic normality of the ESP estimator, and Theorem 2, asymptotic distributions of the Trintiy+1 test statistics. The proof of Theorem 1 builds on the traditional uniform convergence proof technique of Wald (1949). The proof of Theorem 2 adapts the usual way of deriving the trinity tests. The length of the proofs is mainly due to the variance term $|\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}}$ and the high-level of details. The latter should make the proofs more transparent, and should ease the use of the intermediary results in further research.

In addition to the proofs, this appendix contains a table of contents, some formal definitions, the precise assumptions of the paper, a discussion thereof, and additional information regarding the examples.

Content of the paper and the online appendix

Appendix A. Definitions and assumptions

Definition 1 (ESP approximation; Ronchetti and Welsh 1994)**.** *The ESP approximation of the distribution of the solution to the empirical moment conditions* (2) *is*

$$\hat{f}_{\theta_T^*}(\theta) := \exp\left\{T \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\left(\frac{T}{2\pi}\right)^{m/2} |\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}} \tag{11}$$

where $|.|_{\det}$ *denotes the determinant function,* θ_T^* *a solution to the empirical moment conditions* (2)*,* $\psi_t(.) := \psi(X_t, .)$*, and*

$$\Sigma_T(\theta) := \left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}\left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\psi_t(\theta)\psi_t(\theta)'\right]\left[\sum_{t=1}^{T} \hat{w}_{t,\theta}\frac{\partial \psi_t(\theta)'}{\partial \theta}\right]^{-1}, \tag{12}$$

$$\hat{w}_{t,\theta} := \frac{\exp\left[\tau_T(\theta)'\psi_t(\theta)\right]}{\sum_{i=1}^{T}\exp\left[\tau_T(\theta)'\psi_i(\theta)\right]}, \tag{13}$$

$$\tau_T(\theta) \text{ such that (s.t.) } \sum_{t=1}^{T}\psi_t(\theta)\exp\left[\tau_T(\theta)'\psi_t(\theta)\right] = 0_{m\times 1}. \tag{14}$$

Definition 2 (ESP estimator)**.** *The ESP estimator* $\hat{\theta}_T$ *is a maximizer of the ESP approximation* (11)*, i.e.,*

$$\hat{\theta}_T \in \arg\max_{\theta\in\mathbf{\Theta}} \hat{f}_{\theta_T^*}(\theta). \tag{15}$$

We require the following assumption to prove the existence and the consistency of the ESP estimator.

Assumption 1. **(a)** *The data* $(X_t)_{t=1}^{\infty}$ *are a sequence of i.i.d. random vectors of dimension* p *on the complete probability sample space* $(\mathbf{\Omega}, \mathcal{E}, \mathbb{P})$*.* **(b)** *Let the moment function* $\psi : \mathbf{R}^p \times \mathbf{\Theta}^\epsilon \mapsto \mathbf{R}^m$ *be s.t.* $\theta \mapsto \psi(X_1, \theta)$ *is continuously differentiable* $\mathbb{P}$*-a.s., and* $\forall\theta \in \mathbf{\Theta}^\epsilon$*,* $x \mapsto \psi(x, \theta)$ *is* $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R}^m)$*-measurable, where, for* $\epsilon > 0$*,* $\mathbf{\Theta}^\epsilon$ *denotes the* ϵ*-neighborhood of* $\mathbf{\Theta}$*.* **(c)** *In the parameter space* $\mathbf{\Theta}$*, there exists a unique* $\theta_0 \in \mathrm{int}(\mathbf{\Theta})$ *s.t.* $\mathbb{E}\left[\psi(X_1, \theta_0)\right] = 0_{m\times 1}$ *where* $\mathbb{E}$ *denotes the expectation under* $\mathbb{P}$*.* **(d)** *Let the parameter space* $\mathbf{\Theta} \subset \mathbf{R}^m$ *be a compact set, s.t., for all* $\theta \in \mathbf{\Theta}$*, there exists* $\tau(\theta) \in \mathbf{R}^m$ *that solves the equation* $\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1, \theta)\right] = 0$ *for* τ*.* **(e)** $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} \mathrm{e}^{2\tau'\psi(X_1,\theta)}\right] < \infty$ *where* $\mathbf{S} := \{(\theta, \tau) : \theta \in \mathbf{\Theta} \,\&\, \tau \in \mathbf{T}(\theta)\}$ *and* $\mathbf{T}(\theta) := \overline{B_{\epsilon_\mathbf{T}}(\tau(\theta))}$ *with* $\overline{B_{\epsilon_\mathbf{T}}(\tau(\theta))}$ *the closed ball of radius* $\epsilon_\mathbf{T} > 0$ *and center* $\tau(\theta)$*.* **(f)** $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}} |\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|^2\right] < \infty$*, where* $|.|$ *denotes the Euclidean norm.* **(g)** $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1, \theta)\psi(X_1, \theta)'|^2\right] < \infty$*.* **(h)** *For all*

$\theta \in \mathbf{\Theta}$, the matrices $\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]$ *and* $\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ *are invertible, so* $\Sigma(\theta) := \left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]^{-1}$ *is also invertible.*

We require the following additional assumption to prove the asymptotic normality of the ESP estimator.

Assumption 2. **(a)** *The function $\theta \mapsto \psi(X_1,\theta)$ is three times continuously differentiable in a neighborhood $\mathcal{N}$ of θ_0 in $\mathbf{\Theta}$ $\mathbb{P}$-a.s.* **(b)** *There exists a $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R})$-measurable function $b(.)$ satisfying $\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{k_1\tau'\psi(X_1,\theta)}b(X_1)^{k_2}\right] < \infty$ for $k_1 \in [\![1,2]\!]$ and $k_2 \in [\![1,4]\!]$ s.t., for all $j \in [\![0,3]\!]$, $\sup_{\theta\in\mathcal{N}}|\nabla^j\psi(X_1,\theta)| \leqslant b(X_1)$ where $\nabla^j\psi(X_1,\theta)$ denotes a vector of all partial derivatives of $\theta \mapsto \psi(X_1,\theta)$ of order j.*

Assumptions 1 and 2 are stronger than the usual assumptions in the MM literature, but are similar to assumptions used in the entropy literature and related literatures. Assumptions 1 and 2 are essentially adapted from Haberman (1984), Kitamura and Stutzer (1997), and Schennach (2007, Assumption 3). See also Chib et al. (2018) for similar assumptions. The Appendix C.1 (p. 83) contains a detailed discussion of Assumptions 1 and 2.

In addition to Assumptions 1 and 2, we require the following standard and mild assumption to establish the asymptotic distribution of the Wald, LM, ALR, and ET statistics.

Assumption 3 (For the trinity+1)**.** **(a)** *The function $r : \mathbf{\Theta} \to \mathbf{R}^q$ in the null hypothesis (9) is continuously differentiable.* **(b)** *The derivative $R(\theta) := \frac{\partial r(\theta)}{\partial\theta'}$ is full rank at θ_0.*

Appendix B. Proofs

B.1. **Proof of Theorem 1(i): Existence and consistency.** The proof of Theorem 1(i) (i.e., consistency) adapts the Wald's approach to consistency (Wald 1949) along the lines of Kitamura and Stutzer (1997), Schennach (2007), Chib et al. (2018) and others. More precisely, standardizing the logarithm of the ESP approximation, we show that, $\mathbb{P}$-a.s. for T big enough, the ESP estimator maximizes the LogESP function (8) on p. 5, where $\sup_{\theta\in\mathbf{\Theta}}\left|\ln\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right| = o(1)$ and $\sup_{\theta\in\mathbf{\Theta}}\left|\frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det}\right| = O(T^{-1})$. The two main differences between our proof of Theorem 1(i) and the proofs available in the entropy literature are the following. Firstly, we need to ensure that, for T big enough, for all $\theta \in \mathbf{\Theta}$, $|\Sigma_T(\theta)|_{\det}$ is bounded away from zero, so that the LogESP function (8) on p. 5 does not diverge to ∞ on parts of the parameter space. Secondly, we prove that the joint parameter space for θ and τ (i.e., $\mathbf{S}$) is a compact set.

Core of the proof of Theorem 1i. Under Assumption 1(a)(b) and (d)-(h), by Lemma 1 (p. 19), $\mathbb{P}$-a.s. for T big enough, the ESP approximation and the ESP estimator exist. Moreover, under Assumption 1(a)-(b) and (d)-(h), by Lemma 6iv (p. 24), $\mathbb{P}$-a.s. for T big enough, $|\Sigma_T(\theta)|_{\det} > 0$, for all $\theta \in \mathbf{\Theta}$. Thus, we can apply the strictly increasing transformation $x \mapsto \frac{1}{T}[\ln(x) - \frac{m}{2}\ln(\frac{T}{2\pi})]$

to the ESP approximation in equation (11) on p. 17, so that, $\mathbb{P}$-a.s. for T big enough,

$$\begin{aligned}
&\hat{\theta}_T \in \arg\max_{\theta\in\mathbf{\Theta}} \hat{f}_{\theta^*_T}(\theta) \\
\Leftrightarrow\ &\hat{\theta}_T \in \arg\max_{\theta\in\mathbf{\Theta}} \left\{ \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det} \right\}.
\end{aligned} \tag{16}$$

Now, by the triangle inequality,

$$\begin{aligned}
&\sup_{\theta\in\mathbf{\Theta}} \left| \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det} - \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}] \right| \\
\leqslant\ &\sup_{\theta\in\mathbf{\Theta}} \left| \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}] \right| + \sup_{\theta\in\mathbf{\Theta}} \left| \frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det} \right| \\
=\ &o(1)\ \mathbb{P}\text{-a.s. as } T\to\infty
\end{aligned} \tag{17}$$

where the last equality follow from Lemma 2iv (p. 20) and Lemma 6v (p. 24) under Assumption 1(a)-(b) and (d)-(h). Thus, regarding $\hat{\theta}_T$, it is now sufficient to check the assumptions of the standard consistency theorem (e.g. Newey and McFadden 1994, pp. 2121-2122 Theorem 2.1, which is also valid in an almost-sure sense). Firstly, under Assumption 1 (a)-(e) and (g)-(h), by Lemma 10iv (p. 31), $\theta \mapsto \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|$ is uniquely maximized at θ_0, i.e., for all $\theta \in \mathbf{\Theta}\setminus\{\theta_0\}$, $\ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}] < \ln\mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}] = 0$. Secondly, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22), $\theta \mapsto \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]$ is continuous in $\mathbf{\Theta}$. Finally, by Assumption 1(d), the parameter space $\mathbf{\Theta}$ is compact. □

Lemma 1 (Existence of the ESP approximation and estimator)**.** *Under Assumption 1(a)(b) and (d)-(h), $\mathbb{P}$-a.s. for T big enough,*

(i) *the ESP approximation $\hat{f}_{\theta^*_T}(.)$ exists;*
(ii) *$\theta \mapsto \tau_T(\theta)$ is unique and continuously differentiable in $\mathbf{\Theta}$, so that the ESP approximation $\theta \mapsto \hat{f}_{\theta^*_T}(\theta)$ is also unique and continuous in $\mathbf{\Theta}$;*
(iii) *for all $\theta \in \mathbf{\Theta}$, the ESP approximation $\omega \mapsto \hat{f}_{\theta^*_T}(\theta)$ is $\mathcal{E}/\mathcal{B}(\mathbf{R})$-measurable; and*
(iv) *there exists an ESP estimator $\hat{\theta}_T \in \arg\max_{\theta\in\mathbf{\Theta}} \hat{f}_{\theta^*_T}(\theta)$ that is $\mathcal{E}/\mathcal{B}(\mathbf{R}^m)$-measurable.*

Proof. The result follows from Lemmas 2 (p. 20), 3 (p. 22) and 6 (p. 24) and standard arguments. For completeness, a detailed proof is provided.

(i) Under Assumption 1(a)(b), (d)-(e)(g) and (h), by Lemma 2ii (p. 20), $\mathbb{P}$-a.s. there exists a $\mathcal{B}(\mathbf{\Theta})\otimes\mathcal{E}/\mathcal{B}(\mathbf{R}^m)$-measurable function $\tau_T(.)$ s.t., for T big enough, for all $\theta \in \mathbf{\Theta}$, $\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta) = 0_{m\times 1}$ and $\tau_T(\theta) \in \mathrm{int}[\mathbf{T}(\theta)]$. Moreover, under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22) with $\mathrm{P} = \frac{1}{T}\sum_{t=1}^{T}\delta_{X_t}$, for all $T \in [\![1,\infty[\![$, for all $(\theta,\tau) \in \mathbf{S}$, $0 < \frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$, so that, for all $\theta \in \mathbf{\Theta}$, $0 < \frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}$. Thus, the ET term exists. Now, under Assumption 1(a)-(b) and (d)-(h), by Lemma 6iv (p. 24), $\mathbb{P}$-a.s. for T big enough, $\inf_{\theta\in\mathbf{\Theta}}|\Sigma_T(\theta)|_{\det} > 0$, so that the variance term of the ESP approximation exists. Thus, the ESP approximation exists.

(ii) By Assumption 1(b), $\theta \mapsto \psi(X_1,\theta)$ is continuously differentiable in $\mathbf{\Theta}^\epsilon$ $\mathbb{P}$-a.s., so that it is sufficient to show that $\tau_T(.)$ is unique and continuous, which we prove at once with the standard implicit function theorem. Check its assumptions. Firstly, under Assumption 1(a)(b), (d)-(e)(g) and (h), by Lemma 2ii (p. 20), $\mathbb{P}$-a.s. there exists a function $\tau_T(.)$ s.t., for T big

enough, for all $\theta \in \mathbf{\Theta}$, $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta) = 0_{m\times 1}$ and $\tau_T(\theta) \in \operatorname{int}[\mathbf{T}(\theta)]$. Secondly, for all $\dot\theta \in \mathbf{\Theta}$, $\left.\frac{\partial\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\right]}{\partial \tau'}\right|_{(\theta,\tau)=(\dot\theta,\tau_T(\dot\theta))} = \frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\dot\theta)}\psi_t(\dot\theta)\psi_t(\dot\theta)'$, which is full rank $\mathbb{P}$-a.s. for T big enough for all $\theta \in \mathbf{\Theta}$, because under Assumption 1(a)-(b) and (d)-(h), by Lemma 6iv (p. 24), $\mathbb{P}$-a.s. for T big enough, $\inf_{\theta\in\mathbf{\Theta}} |\Sigma_T(\theta)|_{\det} > 0$. Finally, by Assumption 1(b), $(\theta,\tau) \mapsto \frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)$ is continuously differentiable in $\mathbf{S}^{\epsilon}$.

(iii) By Assumption 1(b), for all $\theta \in \mathbf{\Theta}$, $x \mapsto \psi(x,\theta)$ is $\mathcal{B}(\mathbf{R}^p)/\mathcal{B}(\mathbf{R}^m)$-measurable. Moreover, under Assumption 1(a)(b), (d)-(e)(g) and (h), by Lemma 2ii (p. 20), $\mathbb{P}$-a.s. $\tau_T(.)$ is a $\mathcal{B}(\mathbf{\Theta})\otimes\mathcal{E}/\mathcal{B}(\mathbf{R}^m)$-measurable function. Thus, the result follows.

(iv) By Assumption 1(d), $\mathbf{\Theta}$ is compact, so that, by the statements (i)-(iii) of the present lemma, the result follows from the Schmetterer-Jennrich lemma (Schmetterer 1966 Chap. 5 Lemma 3.3; Jennrich 1969 Lemma 2). □

Lemma 2 (Asymptotic limit of the ET term)**.** *Under Assumption 1(a)(b), (d)-(e)(g) and (h),*

(i) $\mathbb{P}$-*a.s. as* $T\to\infty$, $\sup_{(\theta,\tau)\in\mathbf{S}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)} - \mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]\right| = o(1)$, *which implies that* $\mathbb{P}$-*a.s. as* $T\to\infty$, $\sup_{(\theta,\tau)\in\mathbf{S}}\left|\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right] - \ln\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]\right| = o(1)$ *;*

(ii) $\mathbb{P}$-*a.s. there exists a* $\mathcal{B}(\mathbf{\Theta})\otimes\mathcal{E}/\mathcal{B}(\mathbf{R}^m)$*-measurable function* $\tau_T(.)$ *s.t., for* T *big enough, for all* $\theta\in\mathbf{\Theta}$, $\tau_T(\theta)\in\arg\min_{\tau\in\mathbf{R}^m}\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$, $\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta) = 0_{m\times 1}$ *and* $\tau_T(\theta)\in\operatorname{int}[\mathbf{T}(\theta)]$*;*

(iii) $\mathbb{P}$-*a.s. as* $T\to\infty$, $\sup_{\theta\in\mathbf{\Theta}}|\tau_T(\theta)-\tau(\theta)| = o(1)$*;*

(iv) $\mathbb{P}$-*a.s. as* $T\to\infty$, $\sup_{\theta\in\mathbf{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right| = o(1)$, *which implies that* $\mathbb{P}$-*a.s. as* $T\to\infty$, $\sup_{\theta\in\mathbf{\Theta}}\left|\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right| = o(1)$.

Proof. *(i)* Under Assumptions 1 (a)-(b)(d)(e)(g) and (h), by Lemma 4iii (p. 23), $\mathbf{S} := \{(\theta,\tau) : \theta\in\mathbf{\Theta}\wedge\tau\in\mathbf{T}(\theta)\}$ is a compact set.[9] Thus, under Assumption 1(a)-(b), (d) (e) and (h), the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3) yields the first part of the result. Now, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22), $(\theta,\tau)\mapsto\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ is continuous, so that $\{\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}] : (\theta,\tau)\in\mathbf{S}\}$ is a compact set by Assumption 1(d) —continuous mappings preserve compactness (e.g., Rudin 1953, Theorem 4.14). Moreover, $x\mapsto\ln x$ is continuous, and, under Assumptions 1 (a)(b)(d)(e)(g) and (h), again by Lemma 3 (p. 22), $0 < \inf_{(\theta,\tau)\in\mathbf{S}}\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$. Thus, we can choose an $\eta\in\left]0,\inf_{(\theta,\tau)\in\mathbf{S}}\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]\right[$ s.t. $x\mapsto\ln x$ is uniformly continuous on the closed η-neighborhood of $\{\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}] : (\theta,\tau)\in\mathbf{S}\}$ —continuous mappings on a compact set are uniformly continuous (e.g., Rudin 1953, Theorem 4.19). Then, the second part follows from the first part of the result: By the first part, $\mathbb{P}$-a.s. there exists a $\dot T\in\mathbf{N}$ s.t., $\forall T\in[\![\dot T,\infty[\![$, $\sup_{(\theta,\tau)\in\mathbf{S}}|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)} - \mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]| < \eta/2$.

(ii)-(iii) The proof follows the overall strategy of Schennach (2007, Step 1 in the proof of Theorem 10). For completeness and in order to justify our different assumptions, we provide a detailed proof. In particular, note that we formally prove that $0 < \inf_{\theta\in\mathbf{\Theta}}\inf_{\tau\in\mathbf{T}(\theta):|\tau-\tau(\theta)|\geqslant\eta}|\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}] - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|$: See Lemma 5 (p. 24). Let $\eta\in]0,\epsilon_{\mathbf{T}}]$ be a fixed constant. By Assumption 1(a)(b),

[9] Note that, unlike what has been sometimes suggested in the entropy literature, if $\mathbf{T}(\theta)$ is an unspecified compact set, $\{(\theta,\tau):\theta\in\mathbf{\Theta}\wedge\tau\in\mathbf{T}(\theta)\}$ does not need to be a compact set: $\{(\theta,\tau):\theta\in\mathbf{\Theta}\wedge\tau\in\mathbf{T}(\theta)\}$ is not a Cartesian product, but the graph of a correspondence. See Lemma 4 (p. 23) for more details.

$(\theta,\omega)\mapsto \frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$ is continuous w.r.t θ and $\mathcal{E}/\mathcal{B}(\mathbf{R})$-measurable w.r.t to ω, so that it is $\mathcal{B}(\mathbf{\Theta})\otimes\mathcal{E}/\mathcal{B}(\mathbf{R})$-measurable (e.g., Aliprantis and Border 2006/1999, Lemma 4.51). Moreover, under Assumptions 1 (a)-(b)(d)(e)(g) and (h), by Lemma 4ii (p. 23), $\theta\mapsto\mathbf{T}(\theta)$ is a nonempty compact valued measurable correspondence. Then, by a generalization of the Schmetterer-Jennrich lemma (e.g., Aliprantis and Border 2006/1999, Theorem 18.19), we can define a $\mathcal{B}(\mathbf{\Theta})\otimes\mathcal{E}/\mathcal{B}(\mathbf{R})$-measurable function $\tilde\tau_T(\theta)$ s.t., for all $\theta\in\mathbf{\Theta}$, $\tilde\tau_T(\theta)\in\arg\min_{\tau\in\mathbf{T}(\theta)}\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$. For the present proof, put $\varepsilon:=\inf_{\theta\in\mathbf{\Theta}}\inf_{\tau\in\mathbf{T}(\theta):|\tau-\tau(\theta)|\geqslant\eta}|\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|$, which is strictly positive[10] by Lemma 5 (p. 24) under Assumptions 1 (a)(b)(d)(e) and (h).[11] Then, by the definition of ε, whenever $\sup_{\theta\in\mathbf{\Theta}}|\mathbb{E}[\mathrm{e}^{\tilde\tau_T(\theta)'\psi(X_1,\theta)}]-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|<\varepsilon$, then $\sup_{\theta\in\mathbf{\Theta}}|\tilde\tau_T(\theta)-\tau(\theta)|\leqslant\eta$. We now show that it is happening $\mathbb{P}$-a.s. as $T\to\infty$. Under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 10 (p. 31), $\tau(\theta)=\arg\min_{\tau\in\mathbf{R}^m}\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$, so that

$$\begin{aligned}
&\sup_{\theta\in\mathbf{\Theta}}\left|\mathbb{E}[\mathrm{e}^{\tilde\tau_T(\theta)'\psi(X_1,\theta)}]-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right|\\
=&\sup_{\theta\in\mathbf{\Theta}}\left\{\mathbb{E}[\mathrm{e}^{\tilde\tau_T(\theta)'\psi(X_1,\theta)}]-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right\}\\
\overset{(a)}{=}&\sup_{\theta\in\mathbf{\Theta}}\left\{\mathbb{E}[\mathrm{e}^{\tilde\tau_T(\theta)'\psi(X_1,\theta)}]-\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tilde\tau_T(\theta)'\psi_t(\theta)}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tilde\tau_T(\theta)'\psi_t(\theta)}-\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau(\theta)'\psi_t(\theta)}\right.\\
&\left.+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau(\theta)'\psi_t(\theta)}-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right\}\\
\overset{(b)}{\leqslant}&\sup_{\theta\in\mathbf{\Theta}}\left\{\mathbb{E}[\mathrm{e}^{\tilde\tau_T(\theta)'\psi(X_1,\theta)}]-\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tilde\tau_T(\theta)'\psi_t(\theta)}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau(\theta)'\psi_t(\theta)}-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right\}\\
\overset{(c)}{\leqslant}&\sup_{\theta\in\mathbf{\Theta}}\left|\mathbb{E}[\mathrm{e}^{\tilde\tau_T(\theta)'\psi(X_1,\theta)}]-\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tilde\tau_T(\theta)'\psi_t(\theta)}\right|+\sup_{\theta\in\mathbf{\Theta}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau(\theta)'\psi_t(\theta)}-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right|\\
\overset{(d)}{=}&\ o(1)\ \mathbb{P}\text{-a.s. as } T\to\infty.
\end{aligned}\tag{18}$$

(a) Add and subtract $\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tilde\tau_T(\theta)'\psi_t(\theta)}$ and $\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau(\theta)'\psi_t(\theta)}$. *(b)* Note that, under Assumption 1(d) and (e), by definition, $\tau(\theta)\in\mathbf{T}(\theta)$ and $\tilde\tau_T(\theta)\in\arg\min_{\tau\in\mathbf{T}(\theta)}\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$ so that $\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tilde\tau_T(\theta)'\psi_t(\theta)}-\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau(\theta)'\psi_t(\theta)}\leqslant 0$. *(c)* Triangle inequality w.r.t. the uniform norm. *(d)* Under Assumption 1(d)(e), by definition, for all $\theta\in\mathbf{\Theta}$, $\tau(\theta)\in\mathbf{T}(\theta)$ and $\tilde\tau_T(\theta)\in\mathbf{T}(\theta)$ so that the conclusion follows from statement (i).

Inequality (18) implies that $\sup_{\theta\in\mathbf{\Theta}}|\tilde\tau_T(\theta)-\tau(\theta)|=o(1)$ $\mathbb{P}$-a.s. as $T\to\infty$. Moreover, by Assumption 1(e), for all $\theta\in\mathbf{\Theta}$, $\mathbf{T}(\theta)=\overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta))}$ where $\epsilon_{\mathbf{T}}>0$. Thus, $\mathbb{P}$-a.s., for T big enough, for all $\theta\in\mathbf{\Theta}$, $\tilde\tau_T(\theta)\in\operatorname{int}[\mathbf{T}(\theta)]$. Now, for all $\theta\in\mathbf{\Theta}$, $\tau\mapsto\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$ is a convex function (Lemma 29i on p. 86 with $\mathrm{P}=\frac{1}{T}\sum_{t=1}^{T}\delta_{X_t}$ ensures that $\frac{\partial^2[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}]}{\partial\tau\partial\tau'}=\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\geqslant 0$), and the local minimum of a convex function is a global minimum (e.g., Hiriart-Urruty and Lemaréchal 1993/1996, p. 253). Therefore, $\mathbb{P}$-a.s. for T big enough, for

[10]The argument requires $\varepsilon>0$. If $\varepsilon=0$, then the upcoming inequality (18) is not sufficient to show that $\sup_{\theta\in\mathbf{\Theta}}|\mathbb{E}[\mathrm{e}^{\tilde\tau_T(\theta)'\psi(X_1,\theta)}]-\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|<\varepsilon$.

[11]Strict convexity of $\tau\mapsto\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ and compactness of $\mathbf{\Theta}$ are not sufficient to ensure that $\varepsilon>0$: We also need the continuity of the value function of the first infimum, which we obtain through Berge's maximum theorem. See Lemma 5 (p. 24).

all $\theta \in \mathbf{\Theta}$, $\tilde{\tau}_T(\theta)$ minimizes $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}$ not only over $\mathbf{T}(\theta)$, but also over $\mathbf{R}^m$, which means that we can put $\tilde{\tau}_T(\theta) = \tau_T(\theta)$.

(iv) Addition and subtraction of $\mathbb{E}[\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}]$, and the triangle inequality yield $\mathbb{P}$-a.s. for T big enough

$$\begin{aligned}
&\sup_{\theta\in\mathbf{\Theta}} \left| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}] \right| \\
&\leqslant \sup_{\theta\in\mathbf{\Theta}} \left| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} - \mathbb{E}[\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}] \right| + \sup_{\theta\in\mathbf{\Theta}} \left| \mathbb{E}[\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}] - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}] \right| \\
&= o(1) \text{ , as } T \to \infty,
\end{aligned}$$

where the explanations for the last equality are as follows. By the statement (i) of the present lemma, $\mathbb{P}$-a.s. as $T \to \infty$, $\sup_{(\theta,\tau)\in\mathbf{S}} \left| \frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)} - \mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}] \right| = o(1)$. Moreover, by the statement (ii) of the present lemma, $\mathbb{P}$-a.s. for T big enough, $\tau_T(\theta) \in \operatorname{int}[\mathbf{T}(\theta)]$, so that, for all $\theta \in \mathbf{\Theta}$, $(\theta, \tau_T(\theta)) \in \mathbf{S}$. Thus, the first supremum is $o(1)$ as $T \to \infty$. Regarding the second supremum, under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22), $(\theta,\tau) \mapsto \mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ is continuous in $\mathbf{S}$. Now, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 4iii (p. 23), $\mathbf{S}$ is compact, so that $(\theta,\tau) \mapsto \mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ is also uniformly continuous in $\mathbf{S}$ —continuous functions on compact sets are uniformly continuous (e.g., Rudin 1953, Theorem 4.19). Thus, under Assumption 1(a)(b), (d)-(e), (g) and (h), by the statement (iii) of the present lemma, which states that $\sup_{\theta\in\mathbf{\Theta}} |\tau_T(\theta) - \tau(\theta)| = o(1)$ $\mathbb{P}$-a.s. as $T \to \infty$, the second supremum is also $o(1)$ $\mathbb{P}$-a.s. as $T \to \infty$.

The second part of the result follows from the first part as in the proof of the statement (i) of the present lemma. □

Lemma 3. *Let* P *be any probability measure, and* $\mathbb{E}_{\mathrm{P}}$ *denote the expectation under* P*. Under Assumption 1 (a)(b)(d)(e)(g) and (h), if* $\mathbb{E}_{\mathrm{P}}[\sup_{(\theta,\tau)\in\mathbf{S}} \mathrm{e}^{\tau'\psi(X_1,\theta)}] < \infty$*, then* $0 < \inf_{(\theta,\tau)\in\mathbf{S}} \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$*, so that* $0 < \inf_{\theta\in\mathbf{\Theta}} \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]$*. Moreover,* $(\theta,\tau) \mapsto \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ *and* $\theta \mapsto \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]$ *are continuous in* $\mathbf{S}$ *and* $\mathbf{\Theta}$*, respectively. All of these results hold for* $\mathrm{P} = \mathbb{P}$ *under the aforementioned assumptions.*

Proof. Under Assumption 1 (a) and (b), the Lebesgue dominated convergence theorem and the lemma's assumption $\mathbb{E}_{\mathrm{P}}[\sup_{(\theta,\tau)\in\mathbf{S}} \mathrm{e}^{\tau'\psi(X_1,\theta)}] < \infty$ imply that $(\theta,\tau) \mapsto \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ is continuous. Moreover, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 4 (p. 23), $\mathbf{S}$ is compact, and continuous functions over compact sets reach a minimum (e.g., Rudin 1953, Theorem 4.16). Now, if there exist $(\dot{\tau},\dot{\theta}) \in \mathbf{S}$ s.t. $0 = \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\dot{\tau}'\psi(X_1,\dot{\theta})}]$, then $\mathrm{e}^{\dot{\tau}'\psi(X_1,\dot{\theta})} = 0$ P-a.s. (e.g., Kallenberg 2002 (1997, Lemma 1.24), which is impossible by definition of the exponential function. Thus, $0 < \inf_{(\theta,\tau)\in\mathbf{S}} \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$, so that $0 < \inf_{\theta\in\mathbf{\Theta}} \mathbb{E}_{\mathrm{P}}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]$ because by the definition of $\mathbf{S}$ in Assumption 1(e), for all $\theta \in \mathbf{\Theta}$, $(\theta, \tau(\theta)) \in \mathbf{S}$. Regarding the second part of the result, it immediately follows from the Lebesgue dominated convergence theorem, the lemma's assumption that $\mathbb{E}_{\mathrm{P}}[\sup_{(\theta,\tau)\in\mathbf{S}} \mathrm{e}^{\tau'\psi(X_1,\theta)}] < \infty$, and the continuity of $\tau : \mathbf{\Theta} \to \mathbf{R}^m$ by Lemma 10iii (p. 31) under Assumptions 1 (a)(b)(d)(e)(g) and (h). Regarding the third part of the result, it is sufficient to note that, under Assumption 1 (a)(b), by the Cauchy-Schwarz inequality, $\mathbb{E}[\sup_{(\theta,\tau)\in\mathbf{S}} \mathrm{e}^{\tau'\psi(X_1,\theta)}] \leqslant \mathbb{E}[\sup_{(\theta,\tau)\in\mathbf{S}} \mathrm{e}^{2\tau'\psi(X_1,\theta)}]^{1/2} < \infty$, where the last inequality follows from Assumption 1(e). □

Lemma 4 (Compactness of $\mathbf{S}$). *Under Assumptions 1 (a)(b)(d)(e)(g) and (h),*

(i) *The closure of the $\epsilon_{\mathbf{T}}$-neighborhood of $\tau(\mathbf{\Theta})$ (i.e., $\overline{\tau(\mathbf{\Theta})^{\epsilon_{\mathbf{T}}}}$) is compact*

(ii) *For all $\theta \in \mathbf{\Theta}$, the correspondence $\theta \mapsto \mathbf{T}(\theta)$ is nonempty compact-valued and uhc (upper hemi-continuous), and thus measurable;*

(iii) *The set $\mathbf{S} := \{(\theta, \tau) : \theta \in \mathbf{\Theta} \wedge \tau \in \mathbf{T}(\theta)\}$ is compact.*

Proof. (i) Under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 10iii (p. 31), $\tau : \mathbf{\Theta} \to \mathbf{R}^m$ is continuous. Moreover, by Assumption 1(d), $\mathbf{\Theta}$ is compact. Thus, $\tau(\mathbf{\Theta})$ is bounded — continuous mappings preserve compactness (e.g., Rudin 1953, Theorem 4.14). Consequently, $\tau(\mathbf{\Theta})^{\epsilon_{\mathbf{T}}} =: \{\tau \in \mathbf{R}^m : \inf_{\tilde{\tau}\in\tau(\mathbf{\Theta})} |\tau - \tilde{\tau}| < \epsilon_{\mathbf{T}}\}$ is bounded, which means that its closure $\overline{\tau(\mathbf{\Theta})^{\epsilon_{\mathbf{T}}}}$ is closed and bounded, i.e., compact.

(ii) Proof that $\mathbf{T}$ is nonempty and compact valued. By Assumption 1(d), for all $\theta \in \mathbf{\Theta}$, there exists $\tau(\theta)$ s.t. $\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1, \theta)] = 0$. Thus, for all $\theta \in \mathbf{\Theta}$, $\mathbf{T}(\theta) = \overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta))}$ is nonempty. Moreover, by construction, $\overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta))}$ is compact, so that it is nonempty compact valued.

Proof that $\mathbf{T}$ is uhc. Because $\mathbf{T}$ is compact valued, we can use the sequential characterization of upper hemicontinuity (e.g., Aliprantis and Border 2006/1999, Theorem 17.20). Let $((\theta_n, \tau_n))_{n\in\mathbf{N}} \in (\mathbf{S})^{\mathbf{N}}$ be a sequence s.t., for all $n \in \mathbf{N}$, $\tau_n \in \mathbf{T}(\theta_n)$ and $\theta_n \to \bar{\theta} \in \mathbf{\Theta}$ as $n \to \infty$. By construction, for all $n \in \mathbf{N}$, $\tau_n \in \overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta_n))} \subset \overline{\tau(\mathbf{\Theta})^{\epsilon_{\mathbf{T}}}}$. Moreover, by statement (i), $\overline{\tau(\mathbf{\Theta})^{\epsilon_{\mathbf{T}}}}$ is compact, so that there exists a subsequence $(\tau_{\alpha(n)})_{n\in\mathbf{N}}$ s.t. $\tau_{\alpha(n)} \to \bar{\tau} \in \overline{\tau(\mathbf{\Theta})^{\epsilon_{\mathbf{T}}}}$, as $n \to \infty$. Again, by construction, for all $n \in \mathbf{N}$, $(\theta_n, \tau_n) \in \mathbf{S}$, so that $|\tau_{\alpha(n)} - \tau(\theta_{\alpha(n)})| \leqslant \epsilon_{\mathbf{T}}$. Now, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 10iii, $\tau : \mathbf{\Theta} \to \mathbf{R}^m$ is continuous. Thus, $|\tau_{\alpha(n)} - \tau(\theta_{\alpha(n)})| \to |\bar{\tau} - \tau(\bar{\theta})|$ as $n \to \infty$. Thus, $|\bar{\tau} - \tau(\bar{\theta})| \leqslant \epsilon_{\mathbf{T}}$, which means that $\bar{\tau} \in \overline{B_{\epsilon_{\mathbf{T}}}(\tau(\bar{\theta}))} = \mathbf{T}(\bar{\theta})$.

Proof that $\mathbf{T}$ is measurable. Let F be a closed subset of $\mathbf{R}^m$. Then, its complement F^c is an open subset of $\mathbf{R}^m$. Now, a correspondence is uhc iff the upper inverse image of a open set is an open set (e.g., Aliprantis and Border 2006/1999, Lemma 17.4). Thus, by the previous paragraph, $\mathbf{T}^u(F^c) \in \mathcal{B}(\mathbf{\Theta})$, where $\mathbf{T}^u$ denotes the upper inverse of $\mathbf{T}$. Now, denoting the lower inverse of $\mathbf{T}$ with $\mathbf{T}^l$, notice that $\mathbf{T}^u(F^c) = [\mathbf{T}^l(F)]^c$ (e.g., Aliprantis and Border 2006/1999, p. 557), so that $[\mathbf{T}^l(F)]^c \in \mathcal{B}(\mathbf{\Theta})$, which, in turn implies that $\mathbf{T}^l(F) \in \mathcal{B}(\mathbf{\Theta})$ because of the stability of σ-algebras under complementation.

(iii) Note that the compactness of $\mathbf{\Theta}$ and $\mathbf{T}(\theta)$ are not sufficient to ensure the compactness of $\mathbf{S}$ because $\mathbf{S}$ is not a Cartesian product. By the statement (ii) of the present lemma, $\mathbf{T}$ is uhc and closed valued, so that it has a closed graph (e.g., Aliprantis and Border 2006/1999, Theorem 17.10), i.e., $\mathbf{S}$ is closed. Now, by construction, $\mathbf{S}$ is a subset $[\overline{\tau(\mathbf{\Theta})^{\epsilon_{\mathbf{T}}}} \times \mathbf{\Theta}]$, which is compact by statement (i) and Assumption 1(d). Thus, $\mathbf{S}$ is also compact —in metric spaces, closed subsets of compact sets are compact (e.g., Rudin 1953, Theorem 2.35). □

Lemma 5. *Under Assumptions 1 (a)(b)(d)(e) and (h),*

(i) *for any constant $\eta \in]0, \epsilon_{\mathbf{T}}]$, there exists a continuous value function $v : \mathbf{\Theta} \to \mathbf{R}_+$ s.t., for all $\theta \in \mathbf{\Theta}$, $v(\theta) = \inf_{\tau\in\mathbf{T}(\theta):|\tau-\tau(\theta)|\geqslant\eta} |\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}] - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|$;*

(ii) *for any constant $\eta \in]0, \epsilon_{\mathbf{T}}]$, $0 < \inf_{\theta\in\mathbf{\Theta}} \inf_{\tau\in\mathbf{T}(\theta):|\tau-\tau(\theta)|\geqslant\eta} |\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}] - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|$.*

Proof. *(i)* It is a consequence of Berge's maximum theorem (e.g., Aliprantis and Border 2006/1999, Theorem 17.31). Thus, it remains to check its assumptions. For the present proof, define the correspondence $\varphi : \mathbf{\Theta} \twoheadrightarrow \mathbf{R}^m$ s.t. $\varphi(\theta) = \{\tau \in \mathbf{T}(\theta) : |\tau - \tau(\theta)| \geqslant \eta\}$, and the function $f : \mathbf{S} \to \mathbf{R}_+$ s.t. $f(\theta, \tau) = |\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}] - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]|$.

Proof of the continuity of f. Under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22), $(\theta, \tau) \mapsto \mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ and $\theta \mapsto \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]$ are continuous in $\mathbf{S}$ and $\mathbf{\Theta}$, respectively, so that the continuity of f follows immediately.

Proof that φ is nonempty compact valued. By the definition of $\mathbf{T}$ in Assumption 1(e), for all $\theta \in \mathbf{\Theta}$, $\mathbf{T}(\theta) = \overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta))}$, so that, for any $\eta \in]0, \epsilon_{\mathbf{T}}]$, $\varphi(\theta) = \overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta))} \cap \{\tau \in \mathbf{R}^m : \eta \leqslant |\tau - \tau(\theta)|\} \neq \emptyset$, i.e., φ is nonempty valued. Moreover, for all $\theta \in \mathbf{\Theta}$, $\overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta))}$ is a compact set and $\{\tau \in \mathbf{R}^m : \eta \leqslant |\tau - \tau(\theta)|\}$ is a closed set, so that $\varphi(\theta)$, which is their intersection, is compact (e.g., Rudin 1953, Theorem 2.35 and the following Corollary).

