D Notation

For each positive integer $K$ let $q^K(Z) := (q_1(Z), \ldots, q_K(Z))'$ be a $K$-vector of approximating functions. For every $\varepsilon > 0$ and a constant $C > 0$, let $\Theta(C, \varepsilon) := \{\|\theta - \theta_*\| \leq C\varepsilon\}.$
We denote: $W := (X', Z')', W_{1:n} := (W_1, \ldots, W_n)$ the sample of i.i.d. observations of $W$, $g(W, \theta) := \rho(X, \theta) \otimes q^K(Z)$ the expanded moment functions and $g_i(\theta) = g(W_i, \theta)$.

Denote $p(W_{1:n}|\theta) := \prod_{i=1}^n \hat{p}_i(\theta)$,

$$\ell_{n, \theta}(W_i) := \log \hat{p}_i(\theta) = \log \frac{e^{\hat{\lambda}(\theta)' g(W_i, \theta)}}{\sum_{j=1}^n e^{\hat{\lambda}(\theta)' g(W_j, \theta)}}$$

where $\hat{\lambda}(\theta) := \arg\min_{\lambda \in \mathbb{R}^d} \frac{1}{n} \sum_{i=1}^n e^{\lambda' g(W_i, \theta)}$ is the estimated tilting parameter. Moreover, $\tau(\hat{\lambda}, \theta, W) := e^{\hat{\lambda}(\theta)' g(W, \theta)}$ and $\tau_n(\hat{\lambda}, \theta) := \frac{1}{n} \sum_{i=1}^n \tau(\hat{\lambda}, \theta, W_i)$. We use the notation $E_n[\cdot] := \frac{1}{n} \sum_{i=1}^n [\cdot]$ for the empirical mean and $E[\cdot]$ for the population mean with respect to the true distribution $P_\pi$. For a matrix $A$, we denote by $\lambda_{\min}(A)$ and $\lambda_{\max}(A)$ the minimum and maximum eigenvalue of $A$, respectively.

Moreover, $\tilde{g}(\theta) := E_n[g_1(\theta)], \rho_0(X, \theta) := \frac{\partial \tilde{g}(X, \theta)}{\partial \theta}, \tilde{G}(\theta) := E_n[G(W_i, \theta)]$ with $G(W, \theta) := \rho_0(X, \theta) \otimes q^K(Z)$ a $dK \times p$ matrix, $\tilde{G}(\theta) := E_n[\tau(\hat{\lambda}, \theta, W_i) G(W_i, \theta)], \Omega(\theta) := E_n[g(W_i, \theta) g(W_i, \theta)']$ a $dK \times dK$ matrix and $\tilde{\Omega}(\theta) := E_n[\tau(\hat{\lambda}, \theta, W_i) g(W_i, \theta) g(W_i, \theta)']$. Their population counterparts in the correctly specified model are $G_* := E[\rho_0(X, \theta_*) \otimes q^K(Z)]$ and $\Omega_* := E[g(W_i, \theta_*) g(W_i, \theta_*)']$, respectively. In addition $\Sigma(Z) := E[\rho_0(X, \theta_*) \rho(X, \theta_*)'|Z], D(Z) := E[\rho_0(X, \theta_*)|Z], V_{\theta_*}^{-1} := E^P[D(Z)'\Sigma(Z)^{-1}D(Z)]$, and $\rho_{\theta\theta}(x, \theta_*) := \partial^2 \rho_j(x, \theta)/\partial \theta \partial \theta'$.

Finally, let CS, M, and MVT refer to the Cauchy-Schwartz, Markov, and Mean Value Theorem, respectively.

**E.1 Proof of Theorem 3.1**

The main steps of this proof proceed as in the proof of Van der Vaart (1998, Theorem 10.1) while the proofs of the technical theorems and lemmas that we use all along this proof are new. Let us consider a reparametrization of the model centred around the true value $\theta_*$ and define a local parameter $h = \sqrt{n}(\theta - \theta_*)$. Denote by $\pi^h$ and $\pi^h(\cdot|W_{1:n})$ the prior and posterior distribution of $h$, respectively. Denote by $\Phi_n$ the normal distribution $\mathcal{N}_{\Delta_n, \theta_*} V_{\theta_*}$ and by $\phi_n$ its Lebesgue density. For a compact subset $K \subset \mathbb{R}^p$ such that $\pi^h(h \in K|W_{1:n}) > 0$ define, for any Borel set $B \subseteq \Psi$,

$$\pi^h_K(B|W_{1:n}) := \frac{\pi^h(K \cap B|W_{1:n})}{\pi^h(K|W_{1:n})}$$

and let $\Phi^K_n$ be the $\Phi_n$ distribution conditional on $K$. The proof consists of two steps. In the first step we show that the Total Variation (TV) norm of $\pi^h(\cdot|W_{1:n}) - \Phi^K_n$ converges to zero in probability. In the second step we use this result to show that the TV norm of $\pi^h(\cdot|W_{1:n}) - \Phi_n$ converges to zero in probability.
Let Assumptions 3.3 (a) and 3.4 (a) hold. For every open neighborhood $\mathcal{U} \subset \Theta$ of $\theta_*$ and a compact subset $K \subset \mathbb{R}^p$, there exists an $N$ such that for every $n \geq N$:

$$\theta_* + K \frac{1}{\sqrt{n}} \subset \mathcal{U}. \quad \text{(E.1)}$$

Define the function $f_n : K \times K \to \mathbb{R}$ as, $\forall k_1, k_2 \in K$:

$$f_n(k_1, k_2) := \left( 1 - \frac{\phi_n(k_2)s_n(k_1)\pi^h(k_1)}{\phi_n(k_1)s_n(k_2)\pi^h(k_2)} \right)_+$$

where $(a)_+ := \max(a, 0)$, here $\pi^h$ denotes the Lebesgue density of the prior $\pi^h$ for $h$ and $s_n(h) := p(W_{1:n}|\theta_*) + h/\sqrt{n})/p(W_{1:n}|\theta_*)$. The function $f_n$ is well defined for $n$ sufficiently large because of (E.1) and Assumption 3.7. Remark that by (E.1) and since the prior for $\theta$ puts enough mass on $\mathcal{U}$, then $\pi^h$ puts enough mass on $K$ and as $n \to \infty$: $\pi^h(k_1)/\pi^h(k_2) \to 1$. Because of this and by the stochastic LAN expansion in Lemma E.2 below:

$$\log \frac{\phi_n(k_2)s_n(k_1)\pi^h(k_1)}{\phi_n(k_1)s_n(k_2)\pi^h(k_2)} = -\frac{1}{2}(k_2 - \Delta_n,\theta_*)'V^{-1}_{\theta_*}(k_2 - \Delta_n,\theta_*) + \frac{1}{2}(k_1 - \Delta_n,\theta_*)'V^{-1}_{\theta_*}(k_1 - \Delta_n,\theta_*) + k_1'V^{-1}_{\theta_*}\Delta_n,\theta_* - \frac{1}{2}k_1'V^{-1}_{\theta_*}k_1 - k_2'V^{-1}_{\theta_*}\Delta_n,\theta_* + \frac{1}{2}k_2'V^{-1}_{\theta_*}k_2 + o_p(1) = o_p(1). \quad \text{(E.2)}$$

Since, for every $n$, $f_n$ is continuous in $(k_1, k_2)$ and $K \times K$ is compact, then

$$\sup_{k_1, k_2 \in K} f_n(k_1, k_2) \xrightarrow{p} 0, \quad \text{as } n \to \infty. \quad \text{(E.3)}$$

Suppose that the subset $K$ contains a neighborhood of 0 (which guarantees that $\Phi_n(K) > 0$ and then that $\Phi_n^K$ is well defined) and let $\Xi_n := \{\pi^h(K|W_{1:n}) > 0\}$. Moreover, for a given $\eta > 0$ define the event $\Omega_n := \{\sup_{k_1, k_2 \in K} f_n(k_1, k_2) \leq \eta\}$. The TV distance $\|\cdot\|_{TV}$ between two probability measures $P$ and $Q$, with Lebesgue densities $p$ and $q$ respectively, can be expressed as: $\|P - Q\|_{TV} = 2 \int (1 - p/q)_+ dQ$. Therefore, by the Jensen inequality and convexity of the functions $(\cdot)_+$,

$$\frac{1}{2} \mathbb{E}^P\|\pi^h_k(\cdot|W_{1:n}) - \Phi_n^K\|_{TV}\mathbb{1}_{\Omega_n \cap \Xi_n} = \mathbb{E}^P \int_K \left( 1 - \frac{d\Phi_n^K(k_2)}{d\pi^h_K(k_2|W_{1:n})} \right)_+ \pi^h_K(k_2|W_{1:n}) \mathbb{1}_{\Omega_n \cap \Xi_n}
\leq \mathbb{E}^P \int_K \int_K f_n(k_1, k_2) d\Phi_n^K(k_1) d\pi^h_K(k_2|W_{1:n}) \mathbb{1}_{\Omega_n \cap \Xi_n}
\leq \mathbb{E}^P \sup_{k_1, k_2 \in K} f_n(k_1, k_2) \mathbb{1}_{\Omega_n \cap \Xi_n} \leq \eta. \quad \text{(E.4)}$$

Moreover, by remembering that $\|\cdot\|_{TV}$ is upper bounded by 2,

$$\mathbb{E}^P\|\pi^h_k(\cdot|W_{1:n}) - \Phi_n^K\|_{TV}\mathbb{1}_{\Xi_n} \leq \mathbb{E}^P\|\pi^h_k(\cdot|W_{1:n}) - \Phi_n^K\|_{TV}\mathbb{1}_{\Omega_n \cap \Xi_n} + 2P(\Omega_n \cap \Xi_n). \quad \text{(E.5)}$$
where the first term is upper bounded by $2\eta$ by (E.4) and the second term is $o(1)$ by (E.3). In the second step of the proof let $K_n$ be a sequence of closed balls in the parameter space of $h$ centred at 0 with radii $M_n \to \infty$ and redefine $\Xi_n$ accordingly. For each $n \geq 1$, (E.5) holds for these balls. Moreover, by (E.7) in Theorem E.1 below: $P(\Xi_n) \to 1$. Therefore, by the triangular inequality, the TV distance is upper bounded by

$$E_P^P[\|\pi^h(\cdot|W_{1:n}) - \Phi_n\|_{TV}] \leq E_P^P[\|\pi^h(\cdot|W_{1:n}) - \Phi_n\|_{TV1_{\Xi_n}}] + E_P^P[\|\pi^h(\cdot|W_{1:n}) - \Phi_n\|_{TV1_{\Xi'_n}}]$$

$$\leq E_P^P[\|\pi^h(\cdot|W_{1:n}) - \pi^h_{K_n}(\cdot|W_{1:n})\|_{TV}] + E_P^P[\|\pi^h_{K_n}(\cdot|W_{1:n}) - \Phi_n\|_{TV1_{\Xi_n}}]$$

$$+ E_P^P[\|\Phi_{K_n} - \Phi_n\|_{TV}] + P(\Xi'_n)$$

since $E_P^P[\|\pi^h_{K_n}(\cdot|W_{1:n}) - \Phi_n\|_{TV1_{\Xi_n}}] = o_P(1)$ by (E.5) and (E.4), and where in the third line we have used the fact that $E_P^P[\|\pi^h(\cdot|W_{1:n}) - \pi^h_{K_n}(\cdot|W_{1:n})\|_{TV}] = 2E_P^P(\pi^h_{K_n}(\cdot|W_{1:n}))$ and $\|\Phi_{K_n} - \Phi_n\|_{TV} = \|\Phi_{K_n} - \Phi_n\|_{TV} = o_P(1)$ by Kleijn and van der Vaart (2012, Lemma 5.2) since $\Delta_n, \theta_n$ is uniformly tight.

The next theorem establishes that the posterior of $\theta$ concentrates and puts all its mass on $\Theta_n := \{\|\theta - \theta_*\| \leq M_n/\sqrt{n}\}$ as $n \to \infty$.

**Theorem E.1 ((Posterior Consistency)).** Let the Assumptions of Lemma E.2 and Assumption 3.7 hold. Moreover, assume that there exists a constant $C > 0$ such that for any sequence $M_n \to \infty$,

$$P\left(\sup_{\|\theta - \theta_*\| > M_n/\sqrt{n}} \frac{1}{n} \sum_{i=1}^n (\ell_n,\theta(W_i) - \ell_n,\theta_*(W_i)) \leq -CM_n^2/n\right) \to 1,$$  
(6.6)

as $n \to \infty$. Then,

$$\pi(\sqrt{n}\|\theta - \theta_*\| > M_n|W_{1:n}) \overset{P}{\to} 0$$
(6.7)

for any $M_n \to \infty$, as $n \to \infty$.

**Proof.** Define the events $A_{n,1} := \{\sup_{\|\theta - \theta_*\| > M_n/\sqrt{n}} \frac{1}{n} \sum_{i=1}^n (\ell_n,\theta(W_i) - \ell_n,\theta_*(W_i)) \leq -CM_n^2/n\}$ and

$$A_{n,2} := \left\{\int_{\Theta} p(W_{1:n}|\theta)\pi(\theta)d\theta \geq e^{-CM_n^2/2}\right\}.$$  

By (6.6), $P(A_{n,1}^c) \to 0$ and by Lemma E.1 below, $P(A_{n,2}^c) \to 0$. Therefore,

$$E_P^P[\pi(\sqrt{n}\|\theta - \theta_*\| > M_n|W_{1:n})] \leq E_P^P[\pi(\sqrt{n}\|\theta - \theta_*\| > M_n|W_{1:n})|A_{n,1} \cap A_{n,2}] P(A_{n,1} \cap A_{n,2}) + o(1)$$

$$= E_P^P\left[\int_{\Theta} p(W_{1:n}|\theta)\pi(\theta)d\theta \right] A_{n,1} \cap A_{n,2} P(A_{n,1} \cap A_{n,2}) + o(1)$$
and
\[ \sup \pi \] norm) we obtain:
\[ \text{which exists under Assumption 3.7} \]
Then, Lemma E.1.
Let the Assumptions of Lemma E.2 hold and suppose Assumptions 3.7 is satisfied.
\[ \text{for every sequence } a_n \to 0. \]
Proof. For a given \( M > 0 \) define \( \mathcal{C} = \{ h \in \mathbb{R}^p : \| h \| \leq M \} \). Denote by \( h \mapsto \text{Rem}(h) \) the remaining term in (E.12) and remark that \( \sup_{h \in \mathcal{C}} |\text{Rem}(h)| \to 0 \) by the result in Lemma E.2 and compactness of \( \mathcal{C} \). Therefore, for a sequence \( \kappa_n \) that converges to zero slowly enough, the corresponding event \( B_n = \{ \sup_{h \in \mathcal{C}} |\text{Rem}(h)| \leq \kappa_n \} \) has probability \( P(B_n) \to 1 \). Consider the local parameter \( h = \sqrt{n}(\theta - \theta_s) \) and denote by \( \pi^h \) both its prior distribution and prior Lebesgue density which exists under Assumption 3.7 (a). By making the change of variable \( \theta \mapsto \theta_s + h/\sqrt{n} \) so that \( \pi(\theta)d\theta = \pi^h(h)dh \), we upper bound the probability in (E.9) as follows: for a sequence \( K_n \to \infty \),
\[ P \left( \int \frac{p(W_{1:n}|\theta)}{p(W_{1:n}|\theta_s)} \pi(\theta)d\theta < e^{-K_n^2} \right) \leq P \left( \int \frac{p(W_{1:n}|\theta_s + h/\sqrt{n})}{p(W_{1:n}|\theta_s)} \pi^h(h)dh < e^{-K_n^2} \right) \]
\[ = P \left( \left\{ \int_{\mathcal{C}} \sum_{i=1}^{n} (\ell_{i,\theta_s} + h/\sqrt{n}) \pi^h(h)dh < e^{-K_n^2} \right\} \cap B_n \right) + o_p(1). \] (E.10)
By replacing the LAN expansion (E.12) and by noting that for \( n \) sufficiently large, \( \kappa_n \leq \frac{1}{2} K_n^2 \) on \( B_n \) and \( \sup_{h \in \mathcal{C}} h'V_{\theta_s}^{-1}h \leq \sup_{h \in \mathcal{C}} \| h \|^2 \| V_{\theta_s}^{-1} \| \leq M^2\| V_{\theta_s}^{-1} \| \leq \frac{1}{2} K_n^2 \) (where \( \| V_{\theta_s}^{-1} \| \) denotes the operator norm) we obtain:
\[ P \left( \left\{ \int_{\mathcal{C}} \sum_{i=1}^{n} (\ell_{i,\theta_s} + h/\sqrt{n}) \pi^h(h)dh < e^{-K_n^2} \right\} \cap B_n \right) \leq P \left( \int_{\mathcal{C}} e^{h'V_{\theta_s}^{-1}\Delta_{n,\theta_s} \pi^h(h)dh} < e^{-3K_n^2/4} \right) + o_p(1) \]
\[ = P \left( \int_{\mathcal{C}} e^{h'V_{\theta_s}^{-1}\Delta_{n,\theta_s} \pi^h(h)dh} < e^{-10\log \pi^h(\mathcal{C})e^{-K_n^2/4}} \right) + o_p(1) \]
\[ \leq P \left( \exp \left\{ \int_{\mathcal{C}} h'V_{\theta_s}^{-1}\Delta_{n,\theta_s} \pi^h(h)dh \right\} < e^{K_n^2/8}e^{-K_n^2/4} \right) + o_p(1) \]
\[ \leq P \left( \int_{\mathcal{C}} h'V_{\theta_s}^{-1}\Delta_{n,\theta_s} \pi^h(h)dh < -K_n^2/8 \right) + o_p(1) \]
\[ \leq \frac{64}{K_n^4} E^P \left( \int_{\mathcal{C}} (h'V_{\theta_s}^{-1}\Delta_{n,\theta_s})^2 \pi^h(h)dh \right) + o_p(1) \leq \frac{64}{K_n^4} M^2 V_{\theta_s}^{-1} + o_p(1) \to 0 \] (E.11)
where in the second inequality we have used that, for $n$ large enough, $-\log \pi^h(\xi) \leq K_n^2/8$ and the Jensen’s inequality. In the last line we have used the Markov’s inequality and then the Jensen’s inequality. The result follows by Assumption 3.7 (b) and because $V_{\theta_*}^{-1} \Delta_{n,\theta_*} \overset{d}{\to} \mathcal{N}(0, V_{\theta_*}^{-1})$.

