universität wien

DISSERTATION

Titel der Dissertation

Efficiency in Indirect Inference

Verfasser

Mag. rer. nat. Mag. rer. soc. oec. Florian Gach Bakk. rer. soc. oec.

angestrebter akademischer Grad

Doktor der Sozial- und Wirtschaftswissenschaften
(Dr. rer. soc. oec.)

Wien, im Jänner 2010

Studienkennzahl lt. Studienblatt:	A 084 136
Studienrichtung lt. Studienblatt:	Doktorat der Sozial- und Wirtschaftswissenschaften
Betreuer:	O. Univ. Prof. Dr. Benedikt Pötscher

To Benedikt M. Pötscher

Contents

1 Introduction

Let $X_1, \dots, X_n$ be i.i.d. random variables, and let $\mathcal{P}_\Theta$ be a parametric model consisting of probability density functions p_θ that are indexed by some set $\Theta \subseteq \mathbb{R}^m$. Assume, for the moment, that $\mathcal{P}_\Theta$ is correctly specified, meaning that there is a unique parameter $\theta_0 \in \Theta$ such that p_{θ_0} characterizes the distribution of $X_1, \dots, X_n$. Suppose the likelihood function is not explicitly available, but it is possible to simulate from any p_θ (for a typical situation see (1) below). We then may resort to the method of indirect inference as an alternative to maximum likelihood estimation. The idea underlying this simulation-based method can be outlined as follows (see also the survey paper of Jiang & Turnbull (2004) who give a general overview including applications in biostatistics and epidemiology): For each $\theta \in \Theta$, simulate data $X_1(\theta), \dots, X_{k(n)}(\theta)$ according to the density p_θ, and observe that the vectors $(X_1, \dots, X_n)$ and $(X_1(\theta_0), \dots, X_{k(n)}(\theta_0))$ are 'close' with large probability. This suggests to estimate θ_0 through minimizing some distance between $(X_1, \dots, X_n)$ and $(X_1(\theta_0), \dots, X_{k(n)}(\theta_0))$, or statistics thereof. One possibility is to match sample moments of $(X_1, \dots, X_n)$ and $(X_1(\theta_0), \dots, X_{k(n)}(\theta_0))$; see McFadden (1989). Another possibility is to match auxiliary maximum likelihood estimators (i.e., maximum likelihood estimators based on auxiliary parametric models that allow doing maximum likelihood estimation) from the original and the synthetic data. This is the method of indirect inference as introduced in the literature (in the context of time series) by A. Smith in his Ph.D. dissertation (Smith, 1990) and later published in Smith (1993). A slightly more general setup than the one in Smith (1993), allowing for more general objective functions than auxiliary log-likelihood functions (and including exogeneous variables), is given in Gouriéroux, Monfort & Renault (1993). [It is this paper's title, 'Indirect Inference,' that gave the method its name.] There the observations are explained by

$$X_i = g(X_{i-1}, Z_i, U_i, \theta_0) \quad \text{with initial condition } X_0, \tag{1}$$

where the Z_i are observed exogenous variables, the U_i are unobserved disturbances, θ_0 is the unknown true parameter, and g is some known function. Typically, neither an explicit nor computable form of the densities p_θ is available for data as in (1), but simulated data are easily obtained by first simulating disturbances U_i and then computing $X_i(\theta)$ iteratively via $X_i(\theta) = g(X_{i-1}(\theta), Z_i, U_i, \theta)$ with starting condition $X_0(\theta) = X_0$.

To explain in detail how parametric auxiliary models are used, suppose for simplicity that the data $X_1, \dots, X_n$ are i.i.d. and $\mathcal{P}_B$ is a parametric auxiliary model consisting of probability density functions p_β that are indexed by some set $B \subseteq \mathbb{R}^q$. The auxiliary model may or may not be correctly specified. The auxiliary maximum likelihood estimator given $X_1, \dots, X_n$ is defined as the maximizer of the (scaled) auxiliary log-likelihood function

$$L_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \log p_\beta(X_i)$$

over B, and is denoted by $\hat{\beta}_n$. Similarly, the auxiliary maximum likelihood estimator given the simulated data $X_1(\theta), \dots, X_{k(n)}(\theta)$ is defined as the maximizer of

$$L_{k(n)}(\theta, \beta) = \frac{1}{k(n)} \sum_{i=1}^{k(n)} \log p_\beta(X_i(\theta))$$

over B, and is denoted by $\tilde{\beta}_{k(n)}(\theta)$. By an indirect inference estimator $\hat{\theta}_{n,k(n)}$ we mean any minimizer of the weighted distance

$$(\hat{\beta}_n - \tilde{\beta}_{k(n)}(\theta))' W_n (\hat{\beta}_n - \tilde{\beta}_{k(n)}(\theta))$$