Proof of the upper hemicontinuity of φ. Because φ is compact valued, we can use the sequential characterization of upper hemicontinuity (e.g., Aliprantis and Border 2006/1999, Theorem 17.20). Let $((\theta_n, \tau_n))_{n\in\mathbf{N}} \in \mathbf{S}^{\mathbf{N}}$ be a sequence s.t., for all $n \in \mathbf{N}$, $\tau_n \in \varphi(\theta_n)$ and $\theta_n \to \bar{\theta} \in \mathbf{\Theta}$ as $n \to \infty$. Now, under Assumptions 1 (a)(b)(d)(e)(g) and (h), Lemma 4iii (p. 23), $\mathbf{S}$ is a compact set, so that there exists a subsequence $((\theta_{\alpha(n)}, \tau_{\alpha(n)}))_{n\in\mathbf{N}}$ s.t. $(\theta_{\alpha(n)}, \tau_{\alpha(n)}) \to (\bar{\theta}, \bar{\tau}) \in \mathbf{S}$, as $n \to \infty$. The definition of $\mathbf{S}$ implies that $\bar{\tau} \in \mathbf{T}(\bar{\theta})$. Thus, it remains to show that $\eta \leqslant |\bar{\tau} - \tau(\bar{\theta})|$ in order to conclude that $\bar{\tau} \in \varphi(\bar{\theta})$. By construction, for all $n \in \mathbf{N}$, $\eta \leqslant |\tau_{\alpha(n)} - \tau(\theta_{\alpha(n)})|$. Moreover, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 10iii (p. 31), $\tau : \mathbf{\Theta} \to \mathbf{R}^m$ is continuous, so that $|\tau_{\alpha(n)} - \tau(\theta_{\alpha(n)})| \to |\bar{\tau} - \tau(\bar{\theta})|$ as $n \to \infty$, which means that $\eta \leqslant |\bar{\tau} - \tau(\bar{\theta})|$.

Proof of the lower hemicontinuity of φ. Use the sequential characterization of the lower hemicontinuity (e.g., Aliprantis and Border 2006/1999, Theorem 17.21). Let $(\theta_n)_{n\in\mathbf{N}} \in \mathbf{\Theta}^{\mathbf{N}}$ be a sequence s.t. $\theta_n \to \bar{\theta} \in \mathbf{\Theta}$ and $\bar{\tau} \in \varphi(\bar{\theta})$. Define the sequence $(\tau_n)_{n\in\mathbf{N}}$ s.t., for all $n \in \mathbf{N}$, $\tau_n = \tau(\theta_n) + \bar{\tau} - \tau(\bar{\theta})$. By definition of the correspondence φ, for all $n \in \mathbf{N}$, $|\tau_n - \tau(\theta_n)| = |\bar{\tau} - \tau(\bar{\theta})| \in [\eta, \epsilon_{\mathbf{T}}]$, which implies that $\tau_n \in \varphi(\theta_n)$. Moreover, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 10iii (p. 31), $\tau : \mathbf{\Theta} \to \mathbf{R}^m$ is continuous, so that $\lim_{n\to\infty} \tau_n = \lim_{n\to\infty} \tau(\theta_n) + \bar{\tau} - \tau(\bar{\theta}) = \tau(\bar{\theta}) + \bar{\tau} - \tau(\bar{\theta}) = \bar{\tau}$.

(ii) Under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 10 (p. 31), for all $\theta \in \mathbf{\Theta}$, $\tau(\theta)$ is the unique minimum of the strictly convex minimization problem $\inf_{\tau\in\mathbf{R}^m} \mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$. Thus, for all $\theta \in \mathbf{\Theta}$, $v(\theta) > 0$. Moreover, by Assumption 1(d), $\mathbf{\Theta}$ is compact, and by statement (i) of the present lemma, $v(.)$ is continuous. Thus, there exists $\varepsilon_v > 0$ s.t. $\min_{\theta\in\mathbf{\Theta}} v(\theta) > \varepsilon_v$ because a continuous function over a compact set reaches a minimum (e.g., Rudin 1953, Theorem 4.16). □

Lemma 6 (Asymptotic limit of the variance term). *Under Assumption 1(a)-(b) and (d)-(h),*

(i) $\mathbb{P}$-*a.s. for T big enough,* $0 < \inf_{\theta\in\mathbf{\Theta}} \left| \left| \left[\sum_{t=1}^{T} \hat{w}_{t,\theta} \frac{\partial \psi_t(\theta)'}{\partial \theta} \right] \right|_{\det} \right|$;

(ii) $\mathbb{P}$-*a.s. as $T \to \infty$,* $\sup_{\theta\in\mathbf{\Theta}} \left| \Sigma_T(\theta) - \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta) \right| = o(1)$

(iii) $\theta \mapsto \Sigma(\theta)$ *and* $\theta \mapsto \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)$ *are continuous in* $\mathbf{\Theta}$

(iv) $\mathbb{P}$-*a.s. for T big enough,* $\inf_{\theta\in\mathbf{\Theta}} |\Sigma_T(\theta)|_{\det} > 0$;

(v) $\mathbb{P}$-*a.s. as $T \to \infty$,* $\sup_{\theta\in\mathbf{\Theta}} \left| \ln |\Sigma_T(\theta)|_{\det} - \ln \left| \mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta) \right|_{\det} \right| = o(1)$, *so that, for all $\eta > 0$, $\mathbb{P}$-a.s. as $T \to \infty$,* $\sup_{\theta\in\mathbf{\Theta}} |\frac{1}{2T^\eta} \ln |\Sigma_T(\theta)|_{\det}| = o(1)$.

Proof. (i) Under Assumption 1(a)-(b) and (d)-(h), by Lemma 7 (p. 26), $\mathbb{P}$-a.s. as $T \to \infty$, $\sup_{\theta\in\boldsymbol{\Theta}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]-\frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right| = o(1)$, so that it is sufficient to check the invertibility of $\frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ for all $\theta\in\boldsymbol{\Theta}$ and the continuity of $\theta\mapsto\frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ (Lemma 30 on p. 87). Firstly, by Assumption 1(h), for all $\theta\in\boldsymbol{\Theta}$, $\Sigma(\theta):=\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ $\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]^{-1}$ is a positive-definite symmetric matrix, and thus $\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ is invertible. Moreover, under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22), and Assumption 1(e), for all $\theta\in\boldsymbol{\Theta}$, $0<\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]<\infty$, so that $\frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ is invertible for all $\theta\in\boldsymbol{\Theta}$. Secondly, under Assumption 1(a)-(b), (e)-(f), by Lemma 8i (p. 28), $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}|\right]<\infty$, so that the Lebesgue dominated convergence theorem and Assumption 1(b) imply the continuity of $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ in $\mathbf{S}$. Moreover, by definition in Assumption 1(e), for all $\theta\in\boldsymbol{\Theta}$, $(\tau(\theta),\theta)\in\mathbf{S}$, and under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 10iii (p. 31), $\tau:\boldsymbol{\Theta}\to\mathbf{R}^m$ is continuous. Thus, $\theta\mapsto\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta}'\right]$ is continuous. Then, the continuity of $\theta\mapsto\frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ follows from Lemma 3 (p. 22) under Assumption 1 (a)(b)(d)(e)(g) and (h).

(ii) On one hand, by definition, $\Sigma(\theta):=\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ $\left[\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]^{-1}$, which is symmetric positive definite by Assumption 1 (h), and $\Sigma_T(\theta):=$ $\left[\sum_{t=1}^{T}\hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\left[\sum_{t=1}^{T}\hat{w}_{t,\theta}\psi_t(\theta)\psi_t(\theta)'\right]\left[\sum_{t=1}^{T}\hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]^{-1}$, which is well defined $\mathbb{P}$-a.s. for T big enough by the statement (i) of the present lemma. On the other hand, under Assumption 1(a)-(b) and (d)-(h), by Lemma 7iii (p. 26), $\mathbb{P}$-a.s. as $T\to\infty$, $\sup_{\theta\in\boldsymbol{\Theta}}|\left[\frac{1}{T}\sum_{t=1}^{T}\hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]-\frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]|=o(1)$, and, under Assumptions 1(a)-(b), (d)-(e) and (g)-(h), by Lemma 8 (p. 28), $\mathbb{P}$-a.s. as $T\to\infty$, $\sup_{\theta\in\boldsymbol{\Theta}}|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]-\frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]|=o(1)$. Thus, the claim follows from the continuity of the inverse transformation (e.g., Rudin 1953, Theorem 9.8) and the limiting functions, and the compactness of $\boldsymbol{\Theta}$.

(iii) Under Assumption 1(a)-(b), (e)-(g), by Lemma 7i (p. 26) and 8 (p. 28), $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}|\right]<\infty$ and $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$, so that, by the Lebesgue dominated convergence theorem and Assumption 1(b), $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ and $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ are continuous in $\mathbf{S}$. Moreover, by definition in Assumption 1(e), for all $\theta\in\boldsymbol{\Theta}$, $(\theta,\tau(\theta))\in\mathbf{S}$, and under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 10iii (p. 31), $\tau:\boldsymbol{\Theta}\to\mathbf{R}^m$ is continuous. Thus $\theta\mapsto\Sigma(\theta)$ is continuous, which is the first result. Under Assumption 1 (a)(b)(d)(e)(g) and (h), the second result follows from Lemma 3 (p. 22), which states that $\theta\mapsto\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]$ is also continuous.

(iv) By construction, $\Sigma_T(\theta)$ is a symmetric positive semi-definite matrix (Lemma 29i on p. 86 with $\mathrm{P}=\frac{1}{T}\sum_{t=1}^{T}\delta_{X_t}$), so that $|\Sigma_T(\theta)|_{\det}\geqslant 0$. Thus, by the statement (ii) and (iii) of present lemma, it is sufficient to check the invertibility of $\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)$ for all $\theta\in\boldsymbol{\Theta}$ (Lemma 30 on p. 87). By Assumption 1 (h), for all $\theta\in\boldsymbol{\Theta}$,

$\Sigma(\theta) := \left[\mathbb{E}e^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[e^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]\left[\mathbb{E}e^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]^{-1}$ is a positive-definite symmetric matrix, and thus a fortiori invertible. Moreover, under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22), and Assumption 1(e), for all $\theta \in \mathbf{\Theta}$, $0 < \mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}] < \infty$, so that it is also invertible.

(v) Under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22) with $\mathrm{P} = \sum_{t=1}^{T}\delta_{X_t}$ and by the statement (iv) of the present lemma, $\mathbb{P}$-a.s. for T big enough, $\ln|\Sigma_T(\theta)|_{\det}$ is well-defined in $\mathbf{\Theta}$. Similarly, under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22) and Assumption 1 (h), $\ln\left|\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)\right|_{\det}$ is well-defined in $\mathbf{\Theta}$. Then, the first part of the result follows from the statement (ii) of the present lemma. Regarding the second part, by the triangle inequality, $\mathbb{P}$-a.s. as $T \to \infty$,

$$
\begin{aligned}
&\frac{1}{T^{\eta}}\sup_{\theta\in\mathbf{\Theta}}|\ln|\Sigma_T(\theta)|_{\det}| \\
\leqslant\ & \frac{1}{T^{\eta}}\sup_{\theta\in\mathbf{\Theta}}\left|\ln\left[|\Sigma_T(\theta)|_{\det}\right] - \ln\left[\left|\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)\right|_{\det}\right]\right| + \frac{1}{T^{\eta}}\sup_{\theta\in\mathbf{\Theta}}\left|\ln\left[\left|\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)\right|_{\det}\right]\right| \\
=\ & o(1)
\end{aligned}
$$

where the explanations of the last equality are as follows. Under Assumption 1(a)-(b) and (d)-(h), by the statement (iii) of the present lemma $\theta \mapsto \mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)$ is continuous in $\mathbf{\Theta}$, which is a compact set by Assumption 1(d). Now, continuous functions over compact sets are bounded (e.g., Rudin 1953, Theorem 4.16), so that $\sup_{\theta\in\mathbf{\Theta}}\left|\ln\left[\left|\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)\right|_{\det}\right]\right|$ is bounded, which, in turn, implies that $\frac{1}{T^{\eta}}\sup_{\theta\in\mathbf{\Theta}}\left|\ln\left[\left|\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]\Sigma(\theta)\right|_{\det}\right]\right| = o(1)$, as $T \to \infty$. Now the last equality follows from the statement (iv) of the present lemma. $\square$

Lemma 7. *Under Assumptions 1(a)-(b) and (e)-(f),*

(i) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}|\right] < \infty$;

(ii) *under additional Assumption 1(d)(g) and (h), $\mathbb{P}$-a.s. as $T \to \infty$,* $\sup_{(\theta,\tau)\in\mathbf{S}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}e^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right] - \mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right| = o(1)$, *so that* $\sup_{\theta\in\mathbf{\Theta}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}e^{\tau_T(\theta)'\psi_t(\theta)}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right] - \mathbb{E}\left[e^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right| = o(1)$*; and*

(iii) *under additional Assumption 1(d)(g) and (h), $\mathbb{P}$-a.s. as $T \to \infty$,* $\sup_{\theta\in\mathbf{\Theta}}\left|\left[\sum_{t=1}^{T}\hat{w}_{t,\theta}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right] - \frac{1}{\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[e^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right| = o(1)$

Proof. (i) The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}|\right] \\
\leqslant\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|e^{\tau'\psi(X_1,\theta)}|\sup_{(\theta,\tau)\in\mathbf{S}}\left|\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right|\right] \\
\overset{(a)}{\leqslant}\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|e^{\tau'\psi(X_1,\theta)}|^2\right]^{1/2}\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}}\left|\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right|^2\right]^{1/2} \\
\overset{(b)}{<}\ &\infty
\end{aligned} \tag{19}
$$

(a) Firstly, note that the expression in the second supremum does not depend on τ, so that $\sup_{(\theta,\tau)\in\mathbf{S}}\left|\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right| = \sup_{\theta\in\mathbf{\Theta}}\left|\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right|$. Secondly apply the Cauchy-Schwarz inequality. Finally, note that $[\sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|]^2 = \sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|^2$ and $[\sup_{\theta\in\mathbf{\Theta}}\left|\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right|]^2 = \sup_{\theta\in\mathbf{\Theta}}\left|\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right|^2$ because $x\mapsto x^2$ is increasing on $\mathbf{R}_+$. *(b)* Note that $|\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|^2 = \mathrm{e}^{2\tau(\theta)'\psi(X_1,\theta)}$, and then apply Assumption 1(e) to the first term. Then, application of Assumption 1(f) to the second term yields the result.

(ii) By the triangle inequality, as $T\to\infty$ $\mathbb{P}$-a.s.,

$$\begin{aligned}
&\sup_{\theta\in\mathbf{\Theta}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]-\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right|\\
&\leqslant \sup_{\theta\in\mathbf{\Theta}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]-\mathbb{E}\left[\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right|\\
&\qquad+\sup_{\theta\in\mathbf{\Theta}}\left|\mathbb{E}\left[\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]-\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right|\\
&= o(1)
\end{aligned}$$

where the explanations for the last equality are as follows. Regarding the first supremum, under Assumptions 1 (a)-(b)(d)(e)(g) and (h), by Lemma 4iii (p. 23), $\mathbf{S} := \{(\theta,\tau) : \theta\in\mathbf{\Theta}\wedge\tau\in\mathbf{T}(\theta)\}$ is a compact set, so that Assumptions 1(a)-(b), the statement (i) of the present lemma and the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3) imply that, $\mathbb{P}$-a.s. as $T\to\infty$,

$$\sup_{(\theta,\tau)\in\mathbf{S}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]-\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]\right| = o(1).$$

Now, by Assumption 1(e), for all $\theta\in\mathbf{\Theta}$, $\tau(\theta)\in\mathbf{T}(\theta)$, and under Assumption 1(a)(b), (d)-(e), (g) and (h), by Lemma 2ii (p. 20), $\mathbb{P}$-a.s. for T big enough, for all $\theta\in\mathbf{\Theta}$, $\tau_T(\theta)\in\mathbf{T}(\theta)$. Moreover, under Assumption 1(a)(b), (d)-(e), (g) and (h), by Lemma 2iii (p. 20), $\sup_{\theta\in\mathbf{\Theta}}|\tau_T(\theta)-\tau(\theta)| = o(1)$ $\mathbb{P}$-a.s. as $T\to\infty$. Thus, the first supremum is $o(1)$, i.e., $\sup_{\theta\in\mathbf{\Theta}}|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\frac{\partial\psi_t(\theta)'}{\partial\theta} - \mathbb{E}\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}| = o(1)$, as $T\to\infty$ $\mathbb{P}$-a.s. Regarding the second supremum, by Assumption 1(b), $(\theta,\tau)\mapsto\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}$ is continuous. Moreover under Assumptions 1(a)-(b), and (e)-(f), by the statement (i) of the present lemma, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}|\right] < \infty$. Thus, by the Lebesgue dominated convergence theorem and Assumption 1(b), $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ is also continuous in $\mathbf{S}$. Now, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 4iii (p. 23), $\mathbf{S}$ is compact, so that $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]$ is uniformly continuous in $\mathbf{S}$ —continuous functions on compact sets are uniformly continuous (e.g., Rudin 1953, Theorem 4.19). Thus, under Assumption 1(a)(b), (d)-(e), (g) and (h), by Lemma 2iii (p. 20), which states that $\sup_{\theta\in\mathbf{\Theta}}|\tau_T(\theta)-\tau(\theta)| = o(1)$ $\mathbb{P}$-a.s. as $T\to\infty$, the second supremum is also $o(1)$ $\mathbb{P}$-a.s. as $T\to\infty$.

(iii) Under Assumptions 1 (a)(b)(d)(e)(g) and (h), Lemma 3 (p. 22) yields $0 < \inf_{(\theta,\tau)\in\mathbf{S}}\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$ with $\mathrm{P} = \frac{1}{T}\sum_{t=1}^{T}\delta_{X_t}$, and $0 < \inf_{(\theta,\tau)\in\mathbf{S}}\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ with $\mathrm{P} = \mathbb{P}$. Consequently, under Assumption 1(a)(b), (d)-(f), (g) and (h), by Lemma 2iii and iv (p. 20) and

the statement (ii) of the present lemma, as $T \to \infty$, $\mathbb{P}$-a.s., uniformly w.r.t. θ

$$\sum_{t=1}^{T} \hat{w}_{t,\theta} \frac{\partial \psi_t(\theta)'}{\partial \theta} = \frac{1}{\frac{1}{T}\sum_{i=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_i(\theta)}} \frac{1}{T} \sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} \frac{\partial \psi_t(\theta)'}{\partial \theta}$$
$$\to \frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]} \mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)} \frac{\partial \psi(X_1,\theta)'}{\partial \theta}\right].$$

□

Lemma 8. *Under Assumptions 1(a)-(b), (e) and (g),*

(i) $\mathbb{E}\left[\sup_{(\theta,\tau)\in \mathbf{S}^\epsilon} |\mathrm{e}^{\tau'\psi(X_1,\theta)} \psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$

(ii) *under additional Assumption 1(d) and (h), $\mathbb{P}$-a.s. as $T \to \infty$,*
$\sup_{(\theta,\tau)\in \mathbf{S}} \left|\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)' - \mathbb{E}\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right| = o(1)$, *so that*
$\sup_{\theta\in\mathbf{\Theta}} \left|\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)' - \mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right| = o(1)$

(iii) *under additional Assumption 1(d)(f) and (h), $\mathbb{P}$-a.s. as $T \to \infty$,*
$\sup_{\theta\in\mathbf{\Theta}} \left|\sum_{t=1}^{T} \hat{w}_{t,\theta}\psi_t(\theta)\psi_t(\theta)' - \frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right| = o(1).$

Proof. The proof is the same as for Lemma 7 with $\psi(X_1,\theta)\psi(X_1,\theta)'$ and $\psi_t(\theta)\psi_t(\theta)'$ in lieu of $\frac{\partial\psi(X_1,\theta)'}{\partial\theta}$ and $\frac{\partial\psi_t(\theta)'}{\partial\theta}$, respectively. For completeness, we provide a proof.

(i) The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
\leqslant\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} |\mathrm{e}^{\tau'\psi(X_1,\theta)}| \sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} \left|\psi(X_1,\theta)\psi(X_1,\theta)'\right|\right] \\
\overset{(a)}{\leqslant}\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} |\mathrm{e}^{\tau'\psi(X_1,\theta)}|^2\right]^{1/2} \mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} \left|\psi(X_1,\theta)\psi(X_1,\theta)'\right|^2\right]^{1/2} \\
\overset{(b)}{<}\ &\infty.
\end{aligned}$$

(a) Firstly, for any $(\theta,\tau)\in\mathbf{S}^\epsilon$, $\theta\in\mathbf{\Theta}^\epsilon$ because, for all $(\tilde{\tau},\tilde{\theta})\in\mathbf{S}$, $|\theta-\tilde{\theta}| = \sqrt{\sum_{k=1}^{m}(\theta_k-\tilde{\theta}_k)^2} \leqslant \sqrt{\sum_{k=1}^{m}(\theta_k-\tilde{\theta}_k)^2 + \sum_{k=1}^{m}(\tau_k-\tilde{\tau}_k)^2} = |(\theta,\tau)-(\tilde{\tau},\tilde{\theta})| < \epsilon$. Thus, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'| \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|$. Secondly apply the Cauchy-Schwarz inequality. Finally, $[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} |\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|]^2 = \sup_{(\theta,\tau)\in\mathbf{S}^\epsilon} |\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|^2$ and $[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|]^2 = \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2$ because $x \mapsto x^2$ is increasing on $\mathbf{R}_+$. *(b)* Note that $|\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|^2 = \mathrm{e}^{2\tau(\theta)'\psi(X_1,\theta)}$, and then apply Assumption 1(e) to the first term. Then, application of Assumption 1 (g) to the second term yields the result.

(ii) By the triangle inequality, $\mathbb{P}$-a.s. as $T \to \infty$,

$$\begin{aligned}
&\sup_{\theta\in\boldsymbol{\Theta}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]-\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]\right|\\
&\leqslant \sup_{\theta\in\boldsymbol{\Theta}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]-\mathbb{E}\left[\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]\right|\\
&\qquad+\sup_{\theta\in\boldsymbol{\Theta}}\left|\mathbb{E}\left[\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]-\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]\right|\\
&= o(1)
\end{aligned}$$

where the explanations for the last equality are as follows. Regarding the first supremum, under Assumptions 1 (a)-(b)(d)(e)(g) and (h), by Lemma 4iii (p. 23), $\mathbf{S} := \{(\theta,\tau) : \theta \in \boldsymbol{\Theta} \wedge \tau \in \mathbf{T}(\theta)\}$ is a compact set, so that Assumption 1(a)-(b), the statement (i) of the present lemma and the ULLN (uniform law of large numbers) à la Wald yields that (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3), $\mathbb{P}$-a.s. as $T \to \infty$,

$$\sup_{(\theta,\tau)\in\mathbf{S}}\left|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]-\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]\right|=o(1).$$

Now, by Assumption 1(e), for all $\theta \in \boldsymbol{\Theta}$, $\tau(\theta) \in \mathbf{T}(\theta)$, and under Assumption 1(a)(b), (d)-(e), (g) and (h), by Lemma 2ii (p. 20), $\mathbb{P}$-a.s. for T big enough, for all $\theta \in \boldsymbol{\Theta}$, $\tau_T(\theta) \in \mathbf{T}(\theta)$. Moreover, under Assumption 1(a)(b), (d)-(e), (g) and (h), by Lemma 2iii (p. 20), $\sup_{\theta\in\boldsymbol{\Theta}}|\tau_T(\theta)-\tau(\theta)| = o(1)$ $\mathbb{P}$-a.s. as $T \to \infty$. Thus, the first supremum is $o(1)$, i.e., $\sup_{\theta\in\boldsymbol{\Theta}}|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)' - \mathbb{E}\mathrm{e}^{\tau_T(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'| = o(1)$, as $T \to \infty$ $\mathbb{P}$-a.s. Regarding the second supremum, by Assumption 1(b), $(\theta,\tau) \mapsto \mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'$ is continuous in $\mathbf{S}$. Moreover under Assumptions 1(a)-(b), (e) and (g), by the statement (i) of the present lemma, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$. Thus, by the Lebesgue dominated convergence theorem and Assumption 1(b), $(\theta,\tau) \mapsto \mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is also continuous. Now, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 4iii (p. 23), $\mathbf{S}$ is compact, so that $(\theta,\tau) \mapsto \mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is uniformly continuous —continuous functions on compact sets are uniformly continuous (e.g., Rudin 1953, Theorem 4.19). Thus, under Assumption 1(a)(b), (d)-(e), (g) and (h), by Lemma 2iii (p. 20), which states that $\sup_{\theta\in\boldsymbol{\Theta}}|\tau_T(\theta)-\tau(\theta)| = o(1)$ $\mathbb{P}$-a.s. as $T \to \infty$, the second supremum is also $o(1)$ $\mathbb{P}$-a.s. as $T \to \infty$.

(iii) Under Assumptions 1 (a)(b)(d)(e)(g) and (h), Lemma 3 (p. 22) yields $0 < \inf_{(\theta,\tau)\in\mathbf{S}}\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}$ with $\mathrm{P} = \frac{1}{T}\sum_{t=1}^{T}\delta_{X_t}$, and $0 < \inf_{(\theta,\tau)\in\mathbf{S}}\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$ with $\mathrm{P} = \mathbb{P}$. Consequently, under Assumption 1(a)(b), (d)-(e), (g) and (h), by Lemma 2iii and iv (p. 20) and the statement (ii) of the present lemma, as $T \to \infty$, $\mathbb{P}$-a.s., uniformly w.r.t. θ,

$$\begin{aligned}
\sum_{t=1}^{T}\hat{w}_{t,\theta}\psi_t(\theta)\psi_t(\theta)' &= \frac{1}{\frac{1}{T}\sum_{i=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_i(\theta)}}\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\\
&\to \frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right].
\end{aligned}$$

□

Lemma 9. *Under Assumptions 1(a)(b)(g),*

(i) $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^4\right]<\infty$, *so that* $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]<\infty$*; and*

(ii) *under additional Assumption 1(e),* $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)|\right]<\infty$

Proof. *(i)* Put $\psi(X_1,\theta)=:(\psi_1(X_1,\theta)\ \psi_2(X_1,\theta)\ \cdots\ \psi_m(X_1,\theta))'$. Note that $\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^4=\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]^2$ because $x\mapsto x^2$ is an increasing function. Thus, by the Cauchy-Schwarz inequality, $\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]\leqslant\sqrt{\mathbb{E}\left\{\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]^2\right\}}=\sqrt{\mathbb{E}\left\{\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^4\right\}}$, so that it remains to show the first part of the statement. On one hand, by the definition of the Euclidean norm,

$$
\begin{aligned}
\sqrt{\mathbb{E}\left\{\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^4\right\}} &= \sqrt{\mathbb{E}\left\{\sup_{\theta\in\mathbf{\Theta}^\epsilon}\left[\left(\sum_{k=1}^m\psi_k(X_1,\theta)^2\right)^2\right]\right\}}\\
&\leqslant \sqrt{m\mathbb{E}\left\{\sup_{\theta\in\mathbf{\Theta}^\epsilon}\left[\sum_{k=1}^m\psi_k(X_1,\theta)^4\right]\right\}} \qquad (20)
\end{aligned}
$$

where the explanation for the last inequality is as follows. By the Jensen's inequality, $\left(\frac{1}{m}\sum_{k=1}^m a_k\right)^2\leqslant\frac{1}{m}\sum_{k=1}^m a_k^2$, so that $\left(\sum_{k=1}^m a_k\right)^2\leqslant m\sum_{k=1}^m a_k^2$. Apply the later inequality with $\psi_k(X_1,\theta)^2=a_k$.

On the other hand,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}\left|\psi(X_1,\theta)\psi(X_1,\theta)'\right|^2\right]\\
&=\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}\left|\begin{pmatrix}\psi_1(X_1,\theta)^2 & \psi_1(X_1,\theta)\psi_2(X_1,\theta) & \cdots & \psi_1(X_1,\theta)\psi_m(X_1,\theta)\\ \psi_2(X_1,\theta)\psi_1(X_1,\theta) & \psi_2(X_1,\theta)^2 & \cdots & \psi_2(X_1,\theta)\psi_m(X_1,\theta)\\ \vdots & \vdots & \ddots & \vdots\\ \psi_m(X_1,\theta)\psi_1(X_1,\theta) & \psi_m(X_1,\theta)\psi_2(X_1,\theta) & \cdots & \psi_m(X_1,\theta)^2\end{pmatrix}\right|^2\right]\\
&=\mathbb{E}\left\{\sup_{\theta\in\mathbf{\Theta}^\epsilon}\left[\sum_{(i,j)\in[\![1,m]\!]^2}[\psi_i(X_1,\theta)\psi_j(X_1,\theta)]^2\right]\right\}\\
&=\mathbb{E}\left\{\sup_{\theta\in\mathbf{\Theta}^\epsilon}\left[\sum_{k=1}^m\psi_k(X_1,\theta)^4+\sum_{(i,j)\in[\![1,m]\!]^2:i\neq j}[\psi_i(X_1,\theta)\psi_j(X_1,\theta)]^2\right]\right\}
\end{aligned}
$$

Therefore, $\sum_{k=1}^m\psi_k(X_1,\theta)^4\leqslant\sum_{k=1}^m\psi_k(X_1,\theta)^4+\sum_{(i,j)\in[\![1,m]\!]^2:i\neq j}[\psi_i(X_1,\theta)\psi_j(X_1,\theta)]^2$, the later equality and inequality (20) yield

$$
\begin{aligned}
\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right] &\leqslant \sqrt{m\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]}\\
&<\infty
\end{aligned}
$$

where the last inequality follows from Assumption 1(g).

(ii) The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)|\right]\\
\leqslant\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau'\psi(X_1,\theta)}|\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\psi(X_1,\theta)|\right]\\
\overset{(a)}{\leqslant}\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau'\psi(X_1,\theta)}|^2\right]^{1/2}\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]^{1/2}\\
\overset{(b)}{<}\ &\infty
\end{aligned}\tag{21}$$

(a) Firstly, for any $(\theta,\tau)\in\mathbf{S}^\epsilon$, $\theta\in\mathbf{\Theta}^\epsilon$ because, for all $(\tilde{\tau},\tilde{\theta})\in\mathbf{S}$, $|\theta-\tilde{\theta}|=\sqrt{\sum_{k=1}^m(\theta_k-\tilde{\theta}_k)^2}\leqslant\sqrt{\sum_{k=1}^m(\theta_k-\tilde{\theta}_k)^2+\sum_{k=1}^m(\tau_k-\tilde{\tau}_k)^2}=|(\theta,\tau)-(\tilde{\tau},\tilde{\theta})|<\epsilon$. Thus, as the expression in the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\psi(X_1,\theta)\psi(X_1,\theta)'|\leqslant\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)\psi(X_1,\theta)'|$. Secondly apply the Cauchy-Schwarz inequality. Finally, note that $[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|]^2=\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|^2$ and $[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|]^2=\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2$ because $x\mapsto x^2$ is increasing on $\mathbf{R}_+$. *(b)* Note that $|\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}|^2=\mathrm{e}^{2\tau(\theta)'\psi(X_1,\theta)}$, and then apply Assumption 1(e) to the first term. Then, application of the statement (i) of the present lemma to the second term yields the result. □

Remark 1. The first step of the proof shows that even the fourth moment is uniformly bounded. ◇

Lemma 10 (Implicit function $\tau(.)$). *Under Assumption 1 (a)(b)(e)(g) and (h),*

(i) *for all $\theta\in\mathbf{\Theta}$, $\tau\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\right]$ is a strictly convex function s.t. $\frac{\partial\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\right]}{\partial\tau}=\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right]$;*

(ii) *under additional Assumption 1(d), for all $\theta\in\mathbf{\Theta}$, there exists a unique $\tau(\theta)$ such that $\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\right]=0$; and*

(iii) *under additional Assumption 1(d), $\tau:\mathbf{\Theta}\to\mathbf{R}^m$ is continuous; and*

(iv) *under additional Assumption 1(c) and (d), for all $\theta\in\mathbf{\Theta}\setminus\{\theta_0\}$, $\mathbb{E}\left[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}\right]<\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\right]=1$ where $\tau(\theta_0)=0_{m\times 1}$.*

Proof. *(i)* Under Assumption 1(a) and (b), by the Cauchy-Schwarz inequality, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}\mathrm{e}^{\tau'\psi(X_1,\theta)}\right]\leqslant\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}\mathrm{e}^{2\tau'\psi(X_1,\theta)}\right]^{1/2}$, which is finite by Assumption 1(e). Now, by Assumption 1(e), for all $\dot{\theta}\in\mathbf{\Theta}$, $\tau(\dot{\theta})\in\mathrm{int}[\mathbf{T}(\dot{\theta})]$. Then, by a standard result on Laplace's transform (e.g., Monfort (1980, Theorems 3 on p. 183), $\tau\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\dot{\theta})}\right]$ is C^∞ in a neighborhood of $\tau(\dot{\theta})$, and $\tau\mapsto\frac{\partial\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\dot{\theta})}\right]}{\partial\tau}=\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\dot{\theta})}\psi(X_1,\theta)\right]$ and $\tau\mapsto\frac{\partial^2\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\dot{\theta})}\right]}{\partial\tau\partial\tau'}=\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\dot{\theta})}\psi(X_1,\dot{\theta})\psi(X_1,\dot{\theta})'\right]$. Moreover, under Assumptions 1(a)-(b), (e) and (g), Assumption 1(h) implies that,

$\mathbb{E}\left[e^{\tau'\psi(X_1,\dot\theta)}\psi(X_1,\dot\theta)\psi(X_1,\dot\theta)'\right]$ is a symmetric positive-definite matrix because a well-defined covariance matrix is invertible iff it is invertible under an equivalent probability measure (Lemma 29 and Corollary 1i on p. 86).

(ii) Assumption 1(d) ensures existence, while the statement (i) of the present lemma ensures that $\tau(\theta)$ is the solution of a strictly convex problem, so that it is unique.

(iii) Note that, under our assumptions, an application of the standard implicit function (e.g., Rudin 1953, Theorem 9.28) is not directly possible as it requires $(\theta,\tau)\mapsto\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right]$ to be continuously differentiable in $\mathbf{S}^\epsilon$, which, in turn, typically requires to uniformly bound the derivative of the latter in $\mathbf{S}^\epsilon$ (e.g., Davidson 1994, Theorem 9.31). Thus, we apply the sufficiency part of Kumagai's implicit function theorem (Kumagai 1980). Check its assumptions. Firstly, under Assumptions 1(a)(b)(e) and (g), by Lemma 9ii (p. 30) and the Lebesgue dominated convergence theorem, $(\theta,\tau)\mapsto\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right]$ is continuous in $\mathbf{S}^\epsilon$, i.e., in an open neighborhood of every $(\theta,\tau)\in\mathbf{S}$. Secondly, by the inverse function theorem applied to $\tau\mapsto\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right]$ (e.g., Rudin 1953, Theorem 9.24), for all $\theta\in\mathbf{\Theta}^\epsilon$, $\tau\mapsto\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right]$ is locally one-to-one:[12] As explained in the proof of (i), under Assumption 1(a)(b)(e) and (h), $\tau\mapsto\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right]$ is continuously differentiable and, under Assumption 1(a)(b)(e)(g) and (h), for all $\theta\in\mathbf{\Theta}$, $\frac{\partial\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right]}{\partial\tau'}=\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is invertible, so that the assumptions of the inverse function theorem are valid.

(iv) By the statements (i) and (ii) of the present lemma, for all $\theta\in\mathbf{\Theta}$, for all $\tau\neq\tau(\theta)$, $\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]<\mathbb{E}[e^{\tau'\psi(X_1,\theta)}]$. Now, for all $\theta\in\mathbf{\Theta}\setminus\{\theta_0\}$, $\tau(\theta)\neq 0_{m\times1}$: If there existed $\dot\theta\in\mathbf{\Theta}\setminus\{\theta_0\}$ s.t. $\tau(\dot\theta)=0_{m\times1}$, then $0=\mathbb{E}[e^{\tau(\dot\theta)'\psi(X_1,\dot\theta)}\psi(X_1,\dot\theta)]=\mathbb{E}[\psi(X_1,\dot\theta)]$, which would contradict Assumption 1(c). Thus, for all $\theta\in\mathbf{\Theta}\setminus\{\theta_0\}$, $\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]<\mathbb{E}[e^{0_{1\times m}\psi(X_1,\theta)}]=1$. Then, the result follows by the statement (ii) of the present lemma because $0_{m\times1}=\mathbb{E}[\psi(X_1,\theta_0)]=\mathbb{E}[e^{0_{1\times m}\psi(X_1,\theta)}\psi(X_1,\theta_0)]$. □

B.2. **Decomposition and derivatives of the log-ESP $L_T(.,.)$.** In this section, we simplify $L_T(\theta,\tau)$ and study its derivatives. Such results are needed for the proof of Theorem 1ii and other results afterwards.

Lemma 11. *Under Assumption 1(a)-(e) and (g)(h), by Lemma 10 (p. 31), define $\tau(\theta_0)=\tau_0=0_{m\times1}$. Under Assumption 1(a)-(b), (e) and (h),*

(i) *under additional Assumption 1(d) and (g), there exist $(\underline{M}_e,\overline{M}_e)\in\mathbf{R}_+\setminus\{0\}$ s.t. $\mathbb{P}$-a.s. for T big enough, $\underline{M}_e<\inf_{(\theta,\tau)\in\mathbf{S}}\frac{1}{T}\sum_{t=1}^T e^{\tau'\psi_t(\theta)}$ and $\sup_{(\theta,\tau)\in\mathbf{S}}\frac{1}{T}\sum_{t=1}^T e^{\tau'\psi_t(\theta)}<\overline{M}_e$;*

(ii) *under additional Assumption 1 (c)(d) and (g), there exists an open ball $B_r(\theta_0,\tau_0)$ centered at (θ_0,τ_0) of radius $r>0$, which is a subset of $\mathbf{S}$;*

(iii) *under additional Assumption 1(c)(d)(f) and (g), for all (θ,τ) in a closed ball $\overline{B_{r_\partial}(\theta_0,\tau_0)}\subset\mathbf{S}$ centered at (θ_0,τ_0) with radius $r_\partial>0$, $|\mathbb{E}e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|^2_{\det}>0$, so that, $\mathbb{P}$-a.s. for T big enough, $|\frac{1}{T}\sum_{t=1}^T e^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}|^2_{\det}>0$;*

(iv) *under additional Assumption 1(g), $\mathbb{P}$-a.s. for T big enough,*

[12] Here it is necessary to work in an ϵ-neighborhood of $\mathbf{\Theta}$ in order to satisfy the assumption of Kumagai's implicit function theorem (Kumagai 1980). The standard implicit function theorem would also require the existence of open neighborhoods around the parameter values at which the function is zero.

$\inf_{(\theta,\tau)\in\mathbf{S}}|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'|_{\det} > 0.$

Proof. (i) Under Assumption 1(a)(b)(d)(e)(g) and (h), by Lemma 2i (p. 20), which states that, $\mathbb{P}$-a.s. as $T\to\infty$, $\sup_{(\theta,\tau)\in\mathbf{S}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}-\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]\right| = o(1)$, and Lemma 3 (p. 22) with $\mathrm{P}=\mathbb{P}$, which states that $0<\inf_{(\theta,\tau)\in\mathbf{S}}\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}]$, the result follows.