\[ \square \]

**Lemma E.2** (Stochastic LAN). Let Assumptions 3.1, 3.2, 3.3, 3.5 and 3.6 be satisfied and assume $\zeta(K)K^2/\sqrt{n} \to 0$. Let $\mathcal{H}$ denote a compact subset of $\mathbb{R}^p$. Then,

\[
\sup_{h \in \mathcal{H}} \left| \sum_{i=1}^n \ell_{n,\theta_*+h/\sqrt{n}}(W_i) - \sum_{i=1}^n \ell_{n,\theta_*}(W_i) - h'V_{\theta_*}^{-1} \Delta_{n,\theta_*} - \frac{1}{2} h'V_{\theta_*}^{-1} h \right| = o_p(1) \quad (E.12)
\]

where $V_{\theta_*}^{-1} \Delta_{n,\theta_*} \overset{d}{\to} \mathcal{N}(0, V_{\theta_*}^{-1})$ and $\frac{1}{\sqrt{n}} \sum_{i=1}^n \frac{d\ell_{n,\theta_*}(W_i)}{d\theta} - V_{\theta_*}^{-1} \Delta_{n,\theta_*} \overset{p}{\to} 0$.

**Proof.** Define $\tau_i(\lambda, \theta) := \frac{e^{\lambda g_i(\theta)}}{E_n[e^{\lambda g_i(\theta)}]}$, $\hat{\Omega}(\theta, \lambda) := E_n[\tau_i(\lambda, \theta)g_i(\theta)g_i(\theta)^\prime]$ and $\hat{G}(\theta, \lambda) := E_n[\tau_i(\lambda, \theta)G(W_i, \theta)]$. By the MVT expansion of $\sum_{i=1}^n \ell_{n,\theta}(W_i)$ around $\hat{\lambda}(\theta) = 0$, there exists a $\hat{\lambda}_0$ lying on the line between $\hat{\lambda}(\theta)$ and zero such that:

\[
\sum_{i=1}^n \ell_{n,\theta}(W_i) = \sum_{i=1}^n \log \tau_i(\hat{\lambda}, \theta) - n \log(n) = -n \log(n) + n \hat{g}(\theta)^\prime \hat{\lambda}(\theta) - n \hat{g}(\theta)^\prime \hat{\lambda}_0(\theta) - \frac{1}{2} n \hat{\lambda}(\theta)^\prime \hat{\Omega}(\theta, \hat{\lambda}_0) \hat{\lambda}(\theta) + \frac{1}{2} n \left| E_n[\tau_i(\hat{\lambda}_0, \theta)g_i(\theta)] \hat{\lambda}(\theta) \right|^2. \quad (E.13)
\]

By expanding the first order condition for $\hat{\lambda}(\theta)$ around $\hat{\lambda}(\theta) = 0$, there exists a $\hat{\lambda}_0$ lying on the line between $\hat{\lambda}(\theta)$ and zero such that: $\hat{g}(\theta) + \hat{\Omega}(\theta, \hat{\lambda}_0) \hat{\lambda}(\theta) = 0$ which gives $\hat{\lambda}(\theta) = \hat{\Omega}(\theta, \hat{\lambda}_0)^{-1} \hat{g}(\theta)$. By replacing this in (E.13) we obtain:

\[
\sum_{i=1}^n \ell_{n,\theta}(W_i) = -n \log(n) + \frac{1}{2} n \hat{g}(\theta)^\prime \hat{\Omega}(\theta, \hat{\lambda}_0)^{-1} \hat{\Omega}(\theta, \hat{\lambda}_0) \hat{g}(\theta) - \frac{1}{2} n \left| E_n[\tau_i(\hat{\lambda}_0, \theta)g_i(\theta)] \hat{\Omega}(\theta, \hat{\lambda}_0)^{-1} \hat{g}(\theta) \right|^2. \quad (E.14)
\]

Hence, by replacing in $\sum_{i=1}^n \ell_{n,\theta_*+h_n}(W_i)$ the following MVT expansion $\hat{g}(\theta_* + h/\sqrt{n}) = \hat{g}(\theta_*) + \hat{G}(\theta_*) h/\sqrt{n}$ for $\theta$ lying between $\theta_* + h/\sqrt{n}$ and $\theta_*$, and by denoting $\bar{\theta}_1 := \theta_* + h_n$, $\bar{\theta}_n := h/\sqrt{n}$ we get

\[
\sum_{i=1}^n \ell_{n,\theta_*+h_n}(W_i) - \sum_{i=1}^n \ell_{n,\theta_*}(W_i) = \frac{1}{2} n \hat{g}(\theta_*)^\prime \left[ \hat{\Omega}(\theta_*, \hat{\lambda}_0)^{-1} \hat{\Omega}(\theta_*, \hat{\lambda}_0) \hat{\Omega}(\theta_*, \hat{\lambda}_0)^{-1} - \hat{\Omega}(\theta_1, \hat{\lambda}_0_1)^{-1} \hat{\Omega}(\theta_1, \hat{\lambda}_0_1) \hat{\Omega}(\theta_1, \hat{\lambda}_0_1)^{-1} \right] \hat{g}(\theta_*)
\]

\[
- \sqrt{n} \hat{g}(\theta_*)^\prime \hat{\Omega}(\theta_1, \hat{\lambda}_0_1)^{-1} \hat{\Omega}(\theta_1, \hat{\lambda}_0_1) \hat{\Omega}(\theta_1, \hat{\lambda}_0_1)^{-1} \hat{G}(\theta_1) h
\]
\[- \frac{1}{2} h' \hat{G}(\tilde{\theta})' \Omega(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \Omega(\theta_1, \tilde{\lambda}_{\theta_1}) \Omega(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \hat{G}(\tilde{\theta}) h + \frac{1}{2} n \left| \mathbb{E}_n [\tau_i (\tilde{\lambda}_{\theta_1}, \theta_1) g_i (\theta_1)' \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} g(\theta_1) ] \right|^2 - \frac{1}{2} n \left| \mathbb{E}_n [\tau_i (\tilde{\lambda}_{\theta_2}, \theta_2) g_i (\theta_2)' \hat{\Omega}(\theta_2, \tilde{\lambda}_{\theta_2})^{-1} g(\theta_2) ] \right|^2. \]

(E.15)

By using the equality \( A^{-1}BA^{-1} - C^{-1}DC^{-1} = A^{-1}(B-D)A^{-1} + (A^{-1}-C^{-1})DA^{-1} + C^{-1}D(A^{-1}-C^{-1}) \) for matrices \( A, B, C, D \) we can write

\[
\sum_{i=1}^n \ell_{n, \theta_1} (W_i) = \sum_{i=1}^n \ell_{n, \theta_2} (W_i)
\]

\[
= \frac{1}{2} n \hat{g}(\theta_1)' \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \left[ \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) - \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \right] \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \hat{g}(\theta_1) + \frac{1}{2} n \hat{g}(\theta_1)' \left[ \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} - \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \right] \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \hat{g}(\theta_1) + \frac{1}{2} n \hat{g}(\theta_1)' \left[ \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} - \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \right] \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \hat{g}(\theta_1)
\]

\[
- \frac{1}{2} \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \hat{g}(\theta_1)' \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \hat{g}(\theta_1) + \frac{1}{2} n \left| \mathbb{E}_n [\tau_i (\tilde{\lambda}_{\theta_1}, \theta_1) g_i (\theta_1)' \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} g(\theta_1) ] \right|^2 - \frac{1}{2} n \left| \mathbb{E}_n [\tau_i (\tilde{\lambda}_{\theta_2}, \theta_2) g_i (\theta_2)' \hat{\Omega}(\theta_2, \tilde{\lambda}_{\theta_2})^{-1} g(\theta_2) ] \right|^2.
\]

(E.16)

Let us analyse the first three terms in (E.16). Since \( \tilde{\lambda}_{\theta_1}, \tilde{\lambda}_{\theta_2}, \tilde{\lambda}_{\theta_1} \in \Lambda_n \), where \( \Lambda_n \) is as defined in Lemma G.2, in the following we can use the results in Lemma G.4 to get \( \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \leq C \hat{\Omega}(\theta_1)^{-1}, \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \leq C \hat{\Omega}(\theta_1)^{-1} \) and \( \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \leq C \hat{\Omega}(\theta_1)^{-1} \) with probability approaching 1 for any \( 1 < C < \infty \). We start from the first term:

\[
\sup_{h \in \mathcal{H}} R_{n,1}(h) := \frac{1}{2} \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \hat{g}(\theta_1)' \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \left[ \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) - \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \right] \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \hat{g}(\theta_1) \right|^2 \leq \frac{1}{2} \left\| \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \sqrt{n} \hat{g}(\theta_1) \right\| \sup_{h \in \mathcal{H}} \left\| \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) - \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \right\| \leq \left( \min_{1 \leq i \leq n} \tau_i (\tilde{\lambda}_{\theta_1}, \theta_1) \right)^{-2} \left\| \hat{\Omega}(\theta_1)^{-1} \sqrt{n} \hat{g}(\theta_1) \right\|^2 O_p \left( \zeta(K)K/\sqrt{n} \right) = O_p \left( \zeta(K)K^2/\sqrt{n} \right)
\]

by using the first result in Lemma G.7 and because \( \left\| \hat{\Omega}(\theta_1)^{-1} \sqrt{n} \hat{g}(\theta_1) \right\| = \left\| \Omega^{-1}_n \sqrt{n} \hat{g}(\theta_1) \right\| \) with probability approaching 1 by Donald et al. (2003, Lemma A.6) and \( \left\| \Omega^{-1}_n \sqrt{n} \hat{g}(\theta_1) \right\| = O_p(\sqrt{K}) \) by M. For the second term we use the identity \( (A^{-1} - B^{-1}) = A^{-1}(B-A)B^{-1} \) for two matrices \( A, B \), and again the first result in Lemma G.7:

\[
\sup_{h \in \mathcal{H}} R_{n,2}(h) := \frac{1}{2} \sup_{h \in \mathcal{H}} \left| \frac{1}{n} \hat{g}(\theta_1)' \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \left[ \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) - \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \right] \hat{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \right|^2 \]

7
\[
\times \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \tilde{\Omega}(\theta_*, \tilde{\lambda}_{\theta_*})^{-1} \tilde{g}(\theta_*)
\leq \frac{1}{2} \| \tilde{\Omega}(\theta_*, \tilde{\lambda}_{\theta_*})^{-1} \sqrt{n} \tilde{g}(\theta_*) \|^2 O_p(\zeta(K)K/\sqrt{n}) = O_p(\zeta(K)K^2/\sqrt{n}) .
\]

The third term can be treated in a similar way and gives the same rate.

Next, we analyze the last two terms in (E.16). We use again Lemma G.4. Therefore, because \(\| \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{g}(\theta_*) \| = \| \Omega_*^{-1} \sqrt{n} \tilde{g}(\theta_*) \| \) with probability approaching 1 by Donald et al. (2003, Lemma A.6) and \(\| \Omega_*^{-1} \sqrt{n} \tilde{g}(\theta_*) \| = O_p(\sqrt{K})\) by M.

\[
\sup_{h \in \mathcal{H}} \frac{1}{2n} \left| E_n[\tau_i(\tilde{\lambda}_{\theta_1}, \theta_1)g_i(\theta_1)'] \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{g}(\theta_1) \right|^2
\leq \frac{1}{2} CE_n[\tau_i(\tilde{\lambda}_{\theta_1}, \theta_1)g_i(\theta_1)']^2 \| \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{g}(\theta_1) \|^2 = O_p(K^2/n) \quad (E.17)
\]

where we have used the MVT expansion \(\tilde{g}(\theta_1) = \tilde{g}(\theta_*) + \tilde{G}(\tilde{\theta}) h/\sqrt{n}\) for a \(\tilde{\theta}\) lying between \(\theta_1\) and \(\theta_*\) and the result of Lemma G.5. We conclude that

\[
\sup_{h \in \mathcal{H}} \frac{1}{n} \left| \sum_{i=1}^n \ell_{n, \theta_*} h_n(W_i) - \sum_{i=1}^n \ell_{n, \theta_*}(W_i) - \sqrt{n} \tilde{g}(\theta_*)' \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{G}(\tilde{\theta}) h \right|
\leq \frac{1}{2} h' \tilde{G}(\tilde{\theta})' \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{G}(\tilde{\theta}) h
= O_p(\zeta(K)K^2/\sqrt{n}) + O_p(K/\sqrt{n}) = O_p(\zeta(K)K^2/\sqrt{n}) . \quad (E.18)
\]

Under the assumptions of the theorem the term \(O_p(\zeta(K)K^2/\sqrt{n})\) converges to zero. Moreover, by Lemma G.7 and \(\zeta(K)K/\sqrt{n} \rightarrow 0:\)

\[
- \sqrt{n} \tilde{g}(\theta_*)' \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{G}(\tilde{\theta}) h
= -\sqrt{n} \tilde{g}(\theta_*)' \tilde{\Omega}(\theta_*)^{-1} \tilde{G}(\tilde{\theta}) h + o_p(1)
= -\frac{h'}{\sqrt{n}} \sum_{i=1}^n D(Z_i)' \Sigma(Z_i)^{-1} \rho(X_i, \theta_*)' + o_p(1)
\]

where the \(o_p(1)\) term is uniform in \(h \in \mathcal{H}\) and where to get the second equality we have used arguments similar to the ones in Donald et al. (2003, Proof of Theorem 5.6). By the Lindberg-Levy central limit theorem, \(\Delta_{n, \theta_*} := -\frac{1}{\sqrt{n}} \sum_{i=1}^n V_{\theta_*} D(Z_i)' \Sigma(Z_i)^{-1} \rho(X_i, \theta_*)' \xrightarrow{d} \mathcal{N}(0, V_{\theta_*}).\) Similarly as in Donald et al. (2003, Proof of Theorem 5.6) it is possible to show that (by using compactness of \(\mathcal{H}\))

\[
h' \tilde{G}(\tilde{\theta})' \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1}) \tilde{\Omega}(\theta_1, \tilde{\lambda}_{\theta_1})^{-1} \tilde{G}(\tilde{\theta}) h = h' V_{\theta_*} h + o_p(1)
\]

8
where the $o_p(1)$ term is uniform in $h \in H$. Finally, remark that
\[
\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{d\ell_{n,\theta_*(W_i)}}{d\theta} - V_{\theta_*}^{-1} \Delta_{n,\theta_*} \xrightarrow{p} 0.
\]
This establishes the result of the Lemma.

\[\square\]

### E.2 The Misspecified case: Proof of Theorem 3.2

For the misspecified case we use the following notation: $G_i(\theta) := G(W_i, \theta)$, $G_0 := E[G_i(\theta_0)]$, $\tilde{G}_0 := E[\tau_1(\lambda_0(\theta_0), \theta_0)G_i(\theta)]$, $\Omega_0 := E[\tau(\lambda_0(\theta_0), \theta_0, W_i)g_i(\theta_0)g_i(\theta_0)'|\theta_0]$, $\tau(\lambda, \theta) := \frac{e^{X_i g_i(\theta)}}{E_n[e^{X_i g_i(\theta)}]}$, $\tilde{G}(\lambda, \theta) := E_n[\tau(\lambda, \theta)G_i(\theta)]$, $\tilde{\Omega}(\theta, \lambda) := E_n[\tau(\lambda, \theta, W_i)g_i(\theta)g_i(\theta)'].$ We also use standard notation in empirical process theory: $P_n := E_n[\delta_{X_i}]$ where $\delta_x$ is the Dirac measure at $x$, and $G_nf := \sqrt{n}(P_n f - E^P f)$ for every function $f$.

