Online appendix to:

Bayesian Estimation and Comparison of Moment Condition Models

Siddhartha Chib* Minchul Shin† Anna Simoni‡

Olin Business School *University of Illinois* *CREST, CNRS*

This version: July 16, 2017

A Examples

Example 2 (Misspecified model and pseudo-true value)**.** *Let us consider the model* $y_i = \alpha + e_i$, $i = 1, \ldots, n$, *with* e_i *independently drawn from the skewed distribution* P *given in* (2.13)*. We consider the following two moment conditions* $\mathbf{E}^P[y_i - \alpha] = 0$ *and* $\mathbf{E}^P[(y_i - \alpha)^3] = 0$*. This situation is different from the one illustrated in Example 1 in the paper because there are no covariates and the augmented parameter* v *is (incorrectly) forced to be zero. In turn,* α *has to satisfy both the moment restrictions, which is impossible under* P*. Instead, for each* α *the ETEL function is defined by the probability measure* $Q^*(\alpha)$ *which is the closest to the true generating process* P *in terms of KL divergence among the probability measures that are consistent with the given moment restrictions for a given* α*. In Figure 2 (left panel), we present* $\mathbf{E}^P[\log(dQ^*(\alpha)/dP)]$ *which is equal to* $-K(P||Q^*(\alpha))$*. The value that maximizes this function is different from the true value* ($\alpha = 0$) *and it is peaked around* -0.056*. This value is the pseudo-true value. In the right panel of Figure 2, we present the BETEL posterior distribution of* α *for five different sample sizes. The BETEL posterior distribution shrinks*

*Olin Business School, Washington University in St. Louis, Campus Box 1133, 1 Brookings Dr. St. Louis, MO 63130, USA, e-mail: chib@wustl.edu

†Department of Economics, University of Illinois, 214 David Kinley Hall, 1407 W. Gregory, Urbana, IL 61801, e-mail: mincshin@illinois.edu

‡CREST - ENSAE - École Polytechnique, 5, avenue Henry Le Chatelier, 91120 Palaiseau France, e-mail: simoni.anna@gmail.com

and moves toward the pseudo-true value, in conformity with our theoretical result.

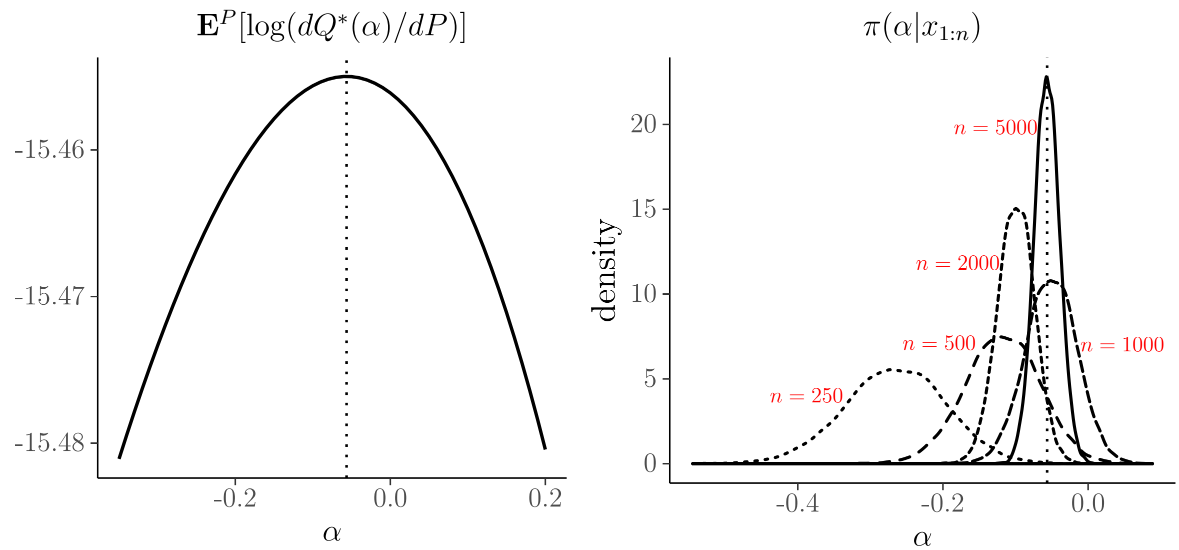

Figure 2: Posterior distributions in Example 2 under misspecification. Left panel presents the function $\alpha \mapsto \mathbf{E}^P[\log(dQ^*(\alpha)/dP)]$ where $Q^*(\alpha) := \text{arginf}_{Q\in\mathcal{P}_\psi} K(Q||P)$ with $\psi := (\alpha, 0)$. For each α, we approximate this function based on the dual representation in (2.16) – which is valid under Assumption 3 – using five million simulation draws from P. In the right panel, we present the BETEL posterior distribution of the location parameter α for $n = 250, 500, 1000, 2000, 5000$ where n is the number of observations. The prior distribution for α is student-t with mean 0 and dispersion 5. Vertical dashed lines are at the pseudo-true parameter value, approximately equal to -0.056. Posterior densities are based on 10,000 draws beyond a burn-in of 1000. The M-H acceptance rate is about 90% for each sample size.

Example 3 (Model selection when all models are correctly specified)**.** *We suppose that for every $i = 1, \ldots, n$, $y_i = \alpha + \beta z_i + e_i$, where $z_i \sim \mathcal{N}(0.5, 1)$ and $e_i \sim \mathcal{N}(0, 1)$ independently of z_i. Let $\theta := (\alpha, \beta)$, $e_i(\theta) := (y_i - \alpha - \beta z_i)$ and the true value of θ be $(0, 1)$. We compare the following models. Model 1: $\mathbf{E}^P[(e_i(\theta), e_i(\theta)z_i, e_i(\theta)^3, e_i(\theta)^2 - 1)'] = 0$, Model 2: $\mathbf{E}^P[(e_i(\theta), e_i(\theta)z_i, e_i(\theta)^2 - 1)'] = 0$ and Model 3: $\mathbf{E}^P[(e_i(\theta), e_i(\theta)^2 - 1)'] = 0$ which, reformulated in terms of an encompassing grand model, become respectively:*

$$\begin{aligned} &M_1 : \mathbf{E}^P[e_i(\theta)] = 0, \quad \mathbf{E}^P[e_i(\theta)z_i] = 0, \quad \mathbf{E}^P[e_i(\theta)^3] = 0, \quad \mathbf{E}^P[e_i(\theta)^2 - 1] = 0 \\ &M_2 : \mathbf{E}^P[e_i(\theta)] = 0, \quad \mathbf{E}^P[e_i(\theta)z_i] = 0, \quad \mathbf{E}^P[e_i(\theta)^3] = v_1, \quad \mathbf{E}^P[e_i(\theta)^2 - 1] = 0, \\ &M_3 : \mathbf{E}^P[e_i(\theta)] = 0, \quad \mathbf{E}^P[e_i(\theta)z_i] = v_2, \quad \mathbf{E}^P[e_i(\theta)^3] = v_1, \quad \mathbf{E}^P[e_i(\theta)^2 - 1] = 0. \end{aligned} \tag{A.1}$$

with $\psi^1 = \theta$, $\psi^2 = (\theta, v_1)$ and $\psi^3 = (\theta, v_1, v_2)$. Note that the last two moment restrictions (which concern the third and second moments) serve as extra information to infer the parameter θ, when they are active. Under the standard normal error distribution, all the three *models are correctly specified: M_1 has four active moment restrictions while M_2 and M_3 have three and two active moment restrictions, respectively.*

In Table 5, we report the percentage of times the marginal likelihood selects each of these models in 500 trials, for different sample sizes. Model M_1, the model with the larger number of valid restrictions, is selected 99% of times by sample size of $n = 500$. The results are virtually indistinguishable for the training sample prior (based on 50 prior samples). Under both priors the proportion of correct selection tends to one.

	Default prior			Training sample prior		
Model	M_1	M_2	M_3	M_1	M_2	M_3
$n = 250$	97.8	1.6	0.6	98.0	1.6	0.4
$n = 500$	99.0	0.8	0.2	99.0	0.8	0.2
$n = 1000$	99.2	0.6	0.2	99.2	0.6	0.0
$n = 2000$	99.2	0.8	0.0	99.2	0.8	0.0

Table 5: Model selection when all models are correctly specified. Frequency (%) of times each of the three models in Example 3 are selected by the marginal likelihood criterion in 500 trials, by sample size, for two different prior distributions.

Example (Count regression: variable selection (continued))**.** *Consider the case where the data are drawn under the Poisson assumption (this information is, of course, not used in the estimation). Specifically, suppose we generate n realizations of $\{y_i, x_i\}$ from the Poisson model*

$$\begin{aligned} y_i|\beta, x_i &\sim Poisson(\mu_i), \qquad i = 1, \ldots, n \\ \log(\mu_i) &= x_i'\beta. \end{aligned} \tag{A.2}$$

where $\beta = (\beta_1, \beta_2, \beta_3)'$ and $x_i = (x_{1,i}, x_{2,i}, x_{3,i})'$, with $\beta_1 = 1$, $\beta_2 = 1$, $\beta_3 = 0$. Thus, $x_{3,i}$ is a redundant regressor. Each explanatory variable $x_{j,i}$ is generated i.i.d. from normal distributions with mean .4 and standard deviation 1/3. Given these data, our goal is to evaluate the finite-sample performance of our marginal likelihood criterion in picking out the correct model. We conduct our MCMC analysis and compute the marginal likelihoods of the four models by the Chib (1995) method under the default student-t prior distribution on β given in (2.11). The results, in Table 6, give the percentage of times in 500 replications that the marginal likelihood criterion picks each model for three different sample sizes. As can be seen, the model with the largest number of overidentifying moment restrictions M_1 is selected by the marginal likelihood criterion with frequency close to one even when $n = 250$.

Model	M_1	M_2	M_3	M_4
$n = 250$	0.99	0.01	0.01	0.00
$n = 500$	0.99	0.00	0.01	0.00
$n = 1000$	0.99	0.00	0.00	0.00

Table 6: Frequency (%) of times each of the four models in (4.3) are selected by the marginal likelihood criterion in 500 trials. The DGP is the Poisson model given in (A.2).

B Assumptions

In this section we state the assumptions that are used to prove the theorems in the paper. For completeness we also report Assumptions 1 – 4 that were already stated in the paper. We start by stating the assumptions that are used in Theorem 2.1 and then the other assumptions relevant for misspecification. As a consequence the numbering is not in the order.

Assumption 1. *Model* (2.2) *is such that* $\psi_* \in \Psi$ *is the unique solution to* $\mathbf{E}^P[g^A(X, \psi)] = 0$.

Assumption 2. *(a)* π *is a continuous probability measure that admits a density with respect to the Lebesgue measure; (b)* π *is positive on a neighborhood of* ψ_*.

The following two assumptions relate to the smoothness of the function $g^A(x, \psi)$, its moments, and the parameter space.

Assumption 5. *(a)* X_i*,* $i = 1, \dots, n$ *are i.i.d. random variables that take values in* $(\mathcal{X}, \mathfrak{B}_{\mathcal{X}})$ *with probability distribution* P*, where* $\mathcal{X} \subseteq \mathbb{R}^{d_x}$*; (b) for every* $0 \leq d_v \leq d - p$*,* $\psi \in \Psi \subset \mathbb{R}^p \times \mathbb{R}^{d_v}$ *where* Θ *and* $\mathcal{V}$ *are compact and connected and* $\Psi := \Theta \times \mathcal{V}$*; (c)* $g(x, \theta)$ *is continuous at each* $\theta \in \Theta$ *with probability one; (d)* $\mathbf{E}^P[\sup_{\psi \in \Psi} \|g^A(X, \psi)\|^\alpha] < \infty$ *for some* $\alpha > 2$*; (e)* Δ *is nonsingular.*

Assumption 6. *(a)* $\psi_* \in int(\Psi)$*; (b)* $g^A(x, \psi)$ *is continuously differentiable in a neighborhood* $\mathfrak{U}$ *of* ψ_* *and* $\mathbf{E}^P[\sup_{\psi \in \mathfrak{U}} \|\partial g^A(X, \psi)/\partial \psi'\|_F] < \infty$*; (c)* $rank(\Gamma) = p$.

Assumptions 5 and 6 are the same as the assumptions of Newey and Smith (2004, Theorem 3.2) and Schennach (2007, Theorem 3).

We now consider misspecified models.

Assumption 3. *For a fixed* $\psi \in \Psi$*, there exists* $Q \in \mathcal{P}_\psi$ *such that* Q *is mutually absolutely continuous with respect to* P*, where* $\mathcal{P}_\psi$ *is defined in Definition 2.1.*

Assumption 4. *The prior distribution* π *is positive on a neighborhood of* $\psi_\circ$ *where* $\psi_\circ$ *is as defined in* (2.16).

In addition to these assumptions, to prove Theorem 2.2 we also use Assumptions 5 *(a)*-*(d)* and 6 *(b)* in the previous section. Finally, in order to guarantee $n^{-1/2}$-convergence of $\widehat{\lambda}$ towards $\lambda_\circ$ and $n^{-1/2}$-contraction of the posterior distribution of ψ around $\psi_\circ$, we introduce Assumptions 7 and 8. These assumptions require the pseudo-true values $\lambda_\circ$ and $\psi_\circ$ to be in the interior of a compact parameter space, and the function $g^A(x,\psi)$ to be sufficiently smooth and uniformly bounded as a function of ψ. These assumptions are not new in the literature and are also required by Schennach (2007, Theorem 10) (adapted to account for the augmented model).

Assumption 7. *(a) There exists a function $M(\cdot)$ such that $\mathbf{E}^P[M(X)] < \infty$ and $\|g^A(x,\psi)\| \leq M(x)$ for all $\psi \in \Psi$; (b) $\lambda_\circ(\psi) \in int(\Lambda(\psi))$ where $\Lambda(\psi)$ is a compact set and $\lambda_\circ$ is as defined in (2.16); (c) it holds $\mathbf{E}^P\left[\sup_{\psi\in\Psi,\lambda\in\Lambda(\psi)} e^{\{\lambda' g^A(X,\psi)\}}\right] < \infty$.*

Assumption 8. *Let $\psi_\circ$ be as defined in (2.16). (a) The pseudo-true value $\psi_\circ \in int(\Psi)$ is the unique maximizer of*

$$\lambda_\circ(\psi)'\mathbf{E}^P[g^A(X,\psi)] - \log \mathbf{E}^P[\exp\{\lambda_\circ(\psi)'g^A(X,\psi)\}],$$

where Ψ is compact; (b) $S_{jl}(x_i,\psi) := \partial^2 g^A(x_i,\psi)/\partial\psi_j\partial\psi_l$ is continuous in ψ for $\psi \in \mathcal{U}_\circ$, where $\mathcal{U}_\circ$ denotes a ball centred at $\psi_\circ$ with radius $n^{-1/2}$; (c) there exists $b(x_i)$ satisfying $\mathbf{E}^P\left[\sup_{\psi\in\mathcal{U}_\circ}\sup_{\lambda\in\Lambda(\psi)}\exp\{\kappa_1\lambda' g^A(X,\psi)\}b(X)^{\kappa_2}\right] < \infty$ for $\kappa_1 = 0,1,2$ and $\kappa_2 = 0,1,2,3,4$ such that $\|g^A(x_i,\psi)\| < b(x_i)$, $\|\partial g^A(x_i,\psi)/\partial\psi'\|_F \leq b(x_i)$ and $\|S_{jl}(x_i,\psi)\| \leq b(x_i)$ for $j,l = 1,\ldots,p$ for any $x_i \in (\mathcal{X},\mathfrak{B}_\mathcal{X})$ and for all $\psi \in \mathcal{U}_\circ$.