(ii) First of all, note that the result is not completely immediate, as $\mathbf{S} := \{(\theta,\tau):\theta\in\mathbf{\Theta}\wedge\tau\in\overline{B_{\epsilon_{\mathbf{T}}}(\tau(\theta))}\}$ is not a Cartesian product. Under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 10iii (p. 31), $\tau:\mathbf{\Theta}\to\mathbf{R}^m$ is continuous. Thus, by the topological definition of continuity, $\tau^{-1}[B_{\epsilon_{\mathbf{T}}/2}(\tau_0)]$ is an open set of $\mathbf{\Theta}$. Moreover, by the definition of τ_0, $\theta_0\in\tau^{-1}[B_{\epsilon_{\mathbf{T}}/2}(\tau_0)]$, and, by Assumption 1(c), $\theta_0\in\mathrm{int}(\mathbf{\Theta})$,[13] so that there exists $r_0>0$ s.t. $B_{r_0}(\theta_0)\subset\tau^{-1}[B_{\epsilon_{\mathbf{T}}/2}(\tau_0)]$ and $B_{r_0}(\theta_0)\subset\mathbf{\Theta}$. Now, for this proof, put $r=\min\{r_0,\epsilon_{\mathbf{T}}/2\}$. Then, it remains to show that $B_r(\theta_0,\tau_0)\subset\mathbf{S}$, i.e., for all $(\dot\theta,\dot\tau)\in B_r(\theta_0,\tau_0)$, $|\dot\tau-\tau(\dot\theta)|\leqslant\epsilon_{\mathbf{T}}$. By the triangle inequality, for any $(\dot\theta,\dot\tau)\in B_r(\theta_0,\tau_0)$,

$$\begin{aligned}|\dot\tau-\tau(\dot\theta)| &\leqslant |\dot\tau-\tau_0|+|\tau_0-\tau(\dot\theta)|\\ &\leqslant \frac{\epsilon_{\mathbf{T}}}{2}+\frac{\epsilon_{\mathbf{T}}}{2}=\epsilon_{\mathbf{T}}\end{aligned}$$

where the explanations for the last inequality are as follows. Firstly, $|\dot\tau-\tau_0|<\sqrt{\sum_{k=1}^{m}(\dot\tau_k-\tau_{0,k})^2}\leqslant\sqrt{\sum_{k=1}^{m}(\dot\theta_k-\theta_{0,k})^2+\sum_{k=1}^{m}(\dot\tau_k-\tau_{0,k})^2}<r\leqslant\frac{\epsilon_{\mathbf{T}}}{2}$ by definition of r. Secondly, and similarly, $|\dot\theta-\theta_0|<\sqrt{\sum_{k=1}^{m}(\dot\theta_k-\theta_{0,k})^2}\leqslant\sqrt{\sum_{k=1}^{m}(\dot\theta_k-\theta_{0,k})^2+\sum_{k=1}^{m}(\dot\tau_k-\tau_{0,k})^2}<r\leqslant r_0$, so that $|\tau_0-\tau(\dot\theta)|<\frac{\epsilon_{\mathbf{T}}}{2}$ because $B_{r_0}(\theta_0)\subset\tau^{-1}[B_{\epsilon_{\mathbf{T}}/2}(\tau_0)]$.

(iii) Under Assumption 1 (a)-(b) and (e)-(f), by Lemma 7i (p. 26), Assumption 1(b) and the Lebesgue dominated convergence theorem, $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]$ is continuous in $\mathbf{S}$, and thus in a neighborhood of (θ_0,τ_0) in $\mathbf{S}$ by Assumption 1(c) and (e). Then, $(\theta,\tau)\mapsto|\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]|_{\det}^2$ is also continuous. Now, by Assumption 1(h), $|\mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}]|_{\det}^2>0$, so that, under Assumption 1(a)-(e) and (g)-(h), by the statement (ii) of the present lemma, there exists a closed ball $\overline{B_{r_\partial}(\theta_0,\tau_0)}\subset\mathbf{S}$ centered at (θ_0,τ_0) with radius $r_\partial>0$, s.t., for all $(\theta,\tau)\in\overline{B_{r_\partial}(\theta_0,\tau_0)}$, $0<|\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]|_{\det}^2$, which is the first part of the result. By Lemma 30 (p. 87), the second part of the result follows from the continuity of $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]$, the invertibility of $\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]$ for all $(\theta,\tau)\in\overline{B_{r_\partial}(\theta_0,\tau_0)}$, and Lemma 7ii (p. 26), which, under Assumption 1(a)-(b) and (e)-(f), implies that
$\sup_{(\theta,\tau)\in\overline{B_{r_\partial}(\theta_0,\tau_0)}}|\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]-\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)'}{\partial\theta}\right]|=o(1)$, $\mathbb{P}$-a.s. as $T\to\infty$.

(iv) It follows from Lemma 30 (p. 87), so that it is sufficient to check its assumptions. Firstly, under Assumptions 1(a)-(b), (e), (g) and (h), by Corollary 1 (p. 86), for all $(\theta,\tau)\in\mathbf{S}$, $\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is a positive definite symmetric matrix, and thus it is invertible. Secondly, under Assumption 1(a)-(b), (e) and (g), by Lemma 8i (p. 28), $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$, so that by the Lebesgue dominated convergence theorem and Assumption 1(b), $(\theta,\tau)\mapsto\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is continuous in $\mathbf{S}$. Finally, under Assumptions 1(a)-(b), (d), (e), (g) and (h), $\mathbb{P}$-a.s. as $T\to\infty$,

[13]This assumption forbids θ_0 to be on the boundary of $\tau^{-1}[B_{\epsilon_{\mathbf{T}}/2}(\tau_0)]$, which is an open set of $\mathbf{\Theta}$, but not necessarily of $\mathbf{R}^m$.

$\sup_{(\theta,\tau)\in\mathbf{S}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'-\mathbb{E}[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)']\right|=o(1).$

□

In order to simplify the analysis of the asymptotic properties of the ESP estimator, we decompose the LogESP into three terms.

Lemma 12 (LogESP decomposition)**.** *Under Assumption 1, $\mathbb{P}$-a.s. for T big enough, define, for all $(\theta,\tau)\in\overline{B_{r_\partial}(\theta_0,\tau_0)}$,* $L_T(\theta,\tau):=\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right]$

$$-\frac{1}{2T}\ln\left\{\left|\left[\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\left[\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\psi_t(\theta)\psi_t(\theta)'\right]\left[\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]^{-1}\right|_{\det}\right\},$$

which exists by Lemma 11 (p. 32), and where $\overline{B_{r_\partial}(\theta_0,\tau_0)}$ is defined as in the aforementioned lemma. Then, under Assumption 1(a)(b) and (d)-(h), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau)\in\overline{B_{r_\partial}(\theta_0,\tau_0)}$,

$$L_T(\theta,\tau)=M_{1,T}(\theta,\tau)+M_{2,T}(\theta,\tau)+M_{3,T}(\theta,\tau)\ \textit{where}$$

$M_{1,T}(\theta,\tau):=\left(1-\frac{m}{2T}\right)\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right]$, $M_{2,T}(\theta,\tau):=\frac{1}{2T}\ln\left[\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right|_{\det}^{2}\right]$, *and*
$M_{3,T}(\theta,\tau):=-\frac{1}{2T}\ln\left[\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right|_{\det}\right]$.

Proof. First of all, note that, under Assumption 1, by Lemma 11 (p. 32), $L_T(.)$ is well-defined $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau)\in\overline{B_{r_\partial}(\theta_0,\tau_0)}$. Thus, under Assumption 1, $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau)\in\overline{B_{r_\partial}(\theta_0,\tau_0)}$.

$$\begin{aligned}
&L_T(\theta,\tau)\\
&=\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right]\\
&\quad-\frac{1}{2T}\ln\Bigg\{\Bigg|\left[\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\left[\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\psi_t(\theta)\psi_t(\theta)'\right]\\
&\qquad\qquad\times\left[\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\frac{\partial\psi_t(\theta)'}{\partial\theta}\right]^{-1}\Bigg|_{\det}\Bigg\}
\end{aligned}\tag{22}$$

$$\begin{aligned}
&\overset{(a)}{=}\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right]+\frac{1}{2T}\ln\left[\left|\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right|_{\det}^{2}\right]\\
&\quad-\frac{1}{2T}\ln\left[\left|\frac{1}{T}\sum_{t=1}^{T}\frac{\mathrm{e}^{\tau'\psi_t(\theta)}}{\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\psi_t(\theta)\psi_t(\theta)'\right|_{\det}\right]\\
&\overset{(b)}{=}\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right]\\
&\quad+\frac{1}{2T}\ln\left[\left(\frac{1}{\frac{1}{T}\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\right)^{2m}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right|_{\det}^{2}\right]\\
&\quad-\frac{1}{2T}\ln\left[\left(\frac{1}{\frac{1}{T}\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}}\right)^{m}\left|\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right|_{\det}\right]\\
&\overset{(c)}{=}\left(1-\frac{m}{2T}\right)\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right]+\frac{1}{2T}\ln\left[\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right|_{\det}^{2}\right]\\
&\quad-\frac{1}{2T}\ln\left[\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right|_{\det}\right].
\end{aligned}\tag{23}$$

(a) Firstly, use that the determinant of the product is the product of the determinants (e.g. Rudin 1953, Theorem 9.35). Secondly, the determinant of an inverse is the inverse of the determinant (e.g. Rudin 1953, p. 233). Finally, use basic properties of the logarithm, and note that we keep the square in the second logarithm in order to ensure the positivity of the argument (then the strict positivity is ensured by Lemma 11 on p. 32). *(b)* Use multilinearity of determinant. *(c)* Note that $1+\frac{-2m}{2T}-\frac{-m}{2T}=1-\frac{m}{2T}$. □

B.2.1. *Derivatives of* $M_{1,T}(\theta,\tau):=\left(1-\frac{m}{2T}\right)\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\right]$. *First derivative* $\frac{\partial M_{1,T}(\theta,\tau)}{\partial\theta_j}$. By Assumption 1(b), $\theta\mapsto\psi(X_1,\theta)$ is differentiable in $\boldsymbol{\Theta}$ $\mathbb{P}$-a.s. Thus, for all $(\theta,\tau)\in\mathbf{S}$, for all $j\in[\![1,m]\!]$,

$$\frac{\partial M_{1,T}(\theta,\tau)}{\partial\theta_j}=\left(1-\frac{m}{2T}\right)\frac{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}}{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}}.\tag{24}$$

Second derivative $\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\theta_\ell \partial\theta_j}$. By Assumption 2(a), $\theta \mapsto \psi(X_1,\theta)$ are three times continuously differentiable in a neighborhood of θ_0 $\mathbb{P}$-a.s. Thus, by equation (24) on p. 35, under Assumptions 1(a)-(e), (g)-(h) and 2(a), by Lemma 11ii (p. 32), $\mathbb{P}$-a.s., for all (θ,τ) in a neighborhood of (θ_0,τ_0), for all $(\ell,j) \in [\![1,m]\!]^2$,

$$
\begin{aligned}
&\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\theta_\ell \partial\theta_j} \\
&= \frac{(1-\frac{m}{2T})}{\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\right]^2} \left\{ \left\{ \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \left[\tau' \frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right] \left[\tau' \frac{\partial\psi_t(\theta)}{\partial\theta_j}\right] + \mathrm{e}^{\tau'\psi_t(\theta)} \left[\tau' \frac{\partial^2\psi_t(\theta)}{\partial\theta_j \partial\theta_\ell}\right] \right\} \right. \\
&\quad \times \left\{ \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \right\} - \left\{ \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \left[\tau' \frac{\partial\psi_t(\theta)}{\partial\theta_j}\right] \right\} \\
&\quad \left. \times \left\{ \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \left[\tau' \frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right] \right\} \right\} \\
&= \frac{(1-\frac{m}{2T})}{\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\right]} \left\{ \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \left[\tau' \frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right] \left[\tau' \frac{\partial\psi_t(\theta)}{\partial\theta_j}\right] + \mathrm{e}^{\tau'\psi_t(\theta)} \left[\tau' \frac{\partial^2\psi_t(\theta)}{\partial\theta_j \partial\theta_\ell}\right] \right\} \\
&\quad - \frac{(1-\frac{m}{2T})}{\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\right]^2} \left\{ \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \tau' \frac{\partial\psi_t(\theta)}{\partial\theta_j} \right\} \times \left\{ \frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \tau' \frac{\partial\psi_t(\theta)}{\partial\theta_\ell} \right\}.
\end{aligned}
\tag{25}
$$

Second derivative $\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\tau_k \partial\theta_j}$. Under Assumption 1(a)-(b), by equation (24) on p. 35, $\mathbb{P}$-a.s., for all $(\theta,\tau) \in \mathbf{S}$, for all $(k,j) \in [\![1,m]\!]^2$,

$$
\begin{aligned}
&\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\tau_k \partial\theta_j} \\
&= \left(1-\frac{m}{2T}\right) \frac{1}{\left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau'\psi_i(\theta)}\right]^2} \\
&\quad \times \left\{ \left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau'\psi_i(\theta)}\right] \frac{1}{T}\sum_{t=1}^T \left\{ \mathrm{e}^{\tau'\psi_t(\theta)} \tau' \frac{\partial\psi_t(\theta)}{\partial\theta_j} \psi_{t,k}(\theta) + \mathrm{e}^{\tau'\psi_t(\theta)} \frac{\partial\psi_{t,k}(\theta)}{\partial\theta_j} \right\} \right. \\
&\qquad \left. - \left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \tau' \frac{\partial\psi_t(\theta)}{\partial\theta_j}\right] \left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau'\psi_i(\theta)} \psi_{i,k}(\theta)\right] \right\}.
\end{aligned}
\tag{26}
$$

First derivative $\frac{\partial M_{1,T}(\theta,\tau)}{\partial\tau_k}$. By definition of $M_{1,T}(\theta,\tau)$ in Lemma 12 (p. 34), for all $(\theta,\tau) \in \mathbf{S}$, for all $k \in [\![1,m]\!]$,

$$
\frac{\partial M_{1,T}(\theta,\tau)}{\partial\tau_k} = \left(1-\frac{m}{2T}\right) \frac{\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)} \psi_{t,k}(\theta)}{\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau'\psi_i(\theta)}}. \tag{27}
$$

Second derivative $\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\tau_h\partial\tau_k}$. By the above equation (27), for all $(\theta,\tau)\in\mathbf{S}$, for all $(h,k)\in[\![1,m]\!]^2$,

$$\begin{aligned}
&\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\tau_h\partial\tau_k}\\
&=\left(1-\frac{m}{2T}\right)\frac{1}{\left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau'\psi_i(\theta)}\right]^2}\times\left\{\left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau'\psi_i(\theta)}\right]\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,h}(\theta)\psi_{t,k}(\theta)\right]\right.\\
&\qquad\left.-\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,h}(\theta)\right]\left[\frac{1}{T}\sum_{i=1}^T \mathrm{e}^{\tau'\psi_i(\theta)}\psi_{i,k}(\theta)\right]\right\}.
\end{aligned}\tag{28}$$

B.2.2. *Derivatives of* $M_{2,T}(\theta,\tau):=\frac{1}{2T}\ln\left[\left|\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right|_{\det}^2\right]$. *First derivative* $\frac{\partial M_{2,T}(\theta,\tau)}{\partial\theta_j}$. If $F(.)$ is a differentiable matrix function s.t. $|F(x)|_{\det}\neq 0$, then $D\ln[|F(x)|_{\det}^2]=2\mathrm{tr}[F(x)^{-1}DF(x)]$ (Lemma 32ii on p. 88) where $DF(x)$ denotes the derivative of $F(.)$ at x. Now, under Assumption 1, by Lemma 11iii (p. 32), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), $\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}$ is invertible. In addition, under Assumption 1(a), by Assumption 2(a), $\theta\mapsto\psi(X_1,\theta)$ is twice differentiable in a neighborhood of θ_0 $\mathbb{P}$-a.s., so that, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $(\theta,\tau)\mapsto\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}$ is also differentiable in a neighborhood of (θ_0,τ_0) $\mathbb{P}$-a.s. Thus, under Assumptions 1 and 2(a), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), for all $j\in[\![1,m]\!]$,

$$\begin{aligned}
\frac{\partial M_{2,T}(\theta,\tau)}{\partial\theta_j}&=\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\right.\\
&\times\left.\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial^2\psi_t(\theta)}{\partial\theta_j\partial\theta'}+\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]\right\}
\end{aligned}\tag{29}$$

Second derivative $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\theta_\ell\partial\theta_j}$. The trace of a derivative is the derivative of the trace because both the trace and derivative operators are linear (e.g., Magnus and Neudecker 1999/1988, chap. 9 sec. 9). Moreover, if $F(.)$ is a differentiable matrix function s.t., for all x in a neighborhood of $\dot{x}$, $|F(x)|_{\det}\neq 0$, then $D\left[F(\dot{x})^{-1}\right]=-F(\dot{x})^{-1}[DF(\dot{x})]F(\dot{x})^{-1}$ (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 4). Now, as explained for the first derivative, under Assumption 1, by Lemma 11iii (p. 32), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), $\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}$ is invertible. In addition, by Assumption 2(a), $\mathbb{P}$-a.s. $\theta\mapsto\psi(X_1,\theta)$ is three times continuously differentiable in a neighborhood of θ_0, so that, under Assumption 1 and 2(a), $\theta\mapsto\frac{\partial M_{2,T}(\theta,\tau)}{\partial\theta_j}$ is differentiable in a neighborhood of (θ_0,τ_0). Thus, under Assumptions 1 and 2(a), by the above equation (29), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of

(θ_0,τ_0), for all $(\ell,j)\in[\![1,m]\!]^2$,

$$\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\theta_\ell\partial\theta_j}=\frac{1}{T}\mathrm{tr}\Bigg\{-\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}$$
$$\left\{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial^2\psi_t(\theta)'}{\partial\theta_\ell\partial\theta'}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right)\frac{\partial\psi_t(\theta)}{\partial\theta'}\right\}$$
$$\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}$$
$$\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial^2\psi_t(\theta)}{\partial\theta_j\partial\theta'}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]$$
$$+\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}$$
$$\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial^3\psi_t(\theta)}{\partial\theta_\ell\partial\theta_j\partial\theta'}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right)\frac{\partial^2\psi_t(\theta)}{\partial\theta_j\partial\theta'}\right.$$
$$+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\frac{\partial^2\psi_t(\theta)}{\partial\theta_\ell\partial\theta'}$$
$$+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial^2\psi_t(\theta)}{\partial\theta_\ell\partial\theta_j}\right)\frac{\partial\psi_t(\theta)}{\partial\theta'}$$
$$\left.+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right)\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]\Bigg\}\tag{30}$$

Second derivative $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\tau_k\partial\theta_j}$. By a reasoning similar to the one for the derivative $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\theta_\ell\partial\theta_j}$, under Assumptions 1 and 2(a), by the above equation (29), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), for all $(k,j)\in[\![1,m]\!]^2$,

$$\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\tau_k\partial\theta_j}\tag{31}$$
$$=\frac{1}{T}\mathrm{tr}\Bigg\{-\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}$$
$$\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\frac{\partial\psi_t(\theta)}{\partial\theta'}\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}+\frac{\partial^2\psi_t(\theta)}{\partial\theta_j\partial\theta'}\right)\right]\Bigg\}$$
$$+\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\left(\frac{\partial\psi_t(\theta)}{\partial\theta'}\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}+\frac{\partial^2\psi_t(\theta)}{\partial\theta_j\partial\theta'}\right)\right]\right\}$$
$$+\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\frac{\partial\psi_t(\theta)}{\partial\theta'}\frac{\partial\psi_{t,k}(\theta)}{\partial\theta_j}\right)\right]\right\}.\tag{32}$$

First derivative $\frac{\partial M_{2,T}(\theta,\tau)}{\partial \tau_k}$. If $F(.)$ is a differentiable matrix function s.t. $|F(x)|_{\det} \neq 0$, then $D\ln[|F(x)|^2_{\det}] = 2\mathrm{tr}[F(x)^{-1}DF(x)]$ (Lemma 32ii on p. 88) where $DF(x)$ denotes the derivative of $F(.)$ at x. Now, under Assumption 1, by Lemma 11iii (p. 32), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial \psi_t(\theta)}{\partial \theta'}$ is invertible. Thus, under Assumption 1, by definition of $M_{2,T}(\theta,\tau)$ in Lemma 12 (p. 34), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), for all $k \in [\![1,m]\!]$,

$$\frac{\partial M_{2,T}(\theta,\tau)}{\partial \tau_k} = \frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]\right\}. \tag{33}$$

Second derivative $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial \tau_h \partial \tau_k}$. The trace of a derivative is the derivative of the trace because both the trace and derivative operators are linear (e.g., Magnus and Neudecker 1999/1988, chap. 9 sec. 9). Moreover, if $F(.)$ is a differentiable matrix function s.t., for all x in a neighborhood of $\dot{x}$, $|F(x)|_{\det} \neq 0$, then $D\left[F(\dot{x})^{-1}\right] = -F(\dot{x})^{-1}[DF(\dot{x})]F(\dot{x})^{-1}$ (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 4). Now, as explained for the first derivative, under Assumption 1, by Lemma 11iii (p. 32), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), $\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial \psi_t(\theta)}{\partial \theta'}$ is invertible. Thus, under Assumption 1, by the above equation (33), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), for all $(h,k) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial \tau_h \partial \tau_k}\\
&= -\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,h}(\theta)\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}\right.\\
&\qquad\left.\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]\right\}\\
&+\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\psi_{t,h}(\theta)\frac{\partial \psi_t(\theta)}{\partial \theta'}\right]\right\}.
\end{aligned} \tag{34}$$

B.2.3. *Derivatives of* $M_{3,T}(\theta,\tau) = -\frac{1}{2T}\ln\left[\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right|_{\det}\right]$. *First derivative* $\frac{\partial M_{3,T}(\theta,\tau)}{\partial \theta_j}$. If $F(.)$ is a differentiable matrix function s.t. $|F(x)|_{\det} > 0$, then $D\ln[|F(x)|_{\det}] = \mathrm{tr}[F(x)^{-1}DF(x)]$ (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 3). Now, under Assumption 1(a)-(b), (e) and (g)(h), by Lemma 11iv (p. 32), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau) \in \mathbf{S}$, $|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'|_{\det} > 0$. In addition, by Assumption 1(b), $\mathbb{P}$-a.s. $\theta \mapsto \psi(X_1,\theta)$ is continuously differentiable in $\boldsymbol{\Theta}$, so that, $\mathbb{P}$-a.s. for T big enough, $\theta \mapsto M_{3,T}(\theta,\tau)$ is differentiable in $\boldsymbol{\Theta}$, for all $(\theta,\tau) \in \mathbf{S}$. Thus, under Assumption 1(a)-(b) and (e)(g)(h), $\mathbb{P}$-a.s.

for T big enough, for all $(\theta,\tau)\in\mathbf{S}$, for all $j\in[\![1,m]\!]$,

$$\begin{aligned}\frac{\partial M_{3,T}(\theta,\tau)}{\partial\theta_j} = -\frac{1}{2T}\mathrm{tr}\Bigg\{&\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\\ &\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left\{\frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)'+\psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_j}\right\}\right.\\ &\left.+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\psi_t(\theta)\psi_t(\theta)'\right]\Bigg\}\end{aligned} \tag{35}$$

Second derivative $\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\theta_\ell\partial\theta_j}$. The trace of a derivative is the derivative of the trace because both the trace and differentiation operators are linear (e.g., Magnus and Neudecker 1999/1988, chap. 9 sec. 9). Moreover, if $F(.)$ is a differentiable matrix function s.t., for all x in a neighborhood of $\dot{x}$, $|F(x)|_{\det}\neq 0$, then $D\left[F(\dot{x})^{-1}\right] = -F(\dot{x})^{-1}[DF(\dot{x})]F(\dot{x})^{-1}$ (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 4). Now, under Assumption 1(a)-(b), (e) and (g)-(h), by Lemma 11iv (p. 32), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau)\in\mathbf{S}$, $\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'$ is invertible. In addition, by Assumption 2(a), $\mathbb{P}$-a.s. $\theta\mapsto\psi(X_1,\theta)$ is three times continuously differentiable in a neighborhood of θ_0, so that, under Assumption 1(a)-(e) and (g)(h), by Lemma 11ii (p. 32), $\theta\mapsto\frac{\partial M_{3,T}(\theta,\tau)}{\partial\theta_j}$ is differentiable in a neighborhood of (θ_0,τ_0). Thus, under Assumptions 1(a)(b), (e) and (g)(h), and 2(a), by the above equation (35), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), for all $(\ell,j)\in[\![1,m]\!]^2$,

$$
\begin{aligned}
\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\theta_\ell\partial\theta_j} = -\frac{1}{2T}\mathrm{tr}\Bigg\{ &- \left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1} \\
&\times\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\psi_t(\theta)'+\psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_\ell}\right)\right. \\
&\qquad \left.+\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right)\psi_t(\theta)\psi_t(\theta)'\right] \\
&\times\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1} \\
&\times\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)'+\psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_j}\right)\right. \\
&\qquad \left.+\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\psi_t(\theta)\psi_t(\theta)'\right] \\
&+\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1} \\
&\times\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\frac{\partial^2\psi_t(\theta)}{\partial\theta_\ell\partial\theta_j}\psi_t(\theta)'+\frac{\partial\psi_t(\theta)}{\partial\theta_j}\frac{\partial\psi_t(\theta)'}{\partial\theta_\ell}+\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\frac{\partial\psi_t(\theta)'}{\partial\theta_j}+\psi_t(\theta)\frac{\partial^2\psi_t(\theta)'}{\partial\theta_\ell\partial\theta_j}\right)\right. \\
&\qquad +\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right)\left(\frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)'+\psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_j}\right) \\
&\qquad +\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\left(\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\psi_t(\theta)'+\psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_\ell}\right) \\
&\qquad +\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial^2\psi_t(\theta)'}{\partial\theta_\ell\partial\theta_j}\right)\psi_t(\theta)\psi_t(\theta)' \\
&\qquad \left.+\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_\ell}\right)\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\right)\psi_t(\theta)\psi_t(\theta)'\right]\Bigg\}
\end{aligned}
\tag{36}
$$

Second derivative $\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\tau_k\partial\theta_j}$. Follow a reasoning similar to the one for the derivative $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\theta_\ell\partial\theta_j}$. The trace of a derivative is the derivative of the trace because both the trace and differentiation operators are linear (e.g., Magnus and Neudecker 1999/1988, chap. 9 sec. 9). Moreover, if $F(.)$ is a differentiable matrix function s.t., for all x in a neighborhood of $\dot{x}$, $|F(x)|_{\det}\neq 0$, then $D\left[F(\dot{x})^{-1}\right] = -F(\dot{x})^{-1}[DF(\dot{x})]F(\dot{x})^{-1}$ (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 4). Now, under Assumption 1(a)-(b), (e) and (g)-(h), by Lemma 11iv (p. 32), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau)\in\mathbf{S}$, $\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'$ is invertible. Thus, under Assumptions 1 (a)-(b), (e), (g)(h), by the above equation (35), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau)\in\mathbf{S}$, for all $(k,j) \in [\![1,m]\!]^2$

$$\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\tau_k \partial\theta_j}$$
$$= \frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\psi_t(\theta)\psi_t(\theta)'\right]\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\right.$$
$$\left.\times\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)\psi_t(\theta)' + \frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)' + \psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_j}\right)\right]\right\}$$
$$-\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\right.$$
$$\left.\times\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\left(\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)\psi_t(\theta)' + \frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)' + \psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_j}\right)\right]\right\}$$
$$-\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_{t,k}(\theta)}{\partial\theta_j}\psi_t(\theta)\psi_t(\theta)'\right]\right\}. \quad (37)$$

First derivative $\frac{\partial M_{3,T}(\theta,\tau)}{\partial\tau_k}$. If $F(.)$ is a differentiable matrix function s.t. $|F(x)|_{\det} > 0$, then $D\ln[|F(x)|_{\det}] = \mathrm{tr}[F(x)^{-1}DF(x)]$ (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 3). Now, under Assumption 1(a)-(b)(e)(g)(h), by Lemma 11iv (p. 32), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau) \in \mathbf{S}$, $|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'|_{\det} > 0$. Thus, under Assumption 1(a)-(b)(e)(g)(h), by definition of $M_{3,T}(\theta,\tau)$ in Lemma 12 (p. 34), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau) \in \mathbf{S}$, for all $k \in [\![1,m]\!]$,

$$\frac{\partial M_{3,T}(\theta,\tau)}{\partial\tau_k} = -\frac{1}{2T}\frac{1}{\left|\frac{1}{T}\sum_{i=1}^{T}\mathrm{e}^{\tau'\psi_i(\theta)}\psi_i(\theta)\psi_i(\theta)'\right|_{\det}}$$
$$\times\left|\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right|_{\det}$$
$$\times\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\psi_t(\theta)\psi_t(\theta)'\right]\right\}$$
$$= -\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\psi_t(\theta)\psi_t(\theta)'\right]\right\} \quad (38)$$

Second derivative $\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\tau_h\partial\tau_k}$. The trace of a derivative is the derivative of the trace because both the trace and differentiation operators are linear (e.g., Magnus and Neudecker 1999/1988, chap. 9 sec. 9). Moreover, if $F(.)$ is a differentiable matrix function s.t., for all x in a neighborhood of $\dot{x}$, $|F(x)|_{\det} \neq 0$, then $D\left[F(\dot{x})^{-1}\right] = -F(\dot{x})^{-1}[DF(\dot{x})]F(\dot{x})^{-1}$ (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 4). Now, under Assumption 1(a)-(b)(e)(g)-(h), by Lemma 11iv (p. 32), $\mathbb{P}$-a.s. for T big enough, for all $(\theta,\tau) \in \mathbf{S}$, $\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'$ is invertible. Thus, under Assumptions 1(a)-(b)(e)(g)(h), by the above equation (38), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0), for all $(h,k)\in[\![1,m]\!]^2$,

$$\begin{aligned}
&\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\tau_h\partial\tau_k}\\
&=\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\psi_t(\theta)\psi_t(\theta)'\right]\right.\\
&\qquad\left.\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,h}(\theta)\psi_t(\theta)\psi_t(\theta)'\right]\right\}\\
&\quad-\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_{t,k}(\theta)\psi_{t,h}(\theta)\psi_t(\theta)\psi_t(\theta)'\right]\right\}. \quad (39)
\end{aligned}$$

B.2.4. *Derivatives of $\theta\mapsto L_T(\theta,\tau)$. First derivative.* Under Assumption 1(a)-(e) and (g)-(h) and 2(a), by Lemma 11ii (p. 32), **S** contains an open neighborhood of (θ_0,τ_0), so that the derivatives derived in **S** also hold in a neighborhood of (θ_0,τ_0). Thus, by equations (24), (29) and (35) on pp. 35-40. Therefore, under Assumptions 1 and 2(a), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of (θ_0,τ_0),

$$\begin{aligned}
&\frac{\partial L_T(\theta,\tau)}{\partial\theta_j}\\
&=\left(1-\frac{m}{2T}\right)\frac{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}}{\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}}\\
&\quad+\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]^{-1}\right.\\
&\quad\left.\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial^2\psi_t(\theta)}{\partial\theta_j\partial\theta'}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\frac{\partial\psi_t(\theta)}{\partial\theta'}\right]\right\}\\
&\quad-\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\right.\\
&\quad\left.\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\left\{\frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)'+\psi_t(\theta)\frac{\partial\psi_t(\theta)'}{\partial\theta_j}\right\}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\tau'\frac{\partial\psi_t(\theta)}{\partial\theta_j}\psi_t(\theta)\psi_t(\theta)'\right]\right\}
\end{aligned}$$

Thus, evaluated at $(\theta_0,\tau(\theta_0))$,

$$\begin{aligned}
&\frac{\partial L_T(\theta_0,\tau_0)}{\partial\theta_j}\\
&=\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial\psi_t(\theta_0)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\frac{\partial^2\psi_t(\theta_0)}{\partial\theta_j\partial\theta'}\right]\right\}\\
&\quad-\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\psi_t(\theta_0)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\partial\psi_t(\theta_0)}{\partial\theta_j}\psi_t(\theta_0)'+\psi_t(\theta_0)\frac{\partial\psi_t(\theta_0)'}{\partial\theta_j}\right\}\right]\right\} \quad (40)
\end{aligned}$$

because $\tau(\theta_0) = 0_{m\times 1}$ by Lemma 10iv (p. 31).

B.3. **Proof of Theorem 1(ii): Asymptotic normality.** The proof of Theorem 1(ii) (i.e., asymptotic normality) adapts the traditional approach of expanding the FOCs (first order conditions). The two main differences w.r.t. the proofs in the entropy literature are the following. Firstly, instead of expanding the FOC $\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial\theta}\right|_{\theta=\hat\theta_T}$, we expand the approximate FOC $\left.\frac{\partial L_T(\theta,\tau)}{\partial\theta}\right|_{(\theta,\tau)=(\hat\theta_T,\tau_T(\hat\theta_T))}$ combined with the FOC (14) for τ on p. 17. Secondly, we need to control the asymptotic behaviour of the derivatives that come from $\ln|\Sigma_T(\theta)|_{\det}$.

Core of the proof of Theorem 1(ii). We prove asymptotic normality adapting the traditional approach of expanding the FOCs (first order conditions). Note that our approximate FOCs are written as a functionof the $2m$ variables θ and τ. In other words, instead of using the implicit function $\tau_T(\theta)$, τ is an estimated parameter and hence the ET equation (14) on p. 17 is also included in the expansion.

Under Assumptions 1 and 2, by Proposition 1 (p. 44), $\mathbb{P}$-a.s. as $T\to\infty$,

$$\begin{aligned}
\sqrt{T}\begin{bmatrix}\left(\hat\theta_T-\theta_0\right)\\ \tau_T(\hat\theta_T)\end{bmatrix} &= -\begin{bmatrix}\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\\ 0_{m\times m}\end{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)+o_{\mathbb{P}}(1)\\
&\underset{(a)}{\xrightarrow{D}} -\begin{bmatrix}\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\\ 0_{m\times m}\end{bmatrix}\mathcal{N}(0,\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right])\\
&\underset{(b)}{\overset{D}{=}}\mathcal{N}\left(0,\begin{bmatrix}\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\\ 0_{m\times m}\end{bmatrix}\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\begin{bmatrix}\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1} & 0_{m\times m}\end{bmatrix}\right)\\
&\overset{D}{=}\mathcal{N}\left(0,\begin{pmatrix}\Sigma(\theta_0) & 0_{m\times m}\\ 0_{m\times m} & 0_{m\times m}\end{pmatrix}\right) \qquad (41)
\end{aligned}$$

where $\Sigma(\theta_0) = \left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1}$. *(a)* Under Assumption 1(a)-(c) and (g), by the Lindeberg-Lévy CLT theorem,
$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\xrightarrow{D}\mathcal{N}(0,\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right])$, as $T\to\infty$. *(b)* Firstly, the minus sign can be discarded because of the symmetry of the Gaussian distribution. Secondly, if X is a random vector and F is a (deterministic) matrix, then $\mathbb{V}(FX) = F\mathbb{V}(X)F'$. □

Proposition 1 (Asymptotic expansion of $\sqrt{T}(\hat\theta_T-\theta_0)$). *Under Assumptions 1 and 2, $\mathbb{P}$-a.s. as $T\to\infty$,*

$$\sqrt{T}\begin{bmatrix}\left(\hat\theta_T-\theta_0\right)\\ \tau_T(\hat\theta_T)\end{bmatrix} = -\begin{bmatrix}\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\\ 0_{m\times m}\end{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)+o_{\mathbb{P}}(1)$$

Proof. The function $L_T(\theta,\tau)$ is well-defined and twice continuously differentiable in a neighborhood of $(\theta_0'\ \tau(\theta_0)')$ $\mathbb{P}$-a.s. for T big enough by subsection B.2 (p. 32), under Assumptions 1 and 2(a). Similarly, let $S_T(\theta,\tau) := \frac{1}{T}\sum_{t=1}^{T}e^{\tau'\psi_t(\theta)}\psi_t(\theta)$, which is continuously differentiable in a neighborhood of $(\theta_0'\ \tau(\theta_0)')$ by Assumption 1(a)(b). Now, under Assumption 1, by Theorem 1i (p. 6), Lemma 2iii (p. 20) and Lemma 10iv (p. 31), $\mathbb{P}$-a.s., $\hat\theta_T\to\theta_0$ and

$\tau_T(\hat\theta_T) \to \tau(\theta_0)$, where $\tau(\theta_0) = 0_{m\times 1}$, so that $\mathbb{P}$-a.s. for T big enough, $(\hat\theta'_T\ \tau_T(\hat\theta_T)')$ is in any arbitrary small neighborhood of $(\theta'_0\ \tau(\theta_0)')$. Therefore, under Assumption 1 and 2(a), a stochastic first-order Taylor-Lagrange expansion (Jennrich 1969, Lemma 3) around $(\theta_0, \tau(\theta_0))$ evaluated at $(\hat\theta_T, \tau_T(\hat\theta_T))$ yields, $\mathbb{P}$-a.s. for T big enough

$$\begin{bmatrix} \frac{\partial L_T(\hat\theta_T,\tau_T(\hat\theta_T))}{\partial\theta} \\ S_T(\hat\theta_T, \tau_T(\hat\theta_T)) \end{bmatrix} = \begin{bmatrix} \frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta} \\ S_T(\theta_0, \tau(\theta_0)) \end{bmatrix} + \begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} \end{bmatrix} \begin{bmatrix} \left(\hat\theta_T - \theta_0\right) \\ \tau_T(\hat\theta_T) \end{bmatrix} \tag{42}$$

where $\bar\theta_T$ and $\bar\tau_T$ are between $\hat\theta_T$ and θ_0, and between $\tau_T(\hat\theta_T)$ and $\tau(\theta_0)$, respectively. Under Assumptions 1 and 2, by Lemma 20 (p. 60) and by definition of $\tau_T(.)$ (equation 14 on p. 17), $\frac{\partial L_T(\hat\theta_T,\tau_T(\hat\theta_T))}{\partial\theta} = O(T^{-1})$ and $S_T(\hat\theta_T, \tau_T(\hat\theta_T)) = 0$, respectively. Moreover, under Assumptions 1 and 2, by Theorem 1i, Lemma 2iii (p. 20) and Lemma 13ii (p. 46), $\mathbb{P}$-a.s. for T big enough, $\begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} \end{bmatrix}$ is invertible. Thus, under Assumptions 1 and 2, $\mathbb{P}$-a.s. for T big enough,

$$\begin{aligned}
&\sqrt{T}\begin{bmatrix} \left(\hat\theta_T - \theta_0\right) \\ \tau_T(\hat\theta_T) \end{bmatrix} \\
=\ & -\begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} \end{bmatrix}^{-1} \sqrt{T}\begin{bmatrix} \frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta} + O(T^{-1}) \\ S_T(\theta_0, \tau(\theta_0)) \end{bmatrix} \\
\overset{(a)}{=}\ & -\begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} \end{bmatrix}^{-1} \begin{bmatrix} O(T^{-\frac{1}{2}}) \\ \sqrt{T}\frac{1}{T}\sum_{t=1}^T \psi_t(\theta_0) \end{bmatrix} \\
\overset{(b)}{=}\ & -\begin{bmatrix} -\Sigma(\theta_0) & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1} \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1} & 0_{m\times m} \end{bmatrix} \begin{bmatrix} O(T^{-\frac{1}{2}}) \\ \sqrt{T}\frac{1}{T}\sum_{t=1}^T \psi_t(\theta_0) \end{bmatrix} \\
& -\left\{\begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} \end{bmatrix}^{-1} - \begin{bmatrix} -\Sigma(\theta_0) & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1} \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1} & 0_{m\times m} \end{bmatrix}\right\} \begin{bmatrix} O(T^{-\frac{1}{2}}) \\ \sqrt{T}\frac{1}{T}\sum_{t=1}^T \psi_t(\theta_0) \end{bmatrix} \\
\overset{(c)}{=}\ & -\begin{bmatrix} \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1} \\ 0_{m\times m} \end{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_t(\theta_0) + o_{\mathbb{P}}(1)
\end{aligned}$$

where $\Sigma(\theta_0) = \left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1} \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right] \left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1}$. *(a)* Firstly, under Assumptions 1 and 2, by Lemma 14i (p. 47), $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j} = O(T^{-1})$, so that $\sqrt{T}\left[\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j} + O(T^{-1})\right] = O(T^{-\frac{1}{2}})$. Secondly, note that $S_T(\theta_0, \tau(\theta_0)) = \frac{1}{T}\sum_{t=1}^T \psi_t(\theta_0)$. *(b)* Add and subtract the matrix $\begin{bmatrix} -\Sigma(\theta_0) & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1} \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1} & 0_{m\times m} \end{bmatrix}$. *(c)* Firstly, the first column of the first square matrix cancels out because the first element of the vector is zero. Secondly, under Assumptions 1 and 2, by Lemma 13iii (p. 46) and Theorem 1i (p. 6), $\mathbb{P}$-a.s. as $T \to \infty$, the curly bracket is $o(1)$, and, under Assumption 1(a)-(c) and (g), by the Lindeberg-Lvy CLT, $\frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_t(\theta_0) = O_{\mathbb{P}}(1)$, as $T \to \infty$. □

Remark 2 (Alternative approximate FOC). In the proof of Theorem 1ii, it is possible to use the approximate FOC $\frac{\partial M_{1,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}=O(T^{-1})$ instead of the approximate FOC $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}=O(T^{-1})$. Under Assumption 1 and 2 (with $k_2\in[\![1,3]\!]$ and $j\in[\![0,2]\!]$ in its part b), by Lemma 12 (p. 34) and 18v-vii,xii-xiv (p. 50) and the ULLN à la Wald, $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}=\frac{\partial M_{1,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}+\frac{\partial M_{2,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}+\frac{\partial M_{3,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}=\frac{\partial M_{1,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}+O(T^{-1})$. The approximate FOC $\frac{\partial M_{1,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}=O(T^{-1})$ would lead to replace expansion (42) on p. 45 with the following expansion

$$\begin{bmatrix}\frac{\partial M_{1,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}\\ S_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))\end{bmatrix}=\begin{bmatrix}\frac{\partial M_{1,T}(\theta_0,\tau(\theta_0))}{\partial\theta}\\ S_T(\theta_0,\tau(\theta_0))\end{bmatrix}+\begin{bmatrix}\frac{\partial^2 M_{1,T}(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 M_{1,T}(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta}\\ \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'} & \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'}\end{bmatrix}\begin{bmatrix}\left(\hat{\theta}_T-\theta_0\right)\\ \tau_T(\hat{\theta}_T)\end{bmatrix}$$

where $\frac{\partial^2 M_{1,T}(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta}$ and $\frac{\partial^2 M_{1,T}(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta}$ can easily be controlled by Lemma 18i-iv (p. 50), Lemma 19i-v (p. 56), Lemma 23i-iii (p. 64) and ULLN à la Wald under Assumptions 1 and 2 (with $k_2\in[\![1,3]\!]$ and $j\in[\![0,2]\!]$ in its part b). The approximate FOC $\frac{\partial M_{1,T}(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}=O(T^{-1})$ requires less assumptions than the approximate FOC $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta}=O(T^{-1})$ because it does not require to control the 2nd derivatives of $M_{2,T}(\theta,\tau)$ and $M_{3,T}(\theta,\tau)$. However, it would not save space and it would require to add one more block of assumptions because our proof of Theorem 2 requires the full Assumption 2. ◇

Lemma 13. *Under Assumptions 1 and 2,*

(i) *for any sequence $(\theta_T,\tau_T)_{T\in\mathbf{N}}$ converging to $(\theta_0,\tau(\theta_0))$, $\mathbb{P}$-a.s. as $T\to\infty$,*
$$\begin{bmatrix}\frac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\tau'\partial\theta}\\ \frac{\partial S_T(\theta_T,\tau_T)}{\partial\theta'} & \frac{\partial S_T(\theta_T,\tau_T)}{\partial\tau'}\end{bmatrix}\to\begin{bmatrix}0_{m\times m} & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]'\\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right] & \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\end{bmatrix};$$

(ii) $\begin{bmatrix}0_{m\times m} & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]'\\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right] & \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\end{bmatrix}$ *is invertible, so that, for any sequence $(\theta_T,\tau_T)_{T\in\mathbf{N}}$ converging to $(\theta_0,\tau(\theta_0))$, $\mathbb{P}$-a.s., for T big enough, the matrix* $\begin{bmatrix}\frac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\tau'\partial\theta}\\ \frac{\partial S_T(\theta_T,\tau_T)}{\partial\theta'} & \frac{\partial S_T(\theta_T,\tau_T)}{\partial\tau'}\end{bmatrix}$ *is invertible; and*

(iii) *for any sequence $(\theta_T,\tau_T)_{T\in\mathbf{N}}$ converging to $(\theta_0,\tau(\theta_0))$, $\mathbb{P}$-a.s. as $T\to\infty$,*
$$\begin{bmatrix}\frac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\tau'\partial\theta}\\ \frac{\partial S_T(\theta_T,\tau_T)}{\partial\theta'} & \frac{\partial S_T(\theta_T,\tau_T)}{\partial\tau'}\end{bmatrix}^{-1}\to\begin{bmatrix}-\Sigma(\theta_0) & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1} & 0_{m\times m}\end{bmatrix},\text{ where}$$
$$\begin{bmatrix}-\Sigma(\theta_0) & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1} & 0_{m\times m}\end{bmatrix}=\begin{bmatrix}0_{m\times m} & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]'\\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right] & \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\end{bmatrix}^{-1}.$$

Proof. *(i)* Under Assumptions 1 and 2, it follows from Lemma 14ii and iii (p. 47) and Lemma 17 (p. 49), given that $\tau(\theta_0)=0_{m\times 1}$ by Lemma 10ii (p. 31) and Assumption 1(c), under Assumption 1(a)(b)(d)(e)(g) and (h).