The proof of Theorem 3.2 proceeds as the proof of Theorem 3.1 and so we omit it. The only differences between the proofs of the two theorems consist in replacing Lemma E.2 with Lemma E.4 and Theorem E.1 with Theorem E.2. In the following of this section we prove Lemma E.4 and state Theorem E.2.

We omit the proof of Theorem E.2 because it proceeds as the proof of Theorem E.1 with Lemma E.1 replaced by Lemma E.3. The proof of Lemma E.3 is the same as the proof of Lemma E.1 with the LAN expansion (E.12) replaced by the LAN expansion (E.22) and then it is also omitted.

**Theorem E.2** (Posterior Consistency). Let the Assumptions of Lemma E.4 and Assumption 3.9 hold. Moreover, assume that there exists a constant $C > 0$ such that for any sequence $M_n \to \infty$,
\[
P \left( \sup_{||\theta - \theta_0|| > M_n/\sqrt{n}} \frac{1}{n} \sum_{i=1}^{n} (\ell_{n,\theta}(W_i) - \ell_{n,\theta_0}(W_i)) \leq -CM_n^2/n \right) \to 1, \tag{E.19}
\]
as $n \to \infty$. Then,
\[
\pi \left( \sqrt{n} ||\theta - \theta_0|| > M_n | W_{1:n} \right) \xrightarrow{p} 0 \tag{E.20}
\]
for any $M_n \to \infty$, as $n \to \infty$.

**Lemma E.3.** Let the Assumptions of Lemma E.4 hold and suppose Assumptions 3.9 is satisfied. Then,
\[
P \left( \int_{\Theta} \frac{p(W_{1:n}|\theta)}{p(W_{1:n}|\theta_0)} \pi(\theta)d\theta < a_n \right) \to 0 \tag{E.21}
\]
for every sequence $a_n \to 0$. 
Lemma E.4 (Stochastic LAN). Let Assumptions 3.1, 3.2, 3.6 and 3.10 hold and assume \( \zeta(K)K^2\sqrt{K/n} \to 0 \). Let \( \mathcal{H} \) denote a compact subset of \( \mathbb{R}^p \) and \( \theta_1 := \theta_0 + h/\sqrt{n} \) with \( h \in \mathcal{H} \). Then,

\[
\sup_{h \in \mathcal{H}} \left| \sum_{i=1}^{n} \ell_{n,\theta}(W_i) - \sum_{i=1}^{n} \ell_{n,\theta_0}(W_i) - h^\prime A_{\theta_0} \Delta_{n,\theta_0} - \frac{1}{2} h^\prime A_{\theta_0} h \right| = o_p(1) 
\] (E.22)

where \( \Delta_{n,\theta_0} \) is a random vector bounded in probability and \( A_{\theta_0} \) is a nonsingular matrix.

Proof. We have to analyse the difference \( \sum_{i=1}^{n} \ell_{n,\theta_1}(W_i) - \sum_{i=1}^{n} \ell_{n,\theta_0}(W_i) \). Because \( W_i \) are i.i.d. then \( \mathbb{E}[g_i(\theta)] = \mathbb{E}[g_j(\theta)] \), and so we can write:

\[
\sum_{i=1}^{n} \ell_{n,\theta}(W_i) = \sum_{i=1}^{n} \log \tau_i(\hat{\lambda}, \theta) - n \log(n) = \sum_{i=1}^{n} \log \frac{e^{\tilde{\lambda}(g_i(\theta) - \mathbb{E}[g_i(\theta)])}}{\mathbb{E}_n[e^{\tilde{\lambda}(g_i(\theta) - \mathbb{E}[g_i(\theta)])}]} - n \log(n) 
\]

\[
= n\tilde{\lambda}(\theta)^\prime \mathbb{E}_n(g_i(\theta) - \mathbb{E}[g_i(\theta)]) - n \log \mathbb{E}_n[e^{\tilde{\lambda}(\theta)^\prime(g_i(\theta) - \mathbb{E}[g_i(\theta)])}] - n \log(n). 
\] (E.23)

Denote \( \overline{g}_i(\theta) := g_i(\theta) - \mathbb{E}[g_i(\theta)] \) and \( \overline{G}_i(\theta) := G_i(\theta) - \mathbb{E}[G_i(\theta)] \), so that \( \sum_{i=1}^{n} \ell_{n,\theta}(W_i) = n\tilde{\lambda}(\theta)^\prime \mathbb{E}_n[\overline{g}_i(\theta)] - n \log \mathbb{E}_n[e^{\tilde{\lambda}(\theta)^\prime \overline{g}_i(\theta)}] - n \log(n) \). By the MVT there exists a \( t \in [0, 1] \) such that \( \tilde{\theta} := \theta_0 + th/\sqrt{n} \) satisfies

\[
\sum_{i=1}^{n} \ell_{n,\theta_1}(W_i) = \sum_{i=1}^{n} \ell_{n,\theta_0}(W_i) + \frac{h'}{\sqrt{n}} \sum_{i=1}^{n} \hat{\ell}_{n,\theta}(W_i) + \frac{1}{2} \frac{h'}{\sqrt{n}} \sum_{i=1}^{n} \hat{\ell}_{n,\theta}(W_i)^2 \frac{h}{\sqrt{n}} 
\] (E.24)

where

\[
\hat{\ell}_{n,\theta}(W_i) := \sum_{i=1}^{n} \frac{d\ell_{n,\theta}(W_i)}{d\theta} \bigg|_{\theta_1=\theta_0} = \frac{d\tilde{\lambda}(\theta_0)^\prime}{d\theta} \overline{g}_i(\theta_0) + \overline{G}_i(\theta_0)^\prime \hat{\lambda}(\theta_0) 
\]

\[
+ \frac{d\hat{\lambda}(\theta_0)^\prime}{d\theta} \mathbb{E}[g_i(\theta_0)] - \mathbb{E}_n \left[ \tau_i(\hat{\lambda}(\theta_0), \theta_0) \overline{G}_i(\theta_0)^\prime \hat{\lambda}(\theta_0) \right] \hat{\lambda}(\theta_0) 
\]

\[
= \frac{d\tilde{\lambda}(\theta_0)^\prime}{d\theta} \overline{g}_i(\theta_0) + \overline{G}_i(\theta_0)^\prime \lambda_0(\theta_0) + \overline{G}_i(\theta_0)^\prime (\hat{\lambda}(\theta_0) - \lambda_0(\theta_0)) + \left( \frac{d\tilde{\lambda}(\theta_0)^\prime}{d\theta} - \frac{d\lambda_0(\theta_0)^\prime}{d\theta} \right) \mathbb{E}[g_i(\theta_0)] 
\]

\[- \mathbb{E}_n \left[ \left( \tau_i(\hat{\lambda}(\theta_0), \theta_0) - \tau_i(\lambda_0, \theta_0) \right) G_i(\theta_0)^\prime \hat{\lambda}(\theta_0) \right] \hat{\lambda}(\theta_0) - \mathbb{E}_n \left[ \tau_i(\lambda_0, \theta_0) G_i(\theta_0)^\prime - G_i^\prime \right] \lambda_0(\theta_0) 
\]

\[- \mathbb{E}_n [\tau_i(\lambda_0, \theta_0) G_i(\theta_0)^\prime - G_i^\prime] (\hat{\lambda}(\theta_0) - \lambda_0(\theta_0)) \] (E.25)

with

\[
\frac{d\tilde{\lambda}(\theta_0)^\prime}{d\theta} = -\mathbb{E}_n \left[ \tau_i(\hat{\lambda}, \theta_0) G_i(\theta_0)^\prime (I + \hat{\lambda}(\theta_0) g_i(\theta_0)^\prime) \right] ^{-1} \left( \mathbb{E}_n [\tau_i(\hat{\lambda}, \theta_0) g_i(\theta_0)^\prime] \right)^{-1}, 
\]

\[
\frac{d\lambda_0(\theta_0)^\prime}{d\theta} = -\mathbb{E} \left[ \tau(\lambda_0, \theta_0, W_i) G_i(\theta_0)^\prime (I + \lambda_0(\theta_0) g_i(\theta_0)^\prime) \right] \Omega^{-1} \] (E.26)
and where we have used the first order condition of the pseudo true value \( \theta_0 \), that is:

\[
\frac{d\lambda_0(\theta_0)'}{d\theta} E[g_i(\theta_0)] + G'_n\lambda_0(\theta_0) - \frac{d\lambda_0(\theta_0)'}{d\theta} E[\tau_i(\lambda_0, \theta_0)g_i(\theta_0)] - E[\tau_i(\lambda_0, \theta_0)G_i(\theta_0)']\lambda_0(\theta_0) = 0
\]

and \( E[\tau_i(\lambda_0, \theta_0)g_i(\theta_0)] = 0 \) because it is the first order condition for \( \lambda_0 \). Moreover,

\[
\ell_{n,\hat{\theta}}(W_i) := \sum_{i=1}^{n} \frac{d^2 \ell_{n,\hat{\theta}}(W_i)}{d\theta d\theta'} \bigg|_{\theta = \hat{\theta}} = \sum_{j=1}^{dK} \frac{d^2 \tilde{\lambda}(\tilde{\theta})'}{d\theta d\theta'} g_{i,j}(\tilde{\theta}) + \frac{d\tilde{\lambda}(\tilde{\theta})'}{d\theta'} - G_i(\tilde{\theta}) + \frac{d\tilde{\lambda}(\tilde{\theta})'}{d\theta'} \tilde{\lambda}(\tilde{\theta}) + \frac{d\tilde{\lambda}(\tilde{\theta})'}{d\theta'} G_i(\tilde{\theta}) + E_n \left[ \frac{d\tau_i(\tilde{\lambda}(\tilde{\theta}), \tilde{\theta})}{d\theta'} \tilde{\lambda}(\tilde{\theta}) + E_n \left[ \frac{d\tau_i(\tilde{\lambda}(\tilde{\theta}), \tilde{\theta}) G_i(\tilde{\theta})}{d\theta'} \right] d\lambda(\tilde{\theta}) \right].
\]

We start with analyzing term (E.25). First, remark that by Lemma G.12 it holds: \( h' \left( \frac{d\lambda(\theta_0)'}{d\theta} \right) \sqrt{n} E_n[\tilde{g}_i(\theta_0)] = o_p(1) \) uniformly in \( h \in \mathcal{H} \). Next, we analyse the third term in (E.25). By a MVT expansion of the first order condition for \( \tilde{\lambda}(\hat{\theta}) \) there exists \( \tau \in [0, 1] \) such that \( \tilde{\lambda}_\tau := \tau(\tilde{\lambda}(\theta_0) - \lambda_0(\theta_0)) + \lambda_0(\theta_0) \) satisfies \( E_n[e^{\tilde{\lambda}_\tau'}g_i(\theta_0)g_i(\theta_0)] = 0 = E_n[e^{\lambda'(\theta_0)'}g_i(\theta_0)g_i(\theta_0)] + \Omega(\theta_0, \tilde{\lambda}_\tau)(\tilde{\lambda}(\theta_0) - \lambda_0(\theta_0)) \) which implies:

\[
(\tilde{\lambda}(\theta_0) - \lambda_0(\theta_0)) = -\Omega(\theta_0, \tilde{\lambda}_\tau)^{-1} E_n[e^{\lambda'(\theta_0)'}g_i(\theta_0)g_i(\theta_0)]. \tag{E.28}
\]

Therefore,

\[
h' \frac{1}{\sqrt{n}} \sum_{i=1}^{n} G_i(\theta_0)'(\tilde{\lambda}(\theta_0) - \lambda_0(\theta_0)) = -\sqrt{n} E_n[e^{\lambda'(\theta_0)'}g_i(\theta_0)g_i(\theta_0)] \Omega(\theta_0, \tilde{\lambda}_\tau)^{-1} E_n[\tilde{G}_i(\theta_0)] h
\]

\[
= -G_n[e^{\lambda'(\theta_0)'}g_i(\theta_0)g_i(\theta_0)] \Omega(\theta_0, \tilde{\lambda}_\tau)^{-1} E_n[\tilde{G}_i(\theta_0)] h = O_p(K/\sqrt{n}).
\]

Here, to get the term \( O_p(K/\sqrt{n}) \) we have used the inequality

\[
sup_{h \in \mathcal{H}} E \left| G_n[e^{\lambda'(\theta_0)'}g_i(\theta_0)g_i(\theta_0)] \Omega(\theta_0, \tilde{\lambda}_\tau)^{-1} E_n[\tilde{G}_i(\theta_0)] h \right|
\]

\[
\leq C^{-2} \sup_{h \in \mathcal{H}} \sqrt{E \| G_n[e^{\lambda'(\theta_0)'}g_i(\theta_0)g_i(\theta_0)] \|^2} \sqrt{E \| E_n[\tilde{G}_i(\theta_0)] h \|^2} = O_p(\sqrt{K} \sqrt{K/\sqrt{n}})
\]

for which we have used Lemma G.9 and Assumption 3.10 (d) and (f). To control the fourth term in (E.25) we use Assumption 3.10 (h). We now control the fifth term in (E.25). For this, we use again (E.28) and a MVT expansion of \( \tau_i(\tilde{\lambda}(\theta_0), \theta_0) \) around \( \lambda_0(\theta_0) \):

\[
\sqrt{n} \tilde{\lambda}(\theta_0)' E_n[\left( \tau_i(\tilde{\lambda}(\theta_0), \theta_0) - \tau_i(\lambda_0(\theta_0), \theta_0) \right) G_i(\theta_0)] h
\]
\[
\begin{align*}
&= \sqrt{n}\left(\hat{\lambda}(\theta_{0}) - \lambda_{0}(\theta_{0})\right)'E_n\left[\frac{\partial \tau_{i}(\lambda_{t}, \theta_{0})}{\partial \lambda}\hat{\lambda}(\theta_{0})'G_{i}(\theta_{0})\right]h \\
&= -\sqrt{n}E_n[e^{\lambda_{0}(\theta_{0})'g_{i}(\theta_{0})}g_{i}(\theta_{0})']\hat{\Omega}(\theta_{0}, \hat{\lambda}_{t})^{-1}E_n\left[\frac{\partial \tau_{i}(\lambda_{t}, \theta_{0})}{\partial \lambda}\lambda_{0}(\theta_{0})'G_{i}(\theta_{0})\right]h \\
&= -G_n[e^{\lambda_{0}(\theta_{0})'g_{i}(\theta_{0})}g_{i}(\theta_{0})']\hat{\Omega}(\theta_{0}, \hat{\lambda}_{t})^{-1}G_n[e^{\lambda_{0}(\theta_{0})'g_{i}(\theta_{0})}g_{i}(\theta_{0})] + o_{p}(1) \quad \text{(E.29)}
\end{align*}
\]

where \( \lambda_{t} = t(\hat{\lambda}(\theta_{0}) - \lambda_{0}(\theta_{0})) + \lambda_{0}(\theta_{0}) \) for some \( t \in [0, 1] \). To control the last term in (E.25) we use again (E.28) to get

\[
- h'\sqrt{n}E_n[\tau_{i}(\lambda_{0}, \theta_{0})G_{i}(\theta_{0})' - G_{0}'(\hat{\lambda}(\theta_{0}) - \lambda_{0}(\theta_{0}))] = h'E_n[\tau_{i}(\lambda_{0}, \theta_{0})G_{i}(\theta_{0})' - G_{0}'\hat{\Omega}(\theta_{0}, \hat{\lambda}_{t})^{-1}G_n[e^{\lambda_{0}(\theta_{0})'g_{i}(\theta_{0})}g_{i}(\theta_{0})] + o_{p}(1) =: h'A_{\theta_{0}}\Delta_{n, \theta_{0}} + o_{p}(1) \quad \text{(E.30)}
\]

where the \( o_{p}(1) \) is uniform in \( h \in \mathcal{H} \), and \( \hat{\mathcal{L}}_{n, \theta_{0}}(W_{i}) := \frac{d}{d\theta}\mathcal{L}_{n, \theta}(W_{i})\big|_{\theta = \theta_{0}} \) with \( \mathcal{L}_{n, \theta_{0}}(W_{i}) := \log(dQ^{*}(\theta_{0})/dP_{\theta})(W_{i}) = \log\frac{e^{\lambda_{0}(\theta_{0})'g_{i}(\theta_{0})}}{E[e^{\lambda_{0}(\theta_{0})'g_{i}(\theta_{0})}]} \). Moreover, as shown above \( \Delta_{n, \theta_{0}} = O_{p}(1) \) and \( A_{\theta_{0}} \) is defined below.