C Proofs for Sections 2.2 and 2.3

In this appendix we prove Theorems 2.1 and 2.2 and Lemma 2.1. It is useful to introduce some notation that will be used hereafter. The estimator $\widehat{\psi} := (\widehat{\theta},\widehat{v})$ denotes Schennach (2007)'ETEL estimator of ψ:

$$\widehat{\psi} := \arg\max_{\psi\in\Psi} \frac{1}{n}\sum_{i=1}^n \left[\widehat{\lambda}(\psi)'g^A(x_i,\psi) - \log\frac{1}{n}\sum_{j=1}^n \exp\{\widehat{\lambda}(\psi)'g^A(x_j,\psi)\}\right] \tag{C.1}$$

where $\widehat{\lambda}(\psi) := \arg\min_{\lambda\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}^n \left[\exp\{\lambda' g^A(x_i,\psi)\}\right]$. The log-likelihood ratio is:

$$l_{n,\psi}(x) - l_{n,\psi_\circ}(x) = \log\frac{e^{\widehat{\lambda}(\psi)'g^A(x,\psi)}}{\frac{1}{n}\sum_{j=1}^n\left[e^{\widehat{\lambda}(\psi)'g^A(x_j,\psi)}\right]} - \log\frac{e^{\widehat{\lambda}(\psi_\circ)'g^A(x,\psi_\circ)}}{\frac{1}{n}\sum_{j=1}^n\left[e^{\widehat{\lambda}(\psi_\circ)'g^A(x_j,\psi_\circ)}\right]}. \tag{C.2}$$

C.1 Proof of Theorem 2.1

Denote by $h := \sqrt{n}(\psi - \psi_*)$ the local parameter and $V_{\psi_*} := \Gamma'\Delta^{-1}\Gamma$. We denote by π^h and $\pi^h(\cdot|x_{1:n})$ the prior and posterior distribution, respectively, of the local parameter h. Therefore, $\pi^h(h) = n^{-d_\psi/2}\pi(\psi_* + h/\sqrt{n})$, where $d_\psi := (p + d_v)$, and

$$\pi^h(h|x_{1:n}) = \frac{\pi(\psi_* + h/\sqrt{n})\exp\{\log\frac{p(x_{1:n}|\psi_*+h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\}}{\int \pi(\psi_* + \tilde{h}/\sqrt{n})\exp\{\log\frac{p(x_{1:n}|\psi_*+\tilde{h}/\sqrt{n})}{p(x_{1:n}|\psi_*)}\}d\tilde{h}}$$

$$=: C_n^{-1}\pi(\psi_* + h/\sqrt{n})\exp\left\{\log\frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\right\}$$

and we need to show (2.15) which is equivalent to

$$\int \left|C_n^{-1}\pi(\psi_* + h/\sqrt{n})\exp\left\{\log\frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\right\} - (2\pi)^{-d_\psi/2}|V_{\psi_*}|^{1/2}e^{\{-h'V_{\psi_*}h/2\}}\right| dh \xrightarrow{p} 0. \tag{C.3}$$

Remark that

$$\int \left|C_n^{-1}\pi(\psi_* + h/\sqrt{n})\exp\left\{\log\frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\right\} - (2\pi)^{-d_\psi/2}|V_{\psi_*}|^{1/2}e^{\{-h'V_{\psi_*}h/2\}}\right| dh$$
$$\leq C_n^{-1}\int \left|\pi(\psi_* + h/\sqrt{n})\exp\left\{\log\frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\right\} - \pi(\psi_*)\exp\{-h'V_{\psi_*}h/2\}\right| dh$$
$$+ C_n^{-1}\int \left|\pi(\psi_*)\exp\{-h'V_{\psi_*}h/2\} - C_n(2\pi)^{-d_\psi/2}|V_{\psi_*}|^{1/2}\exp\{-h'V_{\psi_*}h/2\}\right| dh$$
$$=: \mathcal{I}_1 + \mathcal{I}_2. \tag{C.4}$$

Term $\mathcal{I}_1 \xrightarrow{p} 0$ by Lemma C.1 below. Because Lemma C.1 implies that $C_n \xrightarrow{p} \pi(\psi_*)(2\pi)^{d_\psi/2}|V_{\psi_*}|^{-1/2}$, then term $\mathcal{I}_2 \xrightarrow{p} 0$ and this concludes the proof of the theorem.

□

Lemma C.1. *Under Assumptions 1, 2, 5, 6 and* (2.14)*,*

$$\int \left|\pi(\psi_* + h/\sqrt{n})\exp\left\{\log\frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\right\} - \pi(\psi_*)\exp\{-h'V_{\psi_*}h/2\}\right| dh \xrightarrow{p} 0 \tag{C.5}$$

Proof. Given any $\delta, c > 0$ we break the domain of integration into three regions: (I) $A_1 := \{h; \|h\| < c\log\sqrt{n}\}$; (II) $A_2 := \{h; c\log\sqrt{n} < \|h\| < \delta\sqrt{n}\}$; (III) $A_3 := \{h; \|h\| > \delta\sqrt{n}\}$. We begin with A_3:

$$\int_{A_3} \left|\pi(\psi_* + h/\sqrt{n}) \exp\{\log \frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\} - \pi(\psi_*) \exp\{-h'V_{\psi_*}h/2\}\right| dh$$

$$\leq \int_{A_3} \pi(\psi_* + h/\sqrt{n}) e^{\{\sum_{i=1}^n (l_{n,\psi_*+h/\sqrt{n}}(x_i) - l_{n,\psi_*}(x_i))\}} dh + \int_{A_3} \pi(\psi_*) e^{\{-h'V_{\psi_*}h/2\}} dh.$$

The first integral goes to zero by (2.14). The second integral goes to zero by the properties of the tails of a normal distribution. Let us consider A_1. By (C.9) and (C.7) in Lemma C.2 we have, for a generic constant C:

$$\int_{A_1} \left|\pi(\psi_* + h/\sqrt{n}) \exp\left\{\log \frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\right\} - \pi(\psi_*) \exp\{-h'V_{\psi_*}h/2\}\right| dh$$

$$\leq e^{Cn^{-1/2}\log\sqrt{n}} \int_{A_1} \pi(\psi_* + h/\sqrt{n}) \left|e^{-h'V_{\psi_*}h/2 + Cn^{-1/2}\|h\|^2} - e^{-h'V_{\psi_*}h/2}\right| dh + o_p(\log\sqrt{n}/\sqrt{n})$$

$$+ \int_{A_1} \left|\pi(\psi_* + h/\sqrt{n}) - \pi(\psi_*)\right| e^{-h'V_{\psi_*}h/2} dh.$$

Because π is continuous at ψ_* by Assumption 2, the second integral goes to zero. Because $\left|e^{Cn^{-1/2}\|h\|^2} - 1\right| \leq e^{Cn^{-1/2}\|h\|^2} \left|Cn^{-1/2}\|h\|^2\right|$ for a generic constant C, the first integral is

$$\leq \sup_{h \in A_1} \pi(\psi_* + h/\sqrt{n}) \int_{A_1} e^{-h'V_{\psi_*}h/2} \left|e^{Cn^{-1/2}\|h\|^2} - 1\right|$$

$$\leq \sup_{h \in A_1} \pi(\psi_* + h/\sqrt{n}) \sup_{h \in A_1} e^{Cn^{-1/2}\|h\|^2} \left|Cn^{-1/2}\|h\|^2\right| \int_{A_1} e^{-h'V_{\psi_*}h/2} dh = o_p(1).$$

Next, consider the last region of integration and use (C.8) and (C.9):

$$\int_{A_2} \left|\pi(\psi_* + h/\sqrt{n}) \exp\left\{\log \frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)}\right\} - \pi(\psi_*) \exp\{-h'V_{\psi_*}h/2\}\right| dh$$

$$\leq e^{Cn^{-1/2}\sqrt{n}} \int_{A_2} \pi(\psi_* + h/\sqrt{n}) e^{-h' V_{\psi_*} h/2 + h'\left(\frac{1}{n}\sum_{i=1}^n \ddot{l}_{n,\psi_t}(x_i) - V_{\psi_*}\right)h/2} dh + \int_{A_2} \pi(\psi_*) e^{-h' V_{\psi_*} h/2} dh. \tag{C.6}$$

The second integral can be upper bounded as (for a generic constant $C > 0$):

$$\int_{A_2} \pi(\psi_*) e^{-h' V_{\psi_*} h/2} dh \leq 2\pi(\psi_*) e^{-c\log(\sqrt{n})\rho_{\min}(V_{\psi_*})/2}(\delta\sqrt{n} - c\log\sqrt{n}) \leq C\pi(\psi_*)\frac{\sqrt{n}}{n^{c\rho_{\min}(V_{\psi_*})/4}}$$

so that by choosing $c > 2\rho_{\min}$, the integral goes to zero because, under Assumptions 5 *(e)* and 6 *(c)*, $\rho_{\min}(V_{\psi_*})$ is strictly positive, where $\rho_{\min}(V_{\psi_*})$ denotes the minimum eigenvalue of the matrix V_{ψ_*}. To control the first integral in (C.6), there exists a N such that for all $n \geq N$: $P\left(-h' V_{\psi_*} h/2 + h'\left(\frac{1}{n}\sum_{i=1}^n \ddot{l}_{n,\psi_t}(x_i) - V_{\psi_*}\right)h/2 < -h' V_{\psi_*} h/4 \text{ for all } h \in A_2\right) > 1 - \epsilon$. Therefore, with probability larger than $1 - \epsilon$,

$$\int_{A_2} \pi(\psi_* + \tilde{h}/\sqrt{n}) e^{-h' V_{\psi_*} h/2 + h'\left(\frac{1}{n}\sum_{i=1}^n \ddot{l}_{n,\psi_t}(x_i) - V_{\psi_*}\right)h/2} dh \leq \sup_{h\in A_2} \pi(\psi_* + h/\sqrt{n}) \int_{A_2} e^{-h' V_{\psi_*} h/4} dh$$

which converges to zero as $n \to \infty$. Finally, by putting these three results together we show (C.5).

□

Lemma C.2. *Let Assumptions 1, 2, 5 and 6 hold and denote* $h := \sqrt{n}(\psi - \psi_*)$ *and* $V_{\psi_*} := \Gamma'\Delta^{-1}\Gamma$. *Then,*

$$\log \frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)} = -\frac{1}{2} h' V_{\psi_*} h + O_p((\|h\| + \|h\|^2)n^{-1/2}). \tag{C.7}$$

Proof. Denote $d_\psi := (p + d_v)$, $\tau(\widehat{\lambda}, \psi, x) := e^{\widehat{\lambda}(\psi)' g^A(x,\psi)}$ and $\tau_n(\widehat{\lambda}, \psi) := \frac{1}{n}\sum_{i=1}^n \tau(\widehat{\lambda}, \psi, x_i)$. Moreover, let $G^A(x, \psi_*) := \partial g^A(x, \psi)/\partial\psi'\big|_{\psi=\psi_*}$ be a matrix of dimension $d \times d_\psi$. A first order Taylor expansion of $h \mapsto \log p(x_{1:n}|\psi_* + h/\sqrt{n})$ around $h = 0$, with Lagrange remainder, gives:

$$\log \frac{p(x_{1:n}|\psi_* + h/\sqrt{n})}{p(x_{1:n}|\psi_*)} = \frac{1}{\sqrt{n}} \sum_{i=1}^n \dot{l}_{n,\psi_*}(x_i)' h + \frac{1}{2n} \sum_{i=1}^n h' \ddot{l}_{n,\psi_t}(x_i) h \tag{C.8}$$

where $\dot{l}_{n,\psi_*}(x) := \partial l_{n,\psi}(x)/\partial\psi\big|_{\psi=\psi_*}$, $\ddot{l}_{n,\psi_t}(x) := \partial^2 l_{n,\psi}(x)/(\partial\psi\partial\psi')\big|_{\psi=\psi_t}$ and $\psi_t := \psi_* +$

$th/\sqrt{n}$, $t \in [0,1]^{d_\psi}$. Simple computations give:

$$\begin{aligned}
\frac{1}{\sqrt{n}}\sum_{i=1}^n \dot{l}_{n,\psi_*}(x_i)'h &= \frac{1}{\sqrt{n}}\sum_{i=1}^n \left(1 - \frac{\tau(\widehat{\lambda},\psi_*,x_i)}{\tau_n(\widehat{\lambda},\psi_*)}\right)\left(\widehat{\lambda}(\psi_*)'G^A(x_i,\psi_*) + g^A(x_i,\psi_*)'\frac{d\widehat{\lambda}(\psi_*)}{d\psi'}\right)h \\
&= O_p(n^{-1/2}\|h\|) \qquad (C.9)
\end{aligned}$$

since under Assumption 5, $\widehat{\lambda}(\psi_*) = O_p(n^{-1/2})$ by Newey and Smith (2004, Theorem 3.1) and $\left|1 - \frac{\tau(\widehat{\lambda},\psi,x_i)}{\tau_n(\widehat{\lambda},\psi)}\right| = O_p(n^{-1/2})$ by continuity of $\psi \mapsto \widehat{\lambda}(\psi)$ (due to the Birge's maximum theorem and strict convexity of $\lambda \mapsto \frac{1}{n}\sum_{i=1}\exp\{\lambda' g^A(x_i,\psi)\}$). Denote $\mathcal{A}_1(\widehat{\lambda},\psi,x_i) := \left(\widehat{\lambda}(\psi)'G^A(x_i,\psi) + g^A(x_i,\psi)'\frac{d\widehat{\lambda}(\psi)}{d\psi'}\right)$. Then,

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^n h'\ddot{l}_{n,\psi_t}(x_i)h = \frac{h'}{n}\sum_{i=1}^n &\left(1 - \frac{\tau(\widehat{\lambda},\psi_t,x_i)}{\tau_n(\widehat{\lambda},\psi_t)}\right)\Big(\sum_{j=1}^d \widehat{\lambda}_j(\psi_t)\frac{\partial^2 g_j^A(x_i,\psi_t)}{\partial\psi\partial\psi'} \\
&+ \frac{d\widehat{\lambda}(\psi_t)'}{d\psi}G^A(x_i,\psi_t) + G^A(x_i,\psi_t)'\frac{d\widehat{\lambda}(\psi_t)}{d\psi'} + \sum_{j=1}^d g_j^A(x_i,\psi_t)\frac{d^2\widehat{\lambda}_j(\psi_t)}{d\psi d\psi'}\Big)h \\
-h'\frac{1}{n}\sum_{i=1}^n &\left(\frac{\tau(\widehat{\lambda},\psi_t,x_i)}{\tau_n(\widehat{\lambda},\psi_t)}\mathcal{A}_1(\widehat{\lambda},\psi_t,x_i)' - \frac{\tau(\widehat{\lambda},\psi_t,x_i)}{\tau_n^2(\widehat{\lambda},\psi_t)}\frac{1}{n}\sum_{j=1}^n \tau(\widehat{\lambda},\psi_t,x_j)\mathcal{A}_1(\widehat{\lambda},\psi_t,x_j)'\right)\mathcal{A}_1(\widehat{\lambda},\psi_t,x_i)h \\
&= -h'\frac{1}{n}\sum_{i=1}^n \frac{\tau(\widehat{\lambda},\psi_t,x_i)}{\tau_n(\widehat{\lambda},\psi_t)}\frac{d\widehat{\lambda}(\psi_t)'}{d\psi}g^A(x_i,\psi_t)g^A(x_i,\psi_t)'\frac{d\widehat{\lambda}(\psi_t)}{d\psi'}h + O_p(\|h\|^2 n^{-1/2}) \\
&= -h'\Gamma'\Delta^{-1}\Gamma h + O_p(\|h\|^2 n^{-1/2}) \qquad (C.10)
\end{aligned}$$

because: *(i)* under Assumption 5, $\widehat{\lambda}(\psi_*) = O_p(n^{-1/2})$ by Newey and Smith (2004, Theorem 3.1) so that $\sup_{t\in[0,1]^{d_\psi}}\left|1 - \frac{\tau(\widehat{\lambda},\psi_t,x_i)}{\tau_n(\widehat{\lambda},\psi_t)}\right| = O_p(n^{-1/2})$ by continuity of $\psi \mapsto \widehat{\lambda}(\psi)$ (due to the Birge's maximum theorem and strict convexity of $\lambda \mapsto \frac{1}{n}\sum_{i=1}\exp\{\lambda' g^A(x_i,\psi)\}$); *(ii)* $\frac{1}{n}\sum_{j=1}^n \tau(\widehat{\lambda},\psi_t,x_j)g^A(x_j,\psi_t) = O_p(n^{-1/2})$ by the results in *(i)* and Newey and Smith (2004, Lemma A.3); *(iii)* $\frac{d\widehat{\lambda}(\psi_t)}{d\psi'} = -\Delta^{-1}\Gamma + O_p(n^{-1/2})$ (by Assumptions 5 *(b)* - *(d)* and 6 *(b)* - *(c)*). By replacing (C.9) and (C.10) in (C.8) we get the result of the lemma.