(ii) Assumption 1(h) implies the invertibility of
$\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]=\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]$ and $\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]=\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]$ because $\tau(\theta_0)=0_{m\times 1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e)(g)-(h).

Thus, $\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]$ is also invertible, so that the first part of the statement (ii) follows from Lemma 33ii (p. 88) with $A = 0_{m\times m}$, $B = \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]$, $C = \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]$ and $D = \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]$. Then, the second part of the statement follows from a trivial case of the Lemma 30 (p. 87).

(iii) Under under Assumption 1(a)(b)(c)(d)(e)(g)(h), by the statement (ii) of the present lemma, the limiting matrix is invertible. Thus, by the inverse formula for partitioned matrices (e.g., Magnus and Neudecker 1999/1988, Chap. 1 Sec. 11),

$$\begin{bmatrix} 0_{m\times m} & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]' \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right] & \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\end{bmatrix}^{-1} = \begin{bmatrix} -\Sigma(\theta_0) & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1} \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta}\right]^{-1} & 0_{m\times 1}\end{bmatrix}$$

because $(-M'V^{-1}M)^{-1} = -M^{-1}V(M')^{-1} := -\Sigma(\theta_0)$. Then, the result follows from the continuity of the inverse transformation (e.g., Rudin 1953, Theorem 9.8). □

Lemma 14. *Under Assumptions 1 and 2,*

(i) *$\mathbb{P}$-a.s. as $T\to\infty$,*
$T\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j}\to \mathrm{tr}\left\{\left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\frac{\partial^2\psi_t(\theta_0)}{\partial\theta_j\partial\theta'}\right]\right\}-\frac{1}{2}\mathrm{tr}\Big\{\left[\mathbb{E}\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}$
$\times\left[\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta_j}\psi(X_1,\theta_0)'\right]+\mathbb{E}\left[\psi(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta_j}\right]\right]\Big\}$, *so that* $\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j}=O(T^{-1})$*;*

(ii) *for any sequence $(\theta_T,\tau_T)_{T\in\mathbf{N}}$ converging to $(\theta_0,\tau(\theta_0))$, $\left|\dfrac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\theta_j\partial\theta_\ell}\right| = o(1)$, $\mathbb{P}$-a.s. as $T\to\infty$;*

(iii) *for any sequence $(\theta_T,\tau_T)_{T\in\mathbf{N}}$ converging to $(\theta_0,\tau(\theta_0))$, $\left|\dfrac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\theta'\partial\tau}-\mathbb{E}\left[\dfrac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]\right| = o(1)$, $\mathbb{P}$-a.s. as $T\to\infty$.*

Proof. (i) By equation (40) on p. 43, under Assumptions 1 and 2(a), for all $j\in[\![1,m]\!]$, $\mathbb{P}$-a.s. for T big enough, evaluating $\frac{\partial L_T(\theta,\tau)}{\partial\theta_j}$ at $(\theta_0,\tau(\theta_0))$ yields

$$\begin{aligned}&\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j}\\ &=\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^T\frac{\partial\psi_t(\theta_0)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T\frac{\partial^2\psi_t(\theta_0)}{\partial\theta_j\partial\theta'}\right]\right\}\\ &\quad-\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^T\psi_t(\theta_0)\psi_t(\theta_0)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T\left\{\frac{\partial\psi_t(\theta_0)}{\partial\theta_j}\psi_t(\theta_0)'+\psi_t(\theta_0)\frac{\partial\psi_t(\theta_0)'}{\partial\theta_j}\right\}\right]\right\}. \quad (43)\end{aligned}$$

Now, under Assumption 1(a)(b),

- under additional Assumption 1(h), by the LLN and Lemma 30 (p. 87), $\mathbb{P}$-a.s. for T big enough, $\frac{1}{T}\sum_{t=1}^T\frac{\partial\psi_t(\theta_0)}{\partial\theta'}$ is invertible, so that $\left[\frac{1}{T}\sum_{t=1}^T\frac{\partial\psi_t(\theta_0)}{\partial\theta'}\right]^{-1}\to\left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}$;
- under additional Assumption 2(b), by the LLN, $\frac{1}{T}\sum_{t=1}^T\frac{\partial^2\psi_t(\theta_0)}{\partial\theta_j\partial\theta'}\to\mathbb{E}\left[\frac{\partial^2\psi_t(\theta_0)}{\partial\theta_j\partial\theta'}\right]$;

- under additional Assumption 1(h), by the LLN and Lemma 30 (p. 87), $\mathbb{P}$-a.s. for T big enough, $\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\psi_t(\theta_0)'$ is invertible, so that $\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta_0)\psi_t(\theta_0)'\right]^{-1} \to [\mathbb{E}\psi(X_1,\theta_0)\psi(X_1,\theta_0)']^{-1}$; and
- under additional Assumption 1(f)(g), by the Cauchy-Schwarz inequality and the monotonicity of integration,
$\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta_j}\psi(X_1,\theta_0)\right] \leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|^2\right]\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}}|\psi(X_1,\theta)|^2\right]} < \infty$, so that, by the LLN,
$\frac{1}{T}\sum_{t=1}^{T}\left\{\frac{\partial\psi_t(\theta_0)}{\partial\theta_j}\psi_t(\theta_0)'+\psi_t(\theta_0)\frac{\partial\psi_t(\theta_0)'}{\partial\theta_j}\right\} \to \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta_j}\psi(X_1,\theta_0)'\right]+\mathbb{E}\left[\psi(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta_j}\right]$ $\mathbb{P}$-a.s. as $T\to\infty$.

Thus, under Assumptions 1 and 2, for all $j\in[\![1,m]\!]$, $\mathbb{P}$-a.s. as $T\to\infty$, $T\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j} \to$ $\mathrm{tr}\left\{\left[\mathbb{E}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\frac{\partial^2\psi_t(\theta_0)}{\partial\theta_j\partial\theta'}\right]\right\} - \frac{1}{2}\mathrm{tr}\left\{[\mathbb{E}\psi(X_1,\theta_0)\psi(X_1,\theta_0)']^{-1}\left[\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta_j}\psi(X_1,\theta_0)'\right]\right.\right.$ $\left.\left.+\,\mathbb{E}\left[\psi(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)'}{\partial\theta_j}\right]\right]\right\}$, so that $\frac{\partial L_T(\theta_0,\tau_0)}{\partial\theta_j} = \frac{1}{T}O(1) = O(T^{-1})$.

(ii) Under Assumptions 1 and 2, by Lemma 15 (p. 48) and Lemma 12 (p. 34), $\mathbb{P}$-a.s. as $T\to\infty$, uniformly over a closed ball around $(\theta_0,\tau(\theta_0))$ with strictly positive radius, $\frac{\partial^2 L_T(\theta,\tau)}{\partial\theta_j\partial\theta_\ell} \to$ $\frac{1}{\left[\mathbb{E}e^{\tau'\psi(X_1,\theta)}\right]}\mathbb{E}\left\{e^{\tau'\psi(X_1,\theta)}\left[\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\right]\left[\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\right]+e^{\tau'\psi(X_1,\theta)}\left[\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta_\ell}\right]\right\}$ $-\frac{1}{\left[\mathbb{E}e^{\tau'\psi(X_1,\theta)}\right]^2}\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\right]\times\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\right]$. Now, under Assumption 1(a)(b)(d)(e)(g) and (h), by Lemma 10ii (p. 31) and Assumption 1(c), put $\tau(\theta_0)=0_{m\times1}$, so that the result follows.

(iii) Under Assumptions 1 and 2, by Lemma 16 (p. 49) and Lemma 12 (p. 34), $\mathbb{P}$-a.s. as $T\to\infty$, uniformly over a closed ball around $(\theta_0,\tau(\theta_0))$ with strictly positive radius, $\frac{\partial^2 L_T(\theta,\tau)}{\partial\tau_k\partial\theta_\ell} \to$ $\frac{1}{\left[\mathbb{E}e^{\tau'\psi(X_1,\theta)}\right]^2}\times\left\{\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\right]\mathbb{E}\left\{e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\psi_k(X_1,\theta)+e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_\ell}\right\}\right.$ $\left.-\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\right]\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\right]\right\}$. Now, under Assumption 1(a)(b)(d)(e)(g) and (h), by Lemma 10ii (p. 31) and Assumption 1(c), $\tau(\theta_0)=0_{m\times1}$, so that $\mathbb{P}$-a.s. as $T\to\infty$, $\frac{\partial^2 L_T(\theta_T,\tau_T)}{\partial\tau_k\partial\theta_\ell} \to \mathbb{E}\left[\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_\ell}\right]$. Stack the components together in order to obtain the result.

□

Lemma 15 (Uniform limit of $\frac{\partial^2 L_T(\theta,\tau)}{\partial\theta_j\partial\theta_\ell}$ in a neighborhood of $(\theta_0,\tau(\theta_0))$)**.** *Under Assumptions 1 and 2, for all $(j,\ell)\in[\![1,m]\!]^2$, $\mathbb{P}$-a.s. as $T\to\infty$, uniformly over a closed ball around $(\theta_0,\tau(\theta_0))$ with strictly positive radius,*

(i) $\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\theta_j\partial\theta_\ell} \to \frac{1}{\left[\mathbb{E}e^{\tau'\psi(X_1,\theta)}\right]}\mathbb{E}\left\{e^{\tau'\psi(X_1,\theta)}\left[\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\right]\left[\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\right]+e^{\tau'\psi(X_1,\theta)}\left[\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta_\ell}\right]\right\}$ $-\frac{1}{\left[\mathbb{E}e^{\tau'\psi(X_1,\theta)}\right]^2}\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\right]\times\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\right]$;

(ii) $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\theta_j\partial\theta_\ell} \to 0$;

(iii) $\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\theta_j\partial\theta_\ell} \to 0$.

Proof. (i) Under Assumptions 1 and 2, by Lemma 18i-iv (p. 50), Assumption 1(a) and (b), all the averages in $\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\theta_j\partial\theta_\ell}$ (equation (25) on p. 36) satisfy the assumptions of the ULLN à la Wald. Moreover, under Assumption 1(a)-(b) (d)(e)(g) and (h), by Lemma 11i (p. 32) the averages in the denominators are bounded away from zero. Thus, the result follows from the ULLN à la Wald. Note that the coefficient $\frac{m}{2T}$ vanishes as it goes to zero, as $T\to\infty$.

(ii) Under Assumptions 1 and 2, by Lemma 18v-xi (p. 50), Assumption 1(a) and (b), all the averages in $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\theta_j\partial\theta_\ell}$ (equation (30) on p. 38) satisfy the assumptions of the ULLN à la Wald. Moreover, under Assumption 1, by Lemma 11iii (p. 32) the averages in the inverted matrices are invertible in a neighborhood of $(\theta_0,\tau(\theta_0))$ $\mathbb{P}$-a.s. for T big enough. Thus, the result follows from the ULLN à la Wald, the linearity of the trace operator and the scaling by $\frac{1}{T}$.

(iii) Under Assumptions 1 and 2, by Lemma 18xii-xix (p. 50), Assumption 1(a) and (b), all the averages in $\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\theta_j\partial\theta_\ell}$ (equation (36) on p. 41) satisfy the assumptions of the ULLN à la Wald. Moreover, under Assumption 1(a)(b)(e)(g) and (h), by Lemma 11iv (p. 32) the averages in the inverted matrices are invertible in a neighborhood of $(\theta_0,\tau(\theta_0))$, $\mathbb{P}$-a.s. for T big enough. Thus, the result follows from the ULLN à la Wald, the linearity of the trace operator and the scaling by $\frac{1}{T}$. □

Lemma 16 (Uniform limit of $\frac{\partial^2 L_T(\theta,\tau)}{\partial\tau_k\partial\theta_l}$ in a neighborhood of $(\theta_0,\tau(\theta_0))$)**.** *Under Assumptions 1 and 2, for all $(k,\ell)\in[\![1,m]\!]^2$, $\mathbb{P}$-a.s. as $T\to\infty$, uniformly over a closed ball around $(\theta_0,\tau(\theta_0))$ with strictly positive radius,*

$$\text{(i)}\ \frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\tau_k\partial\theta_\ell}\to\frac{1}{\left[\mathbb{E}e^{\tau'\psi(X_1,\theta)}\right]^2}\times\Big\{\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\right]\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\psi_k(X_1,\theta)+e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_\ell}\right]$$
$$-\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\right]\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\right]\Big\};$$

$$\text{(ii)}\ \frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\tau_k\partial\theta_\ell}\to 0.$$

$$\text{(iii)}\ \frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\tau_k\partial\theta_\ell}\to 0.$$

Proof. The proof is similar to the one of Lemma 15 (p. 48). *(i)* Under Assumptions 1 and 2, by Lemma 19i-v (p. 56), Assumption 1(a) and (b), all the averages in $\frac{\partial^2 M_{1,T}(\theta,\tau)}{\partial\tau_k\partial\theta_\ell}$ (equation (26) on p. 36) satisfy the assumptions of the ULLN à la Wald. Moreover, under Assumption 1(a)-(b)(d)(e)(g) and (h), by Lemma 11i (p. 32) the averages in the denominators are bounded away from zero. Thus, the result follows from the ULLN à la Wald. Note that the coefficient $\frac{m}{2T}$ vanishes as it goes to zero, as $T\to\infty$.

(ii) Under Assumptions 1 and 2, by Lemma 19vi-xii (p. 56), Assumption 1(a) and (b), all the averages in $\frac{\partial^2 M_{2,T}(\theta,\tau)}{\partial\tau_k\partial\theta_\ell}$ (equation (32) on p. 38) satisfy the assumptions of the ULLN à la Wald. Moreover, under Assumption 1, by Lemma 11iii (p. 32) the averages in the inverted matrices are invertible in a neighborhood of $(\theta_0,\tau(\theta_0))$ $\mathbb{P}$-a.s. for T big enough. Thus, the result follows from the ULLN à la Wald, the linearity of the trace operator and the scaling by $\frac{1}{T}$.

(iii) Under Assumptions 1 and 2, by Lemma 19xiii-xix (p. 56), Assumption 1(a) and (b), all the averages in $\frac{\partial^2 M_{3,T}(\theta,\tau)}{\partial\tau_k\partial\theta_\ell}$ (equation (34)on p. 42) satisfy the assumptions of the ULLN à la Wald. Moreover, under Assumption 1(a)(b)(e)(g) and (h), by Lemma 11iv (p. 32) the averages in the inverted matrices are invertible in a neighborhood of $(\theta_0,\tau(\theta_0))$, $\mathbb{P}$-a.s. for T big enough. Thus, the result follows from the ULLN à la Wald, the linearity of the trace operator and the scaling by $\frac{1}{T}$. □

Lemma 17. *Put $S_T(\theta,\tau):=\frac{1}{T}\sum_{t=1}^T e^{\tau'\psi_t(\theta)}\psi_t(\theta)$. Under Assumptions 1 and 2, there exists a closed ball centered at $(\theta_0'\tau(\theta_0)')$ with strictly positive radius s.t., $\mathbb{P}$-a.s. as $T\to\infty$,*

$$\text{(i)}\ \sup_{(\theta,\tau)\in\overline{B_{r_L}((\theta_0,\tau_0))}}\left|\frac{\partial S_T(\theta,\tau)}{\partial\theta'}-\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]-\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]\right|=o(1);$$

$$\text{(ii)}\ \sup_{(\theta,\tau)\in\overline{B_{r_L}((\theta_0,\tau_0))}}\left|\frac{\partial S_T(\theta,\tau)}{\partial\tau'}-\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]\right|=o(1).$$

Proof. *(i)* By definition of $S_T(\theta,\tau)$,

$$\frac{\partial S(\theta,\tau)}{\partial\theta'}=\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\tau'\frac{\partial\psi_t(\theta)}{\partial\theta'}+\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}.$$

Thus, by the triangle inequality,

$$\begin{aligned}&\sup_{(\theta,\tau)\in\overline{B_{r_L}((\theta_0,\tau_0))}}\left|\frac{\partial S_T(\theta,\tau)}{\partial\theta'}-\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]-\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]\right|\\
\leqslant&\sup_{(\theta,\tau)\in\overline{B_{r_L}((\theta_0,\tau_0))}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\tau'\frac{\partial\psi_t(\theta)}{\partial\theta'}-\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]\right|\\
&+\sup_{(\theta,\tau)\in\overline{B_{r_L}((\theta_0,\tau_0))}}\left|\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\frac{\partial\psi_t(\theta)}{\partial\theta'}-\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]\right|\\
=&\ o(1)\ \mathbb{P}\text{-a.s. as } T\to\infty\end{aligned}$$

where the last equality follows from the ULLN à la Wald by Assumption 1(a)(b) and Lemma 18iv-v (p. 50), under Assumptions 1 and 2.

(ii) By definition of $S_T(\theta,\tau)$, $\frac{\partial S_T(\theta,\tau)}{\partial\tau'}=\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'$. Now, under Assumption 1(a)-(b)(e) and (g), by Lemma 8i (p. 28) and Assumption 1(a)(b), the assumptions of the ULLN à la Wald are satisfied, so that the result follows from the latter. □

Lemma 18 (Finiteness of the expectations of the supremum of the terms from $\frac{\partial^2 L_T(\theta,\tau)}{\partial\theta_\ell\partial\theta_j}$)**.** *Under Assumptions 1 and 2, there exists a closed ball $\overline{B_L}\subset\mathbf{S}$ centered at $(\theta_0,\tau(\theta_0))$ with strictly positive radius s.t., for all $(\ell,j)\in[\![1,m]\!]^2$,*

(i) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}\right]<\infty$;

(ii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]<\infty$;

(iii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta_\ell}|\right]<\infty$;

(iv) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}|\right]<\infty$;

(v) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]<\infty$;

(vi) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_l\partial\theta'}|\right]<\infty$;

(vii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]<\infty$;

(viii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^3\psi(X_1,\theta)}{\partial\theta_l\partial\theta_j\partial\theta'}|\right]<\infty$;

(ix) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}|\right]<\infty$;

(x) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta_j}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]<\infty$;

(xi) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]<\infty$;

(xii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$;

(xiii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\psi(X_1,\theta)'|\right]<\infty$;

(xiv) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$;

(xv) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_l\partial\theta_j}\psi(X_1,\theta)'|\right]<\infty$;

(xvi) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]<\infty$;

(xvii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}(\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell})\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)'|\right]<\infty$;

(xviii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}(\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta_j})\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$; *and*

(xix) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}(\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell})(\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j})\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$.

Proof. (i) Under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0,\tau(\theta_0))$, so that, by the Cauchy-Schwarz inequality, for $\overline{B_L}$ of sufficiently small radius, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}e^{\tau'\psi(X_1,\theta)}\right]\leqslant\sqrt{\mathbb{E}\left[(\sup_{(\theta,\tau)\in\overline{B_L}}e^{\tau'\psi(X_1,\theta)})^2\right]}=\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}}e^{2\tau'\psi(X_1,\theta)}\right]}<\infty$ where the equality follows from the fact that supremum of the square of a positive function is the square of the supremum of the function, and the last inequality from Assumption 1(e).

(ii) The norm of a product of matrices is smaller than the product of the norms (e.g., Rudin 1953, Theorem 9.7 and note that all norms are equivalent on finite dimensional spaces). Thus, for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j)\in[\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]\\
&\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}e^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]\\
&\overset{(a)}{\leqslant}(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}e^{\tau'\psi(X_1,\theta)}b(X_1)^2\right]\overset{(b)}{<}\infty.
\end{aligned}$$

(a) Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0,\tau(\theta_0))$ Thus, under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, by definition of $\mathbf{S}$, $\overline{B_L}\subset\{(\theta,\tau):\theta\in\mathcal{N}\wedge\tau\in\mathbf{T}(\theta)\}\subset\mathbf{S}$, because $\mathcal{N}\subset\mathbf{\Theta}$ by Assumption 2(a). Secondly, by Assumption 2(b), $\sup_{\theta\in\mathcal{N}}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}|\leqslant b(X)$ and $\sup_{\theta\in\mathcal{N}}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\leqslant b(X)$. *(b)* Firstly, $\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2<\infty$ because $\overline{B_L}$ is bounded. Secondly, by Assumption 2(b), $\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}e^{\tau'\psi(X_1,\theta)}b(X_1)^2\right]<\infty$.

(iii) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j)\in[\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta_\ell}|\right]\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)$ $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}e^{\tau'\psi(X_1,\theta)}|\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta_\ell}|\right]\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}e^{\tau'\psi(X_1,\theta)}b(X_1)\right]<\infty$, where the two last inequalities follow from Assumption 2(b) and the boundedness of $\overline{B_L}$.

(iv) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $\ell\in[\![1,m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}|\right]\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)$ $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}e^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}|\right]\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}e^{\tau'\psi(X_1,\theta)}b(X_1)\right]<\infty$, where the two last inequalities follow from Assumption 2(b) and the boundedness of $\overline{B_L}$.

(v) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]\leqslant\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}e^{\tau'\psi(X_1,\theta)}b(X_1)\right]<\infty$, where the last inequality follows from Assumption 2(b).

(vi) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $\ell\in[\![1,m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|e^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta'}|\right]$

$\leqslant \mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)\right] < \infty$, where the last inequality follows from Assumption 2(b).

(vii) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $\ell \in [\![1,m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] < (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] \leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)$
$\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^2\right] < \infty$ where the two last inequalities follow from Assumption 2(b) and the boundedness of $\overline{B_L}$.

(viii) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j) \in [\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^3\psi(X_1,\theta)}{\partial\theta_l\partial\theta_j\partial\theta'}|\right] \leqslant \mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)\right] < \infty$, where the last inequality follows from Assumption 2(b).

(ix) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j) \in [\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}|\right] < (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}|\right] \leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)^2$
$\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^2\right] < \infty$ where the two last inequalities follow from Assumption 2(b) and the boundedness of $\overline{B_L}$.

(x) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j) \in [\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta_j}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] < (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta_j}||\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]$
$\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^2\right] < \infty$ where the two last inequalities follow from Assumption 2(b) and the boundedness of $\overline{B_L}$.

(xi) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j) \in [\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] < (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}||\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] \leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)$
$\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^3\right] < \infty$ where the two last inequalities follow from Assumption 2(b) and the boundedness of $\overline{B_L}$.

(xii) Under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0,\tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius,
$\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$
where the last inequality follows from Lemma 8i (p. 28) under Assumption 1(a)-(b)(e)(g).

(xiii) The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small

radius, for all $\ell \in [\![1, m]\!]$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\psi(X_1,\theta)'|\right] \\
\leqslant\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)'|\right] \\
\overset{(a)}{\leqslant}\ &\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)'|^2\right]} \\
\overset{(b)}{\leqslant}\ &\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)'|^2\right]} \overset{(c)}{<} \infty.
\end{aligned}$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0, \tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$ because $\mathcal{N} \subset \mathbf{\Theta}$ and $\mathbf{S} = \{(\theta,\tau) : \theta \in \mathbf{\Theta} \wedge \tau \in \mathbf{T}(\theta)\}$. Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)'|^2$ because $\overline{B_L} \subset \mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* By Assumption 2(b), the first expectation is bounded. Under Assumption 1(a)(b)(g), by Lemma 9i (p. 30), the second expectation is also bounded.

(xiv) Proof similar to the one of statement (xiii). The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $\ell \in [\![1, m]\!]$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
\leqslant\ &(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
\overset{(a)}{\leqslant}\ &(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \\
\overset{(b)}{\leqslant}\ &(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \overset{(c)}{<} \infty.
\end{aligned}$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0, \tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$. Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2$ because $\overline{B_L} \subset \mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* Firstly, because $\overline{B_L}$ is bounded, $(\sup_{(\theta,\tau)\in\overline{B_L}} |\tau|) < \infty$. Secondly, by Assumption 2(b), the first expectation is bounded. Thirdly, by Assumption 1(g), the second expectation is also bounded.

(xv) The proof is the same as for statement (xiii) with $\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_l\partial\theta_j}$ instead of $\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}$. The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j)\in[\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_l\partial\theta_j}\psi(X_1,\theta)'|\right]\\
&\leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_l\partial\theta_j}|\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)'|\right]\\
&\overset{(a)}{\leqslant}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_l\partial\theta_j}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)'|^2\right]}\\
&\overset{(b)}{\leqslant}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)'|^2\right]}\overset{(c)}{<}\infty.
\end{aligned}$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0,\tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L}\subset\{(\theta,\tau):\theta\in\mathcal{N}\wedge\tau\in\mathbf{T}(\theta)\}\subset\mathbf{S}\subset\mathbf{S}^\epsilon$. Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)'|^2\leqslant\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)'|^2$ because $\overline{B_L}\subset\mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* By Assumption 2(b), the first expectation is bounded. Under Assumption 1(a)(b)(g), by Lemma 9i (p. 30), the second expectation is also bounded.

(xvi) Similarly to the proof of statement (ii), under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j)\in[\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]<\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]\leqslant\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^2\right]<\infty$ where the last inequality follows from Assumption 2(b).

(xvii) Proof similar to the one of statement (xiii). The norm of a product of matrices is smaller than the product of the norms (e.g., Rudin 1953, Theorem 9.7 and note that all norms are equivalent on finite dimensional spaces). Moreover, the supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under

Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(\ell, j) \in [\![1, m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}(\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell})\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)'|\right] < \infty \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)|\right] \\
&\overset{(a)}{\leqslant} (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)|^2\right]} \\
&\overset{(b)}{\leqslant} (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^4\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]} \overset{(c)}{<} \infty.
\end{aligned}$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0, \tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$. Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)'|^2$ because $\overline{B_L} \subset \mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* Firstly, because $\overline{B_L}$ is bounded, $(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|) < \infty$. Secondly, by Assumption 2(b), the first expectation is bounded. Thirdly, under Assumption 1(a)(b)(g), by Lemma 9 (p. 30), the second expectation is also bounded.

(xviii) Proof similar to the one of statement (xiii). The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(j, \ell) \in [\![1, m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}(\tau'\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta_j})\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta_j}|\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\overset{(a)}{\leqslant} (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_\ell\partial\theta_j}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \\
&\overset{(b)}{\leqslant} (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \overset{(c)}{<} \infty.
\end{aligned}$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0, \tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$. Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)'|^2$ because $\overline{B_L} \subset \mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* Firstly, because $\overline{B_L}$ is bounded, $(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|) < \infty$. Secondly, by Assumption 2(b), the first expectation is bounded. Thirdly, by Assumption 1(g), the second expectation is also bounded.

(xix) Proof similar to the one of statement (xiii). The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(j,\ell) \in [\![1,m]\!]^2$,

$$\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}(\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell})(\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j})\psi(X_1,\theta)\psi(X_1,\theta)'|\right]$$

$$\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|\right]$$

$$\overset{(a)}{\leqslant} (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]}$$

$$\overset{(b)}{\leqslant} (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^4\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]}\overset{(c)}{<}\infty.$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0, \tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$. Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2$ because $\overline{B_L}\subset\mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* Firstly, because $\overline{B_L}$ is bounded, $(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|^2) < \infty$. Secondly, by Assumption 2(b), the first expectation is bounded. Thirdly, by Assumption 1(g), the second expectation is also bounded. ☐

Lemma 19 (Finiteness of the expectations of the supremum of the terms from $\frac{\partial^2 L_T(\theta,\tau)}{\partial\tau_k\partial\theta_j}$). *Under Assumptions 1 and 2, there exists a closed ball $\overline{B_L}$ centered at $(\theta_0,\tau(\theta_0))$ with strictly positive radius s.t., for all $(k,j)\in[\![1,m]\!]^2$,*

(i) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}\right]<\infty;$

(ii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}\psi_k(X_1,\theta)|\right]<\infty;$

(iii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_\ell}|\right]<\infty;$

(iv) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_\ell}|\right]<\infty;$

(v) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)|\right]<\infty;$

(vi) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]<\infty;$

(vii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]<\infty;$

(viii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]<\infty;$

(ix) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}|\right]<\infty;$

(x) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right]<\infty;$

(xi) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}|\right]<\infty;$

(xii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_j}|\right]<\infty;$

(xiii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty;$

(xiv) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$;

(xv) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$;

(xvi) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)'|\right]<\infty$;

(xvii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$;

(xviii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)'|\right]<\infty$; *and*

(xix) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)\psi(X_1,\theta)'|\right]<\infty$.

Proof. The proofs are similar to the ones of Lemma 18 (p. 50): We only use more often the inequality that states that the norm of a component of a vector is smaller than the norm of the vector (e.g., $|\psi_k(X_1,\theta)|\leqslant\sqrt{\sum_{l=1}^m\psi_l(X_1,\theta)^2}=|\psi(X_1,\theta)|$). Thus, we only provide proof sketches.

(i) See Lemma 18i p. 50.

(ii) For $\overline{B_L}$ of sufficiently small radius, for all $(k,j)\in[\![1,m]\!]^2$,

$$\begin{aligned}&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi_k(X_1,\theta)|\right]\\ &\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\sup_{(\theta,\tau)\in\overline{B_L}}|\psi_k(X_1,\theta)'|\right]\\ &\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi_k(X_1,\theta)'|^2\right]}\\ &\leqslant(\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)'|^2\right]}<\infty,\end{aligned}$$

where the last inequality follows from Assumption 2(b), and Lemma 9i (p. 30), under Assumption 1(a)(b)(g).

(iii) For $\overline{B_L}$ of sufficiently small radius, for all $(k,j)\in[\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_j}|\right]\leqslant\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi_k(X_1,\theta)}{\partial\theta'}|\right]\leqslant\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)\right]<\infty$, where the last inequality follows from Assumption 2(b).

(iv) See Lemma 18iv p. 50.

(v) For $\overline{B_L}$ of sufficiently small radius, for all $k\in[\![1,m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)|\right]\leqslant\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)|\right]\leqslant\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}^\epsilon}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)|\right]<\infty$ where the last inequality follows from Lemma 9ii (p. 30) under Assumption 1(a)(b)(e)(g).

(vi) See Lemma 18v p. 50.

(vii) For $\overline{B_L}$ of sufficiently small radius, for all $k \in [\![1,m]\!]$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]\\
&\leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi_k(X_1,\theta)|\right]\\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi_k(X_1,\theta)|^2\right]}\\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)|^2\right]} < \infty,
\end{aligned}$$

where the last inequality follows from Assumption 2(b) and Lemma 9i (p. 30) under Assumption 1(a)(b)(g).

(viii) For all $j \in [\![1,m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right] < (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)$ $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} \mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta'}||\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] \leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^2\right] < \infty$ where the two last inequalities follow from Assumption 2(b) and the boundedness of $\overline{B_L}$.

(ix) See Lemma 18vi p. 50.

(x) Under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(j,k) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right] < \infty\\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta'}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)\sup_{(\theta,\tau)\in\overline{B_L}}|\psi_k(X_1,\theta)|\right]\\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta'}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi_k(X_1,\theta)|^2\right]}\\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^4\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]} < \infty,
\end{aligned}$$

where the last inequality follows from the boundedness of $\overline{B_L}$, Assumption 2(b) and Lemma 9i (p. 30) under Assumption 1(a)(b)(g) and (e).

(xi) Under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(j,k) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}|\right] \\
&\leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi_k(X_1,\theta)'|\right] \\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial^2\psi(X_1,\theta)}{\partial\theta_j\partial\theta'}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi_k(X_1,\theta)'|^2\right]} \\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)'|^2\right]} < \infty,
\end{aligned}$$

where the last inequality follows from Assumption 2(b) and Lemma 9 (p. 30), under Assumption 1(a)(b)(g) and (e).

(xii) Under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $(j,k) \in [\![1,m]\!]^2$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_j}|\right] < \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} \mathrm{e}^{\tau'\psi(X_1,\theta)}|\frac{\partial\psi(X_1,\theta)}{\partial\theta'}||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right] \leqslant \mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^2\right] < \infty$ where the two last inequalities follow from Assumption 2(b).

(xiii) See Lemma 18xii p. 50.

(xiv) Under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, for all $k \in [\![1,m]\!]$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}} |\tau|^2)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} \left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\psi_k(X_1,\theta)|\right) \sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}} |\tau|^2)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} \left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\psi_k(X_1,\theta)|\right)^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}} |\tau|^2)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} < \infty,
\end{aligned}$$

where the last inequality follows from the boundedness of $\overline{B_L}$, Assumptions 1(g) and 2(b).

(xv) See Lemma 18xiv p. 50.

(xvi) See Lemma 18xiii p. 50.

(xvii) Under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(k,j) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\psi_k(X_1,\theta)||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}\left(\mathrm{e}^{\tau'\psi(X_1,\theta)}|\psi_k(X_1,\theta)||\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\right)^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}}|\tau|)\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X)^4\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \\
&< \infty,
\end{aligned}$$

where the last inequality follows from the boundedness of $\overline{B_L}$, Assumption 2(b) and Assumption 1(g).

(xviii) Under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(k,j) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)'|\right] \\
&\leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)'|\right] \\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta_j}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)|^2\right]} \\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^4\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)|^2\right]} < \infty,
\end{aligned}$$

where the last inequality follows from Assumption 2(b), and Lemma 9i (p. 30), under Assumption 1(a)(b)(g).

(xix) Under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(k,j) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_j}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_j}|\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi_k(X_1,\theta)}{\partial\theta_j}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \\
&\leqslant \sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)}\mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon}|\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} < \infty,
\end{aligned}$$

where the last inequality follows from Assumption 2(b) and Assumption 1(g). □

Lemma 20. *Under Assumptions 1 and 2, $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial L_T(\hat{\theta}_T, \tau_T(\hat{\theta}_T))}{\partial \theta} = O(T^{-1})$.*

Proof. Unlike in most of the rest of the paper, for clarity, in this proof we do not use the potentially ambiguous notation that denotes $\left.\frac{\partial L_T(\theta,\tau)}{\partial \theta}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}$ with $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial \theta}$.[14]

Under Assumptions 1 and 2(a), by subsection B.2 (p. 32), the function $L_T(\theta, \tau)$ is well-defined and twice continuously differentiable in a neighborhood of $(\theta_0' \ \tau(\theta_0)')$ $\mathbb{P}$-a.s. for T big enough. Moreover, under Assumption 1(a)(b) and (d)-(h), by Lemma 21i (p. 61), $\tau_T(.)$ is continuously differentiable in $\mathbf{\Theta}$. Now, under Assumption 1, by Theorem 1i (p. 6) and Lemma 2iii (p. 20), $\mathbb{P}$-a.s., $\hat{\theta}_T \to \theta_0$ and $\tau_T(\hat{\theta}_T) \to \tau(\theta_0)$, so that $\mathbb{P}$-a.s. for T big enough, $(\hat{\theta}_T' \ \tau_T(\hat{\theta}_T)')$ is in any arbitrary small neighborhood of $(\theta_0' \ \tau(\theta_0)')$. Therefore, under Assumption 1 and 2(a), by the chain rule theorem (e.g., Magnus and Neudecker 1999/1988, Chap. 5 sec. 11), $\mathbb{P}$-a.s. for T big enough, $\theta \mapsto L_T(\theta, \tau_T(\theta))$ is continuously differentiable in a neighborhood of $\hat{\theta}_T$, and, for all $j \in [\![1, m]\!]$,

$$
\begin{aligned}
& 0 = \left.\frac{\partial L_T(\theta, \tau_T(\theta))}{\partial \theta_j}\right|_{\theta=\hat{\theta}_T} \\
\overset{(a)}{\Leftrightarrow}\ & 0 = \left.\frac{\partial L_T(\theta,\tau)}{\partial \theta_j}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))} + \left.\frac{\partial L_T(\theta,\tau)}{\partial \tau'}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))} \left.\frac{\partial \tau(\theta)}{\partial \theta_j}\right|_{\theta=\hat{\theta}_T} \\
\Leftrightarrow\ & \left.\frac{\partial L_T(\theta,\tau)}{\partial \theta_j}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))} = - \left.\frac{\partial L_T(\theta,\tau)}{\partial \tau'}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))} \left.\frac{\partial \tau(\theta)}{\partial \theta_j}\right|_{\theta=\hat{\theta}_T} \\
\overset{(b)}{\Leftrightarrow}\ & \left.\frac{\partial L_T(\theta,\tau)}{\partial \theta_j}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))} = O(T^{-1})O(1) = O(T^{-1}).
\end{aligned}
$$

(a) It is an immediate and standard implication of the chain rule (e.g., Magnus and Neudecker 1999/1988, chap. 5, sec. 12, exercise 3). *(b)* Firstly, under Assumptions 1 and 2, by Lemma 22iv (p. 63), $\mathbb{P}$-a.s. as $T \to \infty$, $\left.\frac{\partial L_T(\theta,\tau)}{\partial \tau'}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))} = O(T^{-1})$ because $(\hat{\theta}_T, \tau_T(\hat{\theta}_T)) \to (\theta_0, \tau(\theta_0))$, $\mathbb{P}$-a.s. as $T \to \infty$, by Theorem 1i (p. 6) and Lemma 2iii (p. 20). Secondly, under Assumptions 1 and 2, by Theorem 1i (p. 6) and Lemma 21iii (p. 61), $\mathbb{P}$-a.s. as $T \to \infty$, $\left.\frac{\partial \tau(\theta)}{\partial \theta_j}\right|_{(\theta,\tau)=(\hat{\theta}_T,\tau_T(\hat{\theta}_T))} = O(1)$. □

Lemma 21 (First Derivative of the implicit function $\tau_T(.)$)**.** *Under Assumption 1(a)(b) and (d)-(h),*

(i) *$\mathbb{P}$-a.s. for T big enough, the function $\tau_T : \mathbf{\Theta} \to \mathbf{R}^m$ is continuously differentiable in $\mathbf{\Theta}$ and its first derivative is*

$$
\frac{\partial \tau_T(\theta)}{\partial \theta'} = -\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} \psi_t(\theta)\psi_t(\theta)'\right]^{-1} \left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)} \left(\frac{\partial \psi_t(\theta)}{\partial \theta'} + \psi_t(\theta)\tau_T(\theta)'\frac{\partial \psi_t(\theta)}{\partial \theta'}\right)\right];
$$

(ii) *for any sequence $(\theta_T)_{T\in\mathbf{N}} \in \mathbf{\Theta}^{\mathbf{N}}$ converging to θ_0, $\mathbb{P}$-a.s. for T big enough, there exists $\bar{\theta}_T$ between θ_T and θ_0 s.t. $\sqrt{T}[\tau_T(\theta_T) - \tau_T(\theta_0)] = \frac{\partial \tau_T(\bar{\theta}_T)}{\partial \theta'}\sqrt{T}(\theta_T - \theta_0)$;*

[14]This is a potentially ambiguous notation in the sense that $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial \theta}$ could also denote $\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial \theta}\right|_{\theta=\hat{\theta}_T}$. Except when indicated otherwise, such an ambiguity cannot occur because we never use derivatives of $\theta \mapsto L_T(\theta, \tau_T(\theta))$.