We now analyse the limit of (E.27). For this, we use Lemma G.12, the fact that

\[
\hat{\lambda}(\theta_{0}) - \lambda_{0}(\theta_{0}) = -\hat{\Omega}(\theta_{0}, \hat{\lambda}_{t})^{-1}E_n[e^{\lambda_{0}(\theta_{0})'g_{i}(\theta_{0})}g_{i}(\theta_{0})]
\]

as shown above, and the fact that \( \hat{\lambda}(\theta)' - \hat{\lambda}(\theta_{0})' = (\theta - \theta_{0})\frac{d\hat{\lambda}(\theta)}{d\theta} \) for \( \theta = \theta_{0} + th/\sqrt{n} \) and some \( t \in [0, 1] \), to get

\[
\begin{align*}
&= \frac{h}{\sqrt{n}}\sum_{i=1}^{n} \hat{\ell}_{n, \theta}(W_{i})h = h\frac{1}{n} \sum_{i=1}^{n} \frac{d^{K}\lambda_{0}(\theta_{0})'}{d\theta d\theta''}E[g_{i,j}(\theta)]h + h'\frac{d\lambda_{0}(\theta_{0})'}{d\theta}E[G_{i}(\theta)]h \\
&\quad + h'E\left[G_{i}(\theta)'\lambda_{0}(\theta_{0})\frac{d\tau_{i}(\lambda_{0}(\theta_{0}), \theta_{0})}{d\lambda_{0}(\theta_{0})}\frac{d\lambda_{0}(\theta_{0})}{d\theta''}\right]h + h'E\left[\frac{d\tau_{i}(\lambda_{0}(\theta_{0}), \theta_{0})}{d\theta}\lambda_{0}(\theta_{0})'G_{i}(\theta_{0})\right]h
\end{align*}
\]
\[
+h'E \left[ \tau_i(\lambda_o(\theta_o), \theta_o) \sum_{j=1}^{dK} \frac{d^2 g_{i,j}(\theta_o)}{d\theta d\theta'} \right] + h'E \left[ \tau_i(\lambda_o(\theta_o), \theta_o) G_i(\theta_o) \right] \frac{d\lambda_o(\theta_o)}{d\theta'} h + o_p(1) =: h' \mathcal{A}_{\theta_o} h
\]

(E.31)

where the \(o_p(1)\) is uniform in \(h \in \mathcal{H}\). By replacing (E.30) and (E.31) in (E.24) we get the result of the Lemma.

\[\square\]

F Proofs for Section 4

F.1 Proof of Theorem 4.1

We can write

\[
\log p(W_{1:n}|\theta^\ell; M_\ell) = -n \log n + n \log \widehat{L}(\theta^\ell)
\]

where \(\widehat{L}(\theta^\ell) := \exp\{\lambda(\theta^\ell)'g_i(\theta^\ell)\} \left[ \frac{1}{n} \sum_{i=1}^{n} \exp\{\lambda(\theta^\ell)'g_i(W_i, \theta^\ell)\} \right]^{-1}
\]

and \(L(\theta^\ell) = \exp\{\lambda(\theta^\ell)'E^P[g(W, \theta^\ell)]\} \left( E^P \left[ \exp\{\lambda(\theta^\ell)'g(W, \theta^\ell)\} \right] \right)^{-1}\). Then, we have:

\[
P \left( \log m(W_{1:n}; M_j) > \max_{\ell \neq j} \log m(W_{1:n}; M_\ell) \right) = P \left( n \log \widehat{L}(\theta^\ell_j) + \log \pi(\theta^\ell_j|M_j) - \log \pi(\theta^\ell_j|W_{1:n}, M_j) > \max_{\ell \neq j} \left[ n \log L(\theta^\ell_j) + B_j > \max_{\ell \neq j} \left[ n \log L(\theta^\ell_j) + B_j + n \log \frac{\widehat{L}(\theta^\ell_j)}{L(\theta^\ell_j)} \right] \right] \right) \quad (F.1)
\]

where \(\forall \ell, B_\ell := \log \pi(\theta^\ell_j|M_j) - \log \pi(\theta^\ell_j|W_{1:n}, M_\ell)\) and \(B_\ell = O_p(1)\) under the assumptions of Theorem 3.2. By definition of \(dQ^*(\theta)\) in Section 3.4 we have that:

\[
\log L(\theta^\ell_j) = E^P[\log dQ^*(\theta^\ell_j)/dP] = -E^P[\log dP/dQ^*(\theta^\ell_j)] = -K(P||Q^*(\theta^\ell_j)).
\]

Remark that \(E^P[\log dP/dQ^*(\theta^\ell_j)] > E^P[\log dP/dQ^*(\theta^\ell_j)]\) means that the KL divergence between \(P\) and \(Q^*(\theta^\ell_j)\), is smaller for model \(M_1\) than for model \(M_2\), where \(Q^*(\theta^\ell_j)\) minimizes the KL divergence between \(Q \in \mathcal{P}_{\theta^\ell_j}\) and \(P\) for \(\ell \in \{1, 2\}\) (notice the inversion of the two probabilities).

First, suppose that \(\min_{\ell \neq j} E^P[\log (dP/dQ^*(\theta^\ell_j))] > E^P \left[ \log \left( dP/dQ^*(\theta^\ell_j) \right) \right]\). By (F.1):

\[
P \left( \log m(W_{1:n}; M_j) > \max_{\ell \neq j} \log m(W_{1:n}; M_\ell) \right) \geq
\]

\[
P \left( \log \frac{\widehat{L}(\theta^\ell_j)}{L(\theta^\ell_j)} > \max_{\ell \neq j} \log \frac{\widehat{L}(\theta^\ell_j)}{L(\theta^\ell_j)} + \frac{1}{n} (B_j - \max_{\ell \neq j} B_\ell) > \max_{\ell \neq j} \left[ \log L(\theta^\ell_j) - \log L(\theta^\ell_j) \right] \right) \quad (F.2)
\]

13
This probability converges to 1 because $I_n = K(P||Q^*(\theta^*_1)) - \min_{j \neq j} K(P||Q^*(\theta^*_j)) < 0$ by assumption, and \[
\log \hat{L}(\theta^\ell) - \log L(\theta^\ell) \xrightarrow{P} 0, \text{ for every } \theta^\ell \in \Theta^\ell \text{ and every } \ell \in \{1, 2\} \text{ by Lemma G.10 below and by } K/\sqrt{n} \to 0.
\]

To prove the second direction of the statement, suppose that \(\lim_{n \to \infty} P(\log m(W_{1:n}; M_j) > \max_{\ell \neq j} \log m(W_{1:n}; M_\ell)) = 1\). By (F.1) it holds, \(\forall \ell \neq j\)

\[
P\left(\log m(W_{1:n}; M_j) > \max_{\ell \neq j} \log m(W_{1:n}; M_\ell)\right) \leq P\left(\log \frac{\hat{L}(\theta^\ell_j)}{L(\theta^\ell_j)} - \log \frac{\hat{L}(\theta^\ell_\ell)}{L(\theta^\ell_\ell)} + \frac{1}{n} (B_j - B_\ell) > \log \frac{L(\theta^\ell_j)}{L(\theta^\ell_\ell)}\right). \quad (F.3)
\]

Convergence to 1 of the left hand side implies convergence to 1 of the right hand side which is possible only if $\log L(\theta^\ell_j) - \log L(\theta^\ell_\ell) < 0$. Since this is true for every model $\ell$, then this implies that $K(P||Q^*(\theta^*_j)) < \min_{\ell \neq j} K(P||Q^*(\theta^*_\ell))$ which concludes the proof.

\[\square\]

G Technical Lemmas

**Lemma G.1.** If Assumptions 3.2 and 3.6 are satisfied then

\[
\max_{i \leq n} \sup_{\theta \in \Theta} \|\rho(X_i, \theta)\| = O_p(n^{1/\gamma})
\]

and

\[
\max_{i \leq n} \sup_{\theta \in \Theta} \|g(W_i, \theta)\| = O_p(n^{1/\gamma} \zeta(K)). \quad (G.1)
\]

**Proof.** To get the first conclusion, let $b_i = \sup_{\theta \in \Theta} \|\rho(X_i, \theta)\|$, so for every $\gamma > 1$, $P(\max_{i \leq n} b_i > \varepsilon) = P(\max_{i \leq n} b_i > \varepsilon) \leq P(\sum_{i \leq n} b_i > \varepsilon)$. By the Markov’s inequality this is upper bounded by $E[\sum_{i \leq n} b_i] / \varepsilon$ which, under Assumption 3.6 is bounded by $Cn/\varepsilon^\gamma$ for a constant $0 < C < \infty$. So, $\max_{i \leq n} b_i = O_p(n^{1/\gamma})$. The second result follows from the first result, the inequality

\[
\max_{i \leq n} \sup_{\theta \in \Theta} \|g(W_i, \theta)\| \leq \max_{i \leq n} \sup_{\theta \in \Theta} \|\rho(X_i, \theta)\| \max_{i \leq n} \|q^K(Z_i)\|
\]

and Assumption 3.2.

\[\square\]

**Lemma G.2.** If Assumptions 3.2 and 3.6 is satisfied then for any sequence $\delta_n = o(n^{-1/\gamma} \zeta(K)^{-1})$...
and $\Lambda_n := \{ \lambda \in \mathbb{R}^{dK}; \|\lambda\| \leq \delta_n \}$ we have

$$\max_{i \leq n, \lambda \in \Lambda_n} \sup_{\theta \in \Theta} |\lambda' g(W_i, \theta)| \to 0$$

in $P_\lambda$-probability.

Proof. By using the second result of Lemma G.1 we get

$$\max_{i \leq n, \lambda \in \Lambda_n} \sup_{\theta \in \Theta} |\lambda' g(W_i, \theta)| \leq \max_{i \leq n, \lambda \in \Lambda_n} \sup_{\theta \in \Theta} \|\lambda\| \|g(W_i, \theta)\| = O_p(\delta_n n^{1/\gamma} \zeta(K))$$

which converges to zero.

\[\square\]

G.1 Technical Lemmas for the correctly specified case

Lemma G.3. Assume that $\Sigma(Z)$ is bounded and let $\hat{\lambda}(\theta) = \text{argmin}_{\lambda \in \mathbb{R}^{dK}} E_n[ e^{\lambda' g(W_i, \theta)}]$. If Assumptions 3.2, 3.3, 3.5 (b) and 3.6 are satisfied, then

$$\|\hat{\lambda}(\theta_*)\| = O_p(\sqrt{\frac{K}{n}}) \quad (G.2)$$

and

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^{n} e^{\hat{\lambda}(\theta)' g(W_i, \theta)} \geq \frac{1}{n} \sum_{i=1}^{n} e^{\hat{\lambda}_* (\theta)' g(W_i, \theta_*)} \geq 1 + O_p(K/n). \quad (G.3)$$

Proof. Choose a sequence $\delta_n = o(n^{-1/\gamma} \zeta(K)^{-1})$ and $\sqrt{K/n} = o(\delta_n)$, which is possible by $\zeta(K)^2 K/n^{1-2\gamma} \to 0$. Let $\lambda$ be on the line joining $\hat{\lambda}(\theta)$ and 0. By a second order Mean Value expansion around $\hat{\lambda}(\theta) = 0$, we have $\forall \theta \in \Theta$:

$$1 \geq E_n \left[ e^{\hat{\lambda}(\theta)' g(W_i, \theta)} \right] = 1 + \hat{\lambda}(\theta)' E_n [g(W_i, \theta)] + \frac{1}{2} \hat{\lambda}(\theta)' E_n \left[ e^{\hat{\lambda}(\theta)' g(W_i, \theta)} g(W_i, \theta) g(W_i, \theta)' \right] \hat{\lambda}(\theta)$$

$$\geq 1 - \|\hat{\lambda}(\theta)\| \|E_n [g(W_i, \theta)]\| + \min_{1 \leq i \leq n} e^{-\|\hat{\lambda}(\theta)' g(W_i, \theta)\|} \lambda_{\min}(\hat{\Omega}(\theta)) \|\hat{\lambda}(\theta)\|^2. \quad (G.4)$$

By Donald et al. (2003, Lemma A.6): $\lambda_{\min}(\hat{\Omega}(\theta)) \geq 1/C$ with probability approaching 1. Let $\Lambda_n$ be as defined in Lemma G.2: $\lambda_{\min} \in \Lambda_n \Rightarrow E_n \left[ e^{\lambda' g(W_i, \theta)} \right]$. Let $\hat{\lambda}(\theta_*)$ be on the line joining $\hat{\lambda}(\theta_*)$ and zero. Then, by Lemma G.2: $\min_{1 \leq i \leq n} e^{-\|\hat{\lambda}(\theta_*)' g(W_i, \theta)\|} \geq e^{-\max_{i \leq n, \lambda \in \Lambda_n} \sup_{\theta \in \Theta} |\lambda(\theta)' g(W_i, \theta)|} \leq C$ for some finite constant $C > 0$. Therefore, after simplifica-
tions, (G.4) evaluated at $(\widetilde{\lambda}(\theta_*), \theta_*)$ gives:

$$\|\tilde{\lambda}(\theta_*)\| \leq \frac{C}{C} \|E_n[g(W_i, \theta_*)]\| = O_p(\sqrt{K/n})$$

where the last equality uses the result of Donald et al. (2003, Lemma A.9). Then, $\|\tilde{\lambda}(\theta_*)\| < \delta_n$ and $\tilde{\lambda}(\theta_*) \in \text{int}(\Lambda_n)$. By convexity of $E_n[e^{\lambda'g(W_i, \theta)}]$ it then follows that $\tilde{\lambda}(\theta_*) = \tilde{\lambda}(\theta_*)$, which gives result (G.15).

Moreover, by using (G.4):

$$\sup_{\theta \in \Theta} \frac{1}{n} \sum_{i=1}^n e^{\tilde{\lambda}(\theta)'g(W_i, \theta)} \geq \frac{1}{n} \sum_{i=1}^n e^{\tilde{\lambda}(\theta_*)'g(W_i, \theta_*)} \geq 1 + O_p(K/n)$$

which establishes result (G.3).

\[ \square \]

**Lemma G.4.** Suppose the assumptions of Lemma G.2 are satisfied and let $\Lambda_n$ be as defined in Lemma G.2 and $\tau_i(\lambda, \theta) := \frac{e^{\lambda g_i(\theta)}}{E_n[e^{\lambda g_i(\theta)}]}$. Then, for every bounded constant $C > 1$:

$$P \left( \sup_{\lambda \in \Lambda_n} \tau_i(\lambda, \theta) > C \right) \to 0. \quad \text{(G.5)}$$

$$P \left( \inf_{\lambda \in \Lambda_n} \min_{1 \leq i \leq n} \tau_i(\lambda, \theta) \leq 1/C \right) \to 0. \quad \text{(G.6)}$$

**Proof.** First, remark that $\sup_{\lambda \in \Lambda_n} \tau_i(\lambda, \theta) \leq \sup_{\lambda \in \Lambda_n} \frac{e^{\lambda g_i(\theta)}}{E_n[e^{\lambda g_i(\theta)}]}$. Then, we can upper bound the numerator as

$$\sup_{\lambda \in \Lambda_n} e^{\lambda'g_1(\theta)} \leq e^{\max_{1 \leq i \leq n, \lambda \in \Lambda_n, \theta \in \Theta} |\lambda'g(W_i, \theta)|},$$

and lower bound the denominator as

$$E_n[e^{\min_{\lambda \in \Lambda_n} \lambda'g_1(\theta)}] \geq \exp \{- \max_{1 \leq i \leq n, \lambda \in \Lambda_n, \theta \in \Theta} |\lambda'g(W_i, \theta)|\}$$

so that $\sup_{\lambda \in \Lambda_n} \tau_i(\lambda, \theta) \leq e^{2 \max_{1 \leq i \leq n, \lambda \in \Lambda_n, \theta \in \Theta} |\lambda'g(W_i, \theta)|}$. By Lemma G.2, these two quantities converge to zero in probability. Therefore, for every $1 < C < \infty$ there exists a $N$ such that for every $n \geq N$, $P(\sup_{\lambda \in \Lambda_n} \tau_i(\lambda, \theta) > C) \to 0$.