□

C.2 Proof of Theorem 2.2.

The main steps of this proof proceed as in the proof of Van der Vaart (1998, Theorem 10.1) and Kleijn and van der Vaart (2012, Theorem 2.1) 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 pseudo-true value $\psi_\circ$ and define a local parameter $h = \sqrt{n}(\psi - \psi_\circ)$. Denote by π^h and $\pi^h(\cdot|x_{1:n})$ the prior and posterior distribution of h, respectively. Denote by Φ_n the normal distribution $\mathcal{N}_{\Delta_{n,\psi_\circ}, V_{\psi_\circ}^{-1}}$ and by ϕ_n its Lebesgue density. For a compact subset $K \subset \mathbb{R}^p$ such that $\pi^h(h \in K|x_{1:n}) > 0$ define, for any Borel set $B \subseteq \Psi$,

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

and let Φ_n^K be the Φ_n distribution conditional on K. The proof consists of two steps. In the first step we show that the Total Variation (TV) norm of $\pi_K^h(\cdot|x_{1:n}) - \Phi_n^K$ converges to zero in probability. In the second step we use this result to show that the TV norm of $\pi^h(\cdot|x_{1:n}) - \Phi_n$ converges to zero in probability.

Let Assumption 8 *(a)* hold. For every open neighborhood $\mathcal{U} \subset \Psi$ of $\psi_\circ$ and a compact subset $K \subset \mathbb{R}^p$, there exists an N such that for every $n \geq N$:

$$\psi_\circ + K\frac{1}{\sqrt{n}} \subset \mathcal{U}. \tag{C.11}$$

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 π^h denotes the Lebesgue density of the prior π^h for h and $s_n(h) := p(x_{1:n}|\psi_\circ + h/\sqrt{n})/p(x_{1:n}|\psi_\circ)$. The function f_n is well defined for n sufficiently large because of (C.11) and Assumption 8 *(a)*. Remark that by (C.11) and since the prior for ψ puts enough mass on $\mathcal{U}$, then π^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 (C.16) in Theorem C.1 below:

$$\begin{aligned} \log \frac{\phi_n(k_2)s_n(k_1)\pi^h(k_1)}{\phi_n(k_1)s_n(k_2)\pi^h(k_2)} &= -\frac{1}{2}(k_2 - \Delta_{n,\psi^\circ})'V_{\psi^\circ}(k_2 - \Delta_{n,\psi^\circ}) + \frac{1}{2}(k_1 - \Delta_{n,\psi^\circ})'V_{\psi_\circ}(k_1 - \Delta_{n,\psi_\circ}) \\ &+ k_1'V_{\psi_\circ}\Delta_{n,\psi_\circ} - \frac{1}{2}k_1'V_{\psi_\circ}k_1 - k_2'V_{\psi_\circ}\Delta_{n,\psi_\circ} + \frac{1}{2}k_2'V_{\psi_\circ}k_2 + o_p(1) = o_p(1). \end{aligned} \tag{C.12}$$

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

$$\sup_{k_1, k_2 \in K} f_n(k_1, k_2) \xrightarrow{p} 0, \qquad \text{as } n \to \infty. \tag{C.13}$$

Suppose that the subset K contains a neighborhood of 0 (which guarantees that $\Phi_n(K) > 0$ and then that Φ_n^K is well defined) and let $\Xi_n := \{\pi^h(K|x_{1:n}) > 0\}$. Moreover, for a given $\eta > 0$ define the event $\Omega_n := \left\{\sup_{k_1,k_2 \in K} f_n(k_1,k_2) \leq \eta\right\}$. 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 $(\cdot)_+$,

$$\begin{aligned} \frac{1}{2}\mathbf{E}^P\|\Phi_n^K - \pi_K^h(\cdot|x_{1:n})\|_{TV}1_{\Omega_n\cap\Xi_n} &= \mathbf{E}^P\int_K\left(1 - \frac{d\Phi_n^K(k_2)}{d\pi_K^h(k_2|x_{1:n})}\right)_+ d\pi_K^h(k_2|x_{1:n})1_{\Omega_n\cap\Xi_n} \\ &\leq \mathbf{E}^P\int_K\int_K f_n(k_1,k_2)d\Phi_n^K(k_1)d\pi_K^h(k_2|x_{1:n})1_{\Omega_n\cap\Xi_n} \\ &\leq \mathbf{E}^P\sup_{k_1,k_2\in K} f_n(k_1,k_2)1_{\Omega_n\cap\Xi_n} \quad \text{(C.14)} \end{aligned}$$

that converges to zero by (C.13). By (C.14), it follows that (by remembering that $\|\cdot\|_{TV}$ is upper bounded by 2)

$$\mathbf{E}^P\|\pi_K^h(\cdot|x_{1:n}) - \Phi_n^K\|_{TV}1_{\Xi_n} \leq \mathbf{E}^P\|\pi_K^h(\cdot|x_{1:n}) - \Phi_n^K\|_{TV}1_{\Omega_n\cap\Xi_n} + 2P(\Omega_n^c \cap \Xi_n), \quad \text{(C.15)}$$

where the second term is $o(1)$ by (C.13). 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 Ξ_n accordingly. For each $n \geq 1$, (C.15) holds for these balls. Moreover, by (C.18) in Theorem C.2 below: $P(\Xi_n) \to 1$. Therefore, by the triangular inequality, the TV distance is upper bounded by

$$\begin{aligned} \mathbf{E}^P\|\pi^h(\cdot|x_{1:n}) - \Phi_n\|_{TV} &\leq \mathbf{E}^P\|\pi^h(\cdot|x_{1:n}) - \Phi_n\|_{TV}1_{\Xi_n} + \mathbf{E}^P\|\pi^h(\cdot|x_{1:n}) - \Phi_n\|_{TV}1_{\Xi_n^c} \\ &\leq \mathbf{E}^P\|\pi^h(\cdot|x_{1:n}) - \pi_{K_n}^h(\cdot|x_{1:n})\|_{TV} + \mathbf{E}^P\|\pi_{K_n}^h(\cdot|x_{1:n}) - \Phi_n^{K_n}\|_{TV}1_{\Xi_n} \\ &\qquad + \mathbf{E}^P\|\Phi_n^{K_n} - \Phi_n\|_{TV} + 2P(\Xi_n^c) \\ &\leq 2\mathbf{E}^P(\pi_{K_n^c}^h(\cdot|x_{1:n})) + \mathbf{E}^P\|\pi_{K_n}^h(\cdot|x_{1:n}) - \Phi_n^{K_n}\|_{TV}1_{\Xi_n} + o(1) \xrightarrow{p} 0 \end{aligned}$$

since $\mathbf{E}^P(\pi^h(K_n^c|x_{1:n})) = o(1)$ by (C.18) and $\mathbf{E}^P\|\pi_{K_n}^h(\cdot|x_{1:n}) - \Phi_n^{K_n}\|_{TV}1_{\Xi_n} = o_P(1)$ by (C.15) and (C.14), and where in the third line we have used the fact that: $\mathbf{E}^P\|\pi^h(\cdot|x_{1:n}) - \pi_{K_n}^h(\cdot|x_{1:n})\|_{TV} = 2\mathbf{E}^P(\pi_{K_n^c}^h(\cdot|x_{1:n}))$ and $\|\Phi_n^{K_n} - \Phi_n\|_{TV} = \|\Phi_n^{K_n^c}\|_{TV} = o_p(1)$ by Kleijn and van der Vaart (2012, Lemma 5.2) since Δ_{n,ψ_0} is uniformly tight.

□

The next theorem establishes that the misspecified model satisfies a stochastic Local Asymptotic Normality (LAN) expansion around the pseudo-true value $\psi_\circ$.

Theorem C.1 (Stochastic LAN). *Assume that the matrix $V_{\psi_\circ}$ is nonsingular and that Assumptions 5 (a)-(d), 6 (b), 3, 7, and 8 hold. Then for every compact set $K \subset \mathbb{R}^p$,*

$$\sup_{h\in K}\left|\log\frac{p(x_{1:n}|\psi_\circ + h/\sqrt{n})}{p(x_{1:n}|\psi_\circ)} - h'V_{\psi_\circ}\Delta_{n,\psi_\circ} + \frac{1}{2}h'V_{\psi_\circ}h\right| \xrightarrow{p} 0 \tag{C.16}$$

where $\psi_\circ$ is as defined in (2.16), $V_{\psi_\circ} := -\mathbf{E}^P[\ddot{\mathfrak{L}}_{n,\psi_\circ}]$ and $\Delta_{n,\psi_\circ} := \frac{1}{\sqrt{n}}\sum_{i=1}^n V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)$ is bounded in probability.

Proof. See Appendix E.

□

The next theorem establishes that the posterior of ψ concentrates and puts all its mass on $\Psi_n := \{\|\psi - \psi_\circ\| \leq M_n/\sqrt{n}\}$ as $n \to \infty$.

Theorem C.2 (Posterior Consistency). *Assume that the stochastic LAN expansion (C.16) holds for $\psi_\circ$ defined in (2.16). Moreover, let Assumptions 2 (a), 3 and 4 hold and assume that there exists a constant $C > 0$ such that for any sequence $M_n \to \infty$,*

$$P\left(\sup_{\psi\in\Psi_n^c}\frac{1}{n}\sum_{i=1}^n\left(l_{n,\psi}(x_i) - l_{n,\psi_\circ}(x_i)\right) \leq -\frac{CM_n^2}{n}\right) \to 1 \tag{C.17}$$

as $n \to \infty$ where $\Psi_n := \{\|\psi - \psi_\circ\| \leq M_n/\sqrt{n}\}$. Then,

$$\pi\left(\sqrt{n}\|\psi - \psi_\circ\| > M_n \middle| x_{1:n}\right) \xrightarrow{p} 0 \tag{C.18}$$

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

Proof. See Appendix E.

□

C.3 Proof of Lemma 2.1.

By Theorem 10 of Schennach (2007), which is valid under Assumptions 5 *(a)-(c)*, 3, 7 *(c)*, *(e)* and 8: $\sqrt{n}(\widehat{\psi} - \psi_\circ) = O_p(1)$. Denote $\widehat{h} := \sqrt{n}(\widehat{\psi} - \psi_\circ)$ and $\widetilde{h} := \Delta_{n,\psi_\circ}$. Because of

(C.16), we have:

$$\sum_{i=1}^{n}\left(l_{n,\psi_\circ+\widehat{h}/\sqrt{n}}-l_{n,\psi_\circ}\right)(x_i) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\widehat{h}'\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)-\frac{1}{2}\widehat{h}'V_{\psi_\circ}\widehat{h}+o_p(1) \tag{C.19}$$

$$\sum_{i=1}^{n}\left(l_{n,\psi_\circ+\widetilde{h}/\sqrt{n}}-l_{n,\psi_\circ}\right)(x_i) = \frac{1}{2\sqrt{n}}\sum_{i=1}^{n}\widetilde{h}'\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)+o_p(1). \tag{C.20}$$

By definition of $\widehat{\psi}$ as the maximizer of $\sum_{i=1}^{n} l_{n,\psi}(x_i)$, the left hand side of (C.19) is not smaller than the left hand side of (C.20). It follows that the same relation holds for the right hand sides of (C.19) and (C.20), and by taking their difference we obtain:

$$-\frac{1}{2}\left(\widehat{h}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)\right)' V_{\psi_\circ}\left(\widehat{h}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)\right)+o_p(1)\geq 0. \tag{C.21}$$

Because $-V_{\psi_\circ}$ is negative definite then

$$-\frac{1}{2}\left(\widehat{h}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)\right)' V_{\psi_\circ}\left(\widehat{h}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)\right)\leq 0.$$

This and (C.21) imply that $\left\|V_{\psi_\circ}^{-1/2}\left(\widehat{h}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)\right)\right\|\xrightarrow{p}0$ which in turn implies that

$$\left\|\left(\widehat{h}-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)\right)\right\|\xrightarrow{p}0$$

which establishes the result of the lemma.

□

D Proofs for Section 3.3

In this appendix we prove Theorems 3.1 and 3.2 and Corollary 3.1. The proofs of Theorems 3.1 and 3.2 have already been stated in the Appendix of the paper but for easiness of reading we give them also here. For the same reason we remind the notation already introduced in the Appendix of the paper. Recall the notation $\psi^\ell=(\theta^\ell,v^\ell)$ and the estimator $\widehat{\psi}^\ell:=(\widehat{\theta}^\ell,\widehat{v}^\ell)$ denotes Schennach (2007)'ETEL estimator of ψ^ℓ in model M_ℓ:

$$\widehat{\psi}^\ell:=\arg\max_{\psi^\ell\in\Psi^\ell}\frac{1}{n}\sum_{i=1}^{n}\left[\widehat{\lambda}(\psi^\ell)'g^A(x_i,\psi^\ell)-\log\frac{1}{n}\sum_{j=1}^{n}\exp\{\widehat{\lambda}(\psi^\ell)'g^A(x_j,\psi^\ell)\}\right] \tag{D.1}$$

where $\widehat{\lambda}(\psi^\ell) = \arg\min_{\lambda\in\mathbb{R}^d} \frac{1}{n}\sum_{i=1}^n \left[\exp\{\lambda' g^A(x_i,\psi^\ell)\}\right]$. Denote $\widehat{g}^A(\psi^\ell) := \frac{1}{n}\sum_{i=1}^n g^A(x_i,\psi^\ell)$, $\widehat{g}^A_\ell := \widehat{g}^A(\psi^\ell)$,

$$\widehat{L}(\psi^\ell) := \exp\{\widehat{\lambda}(\psi^\ell)'\widehat{g}^A(\psi^\ell)\}\left[\frac{1}{n}\sum_{i=1}^n \exp\{\widehat{\lambda}(\psi^\ell)' g^A(x_i,\psi^\ell)\}\right]^{-1}$$

and $L(\psi^\ell) = \exp\{\lambda_\circ(\psi^\ell)'\mathbf{E}^P[g^A(x,\psi^\ell)]\}\left(\mathbf{E}^P\left[\exp\{\lambda_\circ(\psi^\ell)'g^A(x,\psi^\ell)\}\right]\right)^{-1}$. Moreover, we use the notation $\Sigma_\ell = \left(\Gamma_\ell'\Delta_\ell^{-1}\Gamma_\ell\right)^{-1}$ where $\Gamma_\ell := \mathbf{E}^P\left[\frac{\partial}{\partial\psi^{\ell\prime}} g^A(X,\psi^\ell_*)\right]$ and $\Delta_\ell := \mathbf{E}^P[g^A(X,\psi^\ell_*)g^A(X,\psi^\ell_*)']$. In the proofs, we omit measurability issues which can be dealt with in the usual manner by replacing probabilities with outer probabilities.