(iii) *under additional Assumptions 1(c) and 2(b), for any sequence* $(\theta_T)_{T\in\mathbf{N}} \in \mathbf{\Theta}^{\mathbf{N}}$ *converging to* θ_0*,* $\mathbb{P}$*-a.s. as* $T \to \infty$*,* $\frac{\partial \tau_T(\theta_T)}{\partial \theta'} \to -\mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']^{-1}\mathbb{E}\left[\frac{\partial \psi(X_1,\theta_0)}{\partial \theta}\right]$*; and*

(iv) *under additional Assumptions 1(c) and 2(b), for any sequence* $(\theta_T)_{T\in\mathbf{N}} \in \mathbf{\Theta}^{\mathbf{N}}$ *converging to* θ_0 *s.t., as* $T \to \infty$*,* $\sqrt{T}(\theta_T - \theta_0) = O_{\mathbb{P}}(1)$ $\mathbb{P}$*-a.s. as* $T \to \infty$*,* $\tau_T(\theta_T) - \tau_T(\theta_0) = -V^{-1}M(\theta_T - \theta_0) + o_{\mathbb{P}}(T^{-1/2})$*, where* $V := \mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']$ *and* $M := \mathbb{E}\left[\frac{\partial \psi(X_1,\theta_0)}{\partial \theta}\right]$.

Proof. (i) Under Assumption 1(a)(b) and (d)-(h), by Lemma 1ii (p. 19) and its proof, $\mathbb{P}$-a.s. for T big enough, the assumptions of the standard implicit function theorem hold and $\tau_T(.)$ is continuously differentiable. Thus, under Assumption 1(a)(b) and (d)-(h), $\mathbb{P}$-a.s. for T big enough, application of the implicit function theorem yields

$$\begin{aligned}
&\frac{\partial \tau_T(\theta)}{\partial \theta'} \\
&= -\left[\frac{\partial\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\right]}{\partial \tau'}\right]^{-1}\left[\frac{\partial\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\right]}{\partial \theta'}\right]\Bigg|_{\tau=\tau_T(\theta)} \\
&= -\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau'\psi_t(\theta)}\left(\frac{\partial \psi_t(\theta)}{\partial \theta'} + \psi_t(\theta)\tau'\frac{\partial \psi_t(\theta)}{\partial \theta'}\right)\right]\Bigg|_{\tau=\tau_T(\theta)} \\
&= -\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\left(\frac{\partial \psi_t(\theta)}{\partial \theta'} + \psi_t(\theta)\tau_T(\theta)'\frac{\partial \psi_t(\theta)}{\partial \theta'}\right)\right].
\end{aligned}$$

(ii) Again, under Assumption 1(a)(b) and (d)-(h), by Lemma 1ii (p. 19), $\mathbb{P}$-a.s. for T big enough, $\tau_T(.)$ is continuously differentiable, so that the result follows from a first-order stochastic Taylor-Lagrange expansion (Jennrich 1969, Lemma 3).

(iii) Firstly, under Assumption 1(a)(b)(d)(e)(g)(h), by Lemma 2iii (p. 20), $\mathbb{P}$-a.s. as $T \to \infty$, $\sup_{\theta\in\mathbf{\Theta}} |\tau_T(\theta) - \tau(\theta)| = o(1)$, so that $\tau_T(\theta_T) \to \tau(\theta_0)$. Secondly, under Assumptions 1 and 2, by Lemma 23iv, vii and x (p. 64), for $\overline{B_L}$ a ball around $(\theta_0, \tau(\theta_0))$ of sufficiently small radius, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial \psi(X_1,\theta)}{\partial \theta'}|\right] < \infty$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$, and $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\tau'\frac{\partial \psi(X_1,\theta)}{\partial \theta'}|\right] < \infty$. Thus, by Assumptions 1(a)(b) and (d), the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3), implies that, for all $k \in [\![1, m]\!]$, $\mathbb{P}$-a.s. as $T \to \infty$,

$$\begin{aligned}
&\frac{\partial \tau_T(\theta_T)}{\partial \theta'} \\
&\to -\mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi(X_1,\theta_0)\psi(X_1,\theta_0)']^{-1}\left\{\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\frac{\partial \psi(X_1,\theta_0)}{\partial \theta}\right]\right. \\
&\qquad \left. + \mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi(X_1,\theta)\tau(\theta_0)'\frac{\partial \psi(X_1,\theta_0)}{\partial \theta}\right]\right\} \\
&= -\mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']^{-1}\mathbb{E}\left[\frac{\partial \psi(X_1,\theta_0)}{\partial \theta}\right]
\end{aligned}$$

because $\tau(\theta_0) = 0_{m\times 1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e) and (g)-(h).

(iv) Under Assumption 1 and 2(b), by the statement (ii) of the present lemma, $\mathbb{P}$-a.s. as $T \to \infty$, there exists $\bar{\theta}_T$ between θ_T and θ_0 s.t.

$$\begin{aligned}\tau_T(\theta_T) - \tau_T(\theta_0) &= \frac{\partial \tau_T(\bar{\theta}_T)}{\partial \theta'}(\theta_T - \theta_0) \\ &\overset{(a)}{=} -V^{-1}M(\theta_T - \theta_0) + \left[\frac{\partial \tau_T(\bar{\theta}_T)}{\partial \theta'} + V^{-1}M\right](\theta_T - \theta_0) \\ &\overset{(b)}{=} -V^{-1}M(\theta_T - \theta_0) + o_{\mathbb{P}}(T^{-1/2})\end{aligned}$$

(a) Add and subtract $V^{-1}M(\theta_T - \theta_0)$. *(b)* Under Assumption 1 and 2(b), by the statement (ii) of the present lemma, $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial \tau_T(\bar{\theta}_T)}{\partial \theta'} + V^{-1}M = o(1)$. Moreover, by assumption, as $T \to \infty$, $\theta_T - \theta_0 = O_{\mathbb{P}}(T^{-1/2})$, so that $\left[\frac{\partial \tau_T(\bar{\theta}_T)}{\partial \theta'} + V^{-1}M\right](\theta_T - \theta_0) = o_{\mathbb{P}}(T^{-1/2})$. □

Remark 3. As notation indicates, $\frac{\partial \tau_T(.)}{\partial \theta'}$ corresponds to a partial derivative as $\tau_T(.)$ is also a function of the data. ◇

Lemma 22 (Asymptotic limit of $\frac{\partial L_T(\theta_T, \tau_T(\theta_T))}{\partial \tau'}$). *Under Assumptions 1 and 2, for any sequence $(\theta_T)_{T \in \mathbf{N}} \in \mathbf{\Theta}^{\mathbf{N}}$ converging to θ_0, for all $k \in [\![1, m]\!]$, $\mathbb{P}$-a.s. as $T \to \infty$,*

(i) $\frac{\partial M_{1,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} = 0$;
(ii) $\frac{\partial M_{2,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} = O(T^{-1})$;
(iii) $\frac{\partial M_{3,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} = O(T^{-1})$; *and*
(iv) $\frac{\partial L_T(\theta_T, \tau_T(\theta_T))}{\partial \tau'} = O(T^{-1})$.

Proof. *(i)* Under Assumption 1(a)(b) and (d)-(h), by Lemma 2ii (p. 20), $\mathbb{P}$-a.s. for T big enough, $\tau_T(\theta_T)$ exists, so that, by equation (27) on p. 36, $\mathbb{P}$-a.s. for T big enough, for all $k \in [\![1, m]\!]$,

$$\begin{aligned}\frac{\partial M_{1,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} &= \left(1 - \frac{m}{2T}\right) \frac{\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta_T)'\psi_t(\theta_T)}\psi_{t,k}(\theta_T)}{\frac{1}{T}\sum_{i=1}^T e^{\tau_T(\theta_T)'\psi_i(\theta_T)}} \\ &= 0\end{aligned}$$

because, by definition of $\tau_T(\theta)$ in equation (14) on p. 17, $\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta_T)'\psi_t(\theta_T)}\psi_{t,k}(\theta_T) = 0$.

(ii) Similarly, under Assumption 1, by equation (33) on p. 39, $\mathbb{P}$-a.s. for T big enough, for all $k \in [\![1, m]\!]$,

$$\frac{\partial M_{2,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} = \frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta_T)'\psi_t(\theta_T)}\frac{\partial \psi_t(\theta_T)}{\partial \theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^T e^{\tau_T(\theta_T)'\psi_t(\theta_T)}\psi_{t,k}(\theta_T)\frac{\partial \psi_t(\theta_T)}{\partial \theta'}\right]\right\}$$

where $\mathbb{P}$-a.s. as $T \to \infty$, $(\theta_T' \quad \tau_T(\theta_T)') \to (\theta_0' \quad \tau(\theta_0)')$ by the lemma's assumption and Lemma 2iii (p. 20). Now, under Assumptions 1 and 2, by Lemma 23iv and v (p. 64), for $\overline{B_L}$ a ball around $(\theta_0, \tau(\theta_0))$ of sufficiently small radius, $\mathbb{E}\left[\sup_{(\theta,\tau) \in \overline{B_L}} |e^{\tau'\psi(X_1,\theta)}\frac{\partial \psi(X_1,\theta)}{\partial \theta'}|\right] < \infty$, and, for all $k \in [\![1, m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau) \in \overline{B_L}} |e^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial \psi(X_1,\theta)}{\partial \theta'}|\right] < \infty$. Thus, by Assumptions 1(a)(b) and (d), the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi

2003, pp. 24-25, Theorem 1.3.3), implies that, for all $k \in [\![1, m]\!]$, $\mathbb{P}$-a.s. as $T \to \infty$,

$$
\begin{aligned}
&T \frac{\partial M_{2,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} \\
&\to \operatorname{tr}\left\{\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\frac{\partial \psi(X_1,\theta_0)}{\partial \theta'}\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_k(X_1,\theta_0)\frac{\partial \psi(X_1,\theta_0)}{\partial \theta'}\right]\right\} \\
&= \operatorname{tr}\left\{\mathbb{E}\left[\frac{\partial \psi(X_1,\theta_0)}{\partial \theta'}\right]^{-1}\mathbb{E}\left[\psi_k(X_1,\theta_0)\frac{\partial \psi(X_1,\theta_0)}{\partial \theta'}\right]\right\},
\end{aligned}
$$

because $\tau(\theta_0) = 0_{m\times 1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e) and (g)-(h). Therefore, $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial M_{2,T}(\theta_T,\tau_T(\theta_T))}{\partial \tau_k} = O(T^{-1})$.

(iii) Under Assumption 1, by equation (38) (p. 42), for all $k \in [\![1, m]\!]$,

$$
\begin{aligned}
&\frac{\partial M_{3,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} \\
&= -\frac{1}{2T}\operatorname{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta_T)'\psi_t(\theta_T)}\psi_t(\theta_T)\psi_t(\theta_T)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta_T)'\psi_t(\theta)}\psi_{t,k}(\theta_T)\psi_t(\theta_T)\psi_t(\theta_T)'\right]\right\}
\end{aligned}
$$

where $\mathbb{P}$-a.s. as $T \to \infty$, $(\theta_T' \quad \tau_T(\theta_T)') \to (\theta_0' \quad \tau(\theta_0)')$ by Theorem 1i (p. 6). Now, under Assumptions 1 and 2, by Lemma 23vii and viii (p. 64), there exists a closed ball $\overline{B_L} \subset \mathbf{S}$ centered at $(\theta_0, \tau(\theta_0))$ with strictly positive radius s.t., for all $k \in [\![1, m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$ and $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$. Thus, under Assumptions 1 and 2, by ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3), for all $k \in [\![1, m]\!]$, $\mathbb{P}$-a.s. as $T \to \infty$,

$$
\begin{aligned}
&T \frac{\partial M_{3,T}(\theta_T, \tau_T(\theta_T))}{\partial \tau_k} \\
&\to -\frac{1}{2}\operatorname{tr}\left\{\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_k(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right\} \\
&= -\frac{1}{2}\operatorname{tr}\left\{\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\psi_k(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right\},
\end{aligned}
$$

because $\tau(\theta_0) = 0_{m\times 1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e) and (g)-(h). Therefore, $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial M_{3,T}(\theta_T,\tau_T(\theta_T))}{\partial \tau_k} = O(T^{-1})$.

(iv) Under Assumption 1(a)-(b) and (d)-(h), by Lemma 12 (p. 34), $L_T(\theta, \tau) = M_{1,T}(\theta, \tau) + M_{2,T}(\theta, \tau) + M_{3,T}(\theta, \tau)$, so that the result follows from the statement (i)-(iii) of the present lemma. □

Remark 4. In the case in which $\theta_T = \hat{\theta}_T$, there exist at least one other way to prove Lemma 22 that do not require Assumption 2. This way follows an approach à la Newey and Smith (2004), which relies on ULLN with $\mathbf{T}_T(\theta) = \{\tau \in \mathbf{R}^m : |\tau| \leqslant T^{-\zeta}\}$ and $\zeta > 0$. We do not follow this ways because (i) Other parts of the proof of Theorem 1ii (p. 6) require the asymptotic normality of $\hat{\theta}_T$ and thus Assumption 2; (ii) It would lengthen the proofs and complicate their logic; (iii) We later use Lemma 22 with $\theta_T = \check{\theta}_T$, where $\check{\theta}_T$ is a constrained estimator. ◇

Lemma 23 (Finiteness of the expectations of supremum of the terms from $\frac{\partial L_T(\theta,\tau)}{\partial \tau}$ and $\frac{\partial^2 L_T(\theta,\tau)}{\partial \tau' \partial \tau}$)**.** *Under Assumptions 1 and 2, there exists a closed ball $\overline{B_L} \subset \mathbf{S}$ centered at $(\theta_0, \tau(\theta_0))$ with strictly positive radius s.t., for all $(h,k) \in [\![1,m]\!]^2$,*

(i) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} \mathrm{e}^{\tau'\psi(X_1,\theta)}\right] < \infty$;

(ii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)|\right] < \infty$;

(iii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)|\right] < \infty$;

(iv) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] < \infty$;

(v) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] < \infty$;

(vi) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] < \infty$;

(vii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$;

(viii) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$;

(ix) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$; *and*

(x) $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] < \infty$.

Proof. (i) Apply Lemma 18i (p. 50) under Assumptions 1 and 2. Note that it does not immediately follow from Assumption 1(e) and the Cauchy-Schwarz inequality because we need additional assumptions to ensures that there exists $\overline{B_L} \subset \mathbf{S}$: See Lemma 11(ii) on p. 32.

(ii) For all $(h,k) \in [\![1,m]\!]^2$, for all $(\theta,\tau) \in \overline{B_L}$, $\mathrm{e}^{\tau'\psi(X_1,\theta)}|\psi_k(X_1,\theta)\psi_h(X_1,\theta)| = \mathrm{e}^{\tau'\psi(X_1,\theta)}$ $\sqrt{[\psi_k(X_1,\theta)\psi_h(X_1,\theta)]^2} \leqslant \mathrm{e}^{\tau'\psi(X_1,\theta)}\sqrt{\sum_{(i,j)\in[\![1,m]\!]^2}[\psi_i(X_1,\theta)\psi_j(X_1,\theta)]^2} = \mathrm{e}^{\tau'\psi(X_1,\theta)}|\psi(X_1,\theta)\psi(X_1,\theta)'|$, so that $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)|\right] \leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)|\right] < \infty$, where the last inequality follows from Lemma 18xii (p. 50) under Assumptions 1 and 2.

(iii) Apply Lemma 19v (p. 56) under Assumptions 1 and 2.

(iv) Apply Lemma 18v (p. 50) under Assumptions 1 and 2.

(v) Apply Lemma 19vii (p. 56) under Assumptions 1 and 2.

(vi) Proof similar to the one of Lemma 18xiii (p. 50). The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under

Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(h,k) \in [\![1,m]\!]^2$,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]\\
\leqslant\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi_k(X_1,\theta)\psi_h(X_1,\theta)|\right]\\
\overset{(a)}{\leqslant}\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|\right]\\
\overset{(b)}{\leqslant}\ &\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]}\\
\overset{(c)}{\leqslant}\ &\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \overset{(d)}{<} \infty.
\end{aligned}
$$

(a) As in the proof of statement (ii), for all $(h,k) \in [\![1,m]\!]^2$, for all $(\theta,\tau) \in \overline{B_L}$, $|\psi_k(X_1,\theta)\psi_h(X_1,\theta)| \leqslant |\psi(X_1,\theta)\psi(X_1,\theta)'|$. *(b)* Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(c)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0,\tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$. Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2$ because $\overline{B_L} \subset \mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(d)* Firstly, by Assumption 2(b), the first expectation is bounded. Secondly, by Assumption 1(g), the second expectation is also bounded.

(vii) Apply Lemma 18xii (p. 50) under Assumptions 1 and 2.

(viii) Proof similar to the one of Lemma 18xiii (p. 50) and to the statement (vi) of the present lemma. The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(h,k) \in [\![1,m]\!]^2$,

$$
\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)|\right]\\
\leqslant\ &\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|\right]\\
\overset{(a)}{\leqslant}\ &\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]}\\
\overset{(b)}{\leqslant}\ &\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^2\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \overset{(c)}{<} \infty.
\end{aligned}
$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0,\tau(\theta_0))$,

so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$. Moreover, for all $k \in [\![1,m]\!]$, for all $\theta \in \mathbf{\Theta}$, $|\psi_k(X_1,\theta)| \leqslant |\psi(X_1,\theta)| \leqslant b(X)$, where the last inequality follows from Assumption 2(b). Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2$ because $\overline{B_L} \subset \mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* Firstly, by Assumption 2(b), the first expectation is bounded. Secondly, by Assumption 1(g), the second expectation is also bounded.

(ix) Proof similar to the one of Lemma 18xiii (p. 50) and to the statement (vi) of the present lemma. The supremum of the absolute value of the product is smaller than the product of the suprema of the absolute values. Thus, under Assumption 1(a)(b), for $\overline{B_L}$ of sufficiently small radius, for all $(h,k) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)|\right] \\
&\leqslant \mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)| \sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|\right] \\
&\overset{(a)}{\leqslant} \sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)|^2\right]}\sqrt{\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \\
&\overset{(b)}{\leqslant} \sqrt{\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{2\tau'\psi(X_1,\theta)}b(X_1)^4\right]}\sqrt{\mathbb{E}\left[\sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2\right]} \overset{(c)}{<} \infty.
\end{aligned}$$

(a) Apply the Cauchy-Schwarz inequality, and note that the supremum of the square of a positive function is the square of the supremum of the function. *(b)* Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0, \tau(\theta_0))$, so that, for $\overline{B_L}$ of sufficiently small radius, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S} \subset \mathbf{S}^\epsilon$. Moreover, for all $k \in [\![1,m]\!]$, for all $\theta \in \mathbf{\Theta}$, $|\psi_k(X_1,\theta)| \leqslant |\psi(X_1,\theta)| \leqslant b(X)$ where the last inequality follows from Assumption 2(b). Secondly, as the second supremum does not depend on τ, $\sup_{(\theta,\tau)\in\overline{B_L}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2 \leqslant \sup_{\theta\in\mathbf{\Theta}^\epsilon} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2$ because $\overline{B_L} \subset \mathbf{S}$, for $\overline{B_L}$ of radius small enough. *(c)* Firstly, by Assumption 2(b), the first expectation is bounded. Secondly, by Assumption 1(g), the second expectation is also bounded.

(x) The norm of a product of matrices is smaller than the product of the norms (e.g., Rudin 1953, Theorem 9.7 and note that all norms are equivalent on finite dimensional spaces). Thus, for $\overline{B_L}$ of sufficiently small radius, for all $(\ell,j) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\tau'\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] \\
&\leqslant (\sup_{(\theta,\tau)\in\overline{B_L}} |\tau|)\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} \mathrm{e}^{\tau'\psi(X_1,\theta)}|\psi(X_1,\theta)||\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right] \\
&\overset{(a)}{\leqslant} (\sup_{(\theta,\tau)\in\overline{B_L}} |\tau|)\mathbb{E}\left[\sup_{\theta\in\mathcal{N}}\sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{\tau'\psi(X_1,\theta)}b(X_1)^2\right] \overset{(b)}{<} \infty.
\end{aligned}$$

(a) Firstly, under Assumption 1(a)-(e) and (g)-(h), by Lemma 11ii (p. 32), $\mathbf{S}$ contains an open ball centered at $(\theta_0, \tau(\theta_0))$ Thus, under Assumption 1(a)-(e) and (g)-(h), for $\overline{B_L}$ of sufficiently small radius, by definition of $\mathbf{S}$, $\overline{B_L} \subset \{(\theta,\tau) : \theta \in \mathcal{N} \wedge \tau \in \mathbf{T}(\theta)\} \subset \mathbf{S}$, because $\mathcal{N} \subset \mathbf{\Theta}$ by Assumption 2(a). Secondly, by Assumption 2(b), $\sup_{\theta\in\mathcal{N}} |\psi(X_1,\theta)| \leqslant b(X)$ and $\sup_{\theta\in\mathcal{N}} |\frac{\partial\psi(X_1,\theta)}{\partial\theta'}| \leqslant b(X)$. *(b)* Firstly, $\sup_{(\theta,\tau)\in\overline{B_L}} |\tau|^2 < \infty$ because $\overline{B_L}$ is bounded. Secondly, by Assumption 2(b), $\mathbb{E}\left[\sup_{\theta\in\mathcal{N}} \sup_{\tau\in\mathbf{T}(\theta)} \mathrm{e}^{\tau'\psi(X_1,\theta)} b(X_1)^2\right] < \infty$. □

B.4. **Proof of Theorem 2: Trinity+1.** The proof of Theorem 2 adapts the traditional way of deriving the trinity along the lines of Smith (2011). As in the proof of Theorem 1, the main difference comes from the complexity of the variance term $|\Sigma_T(\theta)|_{\det}$.

Core of the proof of Theorem 2. Asymptotic distribution of Wald_T. By Assumption 3(a), $r : \mathbf{\Theta} \to \mathbf{R}^q$ is continuously differentiable. Thus, under Assumptions 1 and 2, if the test hypothesis (9) on p. 6 holds, a first-order Taylor-Lagrange expansion at θ_0 evaluated at $\hat{\theta}_T$, ω by ω, yields, $\mathbb{P}$-a.s. as $T \to \infty$,

$$\begin{aligned} r(\hat{\theta}_T) &= r(\theta_0) + R(\bar{\theta}_T)(\hat{\theta}_T - \theta_0) \text{ , where } \bar{\theta}_T \text{ is between } \hat{\theta}_T \text{ and } \theta_0; \\ &\overset{(a)}{=} R(\bar{\theta}_T)(\hat{\theta}_T - \theta_0) \\ &\underset{(b)}{\overset{D}{\to}} R(\theta_0)\mathcal{N}(0, \Sigma(\theta_0)) \\ &\overset{D}{=} \mathcal{N}\left(0, R(\theta_0)\Sigma(\theta_0)R(\theta_0)'\right). \end{aligned}$$

(a) By definition, if the test hypothesis (9) on p. 6 holds, $r(\theta_0) = 0_{q\times 1}$. *(b)* Under Assumptions 1 and 2, by Theorem 1ii (p. 6), $\mathbb{P}$-a.s. as $T \to \infty$, $\sqrt{T}(\hat{\theta}_T - \theta_0) \overset{D}{\to} \mathcal{N}(0, \Sigma(\theta_0))$, which also implies that $\bar{\theta}_T \to \theta_0$. Thus, under Assumptions 1, 2 and 3(a), by continuity of $R(.)$, $\mathbb{P}$-a.s. as $T \to \infty$, $R(\bar{\theta}) \to R(\theta_0)$, so that the result follows by the Slutsky's theorem.

Now, under Assumptions 1, 2 and 3(a), by Lemma 28i (p. 82), $\mathbb{P}$-a.s. as $T \to \infty$, $\check{\theta}_T \to \theta_0$, so that $R(\check{\theta}_T) \to R(\theta_0)$ by Assumption 3(a). Moreover, by the theorem's assumption, as $T \to \infty$, $\widehat{\Sigma(\theta_0)} \overset{\mathbb{P}}{\to} \Sigma(\theta_0)$. In addition, by Assumption 1(h) and 3(b), $\Sigma(\theta_0)$ and $R(\theta_0)$ are full rank, so that $\widehat{\Sigma(\theta_0)}$ and $R(\check{\theta}_T)$ are full rank w.p.a.1 as $T \to \infty$ (Lemma 30 p. 87). Then, the result follows from the Cochran's theorem.

Asymptotic distribution of LM_T. Under Assumptions 1, 2 and 3, by Proposition 2iii (p. 75), if the test hypothesis (9) on p. 6 holds, as $T \to \infty$, $\check{\gamma}_T \overset{D}{\to} \mathcal{N}(0, (R(\theta_0)\Sigma(\theta_0)R(\theta_0)')^{-1})$. Now, under Assumptions 1, 2 and 3(a), by Lemma 28i (p. 82), $\mathbb{P}$-a.s. as $T \to \infty$, $\check{\theta}_T \to \theta_0$, so that $R(\check{\theta}_T) \to R(\theta_0)$ by Assumption 3(a). Moreover, by the theorem's assumption, as $T \to \infty$, $\widehat{\Sigma(\theta_0)} \overset{\mathbb{P}}{\to} \Sigma(\theta_0)$. In addition, by Assumption 1(h) and 3(b), $\Sigma(\theta_0)$ and $R(\theta_0)$ are full rank, so that $\widehat{\Sigma(\theta_0)}$ and $R(\check{\theta}_T)$ are full rank w.p.a.1 as $T \to \infty$ (Lemma 30 p. 87). Then, by the Cochran's theorem, as $T \to \infty$, $T\check{\gamma}_T'[R(\check{\theta}_T)\widehat{\Sigma(\theta_0)}R(\check{\theta}_T)']\check{\gamma}_T \overset{D}{\to} \chi^2_q$. Finally, under Assumptions 1, 2 and 3, by Lemma 28iii (p. 82), $R(\check{\theta}_T)'\check{\gamma}_T = -\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial\theta}\right|_{\theta=\check{\theta}_T}$, so $T\check{\gamma}_T'[R(\check{\theta}_T)\widehat{\Sigma(\theta_0)}R(\check{\theta}_T)']\check{\gamma}_T = T[R(\check{\theta}_T)'\check{\gamma}_T]'\widehat{\Sigma(\theta_0)}[R(\check{\theta}_T)'\check{\gamma}_T] = T\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial\theta}\right|'_{\theta=\check{\theta}_T}\widehat{\Sigma(\theta_0)}\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial\theta}\right|_{\theta=\check{\theta}_T} = \frac{\partial\ln[\hat{f}_{\theta^*_T}(\check{\theta}_T)]}{\partial\theta'}\widehat{\Sigma(\theta_0)}\frac{\partial\ln[\hat{f}_{\theta^*_T}(\check{\theta}_T)]}{\partial\theta}$, where the last equality follows from the definition of the LogESP in Lemma 12 (p. 34), i.e.,
$L_T(\theta,\tau_T(\theta)) := \ln\left[\frac{1}{T}\sum_{t=1}^T \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det} = \frac{1}{T}[\ln(\hat{f}_{\theta^*_T}(\theta)) - \frac{m}{2}\ln(\frac{T}{2\pi})]$.

Asymptotic distribution of ALR_T. Under Assumptions 1,2 and 3, if the test hypothesis (9) on p. 6 holds, by Lemma 24 (p. 70), $\mathbb{P}$-a.s. as $T \to \infty$,

$$\begin{aligned}
&2\{\ln[\hat{f}_{\theta_T^*}(\hat{\theta}_T)] - \ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]\} \\
&= -\left[\sqrt{T}(\hat{\theta}_T - \check{\theta}_T)\right]' \Sigma^{-1} \left[\sqrt{T}(\hat{\theta}_T - \check{\theta}_T)\right] + o_{\mathbb{P}}(1) \\
&\overset{(a)}{=} -\left[\Sigma R'(R\Sigma R')^{-1}RM^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)+o_{\mathbb{P}}(1)\right]'\Sigma^{-1}\left[\Sigma R'(R\Sigma R')^{-1}RM^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)+o_{\mathbb{P}}(1)\right]+o_{\mathbb{P}}(1) \\
&\overset{(b)}{=} -\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]'(M')^{-1}R'(R\Sigma R')^{-1}R\Sigma\Sigma^{-1}\Sigma R'(R\Sigma R')^{-1}RM^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]+o_{\mathbb{P}}(1) \\
&\overset{(c)}{=} -\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]'(M')^{-1}R'(R\Sigma R')^{-1}RM^{-1}\left[\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]+o_{\mathbb{P}}(1) \\
&\overset{(d)}{=} -\left[V^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]'V^{1/2'}(M')^{-1}R'(R\Sigma R')^{-1}RM^{-1}V^{1/2}\left[V^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]+o_{\mathbb{P}}(1) \\
&\overset{(e)}{=} -\left[V^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]'P_{\Sigma^{1/2}R'}\left[V^{-1/2}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]+o_{\mathbb{P}}(1) \\
&\underset{(f)}{\overset{D}{\to}} \chi_q^2
\end{aligned}$$

(a) Under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, by Lemma 24ii (p. 70), $\mathbb{P}$-a.s. as $T \to \infty, \sqrt{T}(\hat{\theta}_T - \check{\theta}_T) = \Sigma R'(R\Sigma R')^{-1}RM^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1)$. *(b)* Transpose the content of the first square bracket, and then note that $\Sigma = \Sigma'$ by symmetry. *(c)* Note that $R\Sigma\Sigma^{-1}\Sigma R'(R\Sigma R')^{-1} = I$. *(d)* Use that $V^{1/2}V^{-1/2} = I$. *(e)* Note that $P_{\Sigma^{1/2}R'} = V^{1/2'}(M')^{-1}R(R'\Sigma R)^{-1}R'M^{-1}V^{1/2}$. *(f)* Under Assumption 1(a)-(c) and (g), by the Lindeberg-Lévy CLT theorem, as $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) \overset{D}{\to} \mathcal{N}(0, V)$ where $V := \mathbb{E}[\psi(X_1, \theta_0)\psi(X_1, \theta_0)']$. Moreover, the orthogonal projection matrix $P_{\Sigma^{1/2}R'}$ has rank q because R is of rank q and Σ has full rank by Assumptions 3(b) and 1(h), respectively. Thus, the result follows from the Cochran's theorem.

Asymptotic distribution of ET_T. Under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, by Proposition 2ii (p. 75), $\mathbb{P}$-a.s. as $T \to \infty$,

$$\begin{aligned}
\sqrt{T}\tau_T(\check{\theta}_T) &= (M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1) \\
&\overset{(a)}{=} (M')^{-1}[V^{1/2}(M')^{-1}]^{-1}P_{\Sigma^{1/2}R'}[M^{-1}V^{1/2'}]^{-1}M^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1) \\
&\overset{(b)}{=} (M')^{-1}M'V^{-1/2}P_{\Sigma^{1/2}R'}V^{-1/2'}MM^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1) \\
&= V^{-1/2}P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1)
\end{aligned}$$

(a) By definition, $M^{-1}V(M')^{-1} =: \Sigma = \Sigma^{1/2'}\Sigma^{1/2}$, so that $\Sigma^{-1/2} := (\Sigma^{1/2})^{-1} = [V^{1/2}(M')^{-1}]^{-1}$ and $\Sigma^{-1/2'} := (\Sigma^{1/2'})^{-1} = [M^{-1}V^{1/2'}]^{-1}$. *(b)* By standard property of inverses, $[V^{1/2}(M')^{-1}]^{-1}$ and $[M^{-1}V^{1/2'}]^{-1} = V^{-1/2'}M$.

Thus, under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, by Proposition 2, $\mathbb{P}$-a.s. as $T \to \infty$,

$$
\begin{aligned}
& T\tau_T(\check{\theta}_T)'\widehat{V}\tau_T(\check{\theta}_T) \\
= & \left[V^{-1/2}P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1)\right]' \widehat{V}_T \left[V^{-1/2}P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1)\right] \\
\overset{(a)}{=} & \left[V^{-1/2}P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' \widehat{V}_T \left[V^{-1/2}P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right] \\
& 2\left[V^{-1/2}P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' \widehat{V}_T\left[o_{\mathbb{P}}(1)\right] + [o_{\mathbb{P}}(1)]'\,\widehat{V}\,[o_{\mathbb{P}}(1)] \\
\overset{(b)}{=} & \left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' V^{-1/2'}\widehat{V}_T V^{-1/2}\left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right] + o_{\mathbb{P}}(1) \\
\overset{(c)}{=} & \left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' \left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right] \\
& \left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' \left(V^{-1/2'}\widehat{V}_T V^{-1/2} - I\right)\left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right] + o_{\mathbb{P}}(1) \\
\overset{(d)}{=} & \left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' \left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right] + o_{\mathbb{P}}(1) \\
\underset{(d)}{\overset{D}{\to}} & \chi^2_q
\end{aligned}
$$

(a) Use the bilinearity and symmetry of the quadratic form defined by the matrix $\widehat{V}_T$, which is symmetric by the theorem's assumption. *(b)* Firstly, by theorem's assumption, as $T \to \infty$, $\widehat{V}_T \overset{\mathbb{P}}{\to} V$, so that $\widehat{V}_T = O_{\mathbb{P}}(1)$. Secondly, under Assumption 1(a)-(c) and (g), by the Lindeberg-Lévy CLT theorem, as $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) \overset{D}{\to} \mathcal{N}(0, V)$ where $V := \mathbb{E}[\psi(X_1, \theta_0)\psi(X_1, \theta_0)']$, so that, as $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) = O_{\mathbb{P}}(1)$. Thus, the second and third terms are $o_{\mathbb{P}}(1)$. *(c)* Add and subtract $\left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]' \left[P_{\Sigma^{1/2}R'}V^{-1/2'}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)\right]$. *(d)* Denoting the convergence in probability with $\overset{\mathbb{P}}{\to}$, by the present theorem assumption, as $T \to \infty$, $\widehat{V}_T \overset{\mathbb{P}}{\to} V$, where V is a positive definite symmetric matrix by Assumption 1(h). Thus, by Lemma 31 (p. 87), w.p.a.1 as $T \to \infty$, $\hat{V}_T$ is p-d.m, so that it has a square root s.t. $\hat{V}_T = \hat{V}_T^{1/2'}\hat{V}_T^{1/2}$, where $\hat{V}_T^{1/2} \overset{\mathbb{P}}{\to} V^{1/2}$.*(d)* Under Assumption 1(a)-(c) and (g), by the Lindeberg-Lévy CLT theorem, as $T \to \infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) \overset{D}{\to} \mathcal{N}(0, V)$ where $V := \mathbb{E}[\psi(X_1, \theta_0)\psi(X_1, \theta_0)']$. Moreover, the orthogonal projection matrix $P_{\Sigma^{1/2}R'}$ has rank q because R is of rank q and Σ has full rank by Assumptions 3(b) and 1(h), respectively. Thus, the result follows from the Cochran's theorem. □

Lemma 24 (Asymptotic expansions for ALR_T)**.** *Under Assumptions 1,2 and 3, if the test hypothesis* (9) *on p. 6 holds, $\mathbb{P}$-a.s. as $T \to \infty$,*

(i) $2\{\ln[\hat{f}_{\theta_T^*}(\hat{\theta}_T)]-\ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]\}=T(\hat{\theta}_T-\check{\theta}_T)'\Sigma^{-1}(\hat{\theta}_T-\check{\theta}_T)+o_{\mathbb{P}}(1)$;
(ii) $\sqrt{T}(\hat{\theta}_T-\check{\theta}_T)=\Sigma R'(R\Sigma R')^{-1}RM^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^T\psi_t(\theta_0)+o_{\mathbb{P}}(1)$

where $\Sigma:=\Sigma(\theta_0):=M^{-1}V(M')^{-1}$, $M:=\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]$, $V:=\mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']$, *and* $R:=\frac{\partial r(\theta_0)}{\partial\theta'}$.