Let us now consider the second result:

$$\inf_{\lambda \in \Lambda_n} \min_{1 \leq i \leq n} \tau_i(\lambda, \theta) \geq \frac{e^{-\max_{1 \leq i \leq n, \lambda \in \Lambda_n, \theta \in \Theta} |\lambda'g(W_i, \theta)|}}{E_n[e^{\max_{1 \leq i \leq n, \lambda \in \Lambda_n, \theta \in \Theta} |\lambda'g(W_i, \theta)|}]} = e^{-2 \max_{1 \leq i \leq n, \lambda \in \Lambda_n, \theta \in \Theta} |\lambda'g(W_i, \theta)|} > 1/C$$

with probability approaching 1 by Lemma G.2.
\[ \| \widehat{G}(\theta) \| = O_p(\sqrt{K}). \]

**Proof.** We have to control \( E\| \widehat{G}(\theta) \|. \) For every \( \theta \in U: \)

\[ E\| \widehat{G}(\theta) \| \leq E\| \rho_\theta(X) \otimes q^K(Z) \| \leq E (\| \rho_\theta(X) \| \| q^K(Z) \|) \leq \sqrt{E\| \rho_\theta(X,\theta) \|^2 \sqrt{E\| q^K(Z) \|^2}} \]

\[ \leq C \sqrt{\text{tr}E[q^K(Z)q^K(Z)']} = O_p(\sqrt{K}) \]

under Assumption 3.4 (b) and by Donald et al. (2003, Lemma A.2). The Markov’s inequality allows to obtain the result of the Lemma.

**Lemma G.6.** Let \( H \) be a compact subset of \( \subset \mathbb{R}^p. \) For \( h \in H \) define \( \theta_1 := \theta_* + h/\sqrt{n}. \) Suppose assumptions 3.2, 3.3 (d) and 3.4 (b) hold. Then, \( \sup_{h\in H} \| \lambda(\theta_1) \| = O_p(\sqrt{K/n}). \)

**Proof.** By expanding the first order condition for \( \lambda(\theta) \) around \( \lambda(\theta) = 0, \) there exists a \( \lambda_0 \) lying between \( \hat{\lambda}(\theta) \) and zero such that: \( \hat{g}(\theta) + \hat{\Omega}(\theta, \hat{\lambda}_0) \hat{\lambda}(\theta) = 0 \) which gives \( \hat{\lambda}(\theta) = -\hat{\Omega}(\theta, \hat{\lambda}_0)^{-1}\hat{g}(\theta). \) By using the expansion \( \hat{g}(\theta_1) = \hat{g}(\theta_*) + \hat{\Omega}(\theta)h/\sqrt{n} \) for \( \theta \) lying between \( \theta_1 \) and \( \theta_* \) we get:

\[ \| \hat{\lambda}(\theta_1) \| \leq \| \hat{\Omega}(\theta_1, \hat{\lambda}(\theta_1))^{-1} \| \left( ||\hat{g}(\theta_*)|| + ||\hat{\Omega}(\theta)h||n^{-1/2} \right). \]

By continuity of \( \theta \mapsto \hat{\lambda}_0, \) due to the implicit function theorem, and continuity of \( \theta \mapsto g_i(\theta), \) due to Assumption 3.3 (d), \( \tau_i(\lambda, \theta_1) \to 0 \) and so \( \hat{\Omega}(\theta_1, \hat{\lambda}_0) \to \hat{\Omega}(\theta_1) \) which then implies by Donald et al. (2003, Lemma A.6) that \( \| \hat{\Omega}(\theta_1, \hat{\lambda}(\theta_1))^{-1} \| = \lambda_{\min}(\hat{\Omega}(\theta_1, \hat{\lambda}(\theta_1))^{-1} \leq C. \) Moreover, by Lemma G.5, there exists \( N > 0 \) such that for every \( n \geq N, \| \hat{G}(\theta) \| = O_p(\sqrt{K}) \) and by Donald et al. (2003, Lemma A.9) \( ||\hat{g}(\theta_*)|| = O_p(\sqrt{K/n}). \) We then conclude that \( \| \hat{\lambda}(\theta_1) \| = O_p(\sqrt{K/n}). \) Compactness of \( H \) and continuity of \( \theta \mapsto \lambda(\theta) \) allow to conclude.

**Lemma G.7.** Let Assumptions 3.2, 3.3, 3.5 and 3.6 be satisfied and assume \( K/n \to 0. \) Define \( \hat{\Omega}(\theta, \lambda) := E_n[\tau_i(\lambda, \theta)g_i(\theta)g_i(\theta)'] \) with \( \tau_i(\lambda, \theta) := \frac{e^{\lambda g_i(\theta)}}{E_n[e^{\lambda g_i(\theta)}]} \). Let \( H \) denote a compact subset of \( \mathbb{R}^p \) and \( \theta_1 := \theta_* + h/\sqrt{n} \) with \( h \in H. \) Then, for all \( t_1, t_2 \in (0, 1) \)

\[ \sup_{h\in H} \| \hat{\Omega}(\theta_1, t_1\hat{\lambda}(\theta_1)) - \hat{\Omega}(\theta_*, t_2\hat{\lambda}(\theta_*)) \| = O_p \left( \zeta(K)K/\sqrt{n} \right) \quad (G.7) \]
and
\[ \sup_{h \in H} \| \tilde{g}(\theta_1) - \tilde{g}(\theta_*) \| = O_p(\sqrt{K/n}). \] (G.8)

**Proof.** Define \( \rho_i(\theta) := \rho(x_i, \theta) \), and \( \tilde{G}(\theta, \lambda) := E_n[\tau_i(\lambda, \theta)G(W_i, \theta)] \). The quantity that we have to control is:

\[
\sup_{h \in H} \| \tilde{\Omega}(\theta_1, t_1 \tilde{\lambda}(\theta_1)) - \tilde{\Omega}(\theta_*, t_2 \tilde{\lambda}(\theta_*)) \| =
\sup_{h \in H} \| E_n[\tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1)g_i(\theta_1)g_i(\theta_1)'] - E_n[\tau_i(t_2 \tilde{\lambda}(\theta_*), \theta_*)g_i(\theta_*)g_i(\theta_*)'] \|
\leq \sup_{h \in H} \| E_n[\left( \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1) - \tau_i(t_2 \tilde{\lambda}(\theta_*), \theta_*) \right) g_i(\theta_1)g_i(\theta_1)'] \|
+ \sup_{h \in H} \| E_n[\tau_i(t_2 \tilde{\lambda}(\theta_*), \theta_*) (g_i(\theta_1)g_i(\theta_1') - g_i(\theta_*)g_i(\theta_*)') ] \| =: A_1 + A_2 \quad (G.9)
\]

We start by considering term \( A_1 \) in \( (G.9) \). By a MVT expansion there exists \( \tilde{\lambda} \) between \( t_1 \tilde{\lambda}(\theta_1) \) and \( t_2 \tilde{\lambda}(\theta_*) \) such that

\[ \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_*) - \tau_i(t_2 \tilde{\lambda}(\theta_*), \theta_*) = \tau_i(\tilde{\lambda}, \theta_*)g_{i,n}(\theta_*)'(t_1 \tilde{\lambda}(\theta_1) - t_2 \tilde{\lambda}(\theta_*)) \] (G.10)

where \( g_{i,n}(\theta_*) := g_i(\theta_1') - E_n[\tau_i(\tilde{\lambda}, \theta_*)g_j(\theta_*)] \). Then, by Lemma G.6, by compactness of \( H \) and by using the condition \( K \zeta(K)^2/n^{1-2/\gamma} \to 0 \) (which holds under Assumption 3.6) it follows that \( \sup_{h \in H} \tilde{\lambda}(\theta_1) \in \Lambda_n \), where \( \Lambda_n \) is as defined in Lemma G.2 and so, \( \tilde{\lambda} \in \Lambda_n \). Hence,

\[
\| E_n[(\tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1) - \tau_i(t_2 \tilde{\lambda}(\theta_*), \theta_*) ) g_i(\theta_1)g_i(\theta_1)'] \|
\leq \| E_n[(\tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1) - \tau_i(t_2 \tilde{\lambda}(\theta_1), \theta_*) ) g_i(\theta_1)g_i(\theta_1)'] \|
+ \| E_n[(\tau_i(\tilde{\lambda}, \theta_*)g_{i,n}(\theta_*)'(t_1 \tilde{\lambda}(\theta_1) - t_2 \tilde{\lambda}(\theta_*)) ) g_i(\theta_1)g_i(\theta_1)'] \| =: A_{11}(h) + A_{12}(h) \quad (G.11)
\]

where we have used first \( \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1) - \tau_i(t_2 \tilde{\lambda}(\theta_*), \theta_*) = \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1) - \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_*) + \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_*) - \tau_i(t_2 \tilde{\lambda}(\theta_*), \theta_*) \) and then \( (G.10) \). We start by analyzing the second term in \( (G.11) \). By CS and the triangle inequality, we have:

\[
\sup_{h \in H} A_{12}(h) \leq \sup_{h \in H} E_n[| \tau_i(\tilde{\lambda}, \theta_*)g_{i,n}(\theta_*)'(t_1 \tilde{\lambda}(\theta_1) - t_2 \tilde{\lambda}(\theta_*))| | g_i(\theta_1)g_i(\theta_1') |]^{1/2}
\leq \left( \sup_{h \in H} \max_{1 \leq i \leq n} \tau_i(\tilde{\lambda}, \theta_*) \right)^{1/2} \sup_{h \in H} \sqrt{t_1 \tilde{\lambda}(\theta_1) - t_2 \tilde{\lambda}(\theta_*)/E_n[\tau_i(\tilde{\lambda}, \theta_*g_i(\theta_*)g_i(\theta_*)'(t_1 \tilde{\lambda}(\theta_1) - t_2 \tilde{\lambda}(\theta_*))]
\times \sqrt{E_n[| g_i(\theta_1)g_i(\theta_1') |^2]}
\leq \sup_{h \in H} \max_{1 \leq i \leq n} \tau_i(\tilde{\lambda}, \theta_*)^{1/2} \sup_{h \in H} \| t_1 \tilde{\lambda}(\theta_1) - t_2 \tilde{\lambda}(\theta_*) \| \sqrt{E_n[| g_i(\theta_1)g_i(\theta_1') |^2]}.
\]
By using Lemma G.4 and continuity of $\tilde{\lambda}$ in $h$ (by the implicit function theorem) we get:

$$\sup_{h \in H} \max_{1 \leq i \leq n} \tau_i(\tilde{\lambda}, \theta_*) < C, \quad \text{for } 1 < C < \infty; \quad (G.12)$$

by Lemmas G.3 and G.6 it holds: $\sup_{h \in H} \left\| (t_1 \tilde{\lambda}(\theta_1) - t_2 \tilde{\lambda}(\theta_2)) \right\| = O_p(\sqrt{K/n})$ and by Donald et al. (2003, Lemma A.6): $\lambda_{max}(\tilde{\Omega}(\theta_*)) = O_p(1)$. Moreover, similarly as in the proof of Donald et al. (2003, Lemma A.16):

$$\sup_{h \in H} E[\|g_i(\theta_1)g_i(\theta_1)\|^2] \leq E \left[ E_{\theta \in U} \|\rho_i(\theta)\|\|Z_i\|q^K(Z_i)\|^2 \right] \leq C\zeta(K)^2K. \quad (G.13)$$

Therefore, by denoting $A_{12} := \{ \sup_{h \in H} A_{12}(h) \leq C_2\zeta(K)K/\sqrt{n} \}$, the previous results and M imply that $P(A_{12}^c) = o(1)$.

Let us now analyze term $A_{11}(h)$ in (G.11). Let $\mathcal{G}_i(\lambda, \theta) := G(W_i, \theta) - E_n[\tau_i(\lambda, \theta)G(W_j, \theta)]$ and consider a MVT expansion around $\theta = \theta_*:

$$\tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1) - \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_*) = \tau_i(t_1 \tilde{\lambda}(\theta_1), \tilde{\theta}) \frac{h'}{\sqrt{n}} \mathcal{G}_i(t_1 \tilde{\lambda}(\theta_1), \tilde{\theta}) t_1 \tilde{\lambda}(\theta_1),$$

where $\tilde{\theta}$ lies between $\theta_1$ and $\theta_*$. By the Cauchy-Schwartz inequality we get:

$$A_{11}(h) := \left\| E_n \left[ \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_1) - \tau_i(t_1 \tilde{\lambda}(\theta_1), \theta_*) \right] g_i(\theta_1)g_i(\theta_1)' \right\|$$

$$= \left\| E_n \left[ \tau_i(t_1 \tilde{\lambda}(\theta_1), \tilde{\theta}) \frac{h'}{\sqrt{n}} \mathcal{G}_i(t_1 \tilde{\lambda}(\theta_1), \tilde{\theta}) t_1 \tilde{\lambda}(\theta_1) \right] \right\|$$

$$\leq \left( E_n \left[ \tau_i(t_1 \tilde{\lambda}(\theta_1), \tilde{\theta}) \frac{h'}{\sqrt{n}} \mathcal{G}_i(\tilde{\theta}) \tilde{\lambda}(\theta_1) \right] \right)^{1/2} \left( E_n \left[ \tau_i(t_1 \tilde{\lambda}(\theta_1), \tilde{\theta}) \| g_i(\theta_1)g_i(\theta_1)\|^2 \right] \right)^{1/2}.$$

Moreover, Lemma G.4 implies (G.12) so that

$$\sup_{h \in H} A_{11}(h) \leq C \sup_{h \in H} \left( E_n \left[ \frac{h'}{\sqrt{n}} \mathcal{G}_i(\tilde{\theta}) \tilde{\lambda}(\theta_1) \right] \right)^{1/2} \sup_{h \in H} \left( E_n \left[ \| g_i(\theta_1)g_i(\theta_1)\|^2 \right] \right)^{1/2}$$

$$= O_p(\sqrt{K/n} \sqrt{K/n}) O_p(\zeta(K) \sqrt{K})$$

for which we have used (G.13) and

$$\sup_{h \in H} E_n \left[ \frac{h'}{\sqrt{n}} \mathcal{G}_i(\tilde{\theta}) \tilde{\lambda}(\theta_1) \right] \leq \sup_{h \in H} \lambda_{max}(\tilde{\lambda}(\theta_1)\tilde{\lambda}(\theta_1)' \frac{h'}{\sqrt{n}} E_n \left[ \mathcal{G}_i(\tilde{\theta}) \mathcal{G}_i(\tilde{\theta}) \right] \frac{h}{\sqrt{n}}$$

$$\leq \sup_{h \in H} \tilde{\lambda}(\theta_1)\| \frac{h'}{\sqrt{n}} E_n \left[ \rho_\theta(X_i, \tilde{\theta}) \rho_\theta(X_i, \tilde{\theta})q^K(Z_i) \right] \|^2 \frac{h}{\sqrt{n}}$$
\[
\sup_{h \in \mathcal{H}} \leq \left\| \hat{\lambda}(\theta) \right\|^2 \left( \frac{h}{\sqrt{n}} \right)^2 E_n \left[ \left| \rho_0(X_i, \bar{\theta}) \right|^2 \left| q^K(Z_i) \right|^2 \right] = O_p\left(\frac{K}{n}\right) = O_p\left(\frac{K}{n}\right)
\]
where the last line holds by Assumption 3.4 (b) and a similar strategy as in the proof of Donald et al. (2003, Lemma A.16), and since \( \lambda_{\text{max}}(\hat{\lambda}(\theta), \hat{\lambda}(\theta))' = \left\| \hat{\lambda}(\theta) \hat{\lambda}(\theta)' \right\| \leq tr(\hat{\lambda}(\theta) \hat{\lambda}(\theta)') = \left\| \hat{\lambda}(\theta) \right\|^2 = O_p(K/n) \) by Lemma G.6. Therefore, by denoting \( A_{11} := \{ \sup_{h \in \mathcal{H}} A_{11}(h) \leq C_1 \zeta(K) K \sqrt{K/n} \} \), the previous results and M imply that \( P(A_{11}^c) = o(1) \).

Next, we analyse term \( A_2 \) in (G.9):

\[
\left\| E_n[\tau_i(t_2\hat{\lambda}(\theta), \theta)] \left[ \rho_i(\theta_1) \rho_i(\theta_1)' - \rho_i(\theta_2) \rho_i(\theta_2)' \right] \otimes q^K(Z_i)q^K(Z_i)' \right\|^2 \\
\leq E_n[\tau_i(t_2\hat{\lambda}(\theta), \theta)] \left\| \rho_i(\theta_1) \rho_i(\theta_1)' - \rho_i(\theta_2) \rho_i(\theta_2)' \right\|^2 \left\| q^K(Z_i) \right\|^2 \\
\leq E_n[\tau_i(t_2\hat{\lambda}(\theta), \theta)] \left( \left\| \rho_i(\theta_1) - \rho_i(\theta_2) \right\|^2 + 2 \left\| \rho_i(\theta_1) - \rho_i(\theta_2) \right\| \left\| \rho_i(\theta_2) \right\| \right) \left\| q^K(Z_i) \right\|^2 \\
\leq C \left\| \theta_1 - \theta_2 \right\| E_n[\tau_i(t_2\hat{\lambda}(\theta), \theta)] M_i(\theta_2) \left\| q^K(Z_i) \right\|^2 = O_p\left(\left\| \theta_1 - \theta_2 \right\| K \right) = O_p(hK/\sqrt{n})
\]
under Assumption 3.5 (b) by noting that \( M_i(\theta_2) := \delta(X_i) + 2\delta(X_i) \left\| \rho_i(\theta_1) \right\| \) has \( E[M_i(\theta_2)|Z_i] \) bounded by Cauchy-Schwartz (where \( \delta(X) \) is defined in Assumption 3.5) so that \( E[M_i(\theta_2)\left\| q^K(Z_i) \right\|^2] = E[E[M_i(\theta_2)|Z_i]\left\| q^K(Z_i) \right\|^2] \leq C E[\left\| q^K(Z_i) \right\|^2] \leq CK \).