D.1 Proof of Theorem 3.1

By (3.5) and Lemmas D.1 and D.2 below we obtain

$$\begin{aligned}P\left(\max_{\ell\neq j}\log m(x_{1:n};M_\ell) < \log m(x_{1:n};M_j)\right) = P\Big(&\max_{\ell\neq j}\Big[-\frac{n}{2}\widehat{g}_\ell^{A\prime}\Delta^{-1}\widehat{g}_\ell^A + \log\pi(\widehat{\psi}^\ell|M_\ell)\\ &-\frac{(p_\ell+d_{v_\ell})}{2}(\log n - \log(2\pi)) + \frac{1}{2}\log|\Sigma_\ell|\Big] + \frac{n}{2}\widehat{g}_j^{A\prime}\Delta^{-1}\widehat{g}_j^A + o_p(1)\\ &< \log\pi(\widehat{\psi}^j|M_j) - \frac{(p_j+d_{v_j})}{2}(\log n - \log(2\pi)) + \frac{1}{2}\log|\Sigma_j|\Big). \quad \text{(D.2)}\end{aligned}$$

Remark that $n\widehat{g}_j^{A\prime}\Delta^{-1}\widehat{g}_j^A \xrightarrow{d} \chi^2_{d-(p_j+d_{v_j})}$, $\forall j$, so that $n\widehat{g}_j^{A\prime}\Delta^{-1}\widehat{g}_j^A = O_p(1)$. Suppose first that $(p_\ell + d_{v_\ell} > p_j + d_{v_j})$, $\forall \ell\neq j$. Since $-n\widehat{g}_\ell^{A\prime}\Delta^{-1}\widehat{g}_\ell^A < 0$ for every ℓ, we lower bound (D.2) as

$$\begin{aligned}P\left(\max_{\ell\neq j}\log m(x_{1:n};M_\ell) < \log m(x_{1:n};M_j)\right) &\geq P\Big(\frac{n}{2}\widehat{g}_j^{A\prime}\Delta^{-1}\widehat{g}_j^A + o_p(1)\\ &< \log n\Big[\frac{\min_{\ell\neq j}(p_\ell+d_{v_\ell}) - p_j - d_{v_j}}{2} - \frac{\min_{\ell\neq j}(p_\ell+d_{v_\ell}) - p_j - d_{v_j})}{2\log n}\log(2\pi)\\ &\quad - \frac{\log[\max_{\ell\neq j}\pi(\widehat{\psi}^\ell|M_\ell)/\pi(\widehat{\psi}^j|M_j)]}{\log n} - \frac{1}{2\log n}\left(\max_{\ell\neq j}\log|\Sigma_\ell| - \log|\Sigma_j|\right)\Big]\Big)\\ &= P\Big(\underbrace{\frac{n}{2}\widehat{g}_j^{A\prime}\Delta^{-1}\widehat{g}_j^A + o_p(1)}_{=:\mathcal{I}_n} < \underbrace{\log n\Big[\frac{\min_{\ell\neq j}(p_\ell+d_{v_\ell}) - p_j - d_{v_j}}{2} + \mathcal{O}_p((\log n)^{-1})\Big]}_{=:\mathcal{II}_n}\Big). \quad \text{(D.3)}\end{aligned}$$

Because $\mathcal{I}_n = \mathcal{O}_p(1)$ (and is asymptotically positive) and $\mathcal{II}_n$ is strictly positive as $n\to\infty$ (since $(p_\ell+d_{v_\ell}) > (p_j+d_{v_j})$, $\forall\ell\neq j$) and converges to $+\infty$, then the probability converges to 1. This proves one direction of the statement.

To prove the second direction of the statement, suppose that $\lim_{n\to\infty} P(\max_{\ell\neq j}\log m(x_{1:n};M_\ell) < \log m(x_{1:n};M_j)) = 1$ and consider the following upper bound (which follows from (D.2) and the fact that $n\widehat{g}_j^{A'}\Delta^{-1}\widehat{g}_j^A > 0$, $\forall n$):

$$
\begin{aligned}
&P\left(\max_{\ell\neq j}\log m(x_{1:n};M_\ell) < \log m(x_{1:n};M_j)\right)\\
&\qquad\qquad \leq P\left(\log m(x_{1:n};M_\ell) < \log m(x_{1:n};M_j)\right), \qquad \forall \ell\neq j\\
&\leq P\Big(-\frac{n}{2}\widehat{g}_\ell^{A'}\Delta^{-1}\widehat{g}_\ell^A + o_p(1) + \log n\Big[\frac{(p_j+d_{v_j})-(p_\ell+d_{v_\ell})}{2} + \mathcal{O}_p\left(\frac{1}{\log n}\right)\Big] < 0\Big), \quad \forall \ell\neq j.
\end{aligned}
\tag{D.4}
$$

Because the probability in the first line of (D.4) converges to 1 as $n\to\infty$ then, necessarily, the probability in the last line of (D.4) converges to 1 which is possible only if $(p_j + d_{v_j}) < (p_\ell + d_{v_\ell})$ because $\log n\left[\frac{(p_j+d_{v_j})-(p_\ell+d_{v_\ell})}{2}\right]$ is the dominating term since $-\frac{n}{2}\widehat{g}_\ell^{A'}\Delta^{-1}\widehat{g}_\ell^A < 0$ and it remains bounded as $n\to\infty$. Since the first inequality in (D.4) holds $\forall\ell\neq j$ then convergence to 1 of the probability in the last line of (D.4) is possible only if $(p_j+d_{v_j}) < (p_\ell+d_{v_\ell})$, $\forall \ell\neq j$.

□

D.2 Proof of Theorem 3.2

We can write $\log p(x_{1:n}|\psi^\ell;M_\ell) = -n\log n + n\log\widehat{L}(\psi^\ell)$. Then, we have:

$$
\begin{aligned}
P\left(\log m(x_{1:n};M_j) > \max_{\ell\neq j}\log m(x_{1:n};M_\ell)\right) &= P\Big(n\log\widehat{L}(\psi_\circ^j) + \log\pi(\psi_\circ^j|M_j) - \log\pi(\psi_\circ^j|x_{1:n},M_j)\\
&\qquad > \max_{\ell\neq j}[n\log\widehat{L}(\psi_\circ^\ell) + \log\pi(\psi_\circ^\ell|M_\ell) - \log\pi(\psi_\circ^\ell|x_{1:n},M_\ell)]\Big)\\
&= P\Big(n\log L(\psi_\circ^j) + n\log\frac{\widehat{L}(\psi_\circ^j)}{L(\psi_\circ^j)} + \mathcal{B}_j > \max_{\ell\neq j}\Big[n\log L(\psi_\circ^\ell) + \mathcal{B}_\ell + n\log\frac{\widehat{L}(\psi_\circ^\ell)}{L(\psi_\circ^\ell)}\Big]\Big)
\end{aligned}
\tag{D.5}
$$

where $\forall\ell$, $\mathcal{B}_\ell := \log\pi(\psi_\circ^\ell|M_\ell) - \log\pi(\psi_\circ^\ell|x_{1:n},M_\ell)$ and $\mathcal{B}_\ell = O_p(1)$ under the assumptions of Theorem 2.2. By definition of $dQ^*(\psi)$ in Section 2.3 we have that: $\log L(\psi_\circ^\ell) = \mathbf{E}^P[\log dQ^*(\psi_\circ^\ell)/dP] = -\mathbf{E}^P[\log dP/dQ^*(\psi_\circ^\ell)] = -K(P||Q^*(\psi_\circ^\ell))$. Remark that $\mathbf{E}^P[\log(dP/dQ^*(\psi_\circ^2))] > \mathbf{E}^P[\log(dP/dQ^*(\psi_\circ^1))]$ means that the KL divergence between P and $Q^*(\psi_\circ^\ell)$, is smaller for model M_1 than for model M_2, where $Q^*(\psi_\circ^\ell)$ minimizes the KL divergence between $Q\in\mathcal{P}_{\psi_\circ^\ell}$ and P for $\ell\in\{1,2\}$ (notice the inversion of the two probabilities).

First, suppose that $\min_{\ell\neq j} \mathbf{E}^P\left[\log\left(dP/dQ^*(\psi_\circ^\ell)\right)\right] > \mathbf{E}^P\left[\log\left(dP/dQ^*(\psi_\circ^j)\right)\right]$. By (D.5):

$$P\left(\log m(x_{1:n};M_j) > \max_{\ell\neq j}\log m(x_{1:n};M_\ell)\right) \geq$$
$$P\left(\log\frac{\widehat{L}(\psi_\circ^j)}{L(\psi_\circ^j)} - \max_{\ell\neq j}\log\frac{\widehat{L}(\psi_\circ^\ell)}{L(\psi_\circ^\ell)} + \frac{1}{n}(\mathcal{B}_j - \max_{\ell\neq j}\mathcal{B}_\ell) > \underbrace{\max_{\ell\neq j}\log L(\psi_\circ^\ell) - \log L(\psi_\circ^j)}_{=:\mathcal{I}_n}\right). \quad (D.6)$$

This probability converges to 1 because $\mathcal{I}_n = K(P||Q^*(\psi_\circ^j)) - \min_{\ell\neq j} K(P||Q^*(\psi_\circ^\ell)) < 0$ by assumption, and $\left[\log\widehat{L}(\psi^\ell) - \log L(\psi^\ell)\right] \xrightarrow{p} 0$, for every $\psi^\ell\in\Psi^\ell$ and every $\ell\in\{1,2\}$ by Lemma D.3 below.

To prove the second direction of the statement, suppose that $\lim_{n\to\infty} P(\log m(x_{1:n};M_j) > \max_{\ell\neq j}\log m(x_{1:n};M_\ell)) = 1$. By (D.5) it holds, $\forall\ell\neq j$

$$P\left(\log m(x_{1:n};M_j) > \max_{\ell\neq j}\log m(x_{1:n};M_\ell)\right) \leq$$
$$P\Big(\log\frac{\widehat{L}(\psi_\circ^j)}{L(\psi_\circ^j)} - \log\frac{\widehat{L}(\psi_\circ^\ell)}{L(\psi_\circ^\ell)} + \frac{1}{n}(\mathcal{B}_j - \mathcal{B}_\ell) > \log\frac{L(\psi_\circ^\ell)}{L(\psi_\circ^j)}\Big). \quad (D.7)$$

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

□

D.3 Proof of Corollary 3.1

We can write $\log p(x_{1:n}|\psi^\ell;M_\ell) = -n\log n + n\log\widehat{L}(\psi^\ell)$. Moreover, denote by $S_m := \{j;\, M_j \text{ does not satisfy Assumption } 1\}$ the set of indices of the models that are misspecified and by S_m^c its complement in $\{1,2,\ldots,J\}$.

First, suppose that $\lim_{n\to\infty} P\left(\log m(x_{1:n};M_1) > \max_{j\neq 1}\log m(x_{1:n};M_j)\right) = 1$. Then, because $\max_{j\neq 1}\log m(x_{1:n};M_j) \geq \max_{j\neq 1; j\in S_m^c}\log m(x_{1:n};M_j)$,

$$P\left(\log m(x_{1:n};M_1) > \max_{j\neq 1}\log m(x_{1:n};M_j)\right)$$
$$\leq P\left(\log m(x_{1:n};M_1) > \max_{j\neq 1;\, j\in S_m^c}\log m(x_{1:n};M_j)\right) \quad (D.8)$$

which implies that the probability on the right hand side converges to 1 as $n \to \infty$. Then by Theorem 3.1, we necessarily have $(p_1 + d_{v_1}) < (p_j + d_{v_j})$, $\forall j \neq 1$, $j \in S_m^c$.

Next, suppose that $(p_1 + d_{v_1}) < (p_j + d_{v_j})$, $\forall j \neq 1$. Define the event

$$\mathcal{A} := \left\{ \max_{j \neq 1;\, j \in S_m^c} \log m(x_{1:n}; M_j) > \max_{j \neq 1; j \in S_m} \log m(x_{1:n}; M_j) \right\}.$$

Because all the models M_j with $j \in S_m^c$ have $K(P||Q^*(\psi^j)) = 0$ (because they are correctly specified) then $\lim_{n\to\infty} P(\mathcal{A}) = 1$ by Theorem 3.2. By the Law of Total Probability we can write

$$\begin{aligned} P\left(\log m(x_{1:n}; M_1) > \max_{j\neq 1} \log m(x_{1:n}; M_j)\right) &= P\Big(\log m(x_{1:n}; M_1) > \\ \max_{j \neq 1} \log m(x_{1:n}; M_j) \Big| \mathcal{A}\Big) P(\mathcal{A}) &+ P\left(\log m(x_{1:n}; M_1) > \max_{j\neq 1} \log m(x_{1:n}; M_j) \Big| \mathcal{A}^c\right) P(\mathcal{A}^c) \\ &\geq P\left(\log m(x_{1:n}; M_1) > \max_{j\neq 1} \log m(x_{1:n}; M_j) \Big| \mathcal{A}\right) P(\mathcal{A}) \\ &= P\left(\log m(x_{1:n}; M_1) > \max_{j\neq 1;\, j \in S_m^c} \log m(x_{1:n}; M_j)\right) P(\mathcal{A}) \quad \text{(D.9)} \end{aligned}$$

which converges to 1 by Theorem 3.1.

□

D.4 Technical Lemmas

Lemma D.1. *Let Assumptions 1, 5 and 6 hold for ψ^ℓ. Then,*

$$\begin{aligned} \log p(x_{1:n}|\widehat{\psi}^\ell; M_\ell) &= -n \log n - \frac{n}{2} \widehat{g}_\ell^{A'} \Delta_\ell^{-1} \widehat{g}_\ell^{A} + o_p(1) \\ &= -n \log n - \frac{\chi^2_{d-(p_\ell + d_{v_\ell}}}{2} + o_p(1) \quad \text{(D.10)} \end{aligned}$$

where $\chi^2_{d-(p_\ell+d_{v_\ell})}$ denotes a chi square distribution with $(d - (p_\ell + d_{v_\ell}))$ degrees of freedom.

Proof. See Appendix F.

□

Lemma D.2. *Let Assumptions 1, 2, 5, 6 and (2.15) hold for ψ^ℓ. Then,*

$$-\log \pi(\widehat{\psi}^\ell | x_{1:n}; M_\ell) = -\frac{(p_\ell + d_{v_\ell})}{2} [\log n - \log(2\pi)] + \frac{1}{2} \log |\Sigma_\ell| + o_p(1).$$

Proof. See Appendix F. ☐

Lemma D.3. *Let M_ℓ be a misspecified model (that is, a model that does not satisfy Assumption 1) and let $g^A(x, \psi^\ell)$ and ψ^ℓ be the corresponding moment functions and parameters. Then, under Assumptions 5 (a)-(d), 3 and 7,*

$$\sup_{\psi^\ell \in \Psi^\ell} \left| \log \frac{\exp\{\widehat{\lambda}(\psi^\ell)' \widehat{g}^A(\psi^\ell)\}}{\frac{1}{n}\sum_{i=1}^n \exp\{\widehat{\lambda}(\psi^\ell)' g^A(x_i, \psi^\ell)\}} - \log \frac{\exp\{\lambda_\circ(\psi^\ell)' \mathbf{E}^P[g^A(x, \psi^\ell)]\}}{\mathbf{E}^P\left[\exp\{\lambda_\circ(\psi^\ell)' g^A(x, \psi^\ell)\}\right]} \right| \xrightarrow{p} 0.$$

Proof. See Appendix F. ☐

E Proof of Theorems C.1 and C.2

E.1 Proof of Theorem C.1

For a vector z and a scalar $\delta > 0$ we denote by $B(z, \delta)$ the closed ball centred on z with radius δ. In this proof we use standard notation in empirical process theory: $\mathbb{P}_n := \frac{1}{n}\sum_{i=1}^n \delta_{x_i}$ where δ_x is the Dirac measure at x, and $\mathbb{G}_n f := \sqrt{n}(\mathbb{P}_n f - \mathbf{E}^P f)$ for every function f. Moreover, we use Van der Vaart (2002, Theorem 6.16) which we report here for convenience (in a slightly modified version):

Theorem E.1 (Theorem 6.16 in Van der Vaart (2002))**.** *Let $\mathcal{F}_n$ be classes of measurable functions (that may change with n) such that $P(\widehat{f}_n \in \mathcal{F}_n) \to 1$ and such that: (i) the bracketing integral $J_{[]}(\delta_n, \mathcal{F}_n, L_2(P)) \to 0$, for every $\delta_n \downarrow 0$, and (ii) its envelope functions satisfy the Lindeberg condition. If $\mathbf{E}^P[(\widehat{f}_n - f_0)^2] \to 0$ in probability for some $f_0 \in L_2(P)$ then $\mathbb{G}_n(\widehat{f}_n - f_0) \xrightarrow{p} 0$.*

The maps $x \mapsto l_{n,\psi_\circ}(x)$ and $x \mapsto l_{n,\psi_\circ + h/\sqrt{n}}(x)$ are random functions, that is, measurable functions that, for a fixed x, are functions of the observations $x_1, \ldots, x_n$. By writing $\mathbb{P}_n[l_{n,\psi_\circ}]$ and $\mathbf{E}^P[l_{n,\psi_\circ}]$ we mean the (empirical and true) expectations of the function $x \mapsto l_{n,\psi_\circ}(x)$ with $(x_1, \ldots, x_n)$ kept fixed (and similarly for $l_{n,\psi_\circ + h/\sqrt{n}}(x)$). Denote by $\dot{l}_{n,\psi_\circ}(x)$ and $\ddot{l}_{n,\psi_\circ}(x)$ the first and second order derivatives of the function $\psi \mapsto l_{n,\psi}(x)$ evaluated at $\psi_\circ$ (where we leave implicit the argument x).