Proof. *(i)* Under Assumption 1, by Lemma 12 (p. 34), $\mathbb{P}$-a.s. for T big enough, for all (θ,τ) in a neighborhood of $(\theta_0,\tau(\theta_0))$, $L_T(\theta,\tau)$ exists. Moreover, under Assumptions 1, 2 and 3(a), if the test hypothesis (9) on p. 6 holds, by Theorem 1i (p. 6), Lemma 28i (p. 82) and Lemma 2iii (p. 20), $\hat{\theta}_T\to\theta_0$, $\check{\theta}_T\to\theta_0$, $\tau_T(\hat{\theta}_T)\to\tau(\theta_0)$ and $\tau_T(\check{\theta}_T)\to\tau(\theta_0)$, $\mathbb{P}$-a.s. as $T\to\infty$. Thus, noting that $\ln[\hat{f}_{\theta_T^*}(\theta)]=L_T(\theta,\tau_T(\theta))$,under Assumptions 1, 2 and 3(a), if the test hypothesis (9) on p. 6 holds, $\mathbb{P}$-a.s. for T big enough,

$$2\{\ln[\hat{f}_{\theta_T^*}(\hat{\theta}_T)]-\ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]\}=-2T[L_T(\check{\theta}_T,\tau_T(\check{\theta}_T))-L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))].$$

Now, under Assumptions 1 and 2(a), by subsection B.2 (p. 32), $L_T(.,.)$ is twice continuously differentiable in a neighborhood of $(\theta_0'\ \tau(\theta_0)')$ $\mathbb{P}$-a.s. for T big enough, so that a stochastic second-order Taylor-Lagrange expansion (e.g., Aliprantis and Border 2006/1999, Theorem 18.18) around $(\hat{\theta}_T,\tau_T(\hat{\theta}_T))$ and evaluated $(\check{\theta}_T,\tau_T(\check{\theta}_T))$ yields, $\mathbb{P}$-a.s. for T big enough,

$$L_T(\check{\theta}_T,\tau_T(\check{\theta}_T))=L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))+\left[\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta'}\ \frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\tau'}\right]\begin{bmatrix}\check{\theta}_T-\hat{\theta}_T\\ \tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T)\end{bmatrix}$$
$$+\frac{1}{2}\left[(\check{\theta}_T-\hat{\theta}_T)'\ (\tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T))'\right]\begin{bmatrix}\frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta}\\ \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta} & \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\tau}\end{bmatrix}\begin{bmatrix}\check{\theta}_T-\hat{\theta}_T\\ \tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T)\end{bmatrix}$$

where $(\bar{\theta}_T,\bar{\tau}_T)$ is between $(\hat{\theta}_T,\tau_T(\hat{\theta}_T))$ and $(\check{\theta}_T,\tau(\check{\theta}_T))$;

$$\Rightarrow L_T(\check{\theta}_T,\tau_T(\check{\theta}_T))-L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))$$
$$=\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta'}(\check{\theta}_T-\hat{\theta}_T)+\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\tau'}(\tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T))$$
$$+\frac{1}{2}(\check{\theta}_T-\hat{\theta}_T)'\frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta}(\check{\theta}_T-\hat{\theta}_T)+\frac{1}{2}(\tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T))'\frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\tau}(\tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T))$$
$$+(\check{\theta}_T-\hat{\theta}_T)'\frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta}(\tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T)),$$

where

- Under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, by Theorem 1ii (p. 6) and Proposition 2iii (p. 75), $\mathbb{P}$-a.s. as $T\to\infty$, $\check{\theta}_T-\hat{\theta}_T=(\check{\theta}_T-\theta_0)-(\hat{\theta}_T-\theta_0)=O_{\mathbb{P}}(T^{-\frac{1}{2}})+O_{\mathbb{P}}(T^{-\frac{1}{2}})=O_{\mathbb{P}}(T^{-\frac{1}{2}})$;
- Under Assumptions 1 and 2, by Lemma 20 (p. 60), $\mathbb{P}$-a.s. as $T\to\infty$, $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta'}=O(T^{-1})$, so that $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial\theta'}(\check{\theta}_T-\hat{\theta}_T)=O_{\mathbb{P}}(T^{-\frac{3}{2}})$ by the first bullet point;
- Under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, by Lemma 21iv (p. 61) and Theorem 1i (p. 6) and Lemma 28i (p. 82), $\mathbb{P}$-a.s. as $T\to\infty$, there exists $\tilde{\theta}_T$ between $\hat{\theta}_T$ and θ_0 s.t. $\sqrt{T}[\tau_T(\check{\theta}_T)-\tau_T(\hat{\theta}_T)]=\sqrt{T}[\tau_T(\check{\theta}_T)-\tau_T(\theta_0)]-\sqrt{T}[\tau_T(\hat{\theta}_T)-\tau_T(\theta_0)]=-V^{-1}M(\check{\theta}_T-\theta_0)+o_{\mathbb{P}}(1)-\left[-V^{-1}M(\hat{\theta}_T-\theta_0)+o_{\mathbb{P}}(T^{-1/2})\right]=-V^{-1}M(\check{\theta}_T-\theta_0)+V^{-1}M(\hat{\theta}_T-\theta_0)+o_{\mathbb{P}}(1)=-V^{-1}M(\check{\theta}_T-\hat{\theta}_T)+o_{\mathbb{P}}(1)$;

- Under Assumptions 1 and 2, by Lemma 22iv (p. 63), $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial L_T(\hat{\theta}_T, \tau_T(\hat{\theta}_T))}{\partial \tau'} = O(T^{-1})$, so that, by the first and third bullet point, $\frac{\partial L_T(\hat{\theta}_T, \tau_T(\hat{\theta}_T))}{\partial \tau'}(\tau_T(\check{\theta}_T) - \tau_T(\hat{\theta}_T)) = O_{\mathbb{P}}(T^{-3/2})$, under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds;
- Under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, by Lemma 21iv (p. 61) and Theorem 1i (p. 6) and Lemma 28i (p. 82), $\mathbb{P}$-a.s. as $T \to \infty$, there exists $\tilde{\theta}_T$ between $\hat{\theta}_T$ and θ_0 s.t. $\sqrt{T}[\tau_T(\check{\theta}_T) - \tau_T(\hat{\theta}_T)] = \sqrt{T}[\tau_T(\check{\theta}_T) - \tau_T(\theta_0)] - \sqrt{T}[\tau_T(\hat{\theta}_T) - \tau_T(\theta_0)] = -V^{-1}M(\check{\theta}_T - \theta_0) + o_{\mathbb{P}}(1) - \left[-V^{-1}M(\hat{\theta}_T - \theta_0) + o_{\mathbb{P}}(T^{-1/2})\right] = -V^{-1}M(\check{\theta}_T - \theta_0) + V^{-1}M(\hat{\theta}_T - \theta_0) + o_{\mathbb{P}}(1) = -V^{-1}M(\check{\theta}_T - \hat{\theta}_T) + o_{\mathbb{P}}(1)$; and
- under Assumptions 1 and 2, by Lemma 14ii (p. 47), $\mathbb{P}$-a.s. as $T \to \infty$ $\left|\frac{\partial^2 L_T(\theta_T, \tau_T)}{\partial \theta_j \partial \theta_\ell}\right| = o(1)$, so that, by Theorem 1ii (p. 6), $(\check{\theta}_T - \hat{\theta}_T)' \frac{\partial^2 L_T(\bar{\theta}_T, \bar{\tau}_T)}{\partial \theta' \partial \theta}(\check{\theta}_T - \hat{\theta}_T) = o_{\mathbb{P}}(T^{-1})$ by the first bullet point.

Therefore, Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds,

$$
\begin{aligned}
&2\{\ln[\hat{f}_{\theta_T^*}(\hat{\theta}_T)] - \ln[\hat{f}_{\theta_T^*}(\check{\theta}_T)]\} = -2T\left[L_T(\check{\theta}_T, \tau_T(\check{\theta}_T)) - L_T(\hat{\theta}_T, \tau_T(\hat{\theta}_T))\right] \\
&= -\left[V^{-1}M\sqrt{T}(\check{\theta}_T - \hat{\theta}_T)\right]' \frac{\partial^2 L_T(\bar{\theta}_T, \bar{\tau}_T)}{\partial \tau' \partial \tau}\left[V^{-1}M\sqrt{T}(\check{\theta}_T - \hat{\theta}_T)\right] \\
&\quad + 2\sqrt{T}(\theta_0 - \hat{\theta}_T)' \frac{\partial^2 L_T(\bar{\theta}_T, \bar{\tau}_T)}{\partial \tau' \partial \theta}\left[V^{-1}M\sqrt{T}(\theta_0 - \hat{\theta}_T)\right] + o_{\mathbb{P}}(1) \\
&= -\sqrt{T}(\check{\theta}_T - \hat{\theta}_T)'\left[M'V^{-1}\frac{\partial^2 L_T(\bar{\theta}_T, \bar{\tau}_T)}{\partial \tau' \partial \tau}V^{-1}M - 2\frac{\partial^2 L_T(\bar{\theta}_T, \bar{\tau}_T)}{\partial \tau' \partial \theta}V^{-1}M\right]\sqrt{T}(\check{\theta}_T - \hat{\theta}_T) + o_{\mathbb{P}}(1) \\
&= \sqrt{T}(\check{\theta}_T - \hat{\theta}_T)'\Sigma(\theta_0)^{-1}\sqrt{T}(\check{\theta}_T - \hat{\theta}_T) + o_{\mathbb{P}}(1),
\end{aligned}
$$

where the explanations for the convergence are as follow. Firstly, under Assumptions 1 and 2, by Lemma 14ii (p. 47), for any sequence $(\theta_T, \tau_T)_{T \in \mathbf{N}}$ converging to $(\theta_0, \tau(\theta_0))$, $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial^2 L_T(\theta_T, \tau_T)}{\partial \theta' \partial \tau} \to \mathbb{E}\left[\frac{\partial \psi(X_1, \theta_0)}{\partial \theta'}\right] =: M$. Secondly, under Assumptions 1 and 2, by Lemma 25iv (p. 73), $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial L_T(\theta_T, \tau_T)}{\partial \tau \partial \tau'} \to \mathbb{E}[\psi(X_1, \theta_0)\psi(X_1, \theta_0)'] =: V$. Therefore, $\mathbb{P}$-a.s. as $T \to \infty$,

$$
\begin{aligned}
&M'(V')^{-1}\frac{\partial^2 L_T(\bar{\theta}_T, \bar{\tau}_T)}{\partial \tau' \partial \tau}V^{-1}M - 2\frac{\partial^2 L_T(\bar{\theta}_T, \bar{\tau}_T)}{\partial \tau' \partial \theta}V^{-1}M \\
&\to M'(V')^{-1}VV^{-1}M - 2M'V^{-1}M \\
&= -M'V^{-1}M = -\Sigma(\theta_0)^{-1}.
\end{aligned}
$$

(ii) Under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, $\mathbb{P}$-a.s. as $T \to \infty$, addition and subtraction of $\sqrt{T}\theta_0$ yield

$$
\begin{aligned}
\sqrt{T}(\hat{\theta}_T - \check{\theta}_T) &= \sqrt{T}(\hat{\theta}_T - \theta_0) - \sqrt{T}(\check{\theta}_T - \theta_0) \\
&= -M^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1) - \left[M^{-1} - \Sigma R'(R\Sigma R')^{-1}RM^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1)\right] \\
&= \Sigma R'(R\Sigma R')^{-1}RM^{-1}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1)
\end{aligned}
$$

where the explanations for the second equality are the following. Firstly, under Assumptions 1 and 2, by Proposition 1 (p. 44), $\mathbb{P}$-a.s. as $T \to \infty$, $\sqrt{T}(\hat{\theta}_T - \theta_0) = -M\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0) + o_{\mathbb{P}}(1)$.

Secondly, under Assumptions 1, 2 and 3, by Proposition 2i (p. 75), if the test hypothesis (9) on p. 6 holds, $\mathbb{P}$-a.s. as $T \to \infty$, $\sqrt{T}(\check{\theta}_T - \theta_0) = M^{-1} - \Sigma R'(R\Sigma R')^{-1} R M^{-1} \frac{1}{\sqrt{T}} \sum_{t=1}^{T} \psi_t(\theta_0) + o_{\mathbb{P}}(1)$. □

Lemma 25 (Asymptotic limit of $\frac{\partial^2 L_T(\hat{\theta}_T, \tau_T(\hat{\theta}_T))}{\partial \tau' \partial \tau}$). *Under Assumptions 1 and 2, for any sequence $(\theta_T, \tau_T)_{T \in \mathbf{N}}$ converging to $(\theta_0, \tau(\theta_0))$, $\mathbb{P}$-a.s. as $T \to \infty$, for all $(h,k) \in [\![1,m]\!]^2$, $\mathbb{P}$-a.s. as $T \to \infty$,*

(i) $\frac{\partial^2 M_{1,T}(\theta_T, \tau_T)}{\partial \tau_h \partial \tau_k} \to \mathbb{E}[\psi_k(X_1,\theta_0)\psi_h(X_1,\theta_0)]$;
(ii) $\frac{\partial^2 M_{2,T}(\theta_T, \tau_T)}{\partial \tau_h \partial \tau_k} = O(T^{-1})$;
(iii) $\frac{\partial^2 M_{3,T}(\theta_T, \tau_T)}{\partial \tau_h \partial \tau_k} = O(T^{-1})$; *and*
(iv) $\frac{\partial^2 L_T(\theta_T, \tau_T)}{\partial \tau \partial \tau'} \to \mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']$.

Proof. (i) By equation (28) on p. 37, for all $(h,k) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\frac{\partial^2 M_{1,T}(\theta_T, \tau_T)}{\partial \tau_h \partial \tau_k} \\
&= \left(1 - \frac{m}{2T}\right) \frac{1}{\left[\frac{1}{T}\sum_{i=1}^{T} \mathrm{e}^{\tau_T' \psi_i(\theta_T)}\right]^2} \left\{ \left[\frac{1}{T}\sum_{i=1}^{T} \mathrm{e}^{\tau_T' \psi_i(\theta_T)}\right] \left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T' \psi_t(\theta_T)} \psi_{t,h}(\theta_T)\psi_{t,k}(\theta_T)\right] \right. \\
&\left. - \left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T' \psi_t(\theta_T)} \psi_{t,h}(\theta_T)\right] \left[\frac{1}{T}\sum_{i=1}^{T} \mathrm{e}^{\tau_T' \psi_i(\theta_T)} \psi_{i,k}(\theta_T)\right] \right\}.
\end{aligned}$$

where, as $T \to \infty$, $(\theta_T \quad \tau_T) \to (\theta_0 \quad \tau(\theta_0))$ by assumption. Now, under Assumptions 1 and 2, by Lemma 23i-iii (p. 64), for $\overline{B_L}$ a ball around $(\theta_0, \tau(\theta_0))$ of sufficiently small radius, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} \mathrm{e}^{\tau'\psi(X_1,\theta)}\right] < \infty$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)|\right] < \infty$, and $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)|\right]$. Thus, by Assumption 1(a)(b) and (d), the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3), implies that, for all $(h,k) \in [\![1,m]\!]^2$, $\mathbb{P}$-a.s. as $T \to \infty$,

$$\begin{aligned}
&\frac{\partial^2 M_{1,T}(\theta_T, \tau_T)}{\partial \tau_k \partial \tau_h} \\
\to\ & \frac{1}{\mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}]^2} \left\{ \mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}] \mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)} \psi_k(X_1,\theta_0)\psi_h(X_1,\theta_0)] \right. \\
&\left. - \mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_h(X_1,\theta_0)] \mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_k(X_1,\theta_0)] \right\} \\
=\ & \mathbb{E}[\psi_k(X_1,\theta_0)\psi_h(X_1,\theta_0)]
\end{aligned}$$

because $\mathbb{E}[\psi(X_1,\theta_0)] = 0_{m\times 1}$ by Assumption 1(c), and $\tau(\theta_0) = 0_{m\times 1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e) and (g)-(h).

(ii) Under Assumptions 1, by equation (34) on p. 39, $\mathbb{P}$-a.s. for T big enough, for all $(h,k) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\frac{\partial^2 M_{2,T}(\theta_T,\tau_T)}{\partial\tau_h\partial\tau_k} \\
&= -\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\frac{\partial\psi_t(\theta_T)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_{t,k}(\theta_T)\frac{\partial\psi_t(\theta_T)}{\partial\theta'}\right]\right. \\
&\qquad\left.\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\frac{\partial\psi_t(\theta_T)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_{t,h}(\theta_T)\frac{\partial\psi_t(\theta_T)}{\partial\theta'}\right]\right\} \\
&+\frac{1}{T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\frac{\partial\psi_t(\theta_T)}{\partial\theta'}\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_{t,k}(\theta_T)\psi_{t,h}(\theta_T)\frac{\partial\psi_t(\theta_T)}{\partial\theta'}\right]\right\}.
\end{aligned}$$

where, as $T\to\infty$, $(\theta_T \quad \tau_T) \to (\theta_0 \quad \tau(\theta_0))$ by assumption. Now, under Assumptions 1 and 2, by Lemma 23iv-vi (p. 64), for $\overline{B_L}$ a ball around $(\theta_0, \tau(\theta_0))$ of sufficiently small radius, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]$, and $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}}|\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi_h(X_1,\theta)\frac{\partial\psi(X_1,\theta)}{\partial\theta'}|\right]$. Thus, by Assumptions 1(a)(b) and (d), the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3), implies that, for all $(h,k)\in[\![1,m]\!]^2$, $\mathbb{P}$-a.s. as $T\to\infty$,

$$\begin{aligned}
&T\frac{\partial^2 M_{2,T}(\theta_T,\tau_T)}{\partial\tau_h\partial\tau_k} \\
&\to \mathrm{tr}\left\{\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_k(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]\right. \\
&\quad\left.\times\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_h(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]\right\} \\
&\quad+\mathrm{tr}\left\{\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_k(X_1,\theta_0)\psi_h(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]\right\} \\
&= \mathrm{tr}\left\{\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\psi_k(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]\right. \\
&\quad\left.\times\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\psi_h(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]\right\} \\
&\quad+\mathrm{tr}\left\{\mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]^{-1}\mathbb{E}\left[\psi_k(X_1,\theta_0)\psi_h(X_1,\theta_0)\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]\right\}
\end{aligned}$$

because $\tau(\theta_0) = 0_{m\times1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e) and (g)-(h). Therefore, $\mathbb{P}$-a.s. as $T\to\infty$, $\frac{\partial^2 M_{2,T}(\theta_T,\tau_T)}{\partial\tau_h\partial\tau_k} = O(T^{-1})$.

(iii) Under Assumptions 1(a)(b)(e)(g)(h), by equation (39) (p. 43), for all $(h,k) \in [\![1,m]\!]^2$,

$$\begin{aligned}
&\frac{\partial^2 M_{3,T}(\theta_T,\tau_T)}{\partial\tau_h\partial\tau_k}\\
&= \frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_t(\theta_T)\psi_t(\theta_T)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_{t,k}(\theta_T)\psi_t(\theta_T)\psi_t(\theta_T)'\right]\right.\\
&\qquad\left.\times\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_t(\theta_T)\psi_t(\theta_T)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_{t,h}(\theta_T)\psi_t(\theta_T)\psi_t(\theta_T)'\right]\right\}\\
&\quad -\frac{1}{2T}\mathrm{tr}\left\{\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_t(\theta_T)\psi_t(\theta_T)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T'\psi_t(\theta_T)}\psi_{t,k}(\theta_T)\psi_{t,h}(\theta_T)\psi_t(\theta_T)\psi_t(\theta_T)'\right]\right\}
\end{aligned}$$

where, as $T\to\infty$, $(\theta_T \quad \tau_T) \to (\theta_0 \quad \tau(\theta_0))$ by assumption. Now, under Assumptions 1 and 2, by Lemma 23vii-ix (p. 64), there exists a closed ball $\overline{B_L} \subset \mathbf{S}$ centered at $(\theta_0, \tau(\theta_0))$ with strictly positive radius s.t., for all $k \in [\![1,m]\!]$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$, $\mathbb{E}\left[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$, and $\mathbb{E}[\sup_{(\theta,\tau)\in\overline{B_L}} |\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi_h(X_1,\theta)$ $\psi_k(X_1,\theta)\psi(X_1,\theta)\psi(X_1,\theta)'|] < \infty$. Thus, by Assumptions 1(a)(b) and (d), the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3), implies that, for all $(h,k)\in[\![1,m]\!]^2$, $\mathbb{P}$-a.s. as $T\to\infty$,

$$\begin{aligned}
&T\frac{\partial^2 M_{3,T}(\theta_T,\tau_T)}{\partial\tau_h\partial\tau_k}\\
&\to \frac{1}{2}\mathrm{tr}\left\{\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_k(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right.\\
&\quad\left.\times\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_h(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right\}\\
&\quad -\frac{1}{2}\mathrm{tr}\left\{\mathbb{E}\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\right.\\
&\qquad\qquad\left.\times E\left[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}\psi_k(X_1,\theta_0)\psi_h(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right\}\\
&= \frac{1}{2}\mathrm{tr}\left\{\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\psi_k(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right.\\
&\quad\left.\times\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\psi_h(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right\}\\
&\quad -\frac{1}{2}\mathrm{tr}\left\{\mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]^{-1}\mathbb{E}\left[\psi_k(X_1,\theta_0)\psi_h(X_1,\theta_0)\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right]\right\}
\end{aligned}$$

because $\tau(\theta_0) = 0_{m\times 1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e) and (g)-(h). Therefore, $\mathbb{P}$-a.s. as $T\to\infty$, $\frac{\partial M_{3,T}(\hat\theta_T,\tau_T(\hat\theta_T))}{\partial\tau_k} = O(T^{-1})$.

(iv) Under Assumption 1(a)-(b) and (d)-(h), by Lemma 12 (p. 34), for all (θ,τ) in a neighborhood of $(\theta_0,\tau(\theta_0))$, $L_T(\theta,\tau) = M_{1,T}(\theta,\tau) + M_{2,T}(\theta,\tau) + M_{3,T}(\theta,\tau)$, so that the result follows from the statement (i)-(iii) of the present lemma. □

Proposition 2 (Asymptotic normality of $\check\theta_T$, $\tau_T(\check\theta_T)$ and $\check\gamma_T$)**.** *Under Assumptions 1, 2 and 3, if the test hypothesis (9) on p. 6 holds, $\mathbb{P}$-a.s. as $T\to\infty$,*

$$\text{(i)}\ \sqrt{T}\begin{bmatrix}\check{\theta}_T-\theta_0\\ \tau_T(\check{\theta}_T)\\ \check{\gamma}_T\end{bmatrix}=\begin{bmatrix}M^{-1}-\Sigma R'(R\Sigma R')^{-1}RM^{-1}\\ (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1}\\ -(R\Sigma R')^{-1}RM^{-1}\end{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)+o_{\mathbb{P}}(1);\ \textit{and}$$

$$\text{(ii)}\ \sqrt{T}\begin{bmatrix}\check{\theta}_T-\theta_0\\ \tau_T(\check{\theta}_T)\\ \check{\gamma}_T\end{bmatrix}=\begin{bmatrix}\Sigma^{1/2'}P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ (M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ -(R\Sigma R')^{-1}RM^{-1}\end{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\psi_t(\theta_0)+o_{\mathbb{P}}(1)$$

$$\text{(iii)}\ \sqrt{T}\begin{bmatrix}\check{\theta}_T-\theta_0\\ \tau_T(\check{\theta}_T)\\ \check{\gamma}_T\end{bmatrix}\xrightarrow{D}\mathcal{N}\left(0,\begin{bmatrix}(\Sigma^{1/2})'P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{1/2} & 0_{m\times m} & 0_{m\times q}\\ 0_{m\times m} & (V^{1/2})^{-1}P_{\Sigma^{1/2}R'}(V^{1/2'})^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1}\\ 0_{q\times m} & -(R\Sigma R')^{-1}RM^{-1} & (R\Sigma R')^{-1}\end{bmatrix}\right),$$

where $\Sigma := \Sigma(\theta_0) := M^{-1}V(M')^{-1}$, $M := \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]$, $V := \mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']$, *and* $R := \frac{\partial r(\theta_0)}{\partial\theta'}$.

Proof. (i)-(ii) The function $L_T(\theta,\tau)$ is well-defined and twice continuously differentiable in a neighborhood of $(\theta_0'\ \tau(\theta_0)')$ $\mathbb{P}$-a.s. for T big enough by subsection B.2 (p. 32), under Assumptions 1 and 2(a). Similarly, the function $S_T(\theta,\tau) := \frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau'\psi_t(\theta)}\psi_t(\theta)$ and $\theta\mapsto r(\theta)$ are continuously differentiable in a neighborhood of $(\theta_0'\ \tau(\theta_0)')$ by Assumption 1(a)(b) and 3(a). Now, under Assumptions 1, 2, and 3(a), by Lemma 28i (p. 82), Lemma 2iii (p. 20), $\mathbb{P}$-a.s., $\check{\theta}_T\to\theta_0$ and $\tau_T(\check{\theta}_T)\to\tau(\theta_0)$, so that $\mathbb{P}$-a.s. for T big enough, $(\check{\theta}_T'\ \tau_T(\check{\theta}_T)')$ is in any arbitrary small neighborhood of $(\theta_0'\ \tau(\theta_0)')$. Therefore, under Assumptions 1, 2 and 3 (a), stochastic first-order Taylor-Lagrange expansions (Jennrich 1969, Lemma 3) around $(\theta_0,\tau(\theta_0))$ evaluated at $(\check{\theta}_T,\tau_T(\check{\theta}_T))$ yield, $\mathbb{P}$-a.s. for T big enough

$$\frac{\partial L_T(\check{\theta}_T,\tau_T(\check{\theta}_T))}{\partial\theta}=\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta}+\frac{\partial^2L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta}\left(\check{\theta}_T-\theta_0\right)+\frac{\partial^2L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta}\tau_T(\check{\theta}_T)$$

$$S_T(\check{\theta}_T,\tau_T(\check{\theta}_T))=S_T(\theta_0,\tau(\theta_0))+\frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'}\left(\check{\theta}_T-\theta_0\right)+\frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'}\tau_T(\check{\theta}_T)$$

$$r(\check{\theta}_T)=r(\theta_0)+\frac{\partial r(\bar{\theta}_T)}{\partial\theta'}\left(\check{\theta}_T-\theta_0\right)$$

because $\tau(\theta_0)=0_{m\times1}$ by Lemma 10iv (p. 31), and where $\bar{\theta}_T$ and $\bar{\tau}_T$ are between $\check{\theta}_T$ and θ_0, and between $\tau_T(\check{\theta}_T)$ and $\tau(\theta_0)$, respectively. Now, under Assumptions 1 and 2, by definition of $\check{\theta}_T$ and by definition of $\tau_T(.)$ (equation 14 on p. 17), $r(\check{\theta}_T)=0_{q\times1}$ and $S_T(\check{\theta}_T,\tau_T(\check{\theta}_T))=0$, respectively. Moreover, under Assumptions 1, 2 and 3, by Lemma 28iv (p. 82), $\mathbb{P}$-a.s. as $T\to\infty$, $\frac{\partial L_T(\check{\theta}_T,\tau_T(\check{\theta}_T))}{\partial\theta}=-\frac{\partial r(\check{\theta}_T)'}{\partial\theta}\check{\gamma}_T+O(T^{-1})$. Therefore, under Assumptions 1, 2 and 3, $\mathbb{P}$-a.s. as $T\to\infty$,

$$\begin{cases}O(T^{-1})=\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta}+\frac{\partial^2L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta}\left(\check{\theta}_T-\theta_0\right)+\frac{\partial^2L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta}\tau_T(\check{\theta}_T)+\frac{\partial r(\check{\theta}_T)'}{\partial\theta}\check{\gamma}_T\\ 0_{m\times1}=S_T(\theta_0,\tau(\theta_0))+\frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'}\left(\check{\theta}_T-\theta_0\right)+\frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'}\tau_T(\check{\theta}_T)\\ 0_{q\times1}=r(\theta_0)+\frac{\partial r(\bar{\theta}_T)}{\partial\theta'}\left(\check{\theta}_T-\theta_0\right)\end{cases},$$

which in matrix form is

$$\begin{bmatrix}O(T^{-1})\\ 0_{m\times1}\\ 0_{q\times1}\end{bmatrix}=\begin{bmatrix}\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta}\\ S_T(\theta_0,\tau(\theta_0))\\ r(\theta_0)\end{bmatrix}+\begin{bmatrix}\frac{\partial^2L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta} & \frac{\partial^2L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\check{\theta}_T)'}{\partial\theta'}\\ \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'} & \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'} & 0\\ \frac{\partial r(\bar{\theta}_T)}{\partial\theta'} & 0 & 0\end{bmatrix}\begin{bmatrix}\check{\theta}_T-\theta_0\\ \tau_T(\check{\theta}_T)\\ \check{\gamma}_T\end{bmatrix}.$$

Now, under Assumptions 1, 2 and 3, by Lemma 26ii (p. 79), $\mathbb{P}$-a.s. for T big enough, the matrix $\begin{bmatrix} \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\breve{\theta}_T)'}{\partial\theta'} \\ \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'} & \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'} & 0 \\ \frac{\partial r(\bar{\theta}_T)}{\partial\theta'} & 0 & 0 \end{bmatrix}$ is invertible. Then, under Assumptions 1, 2 and 3, solving for the parameters and multiplying by $\sqrt{T}$ yield, $\mathbb{P}$-a.s. as $T \to \infty$,

$$\sqrt{T}\begin{bmatrix} \breve{\theta}_T - \theta_0 \\ \tau_T(\breve{\theta}_T) \\ \breve{\gamma}_T \end{bmatrix}$$

$$= -\begin{bmatrix} \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\breve{\theta}_T)'}{\partial\theta} \\ \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'} & \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'} & 0 \\ \frac{\partial r(\bar{\theta}_T)}{\partial\theta'} & 0 & 0 \end{bmatrix}^{-1} \sqrt{T}\begin{bmatrix} \frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta} + O(T^{-1}) \\ S_T(\theta_0,\tau(\theta_0)) \\ r(\theta_0) \end{bmatrix}$$

$$\stackrel{(a)}{=} -\begin{bmatrix} \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\breve{\theta}_T)'}{\partial\theta} \\ \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'} & \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'} & 0 \\ \frac{\partial r(\bar{\theta}_T)}{\partial\theta'} & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} O(T^{-\frac{1}{2}}) \\ \sqrt{T}\frac{1}{T}\sum_{t=1}^T \psi_t(\theta_0) \\ 0 \end{bmatrix}$$

$$\stackrel{(b)}{=} -\begin{bmatrix} -\Sigma + \Sigma R'(R\Sigma R')^{-1}R\Sigma & M^{-1} - \Sigma R'(R\Sigma R')^{-1}RM^{-1} & \Sigma R'(R\Sigma R')^{-1} \\ (M')^{-1} - (M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1} \\ (R\Sigma R')^{-1}R\Sigma & -(R\Sigma R')^{-1}RM^{-1} & (R\Sigma R')^{-1} \end{bmatrix}$$
$$\times \begin{bmatrix} O(T^{-\frac{1}{2}}) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_t(\theta_0) \\ 0 \end{bmatrix}$$
$$+ \left\{ \begin{bmatrix} -\Sigma + \Sigma R'(R\Sigma R')^{-1}R\Sigma & M^{-1} - \Sigma R'(R\Sigma R')^{-1}RM^{-1} & \Sigma R'(R\Sigma R')^{-1} \\ (M')^{-1} - (M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1} \\ (R\Sigma R')^{-1}R\Sigma & -(R\Sigma R')^{-1}RM^{-1} & (R\Sigma R')^{-1} \end{bmatrix} \right.$$
$$\left. - \begin{bmatrix} \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\breve{\theta}_T)'}{\partial\theta} \\ \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\theta'} & \frac{\partial S_T(\bar{\theta}_T,\bar{\tau}_T)}{\partial\tau'} & 0 \\ \frac{\partial r(\bar{\theta}_T)}{\partial\theta'} & 0 & 0 \end{bmatrix}^{-1} \right\} \begin{bmatrix} O(T^{-\frac{1}{2}}) \\ \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_t(\theta_0) \\ 0 \end{bmatrix}$$

$$\stackrel{(c)}{=} \begin{bmatrix} M^{-1} - \Sigma R'(R\Sigma R')^{-1}RM^{-1} \\ (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} \\ -(R\Sigma R')^{-1}RM^{-1} \end{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_t(\theta_0) + o_{\mathbb{P}}(1)$$

$$\stackrel{(d)}{=} \begin{bmatrix} \Sigma^{1/2'}P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1} \\ (M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1} \\ -(R\Sigma R')^{-1}RM^{-1} \end{bmatrix} \frac{1}{\sqrt{T}}\sum_{t=1}^T \psi_t(\theta_0) + o_{\mathbb{P}}(1)$$

(a) Firstly, under Assumptions 1 and 2, by Lemma 14i (p. 47), $\mathbb{P}$-a.s. as $T \to \infty$, $\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j} = O(T^{-1})$, so that $\sqrt{T}\left[\frac{\partial L_T(\theta_0,\tau(\theta_0))}{\partial\theta_j} + O(T^{-1})\right] = O(T^{-\frac{1}{2}})$. Secondly, note that $S_T(\theta_0,\tau(\theta_0)) = \frac{1}{T}\sum_{t=1}^T \psi_t(\theta_0)$ because $\tau(\theta_0) = 0_{m\times 1}$ by Lemma 10iv (p. 31) under Assumption 1(a)-(e) and (g)-(h). Finally, if the test hypothesis (9) on p. 6 holds, then $r(\theta_0) = 0_{q\times 1}$. *(b)* Add and subtract the

matrix $\begin{bmatrix} -\Sigma+\Sigma R'(R\Sigma R')^{-1}R\Sigma & M^{-1}-\Sigma R'(R\Sigma R')^{-1}RM^{-1} & \Sigma R'(R\Sigma R')^{-1} \\ (M')^{-1}-(M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1} \\ (R\Sigma R')^{-1}R\Sigma & -(R\Sigma R')^{-1}RM^{-1} & (R\Sigma R')^{-1} \end{bmatrix}$.

(c) Firstly, the first and third column of the first square matrix cancel out because the first element and third element of the vector are zeros. Secondly, under Assumptions 1, 2 and 3, by Lemma 26iii (p. 79) and Theorem 1i (p. 6), $\mathbb{P}$-a.s. as $T\to\infty$, the curly bracket is $o(1)$, and, under Assumption 1(a)-(c) and (g), by the Lindeberg-Lévy CLT, $\frac{1}{\sqrt{T}}\sum_{t=1}^T\psi_t(\theta_0)=O_{\mathbb{P}}(1)$, as $T\to\infty$. *(d)* By definition $\Sigma=\Sigma^{1/2'}\Sigma^{1/2}$ and $\Sigma^{-1/2'}=[\Sigma^{1/2'}]^{-1}$. Thus,

- $M^{-1}-\Sigma R'(R\Sigma R')^{-1}RM^{-1}=\Sigma^{1/2'}[I-\Sigma^{1/2}R'(R\Sigma R')^{-1}R\Sigma^{1/2'}]\Sigma^{-1/2'}M^{-1}$ $=\Sigma^{1/2'}P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}$ where $P^{\perp}_{\Sigma^{1/2}R'}$ denotes the orthogonal projection on the orthogonal of the space spanned by the columns of $\Sigma^{1/2}R'$.
- $(M')^{-1}R'(R\Sigma R')^{-1}RM^{-1}=(M')^{-1}\Sigma^{-1/2}[\Sigma^{1/2}R'(R\Sigma R')^{-1}R\Sigma^{1/2'}]\Sigma^{-1/2'}M^{-1}$ $=(M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}=(M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}\Sigma\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}$ $=(M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}=(M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}$ $=(M')^{-1}[V^{1/2}(M')^{-1}]^{-1}P_{\Sigma^{1/2}R'}[M^{-1}V^{1/2'}]^{-1}M^{-1}=(V^{1/2})^{-1}P_{\Sigma^{1/2}R'}(V^{1/2'})^{-1}$ because $M^{-1}V(M')^{-1}=:\Sigma=\Sigma^{1/2'}\Sigma^{1/2}$, so that $\Sigma^{-1/2}:=(\Sigma^{1/2})^{-1}=[V^{1/2}(M')^{-1}]^{-1}=M'V^{-1/2}$ and $\Sigma^{-1/2'}:=(\Sigma^{1/2'})^{-1}=[M^{-1}V^{1/2'}]^{-1}=V^{-1/2'}M$.

(iii) Under Assumptions 1, 2 and 3, by the statement (ii) of the present proposition, $\mathbb{P}$-a.s. as $T\to\infty$,

$$
\begin{aligned}
&\sqrt{T}\begin{bmatrix}\check{\theta}_T-\theta_0\\ \tau_T(\check{\theta}_T)\\ \check{\gamma}_T\end{bmatrix}\\
&=\begin{bmatrix}\Sigma^{1/2'}P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ (M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ -(R\Sigma R')^{-1}RM^{-1}\end{bmatrix}\frac{1}{\sqrt{T}}\sum_{t=1}^T\psi_t(\theta_0)+o(1)\\
&\underset{(a)}{\xrightarrow{D}}-\begin{bmatrix}\Sigma^{1/2'}P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ (M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ -(R\Sigma R')^{-1}RM^{-1}\end{bmatrix}\mathcal{N}(0,V)\\
&\underset{(b)}{\overset{D}{=}}\mathcal{N}\left(0,\begin{bmatrix}\Sigma^{1/2'}P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ (M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1}\\ -(R\Sigma R')^{-1}RM^{-1}\end{bmatrix}V\right.\\
&\qquad\left.\times\begin{bmatrix}(M')^{-1}\Sigma^{-1/2}P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{1/2} & (M')^{-1}\Sigma^{-1/2}P_{\Sigma^{1/2}R'}\Sigma^{-1/2'}M^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1}\end{bmatrix}\right)\\
&\underset{(c)}{\overset{D}{=}}\mathcal{N}\left(0,\begin{bmatrix}(\Sigma^{1/2})'P^{\perp}_{\Sigma^{1/2}R'}\Sigma^{1/2} & 0_{m\times m} & 0_{m\times q}\\ 0_{m\times m} & (V^{1/2})^{-1}P_{\Sigma^{1/2}R'}(V^{1/2'})^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1}\\ 0_{q\times m} & -(R\Sigma R')^{-1}RM^{-1} & (R\Sigma R')^{-1}\end{bmatrix}\right)
\end{aligned}
$$

(a) Under Assumption 1(a)-(c) and (g), by the Lindeberg-Lévy CLT theorem, as $T\to\infty$, $\frac{1}{\sqrt{T}}\sum_{t=1}^T\psi_t(\theta_0)\xrightarrow{D}\mathcal{N}(0,V)$ where $V:=\mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']$. *(b)* Firstly, the minus sign can

be discarded because of the symmetry of the Gaussian distribution. Secondly, if X is a random vector and F is a matrix, then $\mathbb{V}(FX) = F\mathbb{V}(X)F'$. *(c)* Denote the final asymptotic variance matrix with Γ, and its (i,j) block components with $\Gamma_{i,j}$. Then,

- $\Gamma_{1,1} = \Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} V (M')^{-1} \Sigma^{-1/2} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{1/2}$ $= \Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{-1/2'} \Sigma \Sigma^{-1/2} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{1/2} = (\Sigma^{1/2})' P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{1/2}$ because $M^{-1}V(M')^{-1} =: \Sigma = \Sigma^{1/2'}\Sigma^{1/2}$, $\Sigma^{-1/2} := (\Sigma^{1/2})^{-1}$, $\Sigma^{-1/2'} := (\Sigma^{1/2'})^{-1}$, and $P^{\perp}_{\Sigma^{1/2}R'} P^{\perp}_{\Sigma^{1/2}R'} = P^{\perp}_{\Sigma^{1/2}R'}$ by idempotence of projections on linear spaces;
- $\Gamma_{2,2} = (M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} V (M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} =$ $(M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} \Sigma \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1}$ $= (M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} = (M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1}$ $= (M')^{-1} [V^{1/2} (M')^{-1}]^{-1} P_{\Sigma^{1/2}R'} [M^{-1} V^{1/2'}]^{-1} M^{-1}$ $= (V^{1/2})^{-1} P_{\Sigma^{1/2}R'} (V^{1/2'})^{-1}$ because $M^{-1}V(M')^{-1} =: \Sigma = \Sigma^{1/2'}\Sigma^{1/2}$, $\Sigma^{-1/2} := (\Sigma^{1/2})^{-1} = [V^{1/2}(M')^{-1}]^{-1} = M' V^{-1/2}$, $\Sigma^{-1/2'} := (\Sigma^{1/2'})^{-1} = [M^{-1}V^{1/2'}]^{-1} = V^{-1/2'}M$, and $P_{\Sigma^{1/2}R'} P_{\Sigma^{1/2}R'} = P^{\perp}_{\Sigma^{1/2}R'}$ by idempotence;
- $\Gamma_{3,3} = (R\Sigma R')^{-1} R M^{-1} V (M')^{-1} R' (R\Sigma R')^{-1} = (R\Sigma R')^{-1} R\Sigma R' (R\Sigma R')^{-1} = (R\Sigma R')^{-1}$ because $M^{-1}V(M')^{-1} =: \Sigma$;
- $\Gamma_{1,2} = \Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} V (M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} =$ $\Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{-1/2'} \Sigma \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} = \Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} = 0$ because $M^{-1}V(M')^{-1} =: \Sigma = \Sigma^{1/2'}\Sigma^{1/2}$, $\Sigma^{-1/2} := (\Sigma^{1/2})^{-1}$, $\Sigma^{-1/2'} := (\Sigma^{1/2'})^{-1}$, and $P^{\perp}_{\Sigma^{1/2}R'} P_{\Sigma^{1/2}R'} = 0_{m\times m}$;
- $\Gamma_{1,3} = -\Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} V (M')^{-1} R' (R\Sigma R')^{-1}$ $= -\Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{-1/2'} \Sigma R' (R\Sigma R')^{-1} = -\Sigma^{1/2'} P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{1/2} R' (R\Sigma R')^{-1} = 0$ because $M^{-1}V(M')^{-1} =: \Sigma = \Sigma^{1/2'}\Sigma^{1/2}$, $\Sigma^{-1/2'} := (\Sigma^{1/2'})^{-1}$, and $P^{\perp}_{\Sigma^{1/2}R'} \Sigma^{1/2} R' = 0_{m\times q}$;
- $\Gamma_{2,3} = -(M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} M^{-1} V (M')^{-1} R' (R\Sigma R')^{-1}$ $= -(M')^{-1} \Sigma^{-1/2} P_{\Sigma^{1/2}R'} \Sigma^{-1/2'} \Sigma R' (R\Sigma R')^{-1}$ $= -(M')^{-1} \Sigma^{-1/2} [\Sigma^{1/2} R' (R\Sigma R')^{-1} R \Sigma^{1/2'}] \Sigma^{-1/2'} \Sigma R' (R\Sigma R')^{-1}$ $= -(M')^{-1} [R' (R\Sigma R')^{-1} R] \Sigma R' (R\Sigma R')^{-1} = -(M')^{-1} R' (R\Sigma R')^{-1}$ because $M^{-1}V(M')^{-1} =: \Sigma$ and $P_{\Sigma^{1/2}R'} = [\Sigma^{1/2} R' (R\Sigma R')^{-1} R \Sigma^{1/2'}]$.