Finally, by using the upper bounds in (G.11) and (G.14) we get:

\[
P(\sup_{h \in \mathcal{H}} \left\| \hat{\Omega}(\theta_1, t_2\hat{\lambda}(\theta)_i) - \hat{\Omega}(\theta_2, t_2\hat{\lambda}(\theta)_i) \right\| \geq \epsilon_n) \leq P(A_1 + A_2 \geq \epsilon_n) \\
\leq P(\sup_{h \in \mathcal{H}}(A_{11}(h) + A_{12}(h)) + A_2 \geq \epsilon_n) A_{11} \cap A_{12} + P(A_{11}^c) + P(A_{12}^c) \\
\leq P\left( A_2 \geq \epsilon_n - \left( C_1 \sqrt{\frac{K}{n}} + C_2 \right) \zeta(K) K/\sqrt{n} \right) + o(1)
\]
which converges to 0 for \( \epsilon = C_3 \zeta(K) K/\sqrt{n} \) with \( C_3 > \left( C_1 \sqrt{\frac{K}{n}} + C_2 \right) \) since \( \sqrt{K/n} \to 0 \), because of (G.38) and because \( P(A_{11}^c) = o(1) \) and \( P(A_{12}^c) = o(1) \).

Next, we show the second result of the lemma. Under Assumption 3.5 (b),

\[
\left\| \tilde{g}(\theta) - g(\theta) \right\| = \left\| E_n[\rho_i(\theta_1) - \rho_i(\theta_2)] \right\| \leq E_n[\left\| \rho_i(\theta_1) - \rho_i(\theta_2) \right\| \left\| q^K(Z_i) \right\|] \\
\leq \left\| \theta_1 - \theta_2 \right\| E_n[\delta(X_i) \left\| q^K(Z_i) \right\|] = O_p\left( \left\| \theta_1 - \theta_2 \right\| \sqrt{E[\delta(X_i)^2|Z_i]q^K(Z_i)^2} \right) \\
= O_p\left( \left\| \theta_1 - \theta_2 \right\| \sqrt{K} \right).
\]
Therefore, because $\mathcal{H}$ is compact we get:

$$\sup_{h \in \mathcal{H}} \| \hat{g}(\theta_1) - \hat{g}(\theta_*) \| = O_p(\sqrt{K/n}).$$

G.2 Technical Lemmas for the misspecified case

Lemma G.8. Let Assumptions 3.2 and 3.10 be satisfied. Then,

$$\| \hat{\lambda}(\theta_0) - \lambda_0(\theta_0) \| = O_p(\sqrt{K/n})$$

(G.15)

Proof. By a MVT expansion of the first order condition for $\hat{\lambda}(\theta_0)$ there exists $\tau \in [0, 1]$ such that $\hat{\lambda}_\tau := \tau(\hat{\lambda}(\theta_0) - \lambda_0(\theta_0)) + \lambda_0(\theta_0)$ satisfies $E_n[e^{\hat{\lambda}_\tau'(\theta_0)g_i(\theta_0)}] = 0 = E_n[e^{\lambda_0(\theta_0)'g_i(\theta_0)}] + \Omega(\theta_0, \hat{\lambda}_\tau(\lambda_0(\theta_0) - \lambda_0(\theta_0))$ which implies:

$$\begin{aligned}
(\hat{\lambda}(\theta_0) - \lambda_0(\theta_0)) &= -\hat{\Omega}(\theta_0, \hat{\lambda}_\tau)^{-1}E_n[e^{\lambda_0(\theta_0)'g_i(\theta_0)}g_i(\theta_0)]. \\
\end{aligned}
$$

(G.16)

By CS and Lemma G.9, it holds

$$\| \hat{\lambda}(\theta_0) - \lambda_0(\theta_0) \| \leq C \left\| E_n[e^{\lambda_0(\theta_0)'g_i(\theta_0)}g_i(\theta_0)] \right\| = O_p(\sqrt{K/n})$$

by using Assumption 3.10 (f).

Lemma G.9. Let Assumptions 3.2 and 3.10 be satisfied. Let $\mathcal{H}$ denote a compact subset of $\mathbb{R}^p$ and $\tilde{\theta}_2 := \theta_0 + th/\sqrt{n}$ with $h \in \mathcal{H}$ and $t \in [0, 1]$. Then, $\lambda_{\min}(\Omega_0) \geq C^{-1}$ and if $\zeta(K)K^{1/2} = o(1)$ then with probability approaching 1, $\lambda_{\min}(\hat{\Omega}(\tilde{\theta}_2, \hat{\lambda})) \geq C^{-1}$ uniformly in $h \in \mathcal{H}$.

Proof. Consider the matrix $\Omega_0$ and note that by Assumption 3.10 (g):

$$E[e^{\lambda_0(\theta_0)'g(W, \theta_0)p_i(\theta_0)p_i(\theta_0)'|Z}] \geq C^{-1}I_d.$$ Hence,

$$\Omega_0 \geq C^{-1}E[I_2 \otimes q^K(Z)q^K(Z)'] = C^{-1}I_{dK}$$

and $\lambda_{\min}(\Omega_0) \geq C^{-1}$. Also, if $\zeta(K)K^{1/2} = o(1)$ then $\sup_{h \in \mathcal{H}} \| \hat{\Omega}(\tilde{\theta}_2, \hat{\lambda}) - \Omega_0 \| \overset{P}{\rightarrow} 0$ by Lemma G.13 and the Markov inequality. Then, by $\| \hat{\lambda}(A) - \lambda(B) \| \leq \| A - B \|$, where $\lambda(A)$ denotes the minimum or maximum eigenvalue, it follows that

$$\sup_{h \in \mathcal{H}} \| \lambda_{\min}(\hat{\Omega}(\tilde{\theta}_2, \hat{\lambda})) - \lambda_{\min}(\Omega_0) \| \overset{P}{\rightarrow} 0.$$
Lemma G.10. Let Assumptions 3.2 and 3.10 be satisfied. Then,

\[ \| \hat{\lambda}(\theta_o)g(\theta_o) - \lambda_0(\theta_o)g(W, \theta_o) \| = O_p(\sqrt{K/n}) \]  

(G.17)

and

\[ \log \frac{1}{n} \sum_{i=1}^{n} \exp \{ \hat{\lambda}(\theta_o)g_i(\theta_o) \} = O_p(K/n). \]

Proof. We start by proving the first result. By the triangular inequality and CS we get:

\[ \| \hat{\lambda}(\theta_o)g(\theta_o) - \lambda_0(\theta_o)\| g(W, \theta_o)\| \| + \| \lambda_0(\theta_o)\| \| g(\theta_o) - E_\theta g(W, \theta_o)\| \| = O_p(\sqrt{K/n}) \]

where we have used Lemmas G.8 and G.11 and Assumption 3.10 (b). To get the second result we use the inequality \( \log(a) \leq a - 1 \) for every \( a > 0 \):

\[ \log \frac{1}{n} \sum_{i=1}^{n} \exp \{ \hat{\lambda}(\theta_o)g_i(\theta_o) \} \leq \frac{1}{n} \sum_{i=1}^{n} \exp \{ \hat{\lambda}(\theta_o)g_i(\theta_o) \} - E^P[\exp \{ \lambda_0(\theta_o)g(W, \theta_o) \}] \]

\[ \leq \frac{E_n[\exp \{ \hat{\lambda}(\theta_o)g_i(\theta_o) \} - E_n[\exp \{ \lambda_0(\theta_o)g_i(\theta_o) \}] + E_n[\exp \{ \lambda_0(\theta_o)g_i(\theta_o) \} - E^P[\exp \{ \lambda_0(\theta_o)g(W, \theta_o) \}] \] \]

By a MVT expansion of \( \lambda \mapsto \exp(\lambda g_i(\theta_o)) \) around \( \lambda_0(\theta_o) \), there exists a \( t \in [0, 1] \) such that \( \tilde{\lambda} = t(\hat{\lambda}(\theta_o) - \lambda_0(\theta_o)) + \lambda_0(\theta_o) \) satisfies 

\[ E_n[\exp \{ \tilde{\lambda}(\theta_o)g_i(\theta_o) \} - E_n[\exp \{ \lambda_0(\theta_o)g_i(\theta_o) \}] = E_n[\exp \{ \tilde{\lambda}(\theta_o)g_i(\theta_o) \} - \lambda_0(\theta_o))] \]

and so

\[ \| E_n[\exp \{ \tilde{\lambda}(\theta_o)g_i(\theta_o) \} - E_n[\exp \{ \lambda_0(\theta_o)g_i(\theta_o) \}] \| \leq e^{\max_{1 \leq i \leq n} \| \tilde{\lambda}(\theta_o) - \lambda_0(\theta_o) \| \| g_i(\theta_o) \|} E_n[\exp \{ \lambda_0(\theta_o)g_i(\theta_o) \} \| \lambda_0(\theta_o) \|] \| \tilde{\lambda}(\theta_o) - \lambda_0(\theta_o) \| = O_p(K/n) \]

by using Lemmas G.2, G.8 and Assumption 3.10 (f). Finally, by M

\[ E_n[\exp \{ \lambda_0(\theta_o)g_i(\theta_o) \} - E^P[\exp \{ \lambda_0(\theta_o)g(W, \theta_o) \}] = O_p(1/\sqrt{n}). \]

Lemma G.11. If Assumptions 3.2 and 3.10 (d) are satisfied, then \( \| \tilde{g}(\theta_o) - E_g(W, \theta_o) \| = O_p(\sqrt{K/n}). \)

Proof. By the Markov’s inequality, for every \( \varepsilon > 0 \)
\[ P \left( \| \tilde{g}(\theta_0) - \mathbf{E}[g(W_i, \theta_0)] \| > \varepsilon \right) \leq \frac{1}{\varepsilon^2} \mathbf{E} \| \tilde{g}(\theta_0) - \mathbf{E}[g(W_i, \theta_0)] \|^2 = \frac{1}{\varepsilon^2} \text{tr}(\text{Var}[\tilde{g}(\theta_0)]) \]

\[ = \frac{1}{n\varepsilon^2} \text{tr}(\text{Var}[g(W_i, \theta_0)]) \leq \frac{1}{n\varepsilon^2} \text{tr}(\mathbf{E} \left[ (\rho(X_i, \theta_0) \otimes q^K(Z_i))(\rho(X_i, \theta_0) \otimes q^K(Z_i))' \right]) \]

\[ = \frac{1}{n\varepsilon^2} \text{tr}(\mathbf{E} \left[ (\rho(X_i, \theta_0) \rho(X_i, \theta_0)' \otimes (q^K(Z_i)q^K(Z_i)') \right] = \frac{1}{n\varepsilon^2} \mathbf{E} \left[ \| \rho(X_i, \theta_0) \|^2 |Z_i| \text{tr}(q^K(Z_i)q^K(Z_i)') \right] \]

\[ \leq \frac{C}{n\varepsilon^2} \text{tr}(I_K) = \frac{C}{n\varepsilon^2} \text{tr}(I_K) = \frac{CK}{\varepsilon^2 n}. \]

where we have used Donald et al. (2003, Lemma A.2) in the last inequality.

\[ \square \]

**Lemma G.12.** Let Assumptions 3.2, 3.6, 3.8 and 3.10 be satisfied. Let \( \mathcal{H} \) denote a compact subset of \( \mathbb{R}^p \) and \( \tilde{\theta}_2 := \theta_0 + \tau h/\sqrt{n} \) with \( h \in \mathcal{H} \) and \( \tau \in [0, 1] \). Then

\[ \sup_{h \in \mathcal{H}} \left| h' \left( \frac{d\tilde{\lambda}(\tilde{\theta}_2)'}{d\theta} \right) - \frac{d\lambda_0(\theta_0)'}{d\theta} \right| \mathbf{E}_n[|\tilde{g}_i(\tilde{\theta}_2)|] = O_p(\zeta(K)K^2 \sqrt{K/n}). \]  

**(G.19)**

**Proof.** First, remark that we can simplify the expression of \( \frac{d\tilde{\lambda}(\tilde{\theta}_2)'}{d\theta} \) (by simplifying the denominator in \( \tau_i(\tilde{\lambda}, \tilde{\theta}) \)):

\[ \frac{d\tilde{\lambda}(\tilde{\theta}_2)'}{d\theta} = -\mathbf{E}_n \left[ \tau(\tilde{\lambda}, \tilde{\theta}_2, W_i)G_i(\tilde{\theta}_2)'(I + \tilde{\lambda}(\tilde{\theta}_2)g_i(\tilde{\theta}_2)') \right] \tilde{\Omega}^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \]

\[ + \mathbf{E} \left[ \tau(\lambda_0, \theta_0, W_i)G_i(\theta_0)'(I + \lambda_0(\theta_0)g_i(\theta_0)') \right] \lambda_0^{-1} \]

\[ = -\mathbf{E}_n \left[ \left( \tau(\tilde{\lambda}, \tilde{\theta}_2, W_i) - \tau(\lambda_0, \theta_0, \tilde{\theta}_2, W_i) \right) G_i(\tilde{\theta}_2)'(I + \tilde{\lambda}(\tilde{\theta}_2)g_i(\tilde{\theta}_2)') \right] \tilde{\Omega}^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \]

\[ - \mathbf{E}_n \left[ \left( \tau(\lambda_0, \theta_0, \tilde{\theta}_2, W_i)G_i(\tilde{\theta}_2)'(I + \tilde{\lambda}(\tilde{\theta}_2)g_i(\tilde{\theta}_2)') \right) \right] \tilde{\Omega}^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \]

\[ + \mathbf{E} \left[ \tau(\lambda_0, \theta_0, W_i)G_i(\theta_0)'(I + \lambda_0(\theta_0)g_i(\theta_0)') \right] \lambda_0^{-1} \right) =: A_1 + A_2. \]  

**(G.20)**

Further, we define

\[ A_{11} := -\mathbf{E}_n \left[ \left( \tau(\tilde{\lambda}, \tilde{\theta}_2, W_i) - \tau(\lambda_0, \theta_0, \tilde{\theta}_2, W_i) \right) G_i(\tilde{\theta}_2)' \right] \tilde{\Omega}^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \]

\[ A_{12} := -\mathbf{E}_n \left[ \left( \tau(\tilde{\lambda}, \tilde{\theta}_2, W_i) - \tau(\lambda_0, \theta_0, \tilde{\theta}_2, W_i) \right) G_i(\tilde{\theta}_2)'g_i(\tilde{\theta}_2) \right] \tilde{\Omega}^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \]

so that \( A_1 = A_{11} + A_{12} \). In the following we use the argument that \( g_i(\tilde{\theta}_2) = g_i(\theta_0) + G_i(\theta_0)'h/\sqrt{n} + o_p(h/\sqrt{n}) \) to replace \( g_i(\tilde{\theta}_2) \) by \( g_i(\theta_0) \).