A second order Taylor expansion of $l_{n,\psi_\circ+h/\sqrt{n}}$ around $h=0$, for a fixed x, gives

$$l_{n,\psi_\circ+h/\sqrt{n}} = l_{n,\psi_\circ} + \frac{h'}{\sqrt{n}}\dot{l}_{n,\psi_\circ} + \frac{1}{2n}h'\ddot{l}_{n,\psi_\circ}h + Rem. \tag{E.1}$$

By continuity of the map $\psi \mapsto l_{n,\psi}$ (which is valid under Assumption 5 *(c)* and by the Birge's maximum theorem and strict convexity of $\mathbb{P}_n \exp\{\lambda' g^A(x,\psi)\}$), the reminder term Rem is of order $o_p(\|h\|^2/n)$ since $\ddot{l}_{n,\psi_\circ} = \ddot{\mathfrak{L}}_{n,\psi_\circ} + o_p(1)$ and $\ddot{\mathfrak{L}}_{n,\psi_\circ} = O_p(1)$ under Assumptions 7 and 8 (see Schennach (2007, proof of Theorem 10)).

We consider the empirical process

$$\begin{aligned}&\mathbb{G}_n\left(\sqrt{n}(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}) - h'\dot{l}_{n,\psi_\circ}\right)\\ &\qquad := n\left(\mathbb{P}_n\left(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}\right) - \mathbf{E}^P\left(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}\right)\right) - h'\mathbb{G}_n\dot{l}_{n,\psi_\circ}\end{aligned} \tag{E.2}$$

where, according to the definition of random functions given just above:

$$\begin{aligned}\mathbb{P}_n[l_{n,\psi_\circ}] &= \frac{1}{n}\sum_{i=1}^{n}\widehat{\lambda}(\psi_\circ)'g^A(x_i,\psi_\circ) - \log\mathbb{P}_n\left[e^{\widehat{\lambda}(\psi_\circ)'g^A(x_j,\psi_\circ)}\right],\\ \text{and}\quad \mathbf{E}^P[l_{n,\psi_\circ}] &= \mathbf{E}^P\left[\widehat{\lambda}(\psi_\circ)'g^A(X,\psi_\circ)\right] - \log\mathbf{E}_n\left[e^{\widehat{\lambda}(\psi_\circ)'g^A(x_j,\psi_\circ)}\right]\end{aligned}$$

(and similarly for the other functions). The Markov's inequality and (E.1) imply that

$$P\left(\left|\mathbb{G}_n\left(\sqrt{n}(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}) - h'\dot{l}_{n,\psi_\circ}\right)\right| > \epsilon\right) \leq \frac{1}{\epsilon\sqrt{n}}\mathbf{E}^P\left|\mathbb{G}_n\left(\frac{1}{2}h'\ddot{\mathfrak{L}}_{n,\psi_\circ}h\right) + o_p(\|h\|^2)\right|$$

that converges to zero since $\ddot{\mathfrak{L}}_{n,\psi_\circ} = O_p(1)$ under Assumptions 7 and 8. This shows that the sequence $\mathbb{G}_n\left(\sqrt{n}(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}) - h'\dot{l}_{n,\psi_\circ}\right)$ (seen as a stochastic process indexed by h) converges in probability and then (marginally) in distribution to zero. Next, we have to make this result uniform in h, that is, we have to show that the sequence of processes $\mathbb{G}_n\left(\sqrt{n}(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}) - h'\dot{l}_{n,\psi_\circ}\right)$ converges weakly in the space $l^\infty(h; h\in K)$ for a compact set $K\subset\mathbb{R}^p$. To show this we intend to apply Theorem E.1 given above. The proof consists of three steps where each step verifies the assumptions of Theorem E.1.

In the first step, we verify *(i)* in Theorem E.1 and we define a suitable class of functions that changes with the sample size n. Denote $\tau_n(\widehat{\lambda},\psi) := \mathbb{P}_n\left[e^{\widehat{\lambda}(\psi)'g^A(x,\psi)}\right]$, $\tau(\widehat{\lambda},\psi_\circ) := \mathbf{E}^P\left[e^{\widehat{\lambda}(\psi_\circ)'g^A(X,\psi_\circ)}\right]$ and consider the class of functions

$$\mathcal{F}_{1/\sqrt{n}} := \Big\{\lambda_1' g^A(x,\psi) - \lambda_2' g^A(x,\psi_\circ) - \log(\tau_1/\tau_2);\ \tau_1 \in B\left(\tau(\lambda_1,\psi), \frac{1}{\sqrt{n}}\right);\ \lambda_1 \in B(\lambda_\circ(\psi), 1/\sqrt{n});$$
$$\tau_2 \in B\left(\tau(\lambda_2,\psi_\circ), \frac{1}{\sqrt{n}}\right);\ \lambda_2 \in B(\lambda_\circ(\psi_\circ), 1/\sqrt{n});\ \|\psi - \psi_\circ\| \leq \frac{1}{\sqrt{n}}\Big\}. \quad \text{(E.3)}$$

By Lemma E.1, the bracketing integral $J_{[]}(\delta_n, \sqrt{n}\mathcal{F}_{1/\sqrt{n}}, L_2(P))$ of the class $\sqrt{n}\mathcal{F}_{1/\sqrt{n}}$ converges to zero as $\delta_n \to 0$. This satisfies assumption (i) of Theorem E.1.

The second step of the proof consists in finding an envelope function for the class $\sqrt{n}\mathcal{F}_{1/\sqrt{n}}$ and in showing that it satisfies the Lindeberg condition (this is assumption *(ii)* in Theorem E.1). Lemma E.2 shows that $F_{2,n}(x)$ is an envelope function for the class $\mathcal{F}_{1/\sqrt{n}}$, where

$$F_{2,n}(x) := \frac{1}{\sqrt{n}}\left(C_{1,n}b(x) + C_{2,n}\|g^A(x,\psi_\circ)\| + \frac{1}{C_{4,n}}C_{3,n}\right)$$

where $b(x)$ is the function defined in Assumption 8 *(c)* and $C_{i,n} > 0$, $i = 1,\ldots,4$ are sequences of positive and bounded constants that depend on $\psi_\circ$ and $\lambda_\circ(\psi_\circ)$. It follows that $\sqrt{n}F_{2,n}(x)$ is an envelope function for the class $\sqrt{n}\mathcal{F}_{1/\sqrt{n}}$. The function $\sqrt{n}F_{2,n}(x)$ satisfies the Lindeberg condition if:

$$\begin{aligned} n\mathbf{E}^P[F_{2,n}(X)^2] &< \infty, \\ n\mathbf{E}^P\left[F_{2,n}(X)^2 1\{\sqrt{n}F_{2,n}(X) > \varepsilon\sqrt{n}\}\right] &\to 0, \qquad \text{for every } \varepsilon > 0. \end{aligned}$$

Under Assumption 8 *(c)*, $n\mathbf{E}^P[F_{2,n}(X)^2] < \infty$ holds true. The second Lindeberg condition is easily satisfied since $n\mathbf{E}^P\left[F_{2,n}(X)^2 1\{\sqrt{n}F_{2,n}(X) > \varepsilon\sqrt{n}\}\right] \leq n\sqrt{\mathbf{E}^P\left[F_{2,n}(X)^4\right]}\sqrt{P\left(F_{2,n} > \varepsilon\right)}$ which converges to zero for every $\varepsilon > 0$ because $P\left(F_{2,n} > \varepsilon\right) \to 0$ as $n \to \infty$ and, under Assumption 8 *(c)*, $\mathbf{E}^P\left[F_{2,n}^4(X)\right] = O(1)$.

Finally, we verify the last requirement of Theorem E.1. Remark that under Assumption 8 *(c)*

$$\mathbf{E}^P[h'\dot{l}_{n,\psi_\circ}]^2 = \mathbf{E}^P[h'\dot{\mathfrak{L}}_{n,\psi_\circ}\dot{\mathfrak{L}}_{n,\psi_\circ}' h] + o(\|h\|) < \infty \quad \text{(E.4)}$$

because $\mathbf{E}^P tr(\dot{\mathfrak{L}}_{n,\psi_\circ}\dot{\mathfrak{L}}_{n,\psi_\circ}')$ can be shown to be bounded under Assumption 8 *(c)* by following the last part of the proof of Schennach (2007, Theorem 10). Moreover, by a first order Taylor expansion of $l_{n,\psi_\circ+h/\sqrt{n}}$ around h, by continuity of the map $\psi \mapsto l_{n,\psi}$, Assumption 6 *(b)* and (E.4), we have: $\mathbf{E}^P\left[\sqrt{n}\left(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}\right) - h'\dot{l}_{n,\psi_\circ}\right]^2 = o(1)$. Therefore, by Theorem E.1 we conclude that

$$\mathbb{G}_n\Big(\sqrt{n}(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ}) - h'\dot{l}_{n,\psi_\circ}\Big) \xrightarrow{p} 0$$

uniformly in h over a bounded set. Hence, by rewriting this as in (E.2) we see that

$$\sum_{i=1}^{n}(l_{n,\psi_\circ+h/\sqrt{n}}-l_{n,\psi_\circ})(x_i)-\mathbb{G}_n h'\dot{l}_{n,\psi_\circ}-n\mathbf{E}^P(l_{n,\psi_\circ+h/\sqrt{n}}-l_{n,\psi_\circ})=o_p(1). \tag{E.5}$$

By using (E.1) we obtain:

$$\begin{aligned}-\mathbb{G}_n h'\dot{l}_{n,\psi_\circ}-n\mathbf{E}^P(l_{n,\psi_\circ+h/\sqrt{n}}-l_{n,\psi_\circ}) &= -\sqrt{n}\mathbb{P}_n h'\dot{l}_{n,\psi_\circ}+\sqrt{n}\mathbf{E}^P h'\dot{l}_{n,\psi_\circ}-n\mathbf{E}^P(l_{n,\psi_\circ+h/\sqrt{n}}-l_{n,\psi_\circ})\\ &= -\sqrt{n}\mathbb{P}_n h'\dot{l}_{n,\psi_\circ}+\sqrt{n}\mathbf{E}^P[h'\dot{l}_{n,\psi_\circ}]-\sqrt{n}\mathbf{E}^P[h'\dot{l}_{n,\psi_\circ}]-\frac{1}{2}h'\mathbf{E}^P[\ddot{l}_{n,\psi_\circ}]h+o_p(1)\\ &= -\sqrt{n}\mathbb{P}_n h'\dot{l}_{n,\psi_\circ}-\frac{1}{2}h'\mathbf{E}^P[\ddot{l}_{n,\psi_\circ}]h+o_p(1) \end{aligned}\tag{E.6}$$

and by replacing this in (E.5) we get:

$$\sum_{i=1}^{n}(l_{n,\psi_\circ+h/\sqrt{n}}-l_{n,\psi_\circ})(x_i)-\frac{1}{\sqrt{n}}\sum_{i=1}^{n}h'\dot{l}_{n,\psi_\circ}(x_i)-\frac{1}{2}h'\mathbf{E}^P[\ddot{l}_{n,\psi_\circ}]h=o_p(1). \tag{E.7}$$

Because the $o_p(1)$ is uniform in h, this establishes (C.16) with $V_{\psi_\circ}=-\mathbf{E}^P[\ddot{\mathfrak{L}}_{n,\psi_\circ}]$ and $\Delta_{n,\psi_\circ}=\frac{1}{\sqrt{n}}\sum_{i=1}^{n}V_{\psi_\circ}^{-1}\dot{\mathfrak{L}}_{n,\psi_\circ}(x_i)$ if $V_{\psi_\circ}$ is nonsingular since $\ddot{l}_{n,\psi_\circ}=\ddot{\mathfrak{L}}_{n,\psi_\circ}+o_p(1)$ and $\dot{l}_{n,\psi_\circ}=\dot{\mathfrak{L}}_{n,\psi_\circ}+o_p(1)$.

□

Lemma E.1. *Denote* $\tau_n(\widehat{\lambda},\psi)=\mathbb{P}_n\left[e^{\widehat{\lambda}(\psi)'g^A(x,\psi)}\right]$, $\tau(\widehat{\lambda},\psi)=\mathbf{E}^P\left[e^{\widehat{\lambda}(\psi)'g^A(X,\psi)}\right]$ *and consider the class of functions* $\mathcal{F}_{1/\sqrt{n}}$ *defined in* (E.3) *where* $\lambda_\circ$ *and* $\psi_\circ$ *are as defined in* (2.16)*. Under Assumptions 5 (a)-(d), 3, 7 (a), (c), and 8, the bracketing integral* $J_{[]}(\delta_n,\sqrt{n}\mathcal{F}_{1/\sqrt{n}},L_2(P))$ *of the class* $\sqrt{n}\mathcal{F}_{1/\sqrt{n}}$ *converges to zero as* $\delta_n\to 0$*:*

$$J_{[]}(\delta_n,\sqrt{n}\mathcal{F}_{1/\sqrt{n}},L_2(P))=\int_0^{\delta_n}\sqrt{\log N_{[]}(\varepsilon\|F\|_{P,2},\sqrt{n}\mathcal{F}_{1/\sqrt{n}},L_2(P))}d\varepsilon\to 0 \tag{E.8}$$

as $\delta_n\to 0$.