□

Lemma 26. *Using the notation of Proposition 2 (p. 75), under Assumptions 1, 2 and 3,*

(i) *for any sequence $(\theta_T, \tau_T)_{T\in\mathbf{N}}$ converging to $(\theta_0, \tau(\theta_0))$, $\mathbb{P}$-a.s. as $T\to\infty$,*

$$\begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\bar\theta_T)'}{\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} & 0 \\ \frac{\partial r(\bar\theta_T)}{\partial\theta'} & 0 & 0 \end{bmatrix} \to \begin{bmatrix} 0_{m\times m} & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]' & \frac{\partial r(\theta_0)'}{\partial\theta} \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right] & \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right] & 0 \\ \frac{\partial r(\theta_0)}{\partial\theta'} & 0 & 0 \end{bmatrix};$$

(ii) $$\begin{bmatrix} 0_{m\times m} & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]' & \frac{\partial r(\theta_0)'}{\partial\theta} \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right] & \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right] & 0 \\ \frac{\partial r(\theta_0)}{\partial\theta'} & 0 & 0 \end{bmatrix}$$ *is invertible, so that, for any sequence $(\theta_T, \tau_T)_{T\in\mathbf{N}}$ converging to $(\theta_0, \tau(\theta_0))$, $\mathbb{P}$-a.s., for T big enough, the matrix*

$$\begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\bar\theta_T)'}{\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} & 0 \\ \frac{\partial r(\bar\theta_T)}{\partial\theta'} & 0 & 0 \end{bmatrix} \textit{ is invertible; and}$$

(iii) *for any sequence* $(\theta_T,\tau_T)_{T\in\mathbf{N}}$ *converging to* $(\theta_0,\tau(\theta_0))$, $\mathbb{P}$*-a.s. as* $T\to\infty$,

$$\begin{bmatrix} \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'\partial\theta} & \frac{\partial^2 L_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'\partial\theta} & \frac{\partial r(\bar\theta_T)'}{\partial\theta} \\ \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\theta'} & \frac{\partial S_T(\bar\theta_T,\bar\tau_T)}{\partial\tau'} & 0 \\ \frac{\partial r(\bar\theta_T)}{\partial\theta'} & 0 & 0 \end{bmatrix}^{-1}$$

$$\to \begin{bmatrix} -\Sigma+\Sigma R'(R\Sigma R')^{-1}R\Sigma & M^{-1}-\Sigma R'(R\Sigma R')^{-1}RM^{-1} & \Sigma R'(R\Sigma R')^{-1} \\ (M')^{-1}-(M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1} \\ (R\Sigma R')^{-1}R\Sigma & -(R\Sigma R')^{-1}RM^{-1} & (R\Sigma R')^{-1} \end{bmatrix}.$$

Proof. (i) Under Assumptions 1, 2 and 3(a), it follows from the continuity of $\frac{\partial r(.)}{\partial\theta'}$, which is implied by Assumption 3(a), and Lemma 14ii and iii (p. 47) and Lemma 17 (p. 49), given that $\tau(\theta_0)=0_{m\times1}$ by Lemma 10ii (p. 31) and Assumption 1(c), under Assumption 1(a)(b)(d)(e)(g) and (h).

(ii) It is sufficient to check the assumptions of Corollary 2i (p. 88) with $A=\begin{bmatrix}0_{m\times m} & M'\\ M & V\end{bmatrix}$ and $B=\begin{bmatrix}R'\\ 0_{m\times q}\end{bmatrix}$ in order to establish the first part of the statement. Firstly, under Assumptions 1 and 2, by Lemma 13iii (p. 46), $A=\begin{bmatrix}0_{m\times m} & M'\\ M & V\end{bmatrix}$ is invertible. Secondly, by Assumptions 1(h) and 3(b), $(B'A^{-1}B)=-(R\Sigma R')$ is also invertible. Then, the second part of the statement follows from a trivial case of the Lemma 30 (p. 87).

(iii) Under Assumption 1(a)(b)(c)(d)(e)(g)(h), by the statement (ii) of the present lemma, the limiting matrix is invertible. Thus, using the notation of Proposition 2 (p. 75),

$$\begin{bmatrix} 0_{m\times m} & \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]' & \frac{\partial r(\theta_0)'}{\partial\theta} \\ \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right] & \mathbb{E}\left[\psi(X_1,\theta_0)\psi(X_1,\theta_0)'\right] & 0 \\ \frac{\partial r(\theta_0)}{\partial\theta'} & 0 & 0 \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} 0_{m\times m} & M' & R' \\ M & V & 0_{m\times q} \\ R & 0_{q\times m} & 0_{q\times q} \end{bmatrix}^{-1}$$

$$= \begin{bmatrix} -\Sigma+\Sigma R'(R\Sigma R')^{-1}R\Sigma & M^{-1}-\Sigma R'(R\Sigma R')^{-1}RM^{-1} & \Sigma R'(R\Sigma R')^{-1} \\ (M')^{-1}-(M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} & -(M')^{-1}R'(R\Sigma R')^{-1} \\ (R\Sigma R')^{-1}R\Sigma & -(R\Sigma R')^{-1}RM^{-1} & (R\Sigma R')^{-1} \end{bmatrix}$$

where the explanation for the last equality is as follows. Apply Corollary 2ii (p. 88) with $A = \begin{bmatrix} 0_{m\times m} & M' \\ M & V \end{bmatrix}$ and $B = \begin{bmatrix} R' \\ 0_{m\times q} \end{bmatrix}$, and note that, by Lemma 27iii, iv and vi (p. 81),

$$
\begin{aligned}
A^{-1} - A^{-1}B(B'A^{-1}B)B'A^{-1} &= \begin{bmatrix} -\Sigma + \Sigma R'(R\Sigma R')^{-1}R\Sigma & M^{-1} - \Sigma R'(R\Sigma R')^{-1}RM^{-1} \\ (M')^{-1} - (M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} \end{bmatrix} \\
A^{-1}B(B'A^{-1}B)^{-1} &= \begin{bmatrix} \Sigma R'(R\Sigma R')^{-1} \\ -(M')^{-1}R'(R\Sigma R')^{-1} \end{bmatrix} \\
(B'A^{-1}B)^{-1} &= -(R\Sigma R')^{-1}.
\end{aligned}
$$

Then, the result follows from the continuity of the inverse transformation (e.g., Rudin 1953, Theorem 9.8). □

Lemma 27. *Let* $A = \begin{bmatrix} 0_{m\times m} & M' \\ M & V \end{bmatrix}$ *and* $B = \begin{bmatrix} R' \\ 0_{m\times q} \end{bmatrix}$ *where* $\Sigma := \Sigma(\theta_0) := M^{-1}V(M')^{-1}$, $M := \mathbb{E}\left[\frac{\partial\psi(X_1,\theta_0)}{\partial\theta'}\right]$, $V := \mathbb{E}[\psi(X_1,\theta_0)\psi(X_1,\theta_0)']$, *and* $R = \frac{\partial r(\theta_0)}{\partial\theta'}$. *Then, under Assumption 1(a)(b)(h) and 3(b), the following equalities hold*

(i) $A^{-1} = \begin{bmatrix} -\Sigma & M^{-1} \\ (M')^{-1} & 0_{m\times m} \end{bmatrix}$;

(ii) $A^{-1}B = \begin{bmatrix} -\Sigma R' \\ (M')^{-1}R' \end{bmatrix}$, *so that* $B'A^{-1} = \begin{bmatrix} -R\Sigma & RM^{-1} \end{bmatrix}$;

(iii) $(B'A^{-1}B)^{-1} = -(R\Sigma R')^{-1}$;

(iv) $A^{-1}B(B'A^{-1}B)^{-1} = \begin{bmatrix} \Sigma R'(R\Sigma R')^{-1} \\ -(M')^{-1}R'(R\Sigma R')^{-1} \end{bmatrix}$;

(v) $A^{-1}B(B'A^{-1}B)B'A^{-1} = \begin{bmatrix} -\Sigma R'(R\Sigma R')^{-1}R\Sigma & \Sigma R'(R\Sigma R')^{-1}RM^{-1} \\ (M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & -(M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} \end{bmatrix}$; *and*

(vi) $A^{-1} - A^{-1}B(B'A^{-1}B)B'A^{-1}$
$= \begin{bmatrix} -\Sigma + \Sigma R'(R\Sigma R')^{-1}R\Sigma & M^{-1} - \Sigma R'(R\Sigma R')^{-1}RM^{-1} \\ (M')^{-1} - (M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & (M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} \end{bmatrix}$.

Proof. *(i)* It corresponds to a part of Lemma 13iii (p. 46) under Assumptions 1 and 2.

(ii)

$$\begin{bmatrix} -\Sigma & M^{-1} \\ (M')^{-1} & 0_{m\times m} \end{bmatrix}\begin{bmatrix} R' \\ 0_{m\times q} \end{bmatrix} = \begin{bmatrix} -\Sigma R' \\ (M')^{-1}R' \end{bmatrix} = A^{-1}B$$

(iii)

$$\begin{bmatrix} R & 0_{q\times m} \end{bmatrix}\begin{bmatrix} -\Sigma R' \\ (M')^{-1}R' \end{bmatrix} = \begin{bmatrix} -R\Sigma R' \end{bmatrix} = B'A^{-1}B$$

(iv)

$$\begin{bmatrix} -\Sigma R' \\ (M')^{-1}R' \end{bmatrix} \begin{bmatrix} -(R\Sigma R')^{-1} \end{bmatrix} = \begin{bmatrix} \Sigma R'(R\Sigma R')^{-1} \\ -(M')^{-1}R'(R\Sigma R')^{-1} \end{bmatrix} = A^{-1}B(B'A^{-1}B)^{-1}$$

(v)

$$\begin{bmatrix} \Sigma R'(R\Sigma R')^{-1} \\ -(M')^{-1}R'(R\Sigma R')^{-1} \end{bmatrix} \begin{bmatrix} -R\Sigma & RM^{-1} \end{bmatrix}$$
$$= \begin{bmatrix} -\Sigma R'(R\Sigma R')^{-1}R\Sigma & \Sigma R'(R\Sigma R')^{-1}RM^{-1} \\ (M')^{-1}R'(R\Sigma R')^{-1}R\Sigma & -(M')^{-1}R'(R\Sigma R')^{-1}RM^{-1} \end{bmatrix} = A^{-1}B(B'A^{-1}B)B'A^{-1}$$

□

Lemma 28 (Constrained estimator and its Lagrangian)**.** *Under Assumptions 1, 2 and 3(a), if the test hypothesis* (9) *on p. 6 holds,* $\mathbb{P}$*-a.s. for* T *big enough,*

(i) *the constrained estimator* $\check{\theta}_T$ *exists, and* $\check{\theta}_T \to \theta_0$*, as* $T \to \infty$*;*

(ii) $\theta \mapsto L_T(\theta, \tau_T(\theta))$ *is continuously differentiable in a neighborhood of* $\check{\theta}_T$*;*

(iii) *under additional Assumption 3(b), there exists a unique vector,* $\check{\gamma}_T$*, called the Lagrangian multiplier, s.t.* $\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial\theta}\right|_{\theta=\check{\theta}_T} + \frac{\partial r(\check{\theta}_T)'}{\partial\theta}\check{\gamma}_T = 0_{m\times 1}$*;*

(iv) *under additional Assumption 3(b),* $\frac{\partial L_T(\check{\theta}_T,\tau_T(\check{\theta}_T))}{\partial\theta} + \frac{\partial r(\check{\theta}_T)'}{\partial\theta}\check{\gamma}_T = O(T^{-1})$*, as* $T \to \infty$*, where* $\frac{\partial L_T(\check{\theta}_T,\tau_T(\check{\theta}_T))}{\partial\theta} := \left.\frac{\partial L_T(\theta,\tau)}{\partial\theta}\right|_{(\theta,\tau)=(\check{\theta}_T,\tau_T(\check{\theta}_T))}$.

Proof. (i) The constrained set $\tilde{\mathbf{\Theta}} := \{\theta \in \mathbf{\Theta} : r(\theta) = 0\}$ is bounded as a subset of the compact (and thus bounded) set $\mathbf{\Theta}$. The constrained set $\tilde{\mathbf{\Theta}}$ is also closed: For all $(\theta_n)_{n\in\mathbf{N}} \in \tilde{\mathbf{\Theta}}^{\mathbf{N}}$ s.t. $\lim_{n\to}\theta_n = \bar{\theta}$, $\bar{\theta} \in \tilde{\mathbf{\Theta}}$ because (i) by compactness of $\mathbf{\Theta}$, $\bar{\theta} \in \mathbf{\Theta}$; and (ii) by the continuity of $r : \mathbf{\Theta} \to \mathbf{R}^q$ (i.e., Assumption 3(a)), $r(\bar{\theta}) = \lim_{n\to\infty} r(\theta_n) = \lim_{n\to\infty} 0 = 0$. Therefore, the constrained set $\tilde{\mathbf{\Theta}}$ is itself compact. Moreover, under Assumption 1(a)(b) and (d)-(h), by Lemma 1ii-iii (p. 19), $\mathbb{P}$-a.s. for T big enough, $\theta \mapsto \hat{f}_{\theta_T^*}(\theta)$ is continuous and, for all $\theta \in \mathbf{\Theta}$, $\omega \mapsto \hat{f}_{\theta_T^*}(\theta)$ is measurable. Thus, the existence and the measurability of the constrained estimator $\check{\theta}_T$ follows from the Schmetterer-Jennrich lemma (Schmetterer 1966 Chap. 5 Lemma 3.3; Jennrich 1969 Lemma 2).

In order to establish the consistency of $\check{\theta}_T$, it remains to check the other assumptions of the standard consistency theorem (e.g. Newey and McFadden 1994, pp. 2121-2122 Theorem 2.1, which is also valid in an almost-sure sense), where the constrained set $\tilde{\mathbf{\Theta}} := \{\theta \in \mathbf{\Theta} : r(\theta) = 0\}$ is the parameter space. Because $\tilde{\mathbf{\Theta}} \subset \mathbf{\Theta}$, $\mathbb{P}$-a.s. as $T \to \infty$,

$$\sup_{\theta\in\tilde{\mathbf{\Theta}}}\left|\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det} - \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right|$$
$$\leqslant \sup_{\theta\in\mathbf{\Theta}}\left|\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right] - \frac{1}{2T}\ln|\Sigma_T(\theta)|_{\det} - \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]\right| \to 0$$

where the convergence to zero follows from equation (17) on p. 19, under Assumption 1. In addition, under Assumption 1 (a)-(e) and (g)-(h), by Lemma 10iv (p. 31), $\theta \mapsto \ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}]$ is uniquely maximized at θ_0, i.e., for all $\theta \in \mathbf{\Theta}\setminus\{\theta_0\}$, $\ln\mathbb{E}[\mathrm{e}^{\tau(\theta)'\psi(X_1,\theta)}] < \ln\mathbb{E}[\mathrm{e}^{\tau(\theta_0)'\psi(X_1,\theta_0)}] = 0$, and, under Assumptions 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22), $\theta \mapsto \ln \mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]$ is continuous in $\tilde{\boldsymbol{\Theta}} \subset \boldsymbol{\Theta}$.

(ii) Under Assumptions 1 and 2(a), by subsection B.2 (p. 32), the function $L_T(\theta,\tau)$ is well-defined and twice continuously differentiable in a neighborhood of $(\theta_0' \ \tau(\theta_0)')$ $\mathbb{P}$-a.s. for T big enough. Moreover, under Assumption 1(a)(b) and (d)-(h), by Lemma 21i (p. 61), $\tau_T(.)$ is continuously differentiable in $\boldsymbol{\Theta}$. Now, under Assumption 1, by the statement (i) of the present lemma and Lemma 2iii (p. 20), $\mathbb{P}$-a.s., $\check{\theta}_T \to \theta_0$ and $\tau_T(\check{\theta}_T) \to \tau(\theta_0)$, so that $\mathbb{P}$-a.s. for T big enough, $(\check{\theta}_T' \ \tau_T(\check{\theta}_T)')$ is in any arbitrary small neighborhood of $(\theta_0' \ \tau(\theta_0)')$. Therefore, under Assumption 1 and 2(a), by the chain rule theorem (e.g., Magnus and Neudecker 1999/1988, Chap. 5 sec. 11), $\mathbb{P}$-a.s. for T big enough, $\theta \mapsto L_T(\theta,\tau_T(\theta))$ is continuously differentiable at $\check{\theta}_T$.

(iii) It is a consequence of the Lagrange theorem (e.g., Magnus and Neudecker 1999/1988, Chap. 7 sec. 12). Check its assumptions. Firstly, under Assumptions 1 and 2, $\mathbb{P}$-a.s. by the statement (i) of the present lemma, $\mathbb{P}$-a.s. for T big enough, the constrained estimator $\check{\theta}_T$ exists and that it is in the interior of $\boldsymbol{\Theta}$ by consistency and Assumption 1(c). Then, we should check the other assumptions of the Lagrange theorem ω by ω on the subset of $\boldsymbol{\Omega}$ where $\check{\theta}_T$ exists. Firstly, by Assumption 3(a), $r : \boldsymbol{\Theta} \to \mathbf{R}^q$ is continuously differentiable. Secondly, under Assumptions 1,2 and 3(a), if the test hypothesis (9) on p. 6 holds, $\mathbb{P}$-a.s. as $T \to \infty$, $\check{\theta}_T \to \theta_0$, and, by Assumption 3(b), $\frac{\partial r(\theta_0)}{\partial \theta'}$ is full rank, Thus, $\mathbb{P}$-a.s. for T big enough, $\frac{\partial r(\check{\theta}_T)}{\partial \theta'}$ is full rank by continuity of the determinant function. Finally, by the statement (iv) of the present lemma $\theta \mapsto L_T(\theta,\tau_T(\theta))$ is differentiable at $\check{\theta}_T$.

(iv) First of all, note that it does *not* immediately follow from the statement (iii) because $\frac{\partial L_T(\hat{\theta}_T,\tau_T(\hat{\theta}_T))}{\partial \theta}$ denotes $\left.\frac{\partial L_T(\theta,\tau)}{\partial \theta}\right|_{(\theta,\tau)=(\check{\theta}_T,\tau_T(\check{\theta}_T))}$ instead of $\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial \theta}\right|_{\theta=\hat{\theta}_T}$ (see footnote 14 on p. 61). Under Assumption 1(a)(b) and (d)-(h), by Lemma 21i (p. 61), $\tau_T(.)$ is continuously differentiable in $\boldsymbol{\Theta}$. Moreover, under Assumptions 1, 2 and 3(a), if by the statement (ii) of the present lemma, $\mathbb{P}$-a.s. for T big enough, $\theta \mapsto L_T(\theta,\tau_T(\theta))$ is continuously differentiable in a neighborhood of $\check{\theta}_T$. Thus, by an immediate and standard implication of the chain rule (e.g., Magnus and Neudecker 1999/1988, chap. 5, sec. 12, exercise 3), $\mathbb{P}$-a.s. for T big enough, for all $j \in [\![1,m]\!]$,

$$\begin{aligned}
\left.\frac{\partial L_T(\theta,\tau_T(\theta))}{\partial \theta_j}\right|_{\theta=\check{\theta}_T} &= \left.\frac{\partial L_T(\theta,\tau)}{\partial \theta_j}\right|_{(\theta,\tau)=(\check{\theta}_T,\tau_T(\check{\theta}_T))} + \left.\frac{\partial L_T(\theta,\tau)}{\partial \tau'}\right|_{(\theta,\tau)=(\check{\theta}_T,\tau_T(\check{\theta}_T))} \left.\frac{\partial \tau(\theta)}{\partial \theta_j}\right|_{\theta=\check{\theta}_T} \\
&= \left.\frac{\partial L_T(\theta,\tau)}{\partial \theta_j}\right|_{(\theta,\tau)=(\check{\theta}_T,\tau_T(\check{\theta}_T))} + O(T^{-1}) \qquad (44)
\end{aligned}$$

where the explanations for the last equality are as follow. Firstly, under Assumptions 1, 2 and 3(a), by Lemma 22iv (p. 63), $\mathbb{P}$-a.s. as $T \to \infty$, $\left.\frac{\partial L_T(\theta,\tau)'}{\partial \tau}\right|_{(\theta,\tau)=(\check{\theta}_T,\tau_T(\check{\theta}_T))} = O(T^{-1})$ because $\check{\theta}_T \to \theta_0$, $\mathbb{P}$-a.s. as $T \to \infty$, by the second part of the statement (i) of the present lemma. Secondly, under Assumptions 1, 2 and 3(a), by the second part of the statement (i) of the present lemma and Lemma 21iii (p. 61), $\mathbb{P}$-a.s. as $T \to \infty$, $\left.\frac{\partial \tau(\theta)}{\partial \theta_j}\right|_{\theta=\check{\theta}_T} = O(1)$.

Now the results follows by plugging the above equation (44) into the Lagrangian FOC of the statement (iii) of the present lemma. □

Appendix C. On the assumptions

C.1. **Discussion.** Assumptions 1 and 2 are mainly adapted from the entropy literature. Assumption 1(a) ensures the basic requirement for inference, that is, data contain different pieces of information (independence) about the same phenomenon (identically distributed). The conditions "independence and identically distributed" are much stronger than needed, and can be relaxed to allow for time dependence along the lines of Kitamura and Stutzer (1997). We restrain ourself to the i.i.d. case for brevity and clarity. Assumption 1(a) also requires completeness of the probability space so that we can define functions only a probability-one subset of $\mathbf{\Omega}$ without generating potential measurability complications. The completeness of the probability space is without significant loss of generality (e.g., Kallenberg 2002 (1997, p. 13), and it is often implicitly or explicitly required in the literature.

Assumption 1(b) mainly requires standard regularity conditions for the moment function $\psi(.,.)$. As usual in nonlinear econometrics, the existence of the estimator relies on such regularity conditions. An alternative would be to rely on empirical process theory, but it seems here inappropriate as the implicit nature of the definition of the ESP approximation requires smooth functions. We require Assumption 1(b), as well as some of the following assumptions, to hold in an ϵ-neighborhood of the parameter space $\mathbf{\Theta}$, so that we can deal with its boundary $\partial\mathbf{\Theta}$ in the same way as with its interior. In particular, it ensures that $\Sigma(\theta)$ is invertible for $\theta \in \partial\mathbf{\Theta}$ under probability measures equivalent to $\mathbb{P}$ (Corollary 1ii on p. 86), and it allows to apply an implicit function theorem to $\tau(\theta)$, also for $\theta \in \partial\mathbf{\Theta}$ (Lemma 10 on p. 31). For the latter reason, the entropy literature often appears to also (implicitly) assume that assumptions hold in an ϵ-neighborhood of the parameter space. In applications, this is often innocuous as the boundary of the parameter space is often loosely specified. However, in some specific situations, which we rule out, this may be problematic (e.g., Andrews 1999, and references therein).

Assumption 1(c) requires global identification, which is a necessary condition to prove the consistency of an estimator. If we were interested in the ESP approximation instead of its maximizer (i.e., the ESP estimator), global identification could be relaxed as Holcblat (2012) and a companion paper show. Assumption 1(c) also requires equality between the dimension of the parameter space and the number of moment conditions, i.e., just-identified moment conditions. We impose the latter for mainly three reasons. Firstly, it appears reasonable to investigate the ESP estimator in the just-identified case before moving to the over-identified case, which requires to generalize the ESP approximation. Secondly, the just-identified case makes clear the difference between the ESP estimator and the existing alternatives, which are all equal in this case (see section 2.2). Thirdly, this is a standard assumption in the saddlepoint literature. However, note that (i) this assumption is less restrictive than it seems at first sight because, in the linear case, over-identified moment conditions correspond to just-identified moment conditions through the FOCs, and, in the nonlinear case, we can transform over-identified estimating equations into just-identified estimating equations through an extension of the parameter space (e.g., Newey and McFadden 1994, p. 2232); (ii) ongoing work show how to generalize the ESP approximation to over-identified moment conditions.

Assumption 1(d) requires the compactness of the parameter space $\mathbf{\Theta}$, and the existence of a solution $\tau(\theta) \in \mathbf{R}^m$ that solves the equation $\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\right] = 0$, for all $\theta \in \mathbf{\Theta}$. Schennach (2005) also makes this assumption. Compactness of the parameter space is a convenient standard mathematical assumption that is often relevant in practice. A computer can only handle a bounded parameter space —finite memory of a computer. Regarding the existence of $\tau(\theta)$, it is necessary to ensure the asymptotic existence of the ESP approximation. From a theoretical point of view, the existence of $\tau(\theta)$ looks like a reasonable assumption: If, for some $\theta \in \mathbf{\Theta}$, $0_{m\times 1}$ is outside the convex hull of the support of $\psi(X_1,\theta)$, there is not such a solution $\tau(\theta)$, which also means that θ cannot be θ_0, so that it should be excluded from the parameter space. However, the existence of $\tau(\theta)$ might be difficult to check in practice. A way to get around this assumption is to (i) assume the existence of $\tau(\theta)$ only in a neighbohood of θ_0; and (ii) to set the ESP approximation to zero for the θ values that do not have a solution to the finite-sample moment conditions (14). Holcblat (2012) follows such an approach. We do not follow such an approach because it significantly complicates the proofs and the presentation.

Assumptions 1(e) and 2(b) rule out fat-tailed distributions. More precisely, they require the existence of exponential moments. They are necessary to apply the the ULLN (uniform law of large numbers) à la Wald (e.g., Ghosh and Ramamoorthi 2003, pp. 24-25, Theorem 1.3.3) to components of the ESP approximation. Assumptions 1(e) and 2(b) are stronger than the moment existence assumption in Hansen (1982), but they are a common type of assumptions in the entropy literature (e.g., Haberman 1984, Kitamura and Stutzer 1997, Schennach 2007), the saddlepoint literature (e.g., Almudevar et al. 2000) and the literature on exponential models (e.g., Berk 1972). In particular, Assumptions 1(d) and 2(b) are a convenient variant of Assumptions 3.4 and 3.5 in Schennach (2007). Both in Schennach (2007) and in the present paper, the successful estimation of the Hall and Horowitz model, which does not satisfy Assumptions 1(e) and 2(b), suggests that the latter can be relaxed. In practice, Assumptions 1(e) and 2(b) are not as strong as it may appear because observable quantities have finite support (finite memory of computers), which, in turn, implies that they have all finite moments. Moreover, in the case in which unboundedness is a concern (e.g., moment conditions derived from a likelihood), Ronchetti and Trojani (2001) provide a way to bound moment functions.

Assumptions 1(f) and (g) play the same role as Assumptions 1(e) and 2(b), although they are less stringent. Assumption 1(h) requires the invertibility of the asymptotic variance of standard estimators (scaled by $\sqrt{T}$) of any solution to the tilted moment condition. In the present paper, this assumption has two main roles. Firstly, it ensures that the determinant term $|\Sigma_T(\theta)|_{\det}^{-\frac{1}{2}}$ in the ESP approximation (11) does not explode, asymptotically. Secondly, it ensures the positive definiteness of the symmetric matrix $\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ for all $(\theta,\tau)\in\mathbf{S}$, so that the $\min_{\tau\in\mathbf{R}^m}\mathbb{E}[e^{\tau'\psi(X_1,\theta)}]$ is a strictly convex problem, which, in turn, implies the unicity of its solution $\tau(\theta)$. In the setup of the present paper, Assumption 1(g) is equivalent to the invertibility of $\mathbb{E}\left[e^{\tau(\theta)'\psi(X_1,\theta)}\frac{\partial\psi(X_1,\theta)}{\partial\theta'}\right]$ and $\mathbb{E}\left[\psi(X_1,\theta)\psi(X_1,\theta)'\right]$, for all $\theta\in\mathbf{\Theta}$ (Lemma 29 on p. 86 with $\mathrm{P}=\mathbb{P}$ and $\frac{\mathrm{dQ}}{\mathrm{dP}}=\frac{1}{e^{\tau(\theta)'\psi(X_1,\theta)}}$). In this way, it is stronger than the Assumption 4 in Kitamura and Stutzer (1997), but it is close to Stock and Wright (2000, Assumption C). Note that Schennach (2007) also implicitly assumes that $\mathbb{E}\left[e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is full rank for all $(\theta,\tau)\in\mathbf{S}$, because Schennach (2007, p. 649) regards $\tau(\theta)$ as a solution to a strictly convex problem (e.g., Hiriart-Urruty and Lemaréchal 1993/1996, chap. 4, Theorem 4.3.1). Assumption 1(g) should often be reasonable because the set of singular matrices has zero Lebesgue measure in the space of square matrices.[15]

C.2. **Implications of Assumption 1(h).**

Lemma 29. *Let $(\mathbf{\Omega}_A, \mathcal{A})$ be a measurable space, $Z : \mathbf{\Omega} \to \mathbf{R}^k$ be a k-dimensional random vectors with $k \in [\![1, \infty[\![$ and P and Q two probability measures on $(\mathbf{\Omega}_A, \mathcal{A})$. Denote the expectation and the variance under P with $\mathbb{E}_{\mathrm{P}}$ and $\mathbb{V}_{\mathrm{P}}$, respectively.*

(i) *For all $\tau \in \mathbf{R}^k$, $\mathbb{E}_{\mathrm{P}}\left(\mathrm{e}^{\tau' Z} Z Z'\right) \geqslant 0$, it is a positive semi-definite symmetric matrix.*

(ii) *If $\mathrm{P} \sim \mathrm{Q}$ (i.e., they are equivalent), $\mathbb{E}_{\mathrm{P}}(|ZZ'|) < \infty$ and $\mathbb{E}_{\mathrm{Q}}(|ZZ'|) < \infty$, then*

$$\mathbb{E}_{\mathrm{P}}(ZZ') \text{ invertible} \Leftrightarrow \mathbb{E}_{\mathrm{Q}}(ZZ') \text{ invertible}$$

Proof. *(i)* Symmetry follows from the invariance under transposition of $\mathbb{E}_{\mathrm{P}}\left(ZZ'\mathrm{e}^{\tau' Z}\right)$. It remains to show positive semi-definiteness. For all $y \in \mathbf{R}^k$,

$$\begin{aligned} \forall \omega \in \mathbf{\Omega}, \quad y'\mathrm{e}^{\tau' Z} Z Z' y &= \mathrm{e}^{\tau' Z}[y'Z]^2 \geqslant 0 \\ \Rightarrow y'\mathbb{E}_{\mathrm{P}}\left[\mathrm{e}^{\tau' Z} Z Z'\right] y &= \mathbb{E}_{\mathrm{P}}\left[y'\mathrm{e}^{\tau' Z} Z Z' y\right] \geqslant 0. \end{aligned}$$

where the implication follows from the monotonicity of the Lebesgue integral (e.g., Monfort 1997, p. 47).

(ii) By contraposition, it is equivalent to prove that $\mathbb{E}_{\mathrm{P}}(ZZ')$ noninvertible iff $\mathbb{E}_{\mathrm{Q}}(ZZ')$ noninvertible. By statement (i),

$$\begin{aligned} & \mathbb{E}_{\mathrm{P}}(ZZ') \text{ noninvertible} \\ \Leftrightarrow\ & \exists y \in \mathbf{R}^k \setminus \{0_{k\times 1}\} : y'\mathbb{E}_{\mathrm{P}}(ZZ')y = 0 \\ \overset{(a)}{\Leftrightarrow}\ & \exists y \in \mathbf{R}^k \setminus \{0_{k\times 1}\} : \mathbb{E}_{\mathrm{P}}[(y'Z)^2] = 0 \\ \overset{(b)}{\Leftrightarrow}\ & \exists y \in \mathbf{R}^k \setminus \{0_{k\times 1}\} : (y'Z)^2 = 0 \text{ P-a.s.} \\ \overset{(c)}{\Leftrightarrow}\ & \exists y \in \mathbf{R}^k \setminus \{0_{k\times 1}\} : (y'Z)^2 = 0 \text{ Q-a.s.} \\ \overset{(d)}{\Leftrightarrow}\ & \exists y \in \mathbf{R}^k \setminus \{0_{k\times 1}\} : \mathbb{E}_{\mathrm{Q}}[(y'Z)^2] = 0 \\ \overset{(e)}{\Leftrightarrow}\ & \exists y \in \mathbf{R}^k \setminus \{0_{k\times 1}\} : y'\mathbb{E}_{\mathrm{Q}}(ZZ')y = 0 \\ \Leftrightarrow\ & \mathbb{E}_{\mathrm{Q}}(ZZ') \text{ noninvertible} \end{aligned}$$

(a) $y'\mathbb{E}_{\mathrm{P}}(ZZ')y = \mathbb{E}_{\mathrm{P}}[y'Z(y'Z)'] = \mathbb{E}_{\mathrm{P}}[(y'Z)^2]$ *(b)* The integral of a positive function w.r.t a measure is null iff the function is null almost-surely (e.g., Kallenberg 2002 (1997, Lemma 1.24). *(c)* By assumption, $\mathrm{P} \sim \mathrm{Q}$. *(d)* Same as (b). *(a)* Same as (a) with Q instead of P.

□

Corollary 1 (Implication of Assumption 1(h))**.** *Under Assumptions 1(a)-(b), (e) and (g), Assumption 1(h) implies that, for all $(\theta, \tau) \in \mathbf{S}$, $\mathbb{E}\left[\mathrm{e}^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is a positive definite symmetric matrix.*

[15]The set of singular matrices corresponds to the set of zeros of the determinant, which is nonzero polynomial in several variables. Moreover, by induction over the number of variables with the fundamental theorem of algebra for the base step, a nonzero polynomials has a finite number of zeros.

Proof. By Lemma 29i (p. 86) with $Z = \psi(X_1, \theta)$, it is a positive semi-definite matrix. Thus, it remains to show that it is invertible, i.e., definite instead of only semi-definite.

Under Assumption 1 (a)(b)(d)(e)(g) and (h), by Lemma 3 (p. 22) and Assumption 1(d)(e), for all $\theta \in \mathbf{\Theta}$, $0 < \mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}] < \infty$. Moreover, by Assumption 1(h), for all $\theta \in \mathbf{\Theta}$, $\mathbb{E}\left[e^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is invertible, so that $\frac{1}{\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]}\mathbb{E}\left[e^{\tau(\theta)'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'\right]$ is also invertible. For every $(\theta, \tau) \in \mathbf{S}$, check the assumptions of Lemma 29ii (p. 86) with $Z = \psi(X_1, \theta)$, $\frac{d\mathrm{P}_\theta}{d\mathbb{P}} = \frac{e^{\tau(\theta)'\psi(X_1,\theta)}}{\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]}$ and $\frac{d\mathrm{Q}_{(\theta,\tau)}}{d\mathrm{P}_\theta} = \frac{e^{\tau'\psi(X_1,\theta)}\mathbb{E}[e^{\tau(\theta)'\psi(X_1,\theta)}]}{\mathbb{E}[e^{\tau'\psi(X_1,\theta)}]e^{\tau(\theta)'\psi(X_1,\theta)}}$, so that $\frac{d\mathrm{Q}_{(\theta,\tau)}}{d\mathbb{P}} = \frac{e^{\tau'\psi(X_1,\theta)}}{\mathbb{E}[e^{\tau'\psi(X_1,\theta)}]}$. Firstly, for all $(\omega, \tau, \theta) \in \mathbf{\Omega} \times \mathbf{T} \times \mathbf{\Theta}$, $0 < \frac{d\mathrm{Q}_{(\theta,\tau)}}{d\mathrm{P}_\theta}$ and $0 < \frac{d\mathrm{P}_\theta}{d\mathbb{P}}$, so that $\mathrm{Q}_{(\theta,\tau)} \sim \mathrm{P}_\theta \sim \mathbb{P}$. Secondly, by monotonicity of integration and the Cauchy-Schwarz inequality, for all $\dot{\theta} \in \mathbf{\Theta}$, $\mathbb{E}[|\psi(X_1,\dot{\theta})\psi(X_1,\dot{\theta})'|] \leqslant \mathbb{E}[\sup_{\theta\in\mathbf{\Theta}} |\psi(X_1,\theta)\psi(X_1,\theta)'|] < \sqrt{\mathbb{E}[\sup_{\theta\in\mathbf{\Theta}} |\psi(X_1,\theta)\psi(X_1,\theta)'|^2]} < \infty$, where the last inequality follows from Assumption 1(g). Thirdly, under Assumption 1 (a)(b)(d)(e)(g) awnd (h), by Lemma 3 (p. 22) and Assumption 1(d)(e), for all $(\theta, \tau) \in \mathbf{S}$, $0 < \mathbb{E}[e^{\tau'\psi(X_1,\theta)}] < \infty$. Moreover, under Assumptions 1(a)-(b), (e) and (g), by Lemma 8i (p. 28), $\mathbb{E}\left[\sup_{(\theta,\tau)\in\mathbf{S}} |e^{\tau'\psi(X_1,\theta)}\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$, so that, for all $(\theta, \tau) \in \mathbf{S}$, $\mathbb{E}\left[\frac{e^{\tau'\psi(X_1,\theta)}}{\mathbb{E}[e^{\tau'\psi(X_1,\theta)}]}|\psi(X_1,\theta)\psi(X_1,\theta)'|\right] < \infty$. Thus, for each $(\theta, \tau) \in \mathbf{S}$, apply Lemma 29ii (p. 86) to show the result. □

Appendix D. Remaining technical results

Lemma 30 (Asymptotic invertibility of sequence of matrix functions)**.** *Let $A(\gamma)$ be a family of invertible matrices indexed by $\gamma \in \mathbf{\Gamma}$ s.t. $\gamma \mapsto A(\gamma)$ is continuous, and where $\mathbf{\Gamma}$ is a compact subset of a Euclidean space. Let $(A_T(\gamma))_{T\in[\![1,\infty[\![}$ be a sequence of square matrices. If, as $T \to \infty$, $\sup_{\gamma\in\mathbf{\Gamma}} |A_T(\gamma) - A(\gamma)| \to 0$, then there exist a constant $\varepsilon_A > 0$ and $T_A \in \mathbf{N}$ s.t. for all $T \in [\![T_A, \infty[\![$, for all $\gamma \in \mathbf{\Gamma}$, $||A_T(\gamma)|_{\det}| \geqslant \varepsilon_A$.*

Proof. The function $A \mapsto ||A|_{\det}|$ is a continuous function. Moreover, by assumption, for all $\gamma \in \mathbf{\Gamma}$, $||A(\gamma)|_{\det}| > 0$. Thus, by continuity of $\gamma \mapsto A(\gamma)$ and compactness of $\mathbf{\Gamma}$, there exists ε_A s.t. $\min_{\gamma\in\mathbf{\Gamma}} ||A(\gamma)|_{\det}| > 2\varepsilon_A$. Now continuity of $A \mapsto ||A|_{\det}|$ on the compact set $\mathbf{\Gamma}$ implies uniform continuity (e.g., Rudin 1953, Theorem 4.19), so that there exists $T_{\varepsilon_A} \in \mathbf{N}$ s.t., for all $T \in [\![T_{\varepsilon_A}, \infty[\![$, $\sup_{\gamma\in\mathbf{\Gamma}} |||A_T(\gamma)|_{\det}| - ||A(\gamma)|_{\det}|| \leqslant \varepsilon_A$. Then, for all $\gamma \in \mathbf{\Gamma}$, the triangle inequality $||A(\gamma)|_{\det}| \leqslant |||A(\gamma)|_{\det}| - ||A_T(\gamma)|_{\det}|| + ||A_T(\gamma)|_{\det}|$ implies that $\varepsilon_A = 2\varepsilon_A - \varepsilon_A \leqslant ||A(\gamma)|_{\det}| - |||A(\gamma)|_{\det}| - ||A_T(\gamma)|_{\det}|| \leqslant ||A_T(\gamma)|_{\det}|$. □

Lemma 31 (Asymptotic positivity and definiteness of matrices)**.** *Let $(A_T)_{T\geqslant1}$ a sequence of square matrices converging to A as $T \to \infty$.*[16] *Then, if $(A_T)_{T\geqslant1}$ is a sequence of symmetric matrices and A is a positive-definite matrix (p-d.m), then there exists $\dot{T} \in \mathbf{N}$ such that $T \geqslant \dot{T}$ implies A_T is p-d.m.*

Proof. On one hand, A_T is a p-d.m. if and only if all its eigenvalues are strictly positive (e.g., Magnus and Neudecker 1999/1988, Ch. 1 Sec. 13 Theorem 8). On the other hand, $\min \mathrm{sp} A_T = \min_{z:\|z\|=1} z'A_T z$, where $\mathrm{sp} A_T$ denotes the set of eigenvalues of A (e.g., Magnus and Neudecker 1999/1988, Ch. 11 Sec. 5). Thus, it is sufficient to prove that $\lim_{T\to\infty} \min_{z:\|z\|=1} z'A_T z =$

[16]Note that we do not need to specify the norm as all norms are equivalent in finite-dimensional spaces.