By a MVT expansion there exists a \( \tau \in [0, 1] \) such that \( \tilde{\lambda} := \tau \tilde{\lambda}(\tilde{\theta}_2) + (1 - \tau)\lambda_0(\theta_0) \) satisfies

\[ e^{\tilde{\lambda}(\tilde{\theta}_2)'}g_i(\tilde{\theta}_2) - e^{\lambda_0(\theta_0)'}g_i(\theta_0) = e^{\tilde{\lambda}(\tilde{\theta}_2)'}g_i(\tilde{\theta}_2)(\tilde{\lambda}(\tilde{\theta}_2) - \lambda_0(\theta_0)). \]  

**(G.21)**
Therefore, $\left| h' A_{12} \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)] \right|$ can be upper bounded as:

$$\sup_{h \in \mathcal{H}} \left| h' A_{12} \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)] \right| \leq \sup_{h \in \mathcal{H}} \left\| E_n \left[ e^{\Delta g_i (\tilde{\theta}_2)} g_i (\tilde{\theta}_2)' \left( \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right) G_i (\tilde{\theta}_2) \right] h \left\| \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)] \right\|$$

$$\leq \sup_{h \in \mathcal{H}} \max_{1 \leq i \leq n} e^{\tau(\lambda(\tilde{\theta}_2) - \lambda_0 (\theta_0)) g_i (\tilde{\theta}_2)} \left\| E_n \left[ e^{\lambda_0 (\theta_0) g_i (\tilde{\theta}_2)} \left\| G_i (\tilde{\theta}_2) \right\|^2 \right] \right\|$$

$$\times \left\| \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right\| \left\| \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \right\| \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)] = O_p(\sqrt{K/n})$$

where we have used: (i) Lemma G.2 (ii) the compactness of $\mathcal{H}$, (iii) the fact that every differentiable map is locally Lipschitz which allows to show that $\left\| \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right\| \leq C_n$, (iv) the inequality

$$\left\| \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right\| \leq \left\| \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right\| - \left\| \lambda_0 (\theta_0) \right\| = O_p(\sqrt{K/n})$$

by Lemma G.8 and (iii), (v) Assumption 3.10 (f), (vi) the fact that $\sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)] = O_p(\sqrt{K})$. To control term $A_{12}$ we use again (G.34) and the Cauchy-Schwartz inequality to get:

$$\left| h' A_{12} \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)] \right| :=$$

$$h' E_n \left[ \left( \tau(\tilde{\lambda}, \tilde{\theta}_2, W_i) - \tau(\lambda_0, \theta_0, W_i) \right) G_i (\tilde{\theta}_2)' \tilde{\lambda}(\tilde{\theta}_2) g_i (\tilde{\theta}_2)' \right] \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)]$$

$$= h' E_n \left[ e^{\Delta g_i (\tilde{\theta}_2)} g_i (\tilde{\theta}_2)' \left( \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right) G_i (\tilde{\theta}_2)' \tilde{\lambda}(\tilde{\theta}_2) g_i (\tilde{\theta}_2)' \right] \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)]$$

$$\leq \sup_{h \in \mathcal{H}} \max_{1 \leq i \leq n} e^{\tau(\lambda(\tilde{\theta}_2) - \lambda_0 (\theta_0)) g_i (\tilde{\theta}_2)} \times$$

$$E_n \left[ e^{\lambda_0 (\theta_0) g_i (\tilde{\theta}_2)} \left\| g_i (\tilde{\theta}_2) \right\|^2 \right] \left\| \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right\| \left\| \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \right\| \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)]$$

$$\leq O_p(\sqrt{K \sqrt{K/n}}) \sqrt{n} E_n \left[ e^{\lambda_0 (\theta_0) g_i (\tilde{\theta}_2)} \left\| g_i (\tilde{\theta}_2) \right\|^2 \right] \left\| \tilde{\lambda}(\tilde{\theta}_2) - \lambda_0 (\theta_0) \right\| \left\| \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) \right\| \sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)]$$

$$= O_p(K^2 \zeta(K) / \sqrt{n}) \quad (G.22)$$

where we have used Lemma G.8, the fact that $\sqrt{n} E_n [\bar{g}_i (\tilde{\theta}_2)] = O_p(\sqrt{K})$ and argument (vi) above to get the penultimate line, and we have used Assumption 3.10 (f) to get the last line. Next, we analyse the term $A_2$ in (G.20), which we decompose further as follows:

$$A_2 = -E_n \left[ \tau(\lambda_0 (\theta_0), \tilde{\theta}_2, W_i) [G_i (\tilde{\theta}_2) - G_i (\theta_0)'] (I + \tilde{\lambda}(\tilde{\theta}_2) g_i (\tilde{\theta}_2)') \right] \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda})$$

$$- E_n \left[ (\tau(\lambda_0 (\theta_0), \tilde{\theta}_2, W_i) - \tau(\lambda_0 (\theta_0, W_i)) G_i (\theta_0) (I + \tilde{\lambda}(\tilde{\theta}_2) g_i (\tilde{\theta}_2)') \right] \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda})$$

$$- E_n \left[ (\tau(\lambda_0 (\theta_0, W_i) G_i (\theta_0) (I + \tilde{\lambda}(\tilde{\theta}_2) g_i (\tilde{\theta}_2)') \right] \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) - \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda})$$

$$- E_n \left[ (\tau(\lambda_0 (\theta_0, W_i) G_i (\theta_0) (I + \tilde{\lambda}(\tilde{\theta}_2) g_i (\tilde{\theta}_2)') \right] \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda}) - \Omega^{-1}(\tilde{\theta}_2, \tilde{\lambda})$$
Next, we analyze term $A_{21}$. By the Cauchy-Schwartz inequality applied twice we get

$$\sup_{h \in H} |h' A_{21} \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$= \sup_{h \in H} |h' E_n\left[\tau(\lambda_0(\theta_0), \tilde{\theta}_2, W_i)(G_i(\tilde{\theta}_2) - G_i(\theta_0))'\right] \Omega^{-1}(\tilde{\theta}_2, \lambda) \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$\leq \sup_{h \in H} \|h\| \left( E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta})} ||G_i(\tilde{\theta}_2) - G_i(\theta_0)|| + ||\lambda(\tilde{\theta}_2)|| \sqrt{n} E_n[g_i(\tilde{\theta}_2)] \right] \right)^{1/2} (\tilde{\theta}_2, \lambda) \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

Because every differentiable function is locally Lipschitz then, by Assumption 3.10 (d), there exists a $\theta_0$ between $\tilde{\theta}_2$ and $\theta$ such that $||G_i(\tilde{\theta}_2) - G_i(\theta_0)|| = ||\rho_{\theta}(X_i, \tilde{\theta}_2) - \rho_{\theta}(X_i, \theta_0)|| \leq q^K(Z_i) = n^{-1/2} (h' \rho_{\theta}(X_i, \tilde{\theta}_0))_{j=1}^d ||q^K(Z_i)||$. Then, by Assumption 3.2

$$\sup_{h \in H} \sqrt{n} E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta})} ||G_i(\tilde{\theta}_2) - G_i(\theta_0)||^2 \right] =$$

$$O_p(n^{-1/2}) \sqrt{E[E\left[\sup_{h \in H} e^{\lambda_0(\theta_0)'g_i(\tilde{\theta})} \left(\rho_{\theta}(X_i, \tilde{\theta}_0))_{j=1}^d ||q^K(Z_i)|| \right) \right]} = O_p(\sqrt{K/n}).$$

By this, the Jensen’s inequality, Assumption 3.10 (f) and the fact that $\sqrt{n} E_n[g_i(\tilde{\theta})] = O_p(\sqrt{K})$ we get:

$$\sup_{h \in H} |h' A_{21} \sqrt{n} E_n[g_i(\tilde{\theta})]| = O_p(K \sqrt{K/n}).$$

(G.26)

Next, we analyze term $A_{22}$. By a MVT expansion around $\theta_0$ there exists a $\tilde{\theta}_0$ between $\tilde{\theta}_2$ and $\theta_0$ such that $\tau(\lambda_0(\theta_0), \tilde{\theta}_2, W_i) - \tau(\lambda_0, \theta_0, W_i) = e^{\lambda_0(\theta_0)'g_i(\tilde{\theta_0})} \lambda_0(\theta_0)'G_i(\tilde{\theta}_0)\tau h/\sqrt{n}$ for $\tau \in [0, 1]$. Therefore,

$$\sup_{h \in H} |h' A_{22} \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$= \sup_{h \in H} |h' E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta_0})} \lambda_0(\theta_0)'G_i(\tilde{\theta}_0)\tau h/\sqrt{n}G_i(\theta_0)'(I + \lambda(\tilde{\theta}_2)g_i(\tilde{\theta}_2))' \right] \Omega^{-1}(\tilde{\theta}_2, \lambda) \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$\leq \sup_{h \in H} \sqrt{E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta_0})} \lambda_0(\theta_0)'G_i(\tilde{\theta}_0)\tau h/\sqrt{n}G_i(\theta_0)'(I + \lambda(\tilde{\theta}_2)g_i(\tilde{\theta}_2))' \right] \Omega^{-1}(\tilde{\theta}_2, \lambda) \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$+ \sup_{h \in H} \sqrt{E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta_0})} \lambda_0(\theta_0)'G_i(\tilde{\theta}_0)\tau h/\sqrt{n}G_i(\theta_0)'(I + \lambda(\tilde{\theta}_2)g_i(\tilde{\theta}_2))' \right] \Omega^{-1}(\tilde{\theta}_2, \lambda) \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$\leq \sup_{h \in H} n^{-1/2} ||\lambda_0(\theta_0)|| ||h|| \sqrt{E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta_0})} ||G_i(\theta_0)||^2 \right] \Omega^{-1}(\tilde{\theta}_2, \lambda) \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$+ ||\lambda(\tilde{\theta}_2)|| \left( E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta_0})} ||G_i(\tilde{\theta}_0)||^4 \right] \right)^{1/4} (E_n\left[e^{\lambda_0(\theta_0)'g_i(\tilde{\theta_0})} ||G_i(\tilde{\theta}_0)||^4 \right]^{1/4}) \Omega^{-1}(\tilde{\theta}_2, \lambda) \sqrt{n} E_n[g_i(\tilde{\theta})]|$$

$$=: A_{21} + A_{22} + A_{23} + A_{24}. \quad (G.23)$$
Therefore, by plugging this result in (G.28) and by using Assumption 3.10 (f), the first result of Lemma G.13 and the result of Lemma G.9 we find the rate for $\|\tilde{\Omega}^{-1}(\tilde{\theta}_2, \tilde{\lambda}) - \Omega^{-1}_o\|

\[
\sup_{h \in H} \|h^\prime A_{23} \sqrt{n} E_n[\tilde{g}_1(\tilde{\theta}_2)]\| = O_p((\sqrt{K} + K)\sqrt{\zeta(K)K/\sqrt{n}}). 
\]

Finally, we analyse term $A_{24}$:

\[
\sup_{h \in H} |h^\prime A_{24} \sqrt{n} E_n[\tilde{g}_1(\tilde{\theta}_2)]| = O_p((\sqrt{K} + K)\sqrt{\zeta(K)K/\sqrt{n}}).
\]

where we have used Assumption 3.10 (f) and the fact that $\sqrt{n} E_n[\tilde{g}_1(\tilde{\theta}_2)] = O_p(\sqrt{K})$. By putting
Lemma G.13. Let Assumptions 3.2 and 3.10 be satisfied. Let \( \mathcal{H} \) denote a compact subset of \( \mathbb{R}^p \) and \( \tilde{\theta}_2 := \theta_0 + th/\sqrt{n} \) with \( h \in \mathcal{H} \) and \( t \in [0,1] \). Then,

\[
\sup_{h \in \mathcal{H}} \| \tilde{\Omega}(\tilde{\theta}_2, \lambda(\tilde{\theta}_2)) - \hat{\Omega}(\theta_0, \lambda(\theta_0)) \| = O_p(\sqrt{K/n})
\]

The quantity that we have to control is:

\[
\sup_{h \in \mathcal{H}} \| \tilde{\Omega}(\tilde{\theta}_2, \lambda(\tilde{\theta}_2)) - \hat{\Omega}(\theta_0, \lambda(\theta_0)) \| = O_p(\sqrt{K/n})
\]

Remark that by a MVT expansion there exists \( \tau \in [0,1] \) such that \( \tilde{\lambda} := \tau(\tilde{\lambda}(\tilde{\theta}_2) - \lambda(\theta_0)) + \lambda(\theta_0) \) satisfies:

\[
e^\tilde{\lambda}(\tilde{\theta}_2)'g_i(\tilde{\theta}_2) - e^{\lambda(\theta_0)'g_i(\theta_0)} = e^{\tilde{\lambda}'g_i(\theta_0)} (\tilde{\lambda}(\tilde{\theta}_2)' - \lambda(\theta_0)).
\]

We start by considering term \( A_1 \) in (G.33):

\[
\left\| \mathbb{E}_n \left( e^{\tilde{\lambda}(\tilde{\theta}_2)'g_i(\tilde{\theta}_2) - e^{\lambda(\theta_0)'g_i(\theta_0)} g_i(\tilde{\theta}_2)g_i(\tilde{\theta}_2)' } \right) \right\|
\]
\[ \left\| E_n \left[ (e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\tilde{\theta}_2)} - e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\theta_o)})g_i(\tilde{\theta}_2)g_i(\theta_o)' \right] \right\| \\
+ \left\| E_n \left[ e^{\lambda(\theta_o)'g_i(\theta_o)'}(\hat{\lambda}(\tilde{\theta}_2) - \lambda_o(\theta_o))g_i(\tilde{\theta}_2)g_i(\theta_o)' \right] \right\| = A_{11}(h) + A_{12}(h) \quad \text{(G.35)} \]

where we have used first \( e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\tilde{\theta}_2)} - e^{\lambda(\theta_o)'g_i(\theta_o)} = e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\theta_o)} + e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\theta_o)} - e^{\lambda(\theta_o)'g_i(\theta_o)} \) and then (G.34). We start by analyzing the second term in (G.35). By the triangle inequality and the Cauchy-Schwarz inequality, we have:

\[
\sup_{h \in H} \ A_{12}(h) \leq \sup_{h \in H} \max_{1 \leq i \leq n} \left| e^{\lambda(\theta_o)'g_i(\theta_o)'} \left( E_n[|e^{\lambda(\theta_o)'g_i(\theta_o)||g_i(\theta_o)|^2] \right)^{1/2} \left\| \hat{\lambda}(\tilde{\theta}_2) - \lambda_o(\theta_o) \right\| \right.
\times \left( E_n[|e^{\lambda(\theta_o)'g_i(\theta_o)||g_i(\tilde{\theta}_2)g_i(\theta_o)'|^2] \right)^{1/2}
\]
\[
= O_p(\sqrt{K})O_p(\sqrt{K/n})O_p(\varepsilon(\sqrt{K}))
\]

where we have used: (i) the continuity and differentiability of \( \theta \mapsto \hat{\lambda}(\theta) \) (by the Implicit Function Theorem), (ii) the compactness of \( H \), (iii) the fact that every differentiable map is locally Lipschitz which allows to show that there exists a \( N \) such that for every \( n > N: \left\| \hat{\lambda}(\tilde{\theta}_2) - \hat{\lambda}(\theta_o) \right\| \leq C \sqrt{n} \), (iv) the inequality

\[
\sup_{h \in H} \left\| \hat{\lambda}(\tilde{\theta}_2) - \lambda_o(\theta_o) \right\| \leq \left\| \hat{\lambda}(\tilde{\theta}_2) - \lambda(\theta_o) \right\| + \left\| \lambda(\theta_o) - \lambda_o(\theta_o) \right\| = O_p(\sqrt{K/n}) \quad \text{(G.36)}
\]

which is valid by Lemma G.8 and (i)-(iii), (v) Lemma G.2 which is valid since \( \hat{\lambda}(\tilde{\theta}_2) - \lambda_o(\theta_o) \in A \) with \( \delta_n = \sqrt{K/n} \) by (i)-(iv), (vi) Assumption 3.10 (f) which implies \( E_n[|e^{\lambda(\theta_o)'g_i(\theta_o)||g_i(\theta_o)|^2] = O_p(\varepsilon(\sqrt{K})) \) and \( \sup_{h \in H} E_n[|e^{\lambda(\theta_o)'g_i(\theta_o)||g_i(\tilde{\theta}_2)|^2] = O_p(\varepsilon(\sqrt{K})) \).

Therefore, by denoting \( A_{12} := \sup_{h \in H} A_{12}(h) \leq C_2 \varepsilon(\sqrt{K})K \sqrt{K/n} \), the previous results and the Markov’s inequality imply that \( P(A_{12}) = o(1) \).