Proof. The class $\mathcal{F}_{1/\sqrt{n}}$ is indexed by a compact subset $B^{\otimes}:=B\left(\tau(\lambda_1,\psi),\frac{1}{\sqrt{n}}\right)\times B(\lambda_\circ(\psi),\frac{1}{\sqrt{n}})\times B\left(\tau(\lambda_2,\psi_\circ),\frac{1}{\sqrt{n}}\right)\times B(\lambda_\circ(\psi_\circ),\frac{1}{\sqrt{n}})\times B(\psi_\circ,\frac{1}{\sqrt{n}})\subset\mathbb{R}^{p+d_v+2d+2}$. Theorem 10 in Schennach (2007), which is valid under Assumptions 5 *(a)-(c)*, 7 *(b)-(c)* and 8 *(a)-(c)*, shows that $\tau_n(\widehat{\lambda},\psi)\xrightarrow{p}\tau(\widehat{\lambda},\psi)$ and $\widehat{\lambda}(\psi)\xrightarrow{p}\lambda_\circ(\psi)$ at the rate $1/\sqrt{n}$, hence we can write $\tau_n(\psi)\in B(\tau(\widehat{\lambda}(\psi),\psi),1/\sqrt{n})$ and $\widehat{\lambda}(\psi)\in B(\lambda_\circ(\psi),1/\sqrt{n})$ with probability approaching 1. Therefore, $P\left(l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ} \in \mathcal{F}_{1/\sqrt{n}}\right) \to 1$. For every $f_a, f_b \in \mathcal{F}_{1/\sqrt{n}}$:

$$\begin{aligned}
&|f_a(x) - f_b(x)| = \\
&\left|\lambda_{1,a}' g^A(x,\psi_a) - \lambda_{2,a}' g^A(x,\psi_\circ) - \log(\tau_{1,a}/\tau_{2,a}) - \lambda_{1,b}' g^A(x,\psi_b) + \lambda_{2,b}' g^A(x,\psi_\circ) + \log(\tau_{1,b}/\tau_{2,b})\right| \\
&\quad \leq \|\lambda_{1,a}\| \, \|g^A(x,\psi_a) - g^A(x,\psi_b)\| + \|\lambda_{1,a} - \lambda_{1,b}\| \, \|g^A(x,\psi_b)\| + \|\lambda_{2,a} - \lambda_{2,b}\| \, \|g^A(x,\psi_\circ)\| \\
&\qquad - |\log\tau_{1,a} - \log\tau_{1,b}| + |\log\tau_{2,a} - \log\tau_{2,b}|.
\end{aligned}$$

The following results hold by compactness of $B^{\otimes}$ and continuity of $\psi \mapsto g^A(x,\psi)$ (under Assumption 5 *(c)*): *(i)* $\|\lambda_{1,a}\| \leq C$ for a generic constant $C > 0$ since $|\lambda_\circ(\psi)| < \infty$; *(ii)* $\|g^A(x,\psi_a) - g^A(x,\psi_b)\| \leq \|\partial g^A(x,\overline{\psi})/\partial\psi\| \, \|\psi_a - \psi_b\|$ for some $\overline{\psi}$ on the line joining ψ_a and ψ_b by the Mean Value theorem; *(iii)* $\|\lambda_{1,a} - \lambda_{1,b}\| \leq 2/\sqrt{n}$ because $\lambda_{1,a}, \lambda_{1,b} \in B(\lambda_\circ(\psi), 1/\sqrt{n})$; *(iv)* $|\log\tau_{1,a} - \log\tau_{1,b}| \leq |\tau_{1,a} - \tau_{1,b}|/\overline{\tau}_1$ for some $\overline{\tau}_1 > 0$ between $\tau_{1,a}$ and $\tau_{1,b}$ by the Mean Value Theorem (and similarly for $|\log\tau_{2,a} - \log\tau_{2,b}|$). By using all these results:

$$\begin{aligned}
|f_a(x) - f_b(x)| &\leq \left[C\|\partial g^A(x,\overline{\psi})/\partial\psi\| + 2\left(\|g^A(x,\psi_b)\| + \|g^A(x,\psi_\circ)\| + \overline{\tau}_1^{-1} + \overline{\tau}_2^{-1}\right)\right]\frac{1}{\sqrt{n}} \\
&\leq \left[(C + 2/\sqrt{n})b(x) + 2\left(\|g^A(x,\psi_\circ)\| + \overline{\tau}_1^{-1} + \overline{\tau}_2^{-1}\right)\right]\frac{1}{\sqrt{n}} =: F(x)\frac{1}{\sqrt{n}}
\end{aligned}$$

where the second inequality follows by the Mean Value Theorem applied to $\|g^A(x,\psi_\circ)\|$ and Assumption 8 *(c)* that implies that $\|\partial g^A(x,\psi)/\partial\psi\| \leq b(x)$ for every $\psi \in B(\psi_\circ, 1/\sqrt{n})$. Remark that under Assumptions 5 *(d)* and 8 *(c)*:

$$\mathbf{E}^P[F(x)^2] \leq 2(C + 2/\sqrt{n})^2\mathbf{E}^P[b(X)^2] + 16\mathbf{E}\|g^A(x,\psi_\circ)\|^2 + 16(\overline{\tau}_1^{-1} + \overline{\tau}_2^{-1}) < \infty.$$

Therefore, by example 19.7 in Van der Vaart (1998), there exists a constant K independent of ε and n such that the bracketing numbers of the class of functions $\mathcal{F}_{1/\sqrt{n}}$ satisfy

$$N_{[]}\left(\varepsilon\frac{1}{\sqrt{n}}\|F\|_{P,2}, \mathcal{F}_{1/\sqrt{n}}, L_2(P)\right) \leq K\left(\frac{diam\widetilde{B}}{\frac{\varepsilon}{\sqrt{n}}}\right)^{p+d_v+2d+2}, \qquad 0 < \varepsilon < \frac{1}{\sqrt{n}}$$

where $L_2(P)$ denotes the L_2 space of square integrable functions with respect to P and $\|\cdot\|_{P,2}$ denotes the norm in this space. Remark that $diam\widetilde{B} = 2/\sqrt{n}$ so that $\left(\frac{diam\widetilde{B}}{\frac{\varepsilon}{\sqrt{n}}}\right)^{p+d_v+2d+2} = (2/\varepsilon)^{p+d_v+2d+2}$. Then, the bracketing numbers of the class of functions $\sqrt{n}\mathcal{F}_{1/\sqrt{n}}$ satisfy

$$N_{[]}(\varepsilon\|F\|_{P,2}, \sqrt{n}\mathcal{F}_{1/\sqrt{n}}, L_2(P)) \leq K\,(2/\varepsilon)^{p+d_v+2d+2}, \qquad 0 < \varepsilon < \frac{1}{\sqrt{n}}. \tag{E.9}$$

Let us compute the bracketing integral of the class $\sqrt{n}\mathcal{F}_{1/\sqrt{n}}$:

$$\begin{aligned} J_{[]}(\delta_n, \sqrt{n}\mathcal{F}_{1/\sqrt{n}}, L_2(P)) &= \int_0^{\delta_n} \sqrt{\log N_{[]}(\varepsilon \|F\|_{P,2}, \sqrt{n}\mathcal{F}_{1/\sqrt{n}}, L_2(P))} d\varepsilon \\ &\leq \int_0^{\delta_n} \sqrt{\log K + \log{(2/\varepsilon)}^{p+d_v+2d+2}} d\varepsilon \to 0 \quad \text{ as } \delta_n \to 0 \end{aligned}$$

where the last inequality follows from (E.9) and proves the lemma.

□

Lemma E.2. *Denote* $\tau_n(\widehat{\lambda}, \psi) = \mathbb{P}_n\left[e^{\widehat{\lambda}(\psi)' g^A(x,\psi)}\right]$, $\tau(\widehat{\lambda}, \psi_\circ) = \mathbf{E}^P\left[e^{\widehat{\lambda}(\psi_\circ)' g^A(X,\psi_\circ)}\right]$ *and consider the class of functions* $\mathcal{F}_{1/\sqrt{n}}$ *defined in* (E.3) *where* $\lambda_\circ$ *and* $\psi_\circ$ *are as defined in* (2.16). *Under Assumptions 5 (a)-(c), 3, 7, and 8, the function*

$$F_{2,n}(x) := \frac{1}{\sqrt{n}}\left(C_{1,n} b(x) + C_{2,n}\|g^A(x,\psi_\circ)\| + \frac{1}{C_{4,n}} C_{3,n}\right).$$

is an envelope function for the class $\mathcal{F}_{1/\sqrt{n}}$*, where* $b(x)$ *is the function defined in Assumption 8 (c) and* $C_{i,n} > 0$*,* $i = 1, \ldots, 4$ *are sequences of positive and bounded constants that depend on* $\psi_\circ$ *and* $\lambda_\circ(\psi_\circ)$*.*

Proof. First, remark that every $f \in \mathcal{F}_{1/\sqrt{n}}$ satisfies

$$\begin{aligned} |f(x)| \leq & \sup_{\|\psi-\psi_\circ\| \leq n^{-1/2}} \sup_{\lambda \in B(\lambda_\circ(\psi), 1/\sqrt{n})} \|\lambda\| \|g^A(x,\psi) - g^A(x,\psi_\circ)\| \\ & + \sup_{\|\psi-\psi_\circ\| \leq n^{-1/2}} \sup_{\lambda_1 \in B(\lambda_\circ(\psi), 1/\sqrt{n})} \sup_{\lambda_2 \in B(\lambda_\circ(\psi_\circ), 1/\sqrt{n})} \|\lambda_1 - \lambda_2\| \|g^A(x,\psi_\circ)\| + \frac{|\tau_1 - \tau_2|}{\tau_2} \end{aligned} \tag{E.10}$$

for $\tau_1 \in B(\tau(\lambda_1, \psi), 1/\sqrt{n})$, $\tau_2 \in B(\tau(\lambda_2, \psi_\circ), 1/\sqrt{n})$, $\lambda_1 \in B(\lambda_\circ(\psi), 1/\sqrt{n})$ and $\lambda_2 \in B(\lambda_\circ(\psi_\circ), 1/\sqrt{n})$ since $\log \tau_2/\tau_1 \leq (\tau_2 - \tau_1)/\tau_1$. Next, we bound each of these terms separately.

Let $\sup_{\|\psi-\psi_\circ\| \leq n^{-1/2}} \sup_{\lambda \in B(\lambda_\circ(\psi), 1/\sqrt{n})} \|\lambda\| =: C_{1,n} < \infty$. Remark that by the implicit function theorem for vector valued functions applied to the first order condition for $\lambda_\circ$, the function $\psi \mapsto \lambda_\circ(\psi)$ is continuously differentiable in a neighborhood of $\psi_\circ$. Therefore, for every $\lambda_1 \in B(\lambda_\circ(\psi), 1/\sqrt{n})$, $\lambda_2 \in B(\lambda_\circ(\psi_\circ), 1/\sqrt{n})$ with $\psi \in B(\psi_\circ, 1/\sqrt{n})$, by the triangular inequality and the continuity of $\lambda_\circ$, there exists a N such that $\forall n \geq N$

$$\|\lambda_1 - \lambda_2\| \leq \|\lambda_1 - \lambda_\circ(\psi)\| + \|\lambda_\circ(\psi) - \lambda_\circ(\psi_\circ)\| + \|\lambda_2 - \lambda_\circ(\psi_\circ)\|$$

$$\leq \frac{2}{\sqrt{n}} + \left\| \frac{\partial \lambda_\circ(\overline{\psi})}{\partial \psi'} \right\| \|\psi - \psi_\circ\|$$

$$\leq \frac{1}{\sqrt{n}} \left(2 + \sup_{\psi \in B(\psi_\circ, \frac{1}{\sqrt{n}})} \left\| \frac{\partial \lambda_\circ(\psi)}{\partial \psi'} \right\| \right) =: \frac{1}{\sqrt{n}} C_{2,n}$$

where the second inequality follows from the Mean Value theorem with $\overline{\psi}$ being on the line joining ψ and $\psi_\circ$. Remark that $C_{2,n}$ is a sequence of positive and bounded constants which depends on $\psi_\circ$. By similar arguments we have that for every $\tau_1 \in B\left(\tau(\lambda_1, \psi), \frac{1}{\sqrt{n}}\right)$ and $\tau_2 \in B\left(\tau(\lambda_2, \psi_\circ), \frac{1}{\sqrt{n}}\right)$ with $\lambda_1 \in B(\lambda_\circ(\psi), 1/\sqrt{n})$, $\lambda_2 \in B(\lambda_\circ(\psi_\circ), 1/\sqrt{n})$ and $\psi \in B(\psi_\circ, 1/\sqrt{n})$:

$$\begin{aligned}
|\tau_1 - \tau_2| &\leq |\tau_1 - \tau(\lambda_1, \psi)| + |\tau(\lambda_1, \psi) - \tau(\lambda_2, \psi_\circ)| + |\tau_2 - \tau(\lambda_2, \psi_\circ)| \\
&\leq \frac{2}{\sqrt{n}} + \mathbf{E}^P \left[e^{\lambda_1' g^A(X,\psi)} \|\lambda_1\| \left\| \frac{\partial g^A(X, \tilde{\psi})}{\partial \psi} \right\| \right] \|\psi - \psi_\circ\| \\
&\quad + \mathbf{E}^P [e^{\tilde{\lambda}' g^A(X,\psi_\circ)} \|g^A(X, \psi_\circ)\|] \|\lambda_1 - \lambda_2\| \\
&\leq \frac{2}{\sqrt{n}} + \frac{C_{1,n}}{\sqrt{n}} \mathbf{E}^P \left[\sup_{\|\psi - \psi_\circ\| \leq 1/\sqrt{n}} \sup_{\lambda_1 \in B(\lambda_\circ(\psi), 1/\sqrt{n})} e^{\lambda_1' g^A(X,\psi)} b(X) \right] \\
&\quad + \mathbf{E}^P \left[\sup_{t \in (0,1)} \sup_{\psi \in B(\psi_\circ, 1/\sqrt{n})} \sup_{\lambda_1 \in B(\lambda_\circ(\psi), 1/\sqrt{n})} \sup_{\lambda_2 \in B(\lambda_\circ(\psi_\circ), 1/\sqrt{n})} e^{\tilde{\lambda}' g^A(X,\psi_\circ)} b(X) \right] \frac{C_{2,n}}{\sqrt{n}} \\
&=: \frac{1}{\sqrt{n}} C_{3,n}
\end{aligned}$$

where $\tilde{\psi}$ is between ψ and $\psi_\circ$, $\tilde{\lambda} = t\lambda_1 + (1-t)\lambda_2$, $t \in (0,1)$ and $C_{3,n}$ is a sequence of positive and bounded constants by Assumption 8 *(c)* which depends on $\psi_\circ$ and $\lambda_\circ$. Therefore, by this result and since $\log \tau_2/\tau_1 \leq (\tau_2 - \tau_1)/\tau_1$, τ_1 is uniformly bounded away from zero over a compact set: $|\log \tau_2/\tau_1| \leq \frac{1}{C_{4,n}\sqrt{n}} C_{3,n}$ for some strictly positive constant $0 < C_{4,n} < \infty$ that lower bounds τ_2 uniformly. Therefore, by replacing everything in (E.10) we get, $\forall f \in \mathcal{F}_{1/\sqrt{n}}$:

$$\begin{aligned}
|f(x)| &\leq C_{1,n} \left\| \partial g^A(x, \tilde{\psi}) / \partial \psi \right\| \|\psi - \psi_\circ\| + \frac{1}{\sqrt{n}} C_{2,n} \|g^A(x, \psi_\circ)\| + \frac{C_{3,n}}{C_{4,n}\sqrt{n}} \\
&\leq \frac{1}{\sqrt{n}} \left(C_{1,n} b(x) + C_{2,n} \|g^A(x, \psi_\circ)\| + \frac{1}{C_{4,n}} C_{3,n} \right) =: F_{2,n}(x)
\end{aligned}$$

where in the last inequality we have used Assumption 8 *(c)* that holds for every ψ in $B(\psi_\circ, 1/\sqrt{n})$. Therefore, $F_{2,n}(x)$ is an envelope function for the class $\mathcal{F}_{1/\sqrt{n}}$ and this concludes the proof.