$\min_{z:\|z\|=1} z'Az$, which in turn implies that it is sufficient to prove that $\sup_{z:\|z\|=1} |z'A_T z - z'Az| \to 0$, as $T \to \infty$. Prove this last result by contradiction.

Assume that $\sup_{z:\|z\|=1} |z'A_T z - z'Az|$ does not converge to 0 as $T \to \infty$. Then, there exists $\varepsilon > 0$ and an increasing function $\alpha_1 : \mathbf{N} \mapsto \mathbf{N}$ defining a subsequence of vectors of norm 1, $\left(z_{\alpha_1(T)}\right)_{T\geqslant 1}$, and a subsequence of matrices, $\left(A_{\alpha_1(T)}\right)_{T\geqslant 1}$, such that

$$\begin{aligned}
\varepsilon &< \left| z'_{\alpha_1(T)} A_{\alpha_1(T)} z_{\alpha_1(T)} - z'_{\alpha_1(T)} A z_{\alpha_1(T)} \right| \\
&= \left| z'_{\alpha_1(T)} \left(A_{\alpha_1(T)} - A\right) z_{\alpha_1(T)} \right| \leqslant \sum_{(k,l)\in[\![1,m]\!]^2} \left| \left[a^{(k,l)}_{\alpha_1(T)} - a^{(k,l)} \right] z^{(k)}_{\alpha_1(T)} z^{(l)}_{\alpha_1(T)} \right| \\
&\leqslant m^2 \times \max_{(k,l)\in[\![1,m]\!]^2} \left| a^{(k,l)}_{\alpha_1(T)} - a^{(k,l)} \right|
\end{aligned}$$

where m is the size of the matrix A and $a^{(k,l)}$ denotes the component of the matrix A in the kth row and lth column. Now, by assumption, using the max norm, $\max_{(k,l)\in[\![1,m]\!]^2} \left| a^{(k,l)}_{\alpha_1(T)} - a^{(k,l)} \right| \to 0$ as $T \to \infty$. Thus, there is a contradiction. $\square$

Lemma 32 (Differential of a log of a squared determinant)**.** *Let* $\mathbf{G}$ *be an open set of* $\mathbf{R}^q$ *with* $q \in [\![1, \infty[\![$, *and* $F : \mathbf{G} \to \mathbf{R}^{m\times m}$ *a differentiable function on* $\mathbf{G}$. *Then* $|F|_{\det} : \mathbf{G} \to \mathbf{R}$ *is also differentiable on* $\mathbf{G}$. *Moreover, if* $|F(x)|_{\det} \neq 0$ *where* $x \in \mathbf{G}$, *then*

(i) $D|F(x)|_{\det} = |F(x)|_{\det}\mathrm{tr}[F(x)^{-1}DF(x)]$;

(ii) $D\ln[|F(x)|^2_{\det}] = 2\mathrm{tr}[F(x)^{-1}DF(x)]$.

Proof. (i) It is a consequence of the so-called Jacobi's formula (e.g., Magnus and Neudecker 1999/1988, chap. 8 sec. 3).

(ii) First of all, note that the logarithm is well-defined as its argument is strictly positive by assumption. Then, by the statement (i) of the present lemma and the chain rule,

$$D\ln[|F(x)|^2_{\det}] = \frac{1}{|F(x)|^2_{\det}} 2|F(x)|_{\det}|F(x)|_{\det}\mathrm{tr}[F(x)^{-1}DF(x)].$$

$\square$

Lemma 33 (Inverse of a 2×2 partitioned matrix)**.** *Let* F *be a square matrix s.t.*

$$F = \begin{bmatrix} A & B \\ C & D \end{bmatrix}$$

where A *and* D *are square matrices. Then, the following statements hold.*

(i) *If* A *is invertible, then* F *invertible* $\Leftrightarrow$ $(D - CA^{-1}B)$ *invertible. Moreover,*

$$F^{-1} = \begin{bmatrix} A^{-1} + A^{-1}B(D - CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D - CA^{-1}B)^{-1} \\ -(D - CA^{-1}B)^{-1}CA^{-1} & (D - CA^{-1}B)^{-1} \end{bmatrix}.$$

(ii) *If* D *is invertible, then* F *invertible* $\Leftrightarrow$ $(A - BD^{-1}C)^{-1}$ *invertible. Moreover,*

$$F^{-1} = \begin{bmatrix} (A - BD^{-1}C)^{-1} & -(A - BD^{-1}C)^{-1}BD^{-1} \\ -D^{-1}C(A - BD^{-1}C)^{-1} & D^{-1} + D^{-1}C(A - BD^{-1}C)^{-1}BD^{-1} \end{bmatrix}.$$

Proof. This is a standard result (e.g., Magnus and Neudecker 1999/1988, Chap. 1 sec. 11). $\square$

Corollary 2 (Inverse of a 2×2 partitioned matrix in a special case)**.** *Let E be a square matrix s.t.*

$$E = \begin{bmatrix} A & B \\ B' & 0 \end{bmatrix}.$$

Then,

(i) *If A and $B'A^{-1}B$ are invertible, then E in invertible; and*

(ii) $$\begin{bmatrix} A & B \\ B' & 0 \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} - A^{-1}B\left(B'A^{-1}B\right)^{-1}B'A^{-1} & A^{-1}B\left(B'A^{-1}B\right)^{-1} \\ \left(B'A^{-1}B\right)^{-1}B'A^{-1} & -\left(B'A^{-1}B\right)^{-1} \end{bmatrix}.$$

Proof. Apply the above Lemma 33i with $F = E$, $C = B'$ and $D = 0$. □

Appendix E. More on the numerical example

The simulations were performed in R. Each model parameterization is simulated 10,000 times. The robustness of the simulation results was checked with different optimization algoritheorems, starting values and tolerance parameter values. The estimation for a single sample is typically performed in less than a few seconds. The calculations were done on a 24 CPU cores of a Dell server with 4 AMD Opteron 8425 HE processors running at 2.1 GHz. We numerically checked that the reported statistics have a converging behaviour as we increase the number of simulated samples to 10,000.

Appendix F. More on the empirical example

In empirical consumption-based asset pricing, the literature has found little common ground about the value of the relative risk aversion (RRA) of the representative agent: In most studies, point estimates from economically similar moment conditions are generally outside of each other's confidence intervals. Section 4.2 (p. 9) and the present appendix revisit the estimation of the RRA. The popularity of moment-based estimation in consumption-based asset pricing, and more generally in economics is due to the fact that moment-based estimation does not necessarily require the specification of a family of distributions for the data (e.g. Hansen 2013, sec. 3). Typically, an economic model does not imply such family of distributions, except for tractability reasons. Imposing a family of distributions makes it difficult to disentangle the part of the inference results due to the empirical relevance of the economic model from the part due to these additional restrictions. Under regularity conditions, assuming a distribution corresponds to imposing an infinite number of extra moment restrictions (e.g., Feller 1971 (1966, 1971/1966, chap. VII, sec. 3).

In Section 4.2 (p. 9) and the present appendix, we rely on the moment condition (10) on p. 9. This moment condition has several advantages. Firstly, it is as consistent with Lucas (1978) as with more recent consumption-based asset-pricing models, such as Barro (2006) or Gabaix (2012). In other words, despite its simplicity it also correspond to sophisticated models, and it allows us to obtain estimates that are robust to different variations of consumption-based asset pricing theory. Secondly, without loss of generality, it does not require to estimate the time discount rate, about which there is little debate: The time discount rate of the representative agent is consistently found to be between .9 and 1. Note also that it has been common to use moment conditions with a separate parameter for the so-called intertemporal elasticity of substitution, i.e., use Epstein-Zin-Weil preferences (e.g. Epstein and Zin 1991). However, Bommier et al. (2017) show that such a specification makes the economic interpretation of the parameters difficult. In particular, they show that an increase of the so-called RRA (relative risk-aversion) parameter does not yield a behaviour that would be considered more risk averse. E.g., All other things being equal, savings can be a decreasing function of the so-called RRA parameter for an agent with Epstein-Zin-Weil preferences (e.g., Bommier et al. 2017, sec. 6). This difficulty of interpretation comes from a violation of the monotonicity axiom according to which an agent does not choose an action if another available action is preferable in every state of the world.

F.1. **Additional empirical evidence.**

Table 3 (p. 92) is the same as Table 2 (p. 10) with the additional Table 3 Figures (A). The latter clearly shows that the normalized ESP is relatively sharp around the ESP estimator. Table 4 (p. 93) is the counterpart of Table 3 (p. 92) for the 1930-2009 data set. The 95% ET ALR confidence region is based on the inversion of the ALR ET statistic $2T\left[\mathrm{LogET}(\hat{\theta}_T) - \mathrm{LogET}(\theta_0)\right]$ $= 2T\mathrm{LogET}(\theta_0) \rightarrow \chi^2_1$ (Kitamura and Stutzer 1997, Theorem 4 with $K = 0$ and $\mathrm{H}_0 : \theta = \theta_0$), where $\mathrm{LogET}(\theta) :=\ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]$ and $\mathrm{LogET}(\hat{\theta}_T) = \ln\left[\frac{1}{T}\sum_{t=1}^{T} \mathrm{e}^{\tau_T(\hat{\theta}_T)'\psi_t(\hat{\theta}_T)}\right] = 0$ because, in the just-identified case, $\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T) = 0$ so that $\tau_T(\hat{\theta}_T) = 0_{m\times 1}$. The ET and ESP support correspond to the parameter values $\theta \in \boldsymbol{\Theta}$ for which there exists a solution $\tau_T(\theta)$ to the equation (14) on p. 17. Table 4 confirms the findings of Table 3 (p. 92) in Section 4.2 : The ESP is sharper than the ET around its maximum, so that the ESP confidence region is also shorter. Note also that the ESP estimate is almost the same as for the data set 1890-2009. These results are in line with the ESP shrinkage-like behaviour documented in the Monte-Carlo simulations of the section 4.1.

Tables 5 (p. 94) and 7 (p. 95) report the MM estimates and the confidence regions based on the inversion of the MM ALR test statistic $T\left[Q_{\mathrm{MM},T}(\theta_0) - Q_{\mathrm{MM},T}(\hat{\theta}_{\mathrm{MM},T})\right] = TQ_{\mathrm{MM},T}(\theta_0) \xrightarrow{D} \chi^2_1$, as $T \rightarrow \infty$, (e.g., Newey and McFadden 1994, Theorem 9.2), where $Q_{\mathrm{MM},T}(\theta) := \left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]'$ $\times\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_{\mathrm{MM},T})\psi_t(\hat{\theta}_{\mathrm{MM},T})'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]$ and $Q_{\mathrm{MM},T}(\hat{\theta}_{\mathrm{MM},T}) = 0$ because $\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T) = 0$ in the just-identified case. The MM objective function is sharper around its minimum for the 1930-2009 data set than for the 1890-2009. However, the former sharpness appears misleading as it yields a confidence region that does not include the MM estimate of the 1890-2009 data set.

Tables 6 (p. 94) and 8 (p. 95) report the CU (continuously updating) MM estimates and the confidence regions based on the inversion of the CU ALR test statistic $T\left[Q_{\mathrm{CU},T}(\theta_0) - Q_{\mathrm{CU},T}(\hat{\theta}_{\mathrm{MM},T})\right] = TQ_{\mathrm{CU},T}(\theta_0) \rightarrow \chi^2_1$, as $T \rightarrow \infty$, where $Q_{\mathrm{CU},T}(\theta) := \left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]'\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\psi_t(\theta)'\right]^{-1}\left[\frac{1}{T}\sum_{t=1}^{T}\psi_t(\theta)\right]$ and $Q_{\mathrm{CU},T}(\hat{\theta}_{\mathrm{CU},T}) = 0$ because $\frac{1}{T}\sum_{t=1}^{T}\psi_t(\hat{\theta}_T) = 0$ in the just-identified case. In the just-identified case, which is the case addressed in the present paper, such confidence regions correspond to the S-sets, which were proposed by Stock and Wright (2000) —following Hansen et al. (1996, (c) Constrained-Minimized)— as a solution to the flatness of GMM objective functions. As previously documented in the literature (e.g., Hansen et al. 1996), CU GMM objective functions tend to be flat and low in the tails. Thus, the CU ALR confidence regions (and S-sets in the just-identified case) are huge, and hardly informative.

F.2. **Data description.** As in Julliard and Ghosh (2012), our data are standard. For the 1890-2009 data set, our source is the Robert Shiller's web site. The prime commercial paper and the S&P stock price index play the role of proxies for the risk-less asset and the market return.

For the 1930-2009 data set, the proxies for the risk-less asset and the market return are the one month Treasury-bill and the Center for Research in Security Prices (CRSP) value-weighted index of all stocks on the NYSE, AMEX, and NASDAQ. The computation of the growth consumption is based per capita real personal consumption expenditures on nondurable goods from the National Income and Product Accounts (NIPA). Quantities are deflated from the inflation.

Tables 9 and 10 indicate that there is no significant autocorrelation for the excess returns, and only a mild clustering effect (Figures (E) and (F) in Table 10 on p. 96). Thus, the i.i.d. assumption (Assumption 1(a)) appears to be a good approximation for the excess returns for both data set. For the growth consumption, the i.i.d. assumption may appear less appropriate. Table 11 indicates a mild autocorrelation for the growth consumption, and, more strikingly, a change of variance at the end of WWII. However, in the moment function, the growth consumption is multiplied by the excess returns, whose variance is several orders of magnitude higher (Table 9 on p. 96), so that the change of variance is dampened.

University of Luxembourg, LSF, 6 Rue Richard Coudenhove-Kalergi, L-1359, Luxembourg.

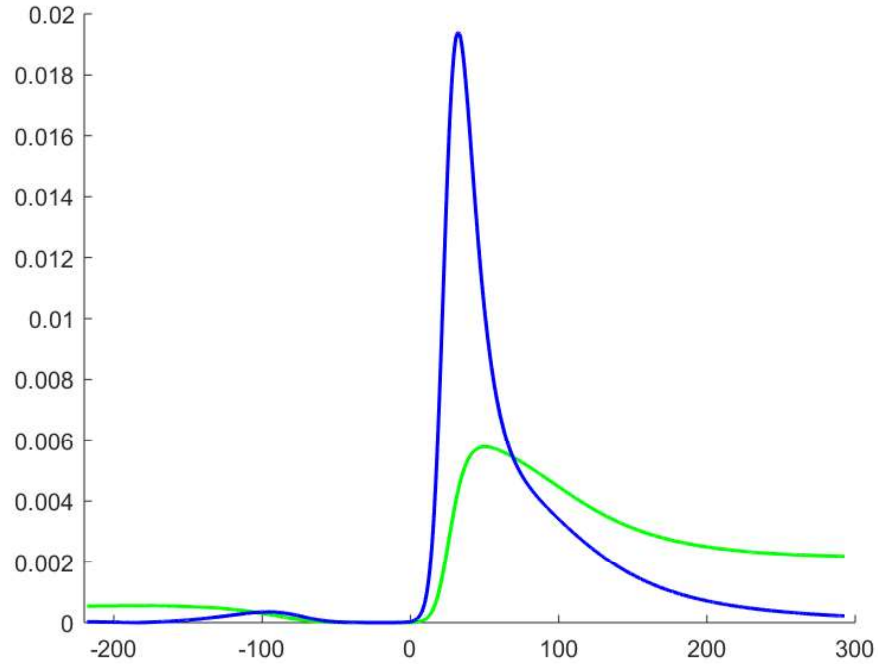
E-mail address: Benjamin.Holcblat@uni.lu

Carnegie Mellon University, Tepper School of Business, 5000 Forbes Ave. Pittsburgh, PA, USA

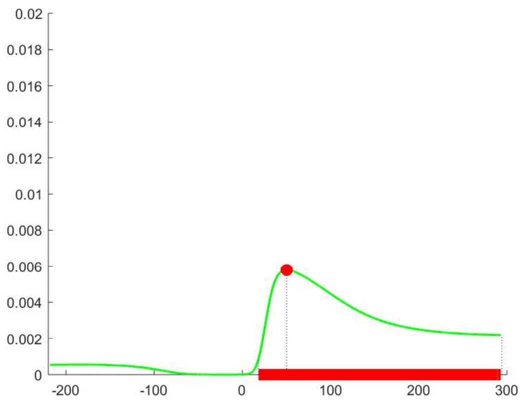
E-mail address: fs0v@andrew.cmu.edu

TABLE 3. **ET vs. ESP inference (1890–2009)**

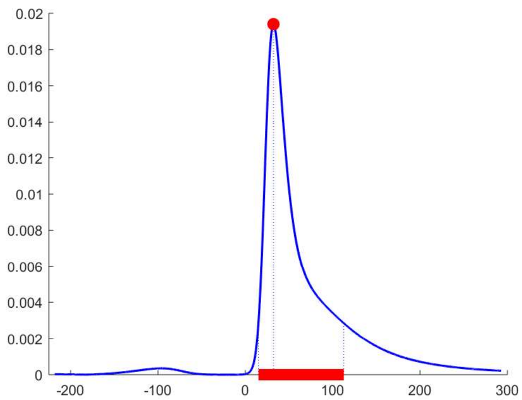
Empirical moment condition: $\frac{1}{2009-1889}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_{m,t}-R_{f,t})\right]=0$, where
$R_{m,t}$:= gross market return, $R_{f,t}$:=risk-free asset gross return, C_t := consumption, and θ :=relative risk aversion;
Normalized ET:=$\exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(.)'\psi_t(.)}\right]\right\}/\int_\Theta \exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\mathrm{d}\theta$;
Normalized ESP:=$\hat{f}_{\theta_T^*}(.)/\int_\Theta \hat{f}_{\theta_T^*}(\theta)\mathrm{d}\theta$;
$\hat{\theta}_{\mathrm{ET},T}=\hat{\theta}_{\mathrm{MM},T}=50.3$ (bullet) and $\hat{\theta}_{\mathrm{ESP},T}=32.21$ (bullet);
ET and ESP support = $[-218.2, 289.0]$; 95% ET ALR conf. region=$[18.3, 289.0]$ (stripe);
95% ESP ALR conf. region=$[15.0, 112.7]$ (stripe).



(A) Normalized ET (light green) vs. normalized ESP (dark blue).



(A) ET est. and ALR conf. region.



(B) ESP est. and ALR conf. region.

Table 4. **ET vs. ESP inference (1930–2009)**

Empirical moment condition: $\frac{1}{2009-1889}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_{m,t}-R_{f,t})\right]=0$, where

$R_{m,t}$:= gross market return, $R_{f,t}$:=risk-free asset gross return, C_t := consumption, and θ :=relative risk aversion;

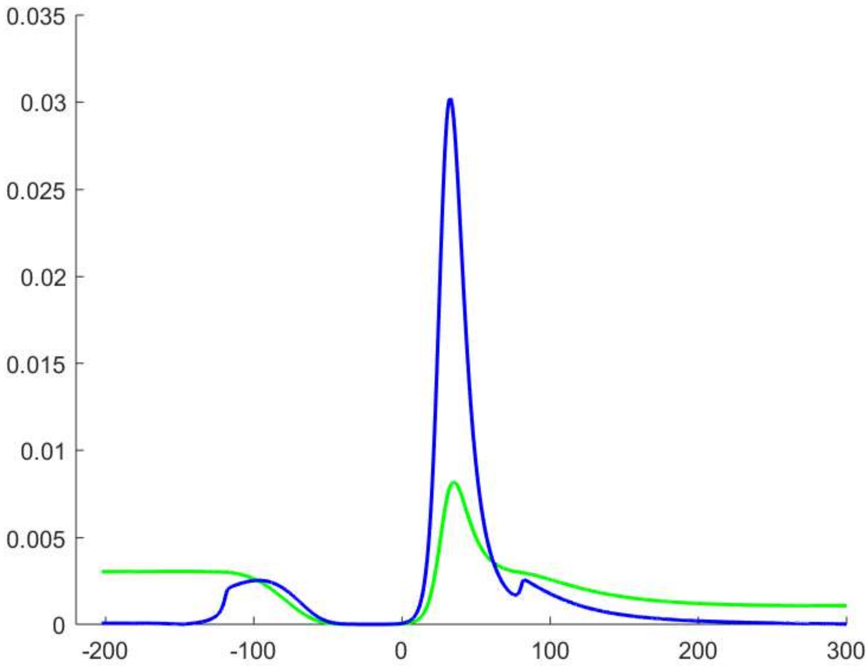
Normalized ET:=$\exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(.)'\psi_t(.)}\right]\right\}/\int_{\Theta}\exp\left\{T\ln\left[\frac{1}{T}\sum_{t=1}^{T}\mathrm{e}^{\tau_T(\theta)'\psi_t(\theta)}\right]\right\}\mathrm{d}\theta$;

Normalized ESP:=$\hat{f}_{\theta_T^*}(.)/\int_{\Theta}\hat{f}_{\theta_T^*}(\theta)\mathrm{d}\theta$;

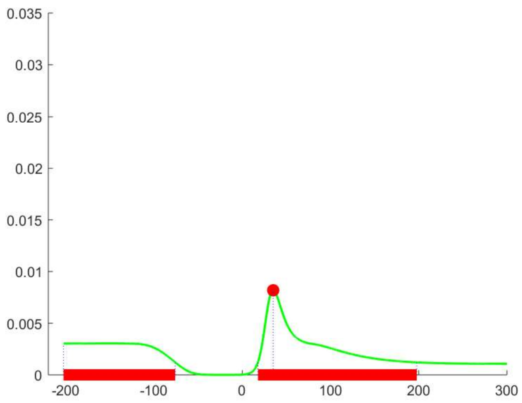
$\hat{\theta}_{\mathrm{ET},T}=35.0$ (bullet) and $\hat{\theta}_{\mathrm{ESP},T}=32.5$ (bullet); ET and ESP support= $[-202.8, 813.3]$

95% ET ALR conf. region=$[-202.8,-76.0]\cup[17.7,197.8]$ (stripe);

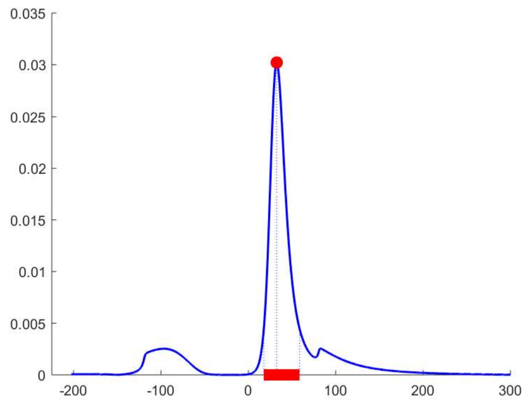
95% ESP ALR conf. region=$[17.7, 58.7]$ (stripe).



(A) Normalized ET (light green) vs. normalized ESP (dark blue).



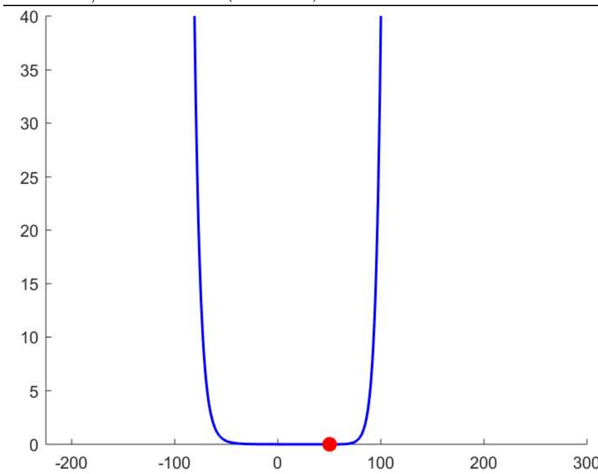
(B) ET est. and ALR conf. region.



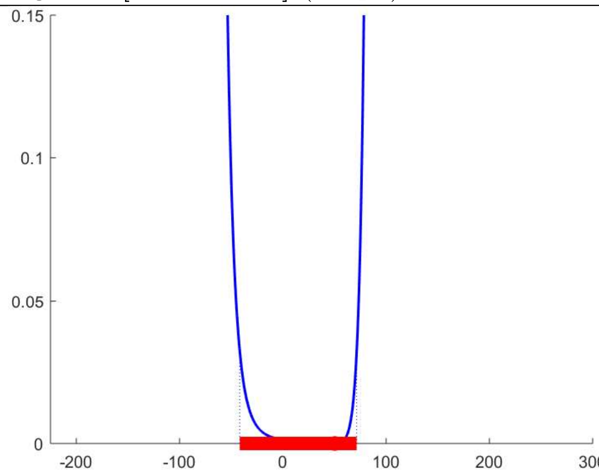
(C) ESP est. and ALR conf. region.

TABLE 5. **MM inference (1890–2009)**

Empirical moment condition: $\frac{1}{2009-1889}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_{m,t}-R_{f,t})\right]=0$, where
$R_{m,t}$:= gross market return, $R_{f,t}$:=risk-free asset gross return, C_t := consumption, and θ :=relative risk aversion.
$\hat{\theta}_{\mathrm{GMM},T}=50.3$ (bullet); 95% ALR confidence region= $[-41.7, 71.5]$ (stripe).



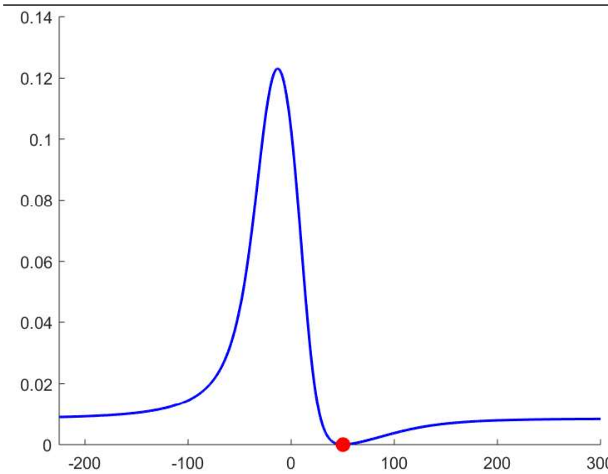
(A) MM objective function and point estimate.



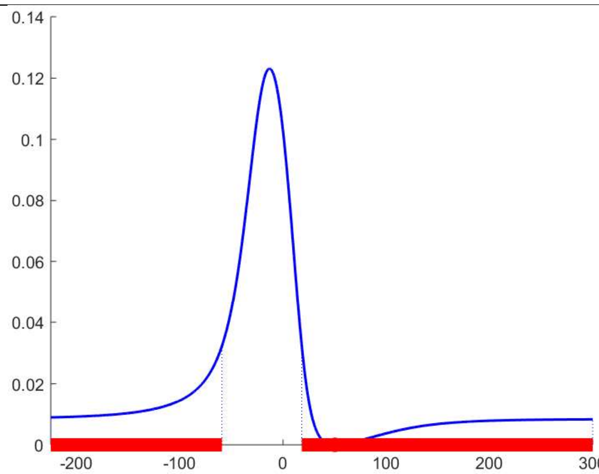
(A zoom) MM obj. function ALR conf. region.

TABLE 6. **Continuously updated (CU) GMM inference (1890–2009)**

Empirical moment condition: $\frac{1}{2009-1890}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_{m,t}-R_{f,t})\right]=0$, where
$R_{m,t}$:= gross market return, $R_{f,t}$:=risk-free asset gross return, C_t := consumption, and θ :=relative risk aversion.
$\hat{\theta}_T^{\mathrm{CU}}$=50.3 (bullet); 95% ALR confidence region (and S-set)=$]\ldots,-59.1]\cup[18.2,\ldots[$ (stripe).
Rk: We constrain the numerical search for point estimate to discard large values of θ.



(A) Objective function and point estimate.



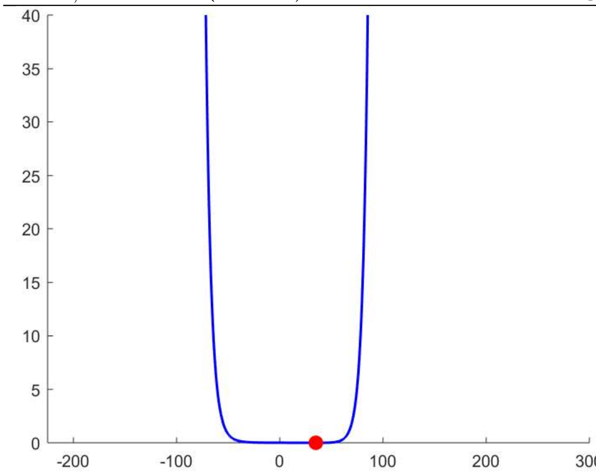
(B) Truncated ALR conf. region (and S-set).

TABLE 7. **MM inference (1930-2009)**

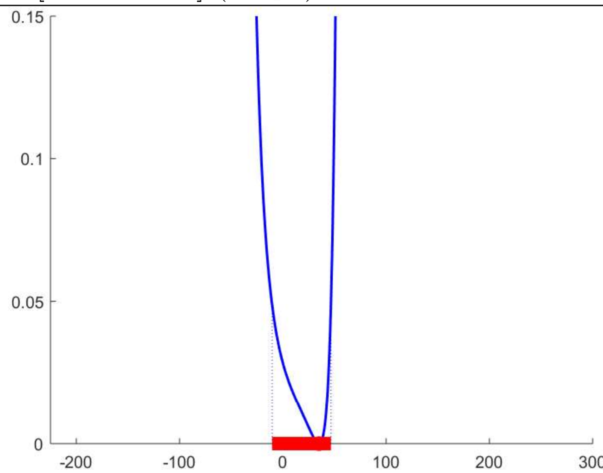
Empirical moment condition: $\frac{1}{2009-1930}\sum_{t=1930}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_{m,t}-R_{f,t})\right]=0$, where

$R_{m,t}$:= gross market return, $R_{f,t}$:=risk-free asset gross return, C_t := consumption,
θ :=relative risk aversion.

$\hat{\theta}_{\mathrm{MM},T}=35.0$ (bullet), ALR confidence region= $[-10.4, 46.5]$ (stripe)



(A) MM objective function and point estimate.



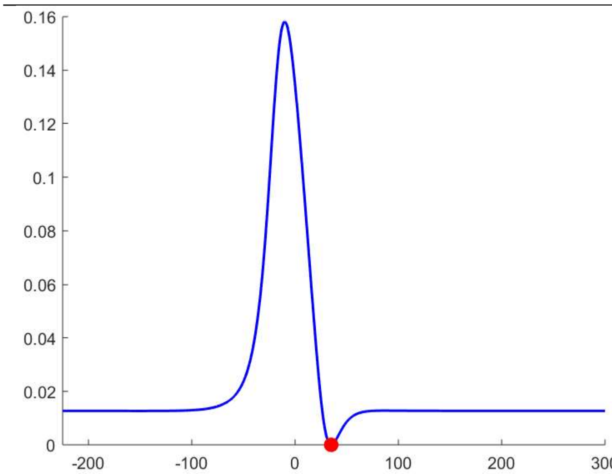
(A zoom) Objective function and point estimate.

TABLE 8. **Continuously updated (CU) GMM inference (1930–2009)**

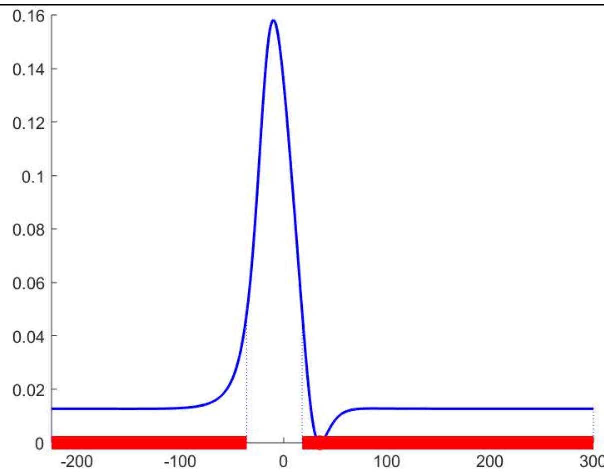
Empirical moment condition: $\frac{1}{2009-1890}\sum_{t=1890}^{2009}\left[\left(\frac{C_t}{C_{t-1}}\right)^{-\theta}(R_{m,t}-R_{f,t})\right]=0$, where

$R_{m,t}$:= gross market return, $R_{f,t}$:=risk-free asset gross return, C_t := consumption,
θ :=relative risk aversion.

$\hat{\theta}_T^{\mathrm{CU}}=50.3$ (bullet); ALR confidence region (and S-set) $=]\ldots,-35.8]\cup[17.9,\ldots[$ (stripe).
Rk: We constrain the numerical search for point estimate to discard large values of θ.



(A) Objective function and point estimate.



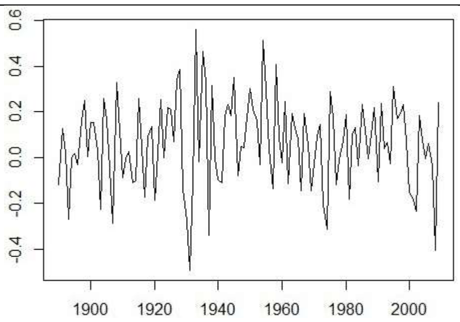
(B) Truncated ALR conf. region (and S-set).

Table 9. **Descriptive statistics.**

Variable	Mean (Variance) 1890-2009	Mean (Variance) 1930-2009
C_t/C_{t-1}	1.0182	1.014
	(.0009)	(.0007)
$R_{m,t} - R_{f,t}$	.0630	.074
	(.0367)	(.0424)

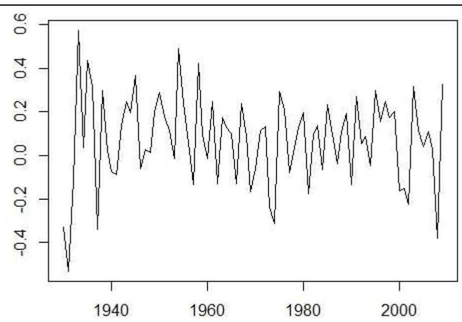
Table 10. **Excess returns:** $R_{m,t} - R_{f,t}$

1890-2009

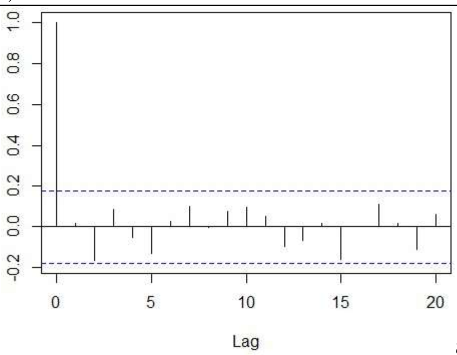


(A) Time series

1930-2009

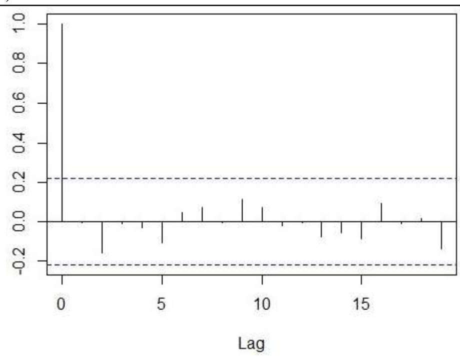


(B) Time series

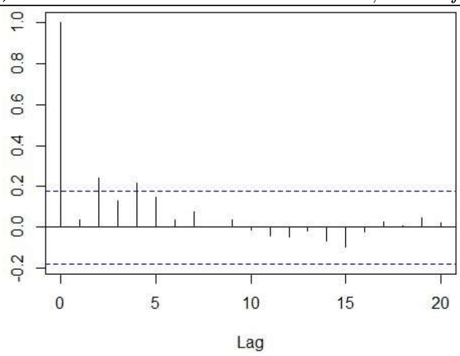


a

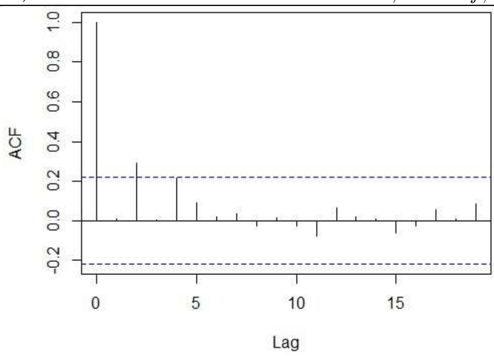
(C) Autocorr. function of $R_{m,t} - R_{f,t}$



(D) Autocorr. function of $R_{m,t} - R_{f,t}$

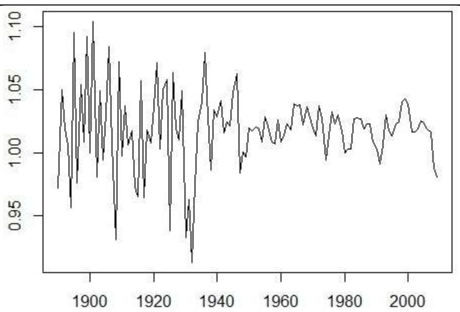
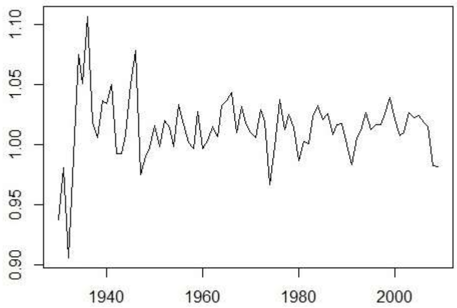
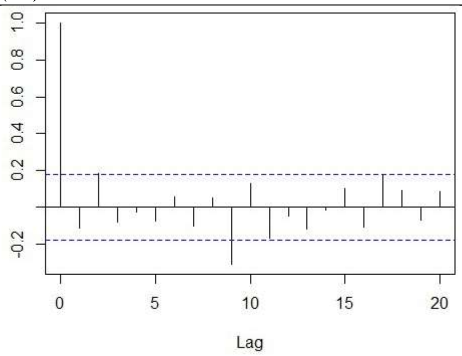
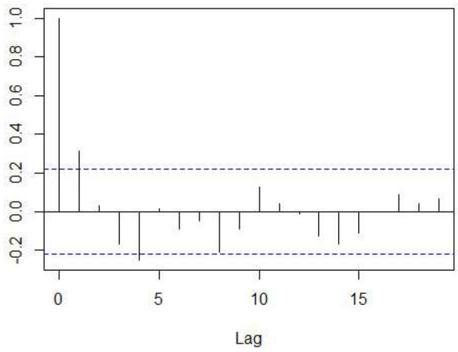


(E) Autocorr. function of $(R_{m,t}-R_{f,t})^2$



(F) Autocorr. function of $(R_{m,t}-R_{f,t})^2$

Table 11. **Growth consumption:** C_t/C_{t-1}.

1890-2009	1930-2009
	
(A) Time series	(B) Time series
	
(C) Autocorr. function of C_t/C_{t-1}	(D) Autocorr. function of C_t/C_{t-1}