Next, we analyse term \( A_{11}(h) \) in (G.35). Consider a MVT expansion around \( \theta_o \):

\[
e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\tilde{\theta}_2)} - e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\theta_o)} = e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\theta_o)'}\hat{\lambda}(\tilde{\theta}_2)'G_i(\hat{\theta})\frac{h}{\sqrt{n}},
\]

where \( \hat{\theta} := \tau(\tilde{\theta}_2 - \theta_o) + \theta_o \) for some \( \tau \in [0,1] \). By this and the triangle inequality and the Cauchy-Schwarz inequality we get:

\[
\sup_{h \in H} \ A_{11}(h) := \sup_{h \in H} \left\| E_n \left[ (e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\tilde{\theta}_2)} - e^{\hat{\lambda}(\tilde{\theta}_2)'g_i(\theta_o)})g_i(\tilde{\theta}_2)g_i(\theta_o)' \right] \right\|
\leq \sup_{h \in H} \max_{1 \leq i \leq n} \left| e^{\lambda(\theta_o)'g_i(\theta_o)'} \left\| \hat{\lambda}(\tilde{\theta}_2) \right\| \left\| h \right\| \frac{1}{\sqrt{n}} E_n \left[ e^{\lambda(\theta_o)'g_i(\theta_o)'} \right\| G_i(\hat{\theta}) \right\| \left\| g_i(\tilde{\theta}_2)g_i(\theta_o)' \right\| \right.
\]

28
\[ \leq O_p(n^{-1/2}) \sup_{h \in \mathcal{H}} \sqrt{\mathbb{E}_n[e^{\lambda_0(\theta_o)g_i(\theta)}\|G_i(\hat{\theta})\|^2]} \sqrt{\mathbb{E}_n[e^{\lambda_0(\theta_o)g_i(\theta)}\|g_i(\hat{\theta}_0)g_i(\hat{\theta}_0)^{'}\|^2]} \]

where we have used Lemma G.2 which is valid since \( \hat{\lambda}(\hat{\theta}_2) - \lambda_0(\theta_o) \in \Delta \) with \( \delta_n = \sqrt{K/n} \) as shown in (i)-(iv) above. Moreover, we have used the fact that \( \|\hat{\lambda}(\hat{\theta}_2)\| = O_p(1) \) by (G.36), continuity of \( \theta \mapsto \lambda_0(\theta) \) and compactness of \( \mathcal{H} \). Therefore, by Assumption 3.10 (f):

\[ \sup_{h \in \mathcal{H}} A_{11}(h) = O_p(K/\sqrt{n}\zeta(K)). \quad (G.37) \]

Therefore, by denoting \( A_{11} := \{\sup_{h \in \mathcal{H}} A_{11}(h) \leq C_1\zeta(K)/\sqrt{n}\} \), the previous results and the Markov’s inequality imply that \( P(A_{11}^{c}) = o(1) \).

Next, we analyse term \( A_2 \) in (G.33):

\[ A_2 = \sup_{h \in \mathcal{H}} \left\| \mathbb{E}_n[e^{\lambda_0(\theta_o)g_i(\theta_o)}(\rho_i(\hat{\theta}_2)\rho_i(\hat{\theta}_2)^{'} - \rho_i(\theta_o)\rho_i(\theta_o)^{'} \otimes [q^K(zi)q^K(zi)^{'}])\right\| \]
\[ \leq \sup_{h \in \mathcal{H}} \mathbb{E}_n[e^{\lambda_0(\theta_o)g_i(\theta_o)}\left\| \rho_i(\hat{\theta}_2)\rho_i(\hat{\theta}_2)^{'} - \rho_i(\theta_o)\rho_i(\theta_o)^{'} \right\| \|q^K(zi)\|^2] \]
\[ \leq \sup_{h \in \mathcal{H}} \|\hat{\theta}_2 - \theta_o\|^2 \mathbb{E}_n[e^{\lambda_0(\theta_o)g_i(\theta_o)}\delta(X_i)^2\|q^K(Z_i)\|^2] \]
\[ + \sup_{h \in \mathcal{H}} \|\hat{\theta}_2 - \theta_o\|^2 \mathbb{E}_n[e^{\lambda_0(\theta_o)g_i(\theta_o)}\delta(X_i)\rho_i(\theta_o)\|q^K(Z_i)\|^2] = O_p(K/\sqrt{n}) \quad (G.38) \]

under Assumption 3.10 (e).

Finally, by using the upper bounds in (G.35) and (G.38) we get:

\[ P(\sup_{h \in \mathcal{H}} \|\hat{\Omega}(\hat{\theta}_2, \hat{\lambda}(\hat{\theta}_2)) - \Omega(\theta_o, \lambda_0(\theta_o))\| \geq \epsilon_n) \leq P(A_1 + A_2 \geq \epsilon_n) \]
\[ \leq P(\{\sup_{h \in \mathcal{H}} (A_{11}(h) + A_{12}(h)) \geq \epsilon_n\} \cap A_{11} \cap A_{12}) + P(A_{11}^{c}) + P(A_{12}^{c}) \]
\[ \leq P \left( A_2 \geq \epsilon_n - \left( \frac{C_1}{\sqrt{K}} + C_2 \right) \zeta(K)K\sqrt{\frac{K}{n}} \right) + o(1) \]

which converges to 0 for \( \epsilon = C_3\zeta(K)K\sqrt{K/n} \) with \( C_3 > (\frac{C_1}{\sqrt{K}} + C_2) \) because of (G.38) and because \( P(A_{11}^{c}) = o(1) \) and \( P(A_{12}^{c}) = o(1) \).

Next, we show the second result of the lemma. Remark that \( \|\hat{\Omega}(\theta_o, \lambda_0(\theta_o)) - \Omega_o\| = O_p(\sqrt{\mathbb{E}[\|\hat{\Omega}(\theta_o, \lambda_0(\theta_o)) - \Omega_o\|^2]}) \) and

\[ \mathbb{E}[\|\hat{\Omega}(\theta_o, \lambda_0(\theta_o)) - \Omega_o\|^2] \leq \sum_{j=1}^{dK} \sum_{l=1}^{dK} \mathbb{E}_n \left| E_n \left( e^{\lambda_0(\theta_o)g_{ij}(\theta_o)g_{il}(\theta_o)} - E_n\left( e^{\lambda_0(\theta_o)g_{ij}(\theta_o)g_{il}(\theta_o)} \right) \right) \right|^2 \]

29
Lemma G.14. Let Assumptions 3.2 and 3.10 (a)-(g) be satisfied and let $\mathcal{H}$ denote a compact subset of $\mathbb{R}^p$. Then,

$$\sup_{h \in \mathcal{H}} h' \left( \frac{d\tilde{\lambda}(\theta)}{d\theta} - \frac{d\lambda_0(\theta)}{d\theta} \right) E[g_i(\theta_o)] = O_p(n^{-1/2}).$$

Proof. By the triangular inequality:

$$(\sqrt{n}h') \left( \frac{d\tilde{\lambda}(\theta)}{d\theta} - \frac{d\lambda_0(\theta)}{d\theta} \right) E[g_i(\theta_o)] = \left| \sqrt{n}h' \left( \frac{d\tilde{\lambda}(\theta)}{d\theta} - \frac{d\lambda_0(\theta)}{d\theta} \right) E[g_i(\theta_o)] \right| \
\leq \left| \sqrt{n}h' \left( \frac{d\tilde{\lambda}(\theta)}{d\theta} - \frac{d\lambda_0(\theta)}{d\theta} \right) E[g_i(\theta_o)] \right| + \left| E_n \left[ e^{\lambda_0(\theta)} g_i(\theta_o) \right] \Omega(\theta_o, \lambda_o)^{-1} E[g_i(\theta_o)] \right| \\+ \left| E_n \left[ e^{\lambda_0(\theta)} g_i(\theta_o) \right] \Omega(\theta_o, \lambda_o)^{-1} \frac{d\lambda_0(\theta)}{d\theta} E[g_i(\theta_o)] \right| \quad (G.39)$$

We start with the analysis of the first term. By the MVT expansion around $\lambda_0(\theta_o)$ there exists a $t \in [0, 1]$ such that $\lambda := t\tilde{\lambda}(\theta_o) + (1 - t)\lambda_0(\theta_o)$ satisfies

$$\sqrt{n}h' \left( \frac{d\tilde{\lambda}}{d\theta} - \frac{d\lambda_0}{d\theta} \right) E[g_i(\theta_o)] = -\sqrt{n}h' E_n \left[ e^{\lambda(\theta)} g_i(\theta_o) \right] \Omega(\theta_o, \lambda_o)^{-1} E[g_i(\theta_o)] \quad (G.39)$$

Finally, we show the last result of the Lemma. Under Assumption 3.10 (e),

$$\| \tilde{g}(\theta_2) - g(\theta_o) \| = \left\| E_n[(\rho_i(\tilde{\theta}_2) - \rho_i(\theta_o)) \otimes q^K(z_i)] \right\| \leq E_n[\|\rho_i(\tilde{\theta}_2) - \rho_i(\theta_o)\| |q^K(z_i)|] \\leq \|\tilde{\theta}_2 - \theta_o\| E_n[\|\delta(x_i)\| \|q^K(z_i)|\] = O_p \left( \|\tilde{\theta}_2 - \theta_o\| \sqrt{E(\|\delta(X_i)^2\|z_i)\|q^K(Z_i)|^2} \right) \\leq O_p(\|\tilde{\theta}_2 - \theta_o\| \sqrt{K}).$$

Therefore, because $\mathcal{H}$ is compact we get:

$$\sup_{h \in \mathcal{H}} \| \tilde{g}(\theta_2) - g(\theta_o) \| = O_p(\sqrt{K/n}).$$

□
By using (G.16) it follows that the last three terms are equal to

\[ h' \left( \mathbb{E}_n \left[ e^{\lambda_0(\theta_i)g_i(\theta_i)'(I + \tilde{\lambda}_0(\theta_i)g_i(\theta_i)')} \right] \right) \Omega(\theta_i, \tilde{\lambda}_i)^{-1} \mathbb{E}_n \left[ e^{\tilde{\lambda}_0(\theta_i)g_i(\theta_i)'(I + \tilde{\lambda}_0(\theta_i)g_i(\theta_i)')} \right] \]

and they are bounded in probability. We now control the second term in (G.39). This is upper bounded by

\[
\left| \sqrt{n}h' \left( \mathbb{E}_n \left[ e^{\lambda_0(\theta_i)g_i(\theta_i)'(I + \lambda_0(\theta_i)g_i(\theta_i)')} \right] \Omega(\theta_i, \lambda_i)^{-1} - \frac{d\lambda_0(\theta_i)'}{d\theta} \right) \mathbb{E}[g_i(\theta_i)] \right|
\leq \left| \sqrt{n}h' \mathbb{E}_n \left[ e^{\lambda_0(\theta_i)g_i(\theta_i)'(I + \lambda_0(\theta_i)g_i(\theta_i)')} \right] \Omega(\theta_i, \lambda_i)^{-1} - \Omega_0^{-1} \right) \mathbb{E}[g_i(\theta_i)] \right|
+ \left| h' \mathbb{E}_n \left[ e^{\lambda_0(\theta_i)g_i(\theta_i)'(I + \lambda_0(\theta_i)g_i(\theta_i)')} \right] \Omega_0^{-1} \mathbb{E}[g_i(\theta_i)] \right|
\]

which, by using Lemma G.13 is bounded in probability.

\[ \square \]

H Additional empirical example

In this section we illustrate an extra application of our techniques in the estimation of causal parameters: the average treatment effect (ATE) estimation under a conditional independence assumption.

Average treatment effect (ATE) estimation. A standard problem in causal inference with a binary treatment \( x_i \in \{0, 1\} \), for control and treated, respectively, and covariates \( z_i : d \times 1 \) assumes that the two potential outcomes \( y_{i0} \) and \( y_{i1} \) for \( n \) randomly chosen subjects satisfy the conditional independence assumption (Rosenbaum and Rubin, 1983)

\[ (y_{i0}, y_{i1}) \perp x_i \mid z_i. \]
If we let $\mathbb{E}^P(y_{i1}|z_i) - \mathbb{E}^P(y_{i0}|z_i)$ denote the ATE conditional on $z_i$ for the $i$th subject, then the goal of the analysis is to calculate the ATE given by

$$\text{ATE} = \frac{1}{n} \sum_{i=1}^{n} \left( \mathbb{E}^P(y_{i1}|z_i) - \mathbb{E}^P(y_{i0}|z_i) \right).$$

We show that the conditional moment technique developed above is ideally suited for calculating the posterior distribution of this quantity, under minimal assumptions. We just need to make assumptions about the conditional expectations $\mathbb{E}^P(y_{ij}|z_i)$ ($j = 0, 1$) without specifying (or restricting) the conditional distributions of $y_{ij}|z_i$ in any further way. For illustration, suppose that

$$\mathbb{E}^P(y_{ij}|z_i) = z_i'\beta_j, \ j = 0, 1.$$

Also suppose that there are $n_0$ control subjects, and that the data are organized such that the observations $i \leq n_0$ are the data on the controls, and the observations $i > n_0$ are the data on the treated. Then, the latter conditional expectations imply that estimation of $\beta_0$ can be based on the conditional moment conditions

$$\mathbb{E}^P((y_{i0} - z_i'\beta_0)|z_i) = 0 \ (i \leq n_0)$$

since $y_{i0}$ is observed for such subjects, and that independently, estimation of $\beta_1$ can be based on the conditional moment conditions

$$\mathbb{E}^P((y_{i1} - z_i'\beta_1)|z_i) = 0 \ (i > n_0)$$

since $y_{i1}$ is observed for these subjects. Now, suppose that our prior-posterior analysis is applied to these sets of moment conditions to produce the MCMC samples

$$\{\beta_0^g\}_{g=1}^{M} \sim \pi(\beta_0 | \{y_{i0}, z_i\}_{i=1}^{n_0}) \text{ and } \{\beta_1^g\}_{g=1}^{M} \sim \pi(\beta_1 | \{y_{i1}, z_i\}_{i>n_0}).$$

Then, the Bayes posterior sample of the ATE is given by the sequence of values

$$\text{ATE}^{(g)} = \frac{1}{n} \sum_{i=1}^{n} \left( z_i'\beta_1^{(g)} - z_i'\beta_0^{(g)} \right), \ g = 1, 2, \ldots, M.$$
conditional on \((z_1, z_2)\), suppose that \(x\) is generated as independent Bernoulli

\[ x \sim B(p) \]

where the propensity score, the probability \(p\) of being treated, is given by

\[ p = \Phi \left( 0.5(\sqrt{0.3}z_{1,i} + \sqrt{0.7}z_{2,i})^3(1 - \sqrt{0.3}z_{1,i} - \sqrt{0.7}z_{2,i}) \right) \]

and \(\Phi(\cdot)\) is the cdf of the standard normal. Finally, suppose that the potential outcomes for each individual in the sample are given by

\[
\begin{align*}
y_0 &= 10 + z_1 + 1.5z_2 + \varepsilon_0, \\
y_1 &= 10 + 1.5z_1 - z_2 + \varepsilon_1
\end{align*}
\]

where the conditional distribution of \(\varepsilon_j\) is skewed normal with conditional variance and conditional skewness depending on \(z = (z_1, z_2)\). In particular,

\[ \varepsilon_j \sim SN(m_j(z), s_j(z), w_j(z)) \] (H.2)

where

\[ s_0(z) = \exp\left(0.5\left(1 + .5z_1 + 0.1z_1^2 + .3z_2\right)\right), \quad w_0(z) = 1 + z_1^2 + .5z_2 \]

and

\[ s_1(z) = \exp\left(0.5\left(1 + z_1 + .2z_1^2 + .3z_2\right)\right), \quad w_1(z) = 1 + z_1^2 + z_2 \]

and \(m_j(z)\) is fixed based on these functions to ensure that \(E(\varepsilon_j|z) = 0\). The observed data is \(y = xy_1 + (1 - x)y_0\).

There are approximately 42 percent treated subjects that emerge from this design. Also note that, because of the extreme nonlinearity of the propensity score function, standard propensity score matching does not perform well with data generated from this design. In addition, any method that is based on direct modeling of the outcome distributions that is not robust to covariate dependent heteroskedasticity, or to covariate dependent skewness, would also not perform well.

Our results in Table 1, which are based on 5 knots for the \(n = 500\) case (implying 13 expanded moment conditions created from \(z_1, z_2, \) and \(z_1z_2\)) and 7 knots for the larger sample sizes (implying 19 expanded moment conditions), show that the ATE is well inferred in this problem.

References

<table>
<thead>
<tr>
<th></th>
<th>True</th>
<th>Mean</th>
<th>SD</th>
<th>Median</th>
<th>Lower</th>
<th>Upper</th>
<th>Ineff</th>
</tr>
</thead>
<tbody>
<tr>
<td>$n = 500$</td>
<td>0.19</td>
<td>0.12</td>
<td>0.12</td>
<td>0.12</td>
<td>-0.11</td>
<td>0.35</td>
<td>1.28</td>
</tr>
<tr>
<td>$n = 1000$</td>
<td>0.20</td>
<td>0.14</td>
<td>0.10</td>
<td>0.14</td>
<td>-0.06</td>
<td>0.33</td>
<td>1.15</td>
</tr>
<tr>
<td>$n = 2000$</td>
<td>0.23</td>
<td>0.25</td>
<td>0.08</td>
<td>0.25</td>
<td>0.10</td>
<td>0.40</td>
<td>1.10</td>
</tr>
</tbody>
</table>

Table 1: Posterior summary for ATE estimation with three data sets. True ATE for each sample size is indicated by True. The summaries are based on 10,000 MCMC draws beyond a burn-in of 1000. The M-H acceptance rate is around 90% in the estimation of the control and treated models.