$\square$

E.2 Proof of Theorem C.2

Define the events $A_{n,1} := \left\{\sup_{\psi\in\Psi_n^c} \frac{1}{n}\sum_{i=1}^n (l_{n,\psi}(x_i) - l_{n,\psi_\circ}(x_i)) \le -CM_n^2/n\right\}$ and

$$A_{n,2} := \left\{\int_\Psi \frac{p(x_{1:n}|\psi)}{p(x_{1:n}|\psi_\circ)}\pi(\psi)d(\psi) \ge e^{-CM_n^2/2}\right\}.$$

By (C.17), $P(A_{n,1}^c) \to 0$ and by Lemma E.3 below, $P(A_{n,2}^c) \to 0$. Therefore,

$$\begin{aligned}
\mathbf{E}^P\left[\pi\left(\sqrt{n}\|\psi-\psi_*\| > M_n \middle| x_{1:n}\right)\right] &\le \mathbf{E}^P\left[\pi\left(\sqrt{n}\|\psi-\psi_*\| > M_n \middle| x_{1:n}\right)\middle| A_{n,1}\cap A_{n,2}\right] \\
&\qquad \times P(A_{n,1}\cap A_{n,2}) + o(1) \\
&= \mathbf{E}^P\left[\left.\frac{\int_{\Psi_n^c}\frac{p(x_{1:n}|\psi)}{p(x_{1:n}|\psi_*)}\pi(\psi)d\psi}{\int_\Psi \frac{p(x_{1:n}|\psi)}{p(x_{1:n}|\psi_*)}\pi(\psi)d\psi}\right| A_{n,1}\cap A_{n,2}\right] P(A_{n,1}\cap A_{n,2}) + o(1) \\
&\le e^{-CM_n^2}\pi(\Psi_n^c)\mathbf{E}^P\left[\left.\left(\int_\Psi \frac{p(x_{1:n}|\psi)}{p(x_{1:n}|\psi_*)}\pi(\psi)d\psi\right)^{-1}\right| A_{n,1}\cap A_{n,2}\right] P(A_{n,1}\cap A_{n,2}) + o(1) \\
&\le e^{-CM_n^2}e^{CM_n^2/2}\pi(\Psi_n^c)P(A_{n,1}\cap A_{n,2}) + o(1) = o(1) \qquad \text{(E.11)}
\end{aligned}$$

which proves the result of the theorem.

$\square$

Lemma E.3. *Assume that the stochastic LAN expansion* (C.16) *holds for* $\psi_\circ$ *defined in* (2.16) *and that Assumptions 2 (a), 3, 4 and 8 are satisfied. Then,*

$$P\left(\int_\Psi \frac{p(x_{1:n}|\psi)}{p(x_{1:n}|\psi_\circ)}\pi(\psi)d\psi < a_n\right) \to 0 \qquad \text{(E.12)}$$

for every sequence $a_n \to 0$.

Proof. For a given $M > 0$ define $\mathfrak{C} = \{h \in \mathbb{R}^{d_v+p} : \|h\| \le M\}$. Denote by $h \mapsto Rem(h)$ the remaining term in (C.16) and remark that $\sup_{h\in\mathfrak{C}} Rem(h) \xrightarrow{p} 0$ by (C.16) and compactness of $\mathfrak{C}$. Therefore, for a sequence κ_n that converges to zero slowly enough, the event $B_n := \{\sup_{h\in\mathfrak{C}} Rem(h) \le \kappa_n\}$ has probability $P(B_n) \to 1$. Let $K_n \to \infty$. By considering the local parameter $h = \sqrt{n}(\psi - \psi_\circ)$ and by denoting by π^h both its prior distribution and prior

Lebesgue density (under Assumption 2 *(a)*), we upper bound the probability in (E.12) as follows:

$$P\left(\int_{\Psi}\frac{p(x_{1:n}|\psi)}{p(x_{1:n}|\psi_\circ)}\pi(\psi)d\psi < e^{-K_n^2}\right) \leq P\left(\int_{\mathfrak{C}}\frac{p(x_{1:n}|\psi_\circ + h/\sqrt{n})}{p(x_{1:n}|\psi_\circ)}\pi^h(h)dh < e^{-K_n^2}\right)$$
$$= P\left(\left\{\int_{\mathfrak{C}} e^{\sum_{i=1}^n (l_{n,\psi_\circ+h/\sqrt{n}} - l_{n,\psi_\circ})}\pi^h(h)dh < e^{-K_n^2}\right\}\cap B_n\right) + o_p(1). \quad \text{(E.13)}$$

By replacing the LAN expansion (C.16) and by noting that for n sufficiently large, $\kappa_n \leq \frac{1}{2}K_n^2$ on B_n and $\sup_{h\in\mathfrak{C}} h'V_{\psi_\circ}h \leq \sup_{h\in\mathfrak{C}}\|h\|^2\|V_{\psi_\circ}\| \leq M^2\|V_{\psi_\circ}\| \leq \kappa_n \leq \frac{1}{2}K_n^2$ (since $M^2\|V_{\psi_\circ}\|$ has the same order as $Rem(h)$ and where $\|V_{\psi_\circ}\|$ denotes the operator norm) we obtain:

$$P\left(\int_{\Psi}\frac{p(x_{1:n}|\psi)}{p(x_{1:n}|\psi_\circ)}\pi(\psi)d\psi < e^{-K_n^2}\right) \leq P\left(\int_{\mathfrak{C}} e^{h'V_{\psi_\circ}\Delta_{n,\psi_\circ}}\pi^h(h)dh < e^{-3K_n^2/4}\right) + o_p(1)$$
$$= P\left(\int_{\mathfrak{C}} e^{h'V_{\psi_\circ}\Delta_{n,\psi_\circ}}\pi^h(h|\mathfrak{C})dh < e^{-\log\pi^h(\mathfrak{C})}e^{-3K_n^2/4}\right) + o_p(1)$$
$$\leq P\left(\exp\left\{\int_{\mathfrak{C}} h'V_{\psi_\circ}\Delta_{n,\psi_\circ}\pi^h(h|\mathfrak{C})dh\right\} < e^{K_n^2/8}e^{-3K_n^2/4}\right) + o_p(1)$$
$$\leq P\left(\int_{\mathfrak{C}} h'V_{\psi_\circ}\Delta_{n,\psi_\circ}\pi^h(h|\mathfrak{C})dh < -5K_n^2/8\right) + o_p(1)$$
$$\leq \frac{64}{25K_n^4}E^P\left(\int_{\mathfrak{C}}\left(h'V_{\psi_\circ}\Delta_{n,\psi_\circ}\right)^2\pi^h(h|\mathfrak{C})dh\right) + o_p(1) \to 0 \quad \text{(E.14)}$$

where in the third line we have used that, for n large enough, $-\log\pi^h(\mathfrak{C}) \leq K_n^2/8$ 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 (E.4) and Assumption 4.

□

F Proof of the technical Lemmas for the proof of Theorems 3.1-3.2

For a vector z and a scalar $\delta > 0$ we denote by $B(z, \delta)$ the closed ball centred on z with radius δ. When the Mean Value theorem is applied to a vector of functions it must be understood that it is applied componentwise.

F.1 Proof of Lemma D.1

Let us consider the expression for the likelihood given in (2.6)-(2.9) and evaluated at $\widehat{\psi}^\ell$:

$$\log p(x_{1:n}|\widehat{\psi}^\ell; M_\ell) \;=\; -n\log n \,+\, \sum_{i=1}^n \widehat{\lambda}(\widehat{\psi}^\ell)' g^A(x_i,\widehat{\psi}^\ell) \,-\, n\log\frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\widehat{\psi}^\ell)' g^A(x_j,\widehat{\psi}^\ell)}. \quad \text{(F.1)}$$

To shorten notation, in the rest of this proof we eliminate the superscripts and subscripts and just write: g, $\widehat{\psi}$ instead of g^A and $\widehat{\psi}^\ell$.

Let $\tilde{\lambda}$ be on the line joining 0 and $\widehat{\lambda}(\widehat{\psi})$, then a second order Taylor expansion around $\widehat{\lambda} = 0$ gives

$$\frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\widehat{\psi})' g(x_j,\widehat{\psi})} = 1 + \frac{1}{n}\sum_{i=1}^n \widehat{\lambda}(\widehat{\psi})' g\left(x_i,\widehat{\psi}\right) + \frac{1}{2}\widehat{\lambda}(\widehat{\psi})'\frac{1}{n}\sum_{j=1}^n e^{\tilde{\lambda}' g(x_j,\widehat{\psi})} g\left(x_j,\widehat{\psi}\right) g\left(x_j,\widehat{\psi}\right)' \widehat{\lambda}(\widehat{\psi}). \quad \text{(F.2)}$$

Under Assumption 5 and because $\tilde{\lambda} = \mathcal{O}_p(n^{-1/2}) = o_p(n^{-\zeta})$ for any $\zeta < 1/2$ (since by Newey and Smith (2004, Lemma A.2) $\widehat{\lambda}(\widehat{\psi}) = \mathcal{O}_p(n^{-1/2}) = o_p(n^{-\zeta})$ for any $\zeta < 1/2$ and $\tilde{\lambda}$ is between 0 and $\widehat{\lambda}(\widehat{\psi})$) we can apply Newey and Smith (2004, Lemma A.1) that implies: $\max_{1\le i\le n} |\tilde{\lambda}' g(x_i,\widehat{\psi})| \xrightarrow{p} 0$. Therefore, $\max_{1\le i\le n}\left|-e^{\tilde{\lambda}' g(x_i,\widehat{\psi})} + 1\right| \xrightarrow{p} 0$ which in turn implies:

$$\frac{1}{n}\sum_{j=1}^n e^{\tilde{\lambda}' g(x_j,\widehat{\psi})} g\left(x_j,\widehat{\psi}\right) g\left(x_j,\widehat{\psi}\right)' \xrightarrow{p} \Delta. \quad \text{(F.3)}$$

By replacing this in (F.2) we obtain:

$$\frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\widehat{\psi})' g(x_j,\widehat{\psi})} = 1 + \frac{1}{n}\sum_{i=1}^n \widehat{\lambda}(\widehat{\psi})' g\left(x_i,\widehat{\psi}\right) + \frac{1}{2}\widehat{\lambda}(\widehat{\psi})'\Delta\widehat{\lambda}(\widehat{\psi}) + o_p\left(n^{-1}\right). \quad \text{(F.4)}$$

We now use the first order Taylor expansion of the function $\log(u)$ around $u = 1$: $\log(u) = u - 1 + o(|u-1|)$, and plug (F.4) in it to obtain:

$$\log\left(\frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\widehat{\psi})' g(x_j,\widehat{\psi})}\right) = \frac{1}{n}\sum_{i=1}^n \widehat{\lambda}(\widehat{\psi})' g\left(x_i,\widehat{\psi}\right) + \frac{1}{2}\widehat{\lambda}(\widehat{\psi})'\Delta\widehat{\lambda}(\widehat{\psi}) + o_p\left(n^{-1}\right) + o\left(\left|\frac{1}{n}\sum_{j=1}^n e^{\widehat{\lambda}(\widehat{\psi})' g(x_j,\widehat{\psi})} - 1\right|\right). \quad \text{(F.5)}$$

In order to simplify (F.5) further and to find the rate of the last term in the right hand side of (F.5) we approximate $\widehat{g}(\psi) := \frac{1}{n}\sum_{i=1}^{n} g(x_i, \psi)$ as follows:

$$\begin{aligned}\widehat{g}(\widehat{\psi}) =& \widehat{g}(\psi_*) + \frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(x_i,\psi_*)}{\partial \psi'}(\widehat{\psi}-\psi_*) + o(\|(\widehat{\psi}-\psi_*)\|) \\ =& \widehat{g}(\psi_*) + \Gamma_\ell(\widehat{\psi}-\psi_*) + o_p(n^{-1/2}) \qquad \text{(F.6)} \\ =& \widehat{g}(\psi_*) - \Gamma_\ell H \widehat{g}(\psi_*) + o_p(n^{-1/2}) \qquad \text{(F.7)} \\ =& -\Delta_\ell \widehat{\lambda}(\widehat{\psi}) + o_p(n^{-1/2}) \qquad \text{(F.8)}\end{aligned}$$

where to get (F.7) we have used the fact that, under Assumptions 1, 5 and 6: $\sqrt{n}\left(\widehat{\psi}-\psi_*\right) = -H\sqrt{n}\widehat{g}(\psi_*) + o_p(1)$ with $H := (\Gamma_\ell'\Delta_\ell^{-1}\Gamma_\ell)^{-1}\Gamma_\ell'\Delta_\ell^{-1}$ (see Schennach (2007, Proof of Theorem 3)) and to get (F.6) we have used the fact that $\|\frac{1}{n}\sum_{i=1}^{n}\frac{\partial g(x_i,\psi_*)}{\partial\psi'} - \Gamma_\ell\| = \mathcal{O}_p(n^{-1/2})$ under Assumption 6 (b) by the Markov's inequality. Finally, (F.8) is obtained by using the fact that $I - \Gamma_\ell H = \Delta_\ell \Phi_\ell$ where $\Phi_\ell := \Delta_\ell^{-1} - \Delta_\ell^{-1}\Gamma_\ell\Sigma_\ell\Gamma_\ell'\Delta_\ell^{-1}$ and $\Sigma_\ell = (\Gamma_\ell'\Delta_\ell^{-1}\Gamma_\ell)^{-1}$, and that, under Assumptions 1, 5 and 6, $\sqrt{n}\widehat{\lambda}(\widehat{\psi}) = -\Phi_\ell\sqrt{n}\widehat{g}(\psi_*) + o_p(1)$ (see Schennach (2007, Proof of Theorem 3)). By substituting this result in (F.4) we obtain:

$$\begin{aligned}\left|\frac{1}{n}\sum_{j=1}^{n} e^{\widehat{\lambda}(\widehat{\psi})' g(x_j,\widehat{\psi})} - 1\right| &= \left|-\widehat{\lambda}(\widehat{\psi})\Delta_\ell\widehat{\lambda}(\widehat{\psi}) + \frac{1}{2}\widehat{\lambda}(\widehat{\psi})'\Delta_\ell\widehat{\lambda}(\widehat{\psi})\right| + o_p(n^{-1}) \\ &= \left|-\frac{1}{2}\widehat{\lambda}(\widehat{\psi})'\Delta_\ell\widehat{\lambda}(\widehat{\psi})\right| + o_p(n^{-1})\end{aligned}$$

which is $\mathcal{O}_p(n^{-1})$ since $\left\|\widehat{\lambda}(\widehat{\psi})\right\|^2 = \mathcal{O}_p(n^{-1})$. By replacing this result and (F.8) in (F.5) we obtain:

$$\log\left(\frac{1}{n}\sum_{j=1}^{n} e^{\widehat{\lambda}(\widehat{\psi})' g(x_j,\widehat{\psi})}\right) = \widehat{\lambda}(\widehat{\psi})'\widehat{g}(\widehat{\psi}) + \frac{1}{2}\widehat{g}(\widehat{\psi})'\Delta_\ell^{-1}\widehat{g}(\widehat{\psi}) + o_p\left(n^{-1}\right). \qquad \text{(F.9)}$$

Next, we replace (F.9) in (F.1) to get:

$$\begin{aligned}\log p(x_{1:n}|\widehat{\psi}; M_\ell) = -n\log n + \sum_{i=1}^{n}\widehat{\lambda}(\widehat{\psi})' g(x_i,\widehat{\psi}) - \sum_{i=1}^{n}\widehat{\lambda}(\widehat{\psi})' g\left(x_i,\widehat{\psi}\right) \\ - n\frac{1}{2}\widehat{g}(\widehat{\psi})'\Delta_\ell^{-1}\widehat{g}(\widehat{\psi}) + o_p(1) \\ = -n\log n - \frac{\chi^2_{d-(p_\ell+d_{v_\ell})}}{2} + o_p(1) \qquad \text{(F.10)}\end{aligned}$$

where the last equality follows from standard arguments as in Hansen (1982) that show that $n\widehat{g}(\widehat{\psi})'\Delta_\ell^{-1}\widehat{g}(\widehat{\psi}) \to^d \chi^2_{d-(p_\ell+d_{v_\ell})}$.

$\square$

F.2 Proof of Lemma D.2

Result (2.15) and consistency of the ETEL estimator $\widehat{\psi}^\ell$ (which is guaranteed under Assumptions 1, 5 and 6, see Schennach (2007, Theorem 3)) implies that the posterior $\pi(\psi^\ell|x_{1:n}; M_\ell)$ of ψ^ℓ converges in total variation towards a $\mathcal{N}_{\widehat{\psi}^\ell, n^{-1}\Sigma_\ell}$ distribution, where $\Sigma_\ell = (\Gamma_\ell'\Delta^{-1}\Gamma_\ell)^{-1}$. Hence, the negative logarithm of the posterior density evaluated at $\widehat{\psi}^\ell$ is

$$-\log\pi(\widehat{\psi}^\ell|x_{1:n}; M_\ell) = -\frac{(p_\ell+d_\ell)}{2}[\log n - \log(2\pi)] + \frac{1}{2}\log|\Sigma_\ell| + o_p(1).$$

$\square$

F.3 Proof of Lemma D.3

In the proof we eliminate the superscript ℓ for simplicity. The proof proceeds in two steps. In the first step we show uniform convergence of $\widehat{g}^A(\psi)$ and $\widehat{\lambda}(\psi)$. By the uniform strong Law of Large Numbers, which is valid under Assumptions 5 *(a)-(b)* and 7 *(a)* (see *e.g.* Newey and McFadden (1994, Lemma 2.4)), it holds:

$$\sup_{\psi\in\Psi}\left\|\widehat{g}^A(\psi) - \mathbf{E}^P[g^A(x,\psi)]\right\| \xrightarrow{p} 0. \tag{F.11}$$

Let $\overline{\lambda}(\psi) = \arg\min_{\lambda\in\Lambda(\psi)}\frac{1}{n}\sum_{i=1}^n\exp\{\lambda' g^A(x_i,\psi)\}$. Schennach (2007, page 668) shows that $\sup_{\psi\in\Psi}\|\overline{\lambda}(\psi) - \lambda_\circ(\psi)\| \xrightarrow{p} 0$ (under Assumptions 5 *(a)-(b)* and 7). Moreover, if $\overline{\lambda}(\psi)$ lies in the interior of $\Lambda(\psi)$ then the minimum of $\frac{1}{n}\sum_{i=1}^n\exp\{\lambda' g^A(x_i,\psi)\}$ is unique by strict convexity of the latter function. As $\overline{\lambda}(\psi)\xrightarrow{p}\lambda_\circ(\psi)$ uniformly in $\psi\in\Psi$ and $\lambda_\circ(\psi)\in int(\Lambda(\psi))$ by Assumption 7 (b), it follows that $\widehat{\lambda}(\psi) - \overline{\lambda}(\psi)\xrightarrow{p} 0$ as $n\to\infty$. By continuity of both $\overline{\lambda}(\psi)$ and $\widehat{\lambda}(\psi)$ in (ψ) (due to the Birge's maximum theorem and strict convexity of $\frac{1}{n}\sum_{i=1}^n\exp\{\lambda' g^A(x_i,\psi)\}$ in λ) and compactness of Ψ we conclude that

$$\sup_{\psi\in\Psi}\|\widehat{\lambda}(\psi) - \lambda_\circ(\psi)\| \xrightarrow{p} 0. \tag{F.12}$$

In the second step of the proof, we use the results of the first step to show the result of the lemma. By the triangular inequality and the Cauchy-Schwartz inequality

$$\sup_{\psi\in\Psi}\left|\log\frac{\exp\{\widehat{\lambda}(\psi)'\widehat{g}^A(\psi)\}}{\frac{1}{n}\sum_{i=1}^n\exp\{\widehat{\lambda}(\psi)'g^A(x_i,\psi)\}}-\log\frac{\exp\{\lambda_\circ(\psi)'\mathbf{E}^P[g^A(x,\psi)]\}}{\mathbf{E}^P\left[\exp\{\lambda_\circ(\psi)'g^A(x,\psi)\}\right]}\right|$$
$$\leq\sup_{\psi\in\Psi}\|\widehat{\lambda}(\psi)-\overline{\lambda}(\psi)\|\sup_{\psi\in\Psi}\|\widehat{g}^A(\psi)\|$$
$$+\sup_{\psi\in\Psi}\|\overline{\lambda}(\psi)\|\sup_{\psi\in\Psi}\|\widehat{g}^A(\psi)-\mathbf{E}^P[g^A(x,\psi)]\|$$
$$+\sup_{\psi\in\Psi}\|\overline{\lambda}(\psi)-\lambda_\circ(\psi)\|\sup_{\psi\in\Psi}\|\mathbf{E}^P[g^A(x,\psi)]\|$$
$$+\sup_{\psi\in\Psi}\left|\log\frac{1}{n}\sum_{i=1}^n\exp\{\widehat{\lambda}(\psi)'g^A(x_i,\psi)\}-\log\mathbf{E}^P\left[\exp\{\lambda_\circ(\psi)'g^A(x,\psi)\}\right]\right|$$
$$=:\mathcal{A}_1+\mathcal{A}_2+\mathcal{A}_3+\mathcal{A}_4. \quad \text{(F.13)}$$

By continuity of $\psi\mapsto\overline{\lambda}(\psi)$ and compactness of Ψ and of $\Lambda(\psi)$ for all $\psi\in\Psi$: $\sup_{\psi\in\Psi}\|\overline{\lambda}(\psi)\|<\infty$. By Assumption 7 (b): $\sup_{\psi\in\Psi}\|\mathbf{E}^P[g^A(x,\psi)]\|\leq\mathbf{E}^P[M(x)]<\infty$ and $\sup_{\psi\in\Psi}\|\widehat{g}^A(\psi)\|<\infty$ (by using (F.11)). Therefore, $\mathcal{A}_i\xrightarrow{p}0$ for $i=1,2,3$ by (F.11) and (F.12).
In order to show the convergence to zero of $\mathcal{A}_4$ remark that because $\log(a)\leq a-1$ for every $a>0$, then,

$$\mathcal{A}_4\leq\sup_{\psi\in\Psi}\frac{\left|\frac{1}{n}\sum_{i=1}^n\exp\{\widehat{\lambda}(\psi)'g^A(x_i,\psi)\}-\mathbf{E}^P\left[\exp\{\lambda_\circ(\psi)'g^A(X,\psi)\}\right]\right|}{\mathbf{E}^P\left[\exp\{\lambda_\circ(\psi)'g^A(X,\psi)\}\right]}$$

which converges to zero by the result in Lemma F.1 below.

□

Lemma F.1. *Let M_ℓ be a misspecified model (that is, a model that does not satisfy Assumption 1) and let $g^A(x,\psi^\ell)$ and ψ^ℓ be the corresponding moment functions and parameters. Then, under Assumptions 5 (a)-(d), 3, and 7:*

$$\sup_{\psi\in\Psi}\left|\frac{1}{n}\sum_{i=1}^n\exp\{\widehat{\lambda}(\psi)'g^A(x_i,\psi)\}-\mathbf{E}^P\left[\exp\{\lambda_\circ(\psi)'g^A(X,\psi)\}\right]\right|\xrightarrow{p}0.$$

Proof. In this proof, let A_n denote the following event: $A_n:=\{\sup_{\psi\in\Psi}\|\widehat{\lambda}(\psi)-\lambda_\circ(\psi)\|\leq\delta_n\}$, for a $\delta_n>0$ converging to zero as $n\to\infty$, and let $B(\lambda_\circ,\delta_n)$ be the closed ball around $\lambda_\circ(\psi)$ with radius δ_n. Then, under Assumption 7 *(b)* there exists an $N>0$ such that $\forall n>N$:

$\widehat{\lambda}(\psi) \in \Lambda(\psi)$ on the event A_n.

Next, we prove the intermediate result

$$\sup_{\psi\in\Psi,\lambda\in\Lambda(\psi)} \left| \frac{1}{n}\sum_{i=1}^{n} e^{\lambda' g^A(x_i,\psi)} - \mathbf{E}^P\left[e^{\lambda' g^A(X,\psi)}\right]\right| \overset{a.s.}{\to} 0. \tag{F.14}$$

Consider the class of functions (on $\mathcal{X}$) $\mathcal{F} := \left\{\exp\{\lambda' g^A(\cdot,\psi)\}; \lambda\in\Lambda(\psi),\ \psi\in\Psi\right\}$. Since: (I) the function $(\lambda,\psi)\mapsto \exp\{\lambda' g^A(x,\psi)\}$ is continuous for P-almost all x (under Assumption 5 *(c)*); (II) Ψ is compact and $\Lambda(\psi)$ is compact for every $\psi\in\Psi$ (by Assumptions 5 *(b)* and 7 *(b)*); (III) the envelope of $\mathcal{F}$, $\sup_{\psi\in\Psi,\lambda\in\Lambda(\psi)}\exp\{\lambda' g^A(x,\psi)\}$ is in $L_1(P)$ (by Assumption 7 *(c)*), then (F.14) holds (see van de Geer (2010, Lemma 3.10, page 38)).

With this result, and the fact that $P(A_n^c) = o(1)$ by (F.12), we are now ready to show the result of the lemma. Let $h(\widehat{\lambda},\psi) := \mathbf{E}_X^P\left[e^{\widehat{\lambda}(\psi)' g^A(X,\psi)}\right]$, where $\mathbf{E}_X^P[\cdot]$ denotes the expectation taken with respect to the distribution of X only (so, we do not integrate out $\widehat{\lambda}$). Moreover, let $\eta > 0$ and denote by B_n the event $B_n := \{\sup_{\psi\in\Psi}\left|h(\widehat{\lambda},\psi) - \mathbf{E}^P\left[e^{\lambda_\circ(\psi)' g^A(X,\psi)}\right]\right| \leq \eta/2\}$. To upper bound $P(B_n^c)$ we use the Markov's inequality and the Mean value theorem applied to the function $h(\cdot,\psi)$ which is defined on $B(\lambda_\circ(\psi),\delta_n)$ on the event A_n:

$$\begin{aligned} P(B_n^c) &= P(B_n^c|A_n)P(A_n) + P(B_n^c|A_n^c)P(A_n^c) \\ &\leq \frac{2}{\eta}\mathbf{E}^P\left(\sup_{\psi\in\Psi}\left|\mathbf{E}_X^P\left[e^{\{\tilde{\lambda}(\psi)' g^A(X,\psi)\}} g^A(X,\psi)'\right](\widehat{\lambda}(\psi) - \lambda_\circ(\psi))\right| \,\Big|\, A_n\right) P(A_n) + P(A_n^c) \\ &= \frac{2\delta_n}{\eta}(\mathbf{E}^P\left(\left(\sup_{\psi\in\Psi}\left\|\mathbf{E}_X^P\left[e^{\{\tilde{\lambda}(\psi)' g^A(X,\psi)\}} g^A(X,\psi)'\right]\right\|\right)^2 1_{A_n}\right))^{1/2} + o(1) \end{aligned} \tag{F.15}$$

where $\tilde{\lambda}(\psi)$ is on the line joining $\widehat{\lambda}(\psi)$ and $\lambda_\circ(\psi)$ and 1_{A_n} is the indicator function of the event A_n. By applying the Cauchy-Schwartz inequality we get

$$\begin{aligned} &\mathbf{E}^P\left(\sup_{\psi\in\Psi}\left\|\mathbf{E}_X^P\left[e^{\tilde{\lambda}(\psi)' g^A(X,\psi)} g^A(X,\psi)\right]\right\|^2 1_{A_n}\right) \\ &\qquad \leq \mathbf{E}^P\left(\sup_{\psi\in\Psi}\mathbf{E}_X^P\left[e^{2\tilde{\lambda}(\psi)' g^A(X,\psi)}\right] 1_{A_n}\right)\sup_{\psi\in\Psi}\mathbf{E}^P\left\|g^A(X,\psi)\right\|^2. \end{aligned} \tag{F.16}$$

The first term in the product in the right hand side is bounded by uniform convergence of $\tilde{\lambda}(\psi)$ towards $\lambda_\circ(\psi)$ on A_n, uniform continuity of the function $\psi\mapsto e^{\lambda' g^A(\psi)}$ on $B(\lambda_\circ(\psi),2\delta_n)\times\Psi$, Assumption 7 *(c)* and by the Dominated Convergence theorem. The second term in the product is bounded by Assumption 5 *(d)* because

$$\sup_{\psi\in\Psi}\mathbf{E}_X^P\left\|g^A(X,\psi)\right\|^2 \leq \mathbf{E}^P\left[\sup_{\psi\in\Psi}\left\|g^A(X,\psi)\right\|^2\right] \leq \mathbf{E}^P\left[\sup_{\psi\in\Psi}\left\|g^A(X,\psi)\right\|^\alpha\right], \qquad \forall\alpha > 2.$$

Therefore, (F.15) and (F.16) and the fact that $\delta_n \to 0$ show that $P(B_n^c) = o(1)$.

Next,

$$P\left(\sup_{\psi\in\Psi}\left|\frac{1}{n}\sum_{i=1}^n e^{\widehat{\lambda}(\psi)' g^A(x_i,\psi)} - \mathbf{E}^P\left[e^{\lambda_\circ(\psi)' g^A(X,\psi)}\right]\right| > \eta\right) \leq$$

$$P\left(\sup_{\psi\in\Psi}\left|\frac{1}{n}\sum_{i=1}^n e^{\widehat{\lambda}(\psi)' g^A(x_i,\psi)} - h(\widehat{\lambda},\psi)\right| > \eta - \sup_{\psi\in\Psi}\left|h(\widehat{\lambda},\psi) - \mathbf{E}^P\left[e^{\widetilde{\lambda}(\psi)' g^A(X,\psi)}\right]\right|\right)$$

$$\leq P\left(\sup_{\psi\in\Psi}\left|\frac{1}{n}\sum_{i=1}^n e^{\widehat{\lambda}(\psi)' g^A(x_i,\psi)} - h(\widehat{\lambda},\psi)\right| > \eta - \frac{\eta}{2}\right) P(B_n) + P(B_n^c)$$

$$\leq P\left(\sup_{\psi\in\Psi}\left|\frac{1}{n}\sum_{i=1}^n e^{\widehat{\lambda}(\psi)' g^A(x_i,\psi)} - h(\widehat{\lambda},\psi)\right| > \frac{\eta}{2}\middle| A_n\right) P(A_n)P(B_n) + P(A_n^c)P(B_n) + P(B_n^c)$$

$$\leq P\left(\sup_{\psi\in\Psi,\lambda\in\Lambda(\psi)}\left|\frac{1}{n}\sum_{i=1}^n e^{\lambda' g^A(x_i,\psi)} - \mathbf{E}^P\left[e^{\lambda' g^A(X,\psi)}\right]\right| > \frac{\eta}{2}\right) + o(1) \quad \text{(F.17)}$$

where to get the last line we have used the fact that the probability is conditional on the event A_n and $P(A_n^c) = o(1)$. Finally, this probability goes to zero by (F.14).

□

References

Chib, S. (1995), 'Marginal Likelihood from the Gibbs Output', *Journal of the American Statistical Association* **90**(432), 1313–1321.

Kleijn, B. and van der Vaart, A. (2012), 'The Bernstein-Von-Mises Theorem Under Misspecification', *Electronic Journal Statistics* **6**, 354–381.

Newey, W. K. and McFadden, D. (1994), Chapter 36: Large sample estimation and hypothesis testing, Vol. 4 of *Handbook of Econometrics*, Elsevier, pp. 2111 – 2245.

Newey, W. K. and Smith, R. J. (2004), 'Higher Order Properties of GMM and Generalized Empirical Likelihood Estimators', *Econometrica* **72**(1), 219–255.

Schennach, S. M. (2007), 'Point Estimation with Exponentially Tilted Empirical Likelihood', *Annals of Statistics* **35**(2), 634–672.

van de Geer, S. A. (2010), *Empirical Processes in M-Estimation*, Cambridge Series in Statistical and Probabilistic Mathematics.

Van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge Series in Statistical and Probabilistic Mathematics.

Van der Vaart, A. W. (2002), Semiparametric Statistics, *in* P. Bernard, ed., 'Lectures on Probability Theory and Statistics - Ecole d'Eté de Probabilités de Saint-Flour XXIX - 1999', Vol. 1781 of *Lecture Notes in Mathematics*, Springer-Verlag Berlin Heidelberg.