over Θ with appropriate weighting matrix W_n. The following asymptotic properties of $\hat{\theta}_{n,k(n)}$ are well-known in the literature (see, e.g., Gouriéroux & Monfort, 1996): First, $\hat{\theta}_{n,k(n)}$ is consistent if the binding function is injective in a neighbourhood of the true parameter θ_0. The binding function is defined by $\beta(\theta) = \arg\min_{\beta \in B} d_{\mathrm{KL}}(p_\theta, p_\beta)$, where d_{KL} denotes the Kullback-Leibler distance, and relates Θ to B. [In applications it is typically hard to verify the injectivity of the binding function; see Lombardi & Calzolari (2008) who address this issue in their paper.] Second, if $\hat{\theta}_{n,k(n)}$ is consistent and $k(n)/n$ converges to some κ with $0 < \kappa \leq \infty$, then (under certain regularity conditions) $\sqrt{n}(\hat{\theta}_{n,k(n)} - \theta_0)$ converges weakly to the law $N(0, (1+\kappa^{-1})\Sigma(\theta_0))$, where $\Sigma(\theta_0)$ is a positive definite matrix. Furthermore, $\Sigma(\theta_0)$ coincides with the Cramér-Rao bound under the assumption that the parametric auxiliary model is correctly specified. This central result can be found as Proposition 3 in Smith (1993) and Proposition 3 in Gouriéroux, Monfort & Renault (1993). The proof for the case $\kappa = \infty$ is included, although not made explicit, in both papers. We see that $\hat{\theta}_{n,k(n)}$ is asymptotically efficient if the number of simulations satisfies $\lim_{n\to\infty} k(n)/n = \infty$ *and* the parametric auxiliary model is *correctly specified*. Since $k(n)$ can be chosen freely, the former condition can easily be satisfied. The latter one, though, is restrictive and led Gallant & Long (1997) to consider a sieve of parametric auxiliary models whose union is assumed to be correctly specified, an assumption that seems to be more plausible. They claim that their simulation-based estimator is asymptotically efficient, but they leave aside the simulation step in the formulation as well as in the proof of their main result (cf. Theorem 2 there). This is an unjustified simplification since choosing the number of simulations to depend on the sample size in the right way is essential to obtain asymptotic normality results as stated above.

In this thesis, we will develop the theory of indirect inference using *non-parametric* auxiliary models. The underlying idea is to consider a non-parametric auxiliary model that is large enough insofar that it is plausible to assume it contains $\mathcal{P}_\Theta$. This has two important consequences: First, the binding function will coincide with the parametrization $\theta \mapsto p_\theta$, and its injectivity around θ_0 then follows from the usual identifiability condition for parametric models. Second, and more importantly, the auxiliary model will be correctly specified (since the given model is so by hypothesis), allowing us to show that the resulting indirect inference estimator is asymptotically efficient. For related results that have been obtained in a parallel development to this thesis, and that make use of spline density estimators instead of non-parametric maximum likelihood estimators as auxiliary statistics, we refer the reader to Nickl & Pötscher (2009). Related results using empirical characteristic functions are given in Carrasco, Chernov, Florens & Ghysels (2007), whereas Altissimo & Mele (2009) use kernel density estimators. We note, however, that the proofs of the main results in Carrasco, Chernov, Florens & Ghysels (2007) are highly incomplete and those in Altissimo & Mele (2009) are fundamentally incorrect.

The thesis is organized as follows. Section 2 introduces notation that is used throughout the text and defines Sobolev spaces upon which the non-parametric auxiliary models are based. Section 3 provides the general setup for all subsequent investigations. Section 4 lists a catalogue of assumptions on the parametric model $\mathcal{P}_\Theta$, on the way of simulating data, and on the auxiliary model $\mathcal{P}$. In Section 5 we follow the basic idea of indirect inference outlined above. We define the auxiliary maximum likelihood estimator given $X_1, \ldots, X_n$ as the maximizer of the (scaled) auxiliary log-likelihood function

$$L_n(p) = \frac{1}{n} \sum_{i=1}^{n} \log p(X_i)$$

over $\mathcal{P}$, and denote it by $\hat{p}_n$. Similarly, we define the auxiliary maximum likelihood estimator given the simulated data $X_1(\theta), \ldots, X_{k(n)}(\theta)$ as the maximizer of

$$L_{k(n)}(\theta, p) = \frac{1}{k(n)} \sum_{i=1}^{k(n)} \log p(X_i(\theta))$$

over $\mathcal{P}$, and denote it by $\tilde{p}_{k(n)}(\theta)$. Theorem 6 shows that these maximizers exist and are unique. They converge at rates slower than $k(n)^{-1/2}$ in certain Sobolev norms (Theorem 16) and at rate $k(n)^{-1/2}$ in the weak sense (Theorem 17).

By an indirect inference estimator $\hat{\theta}_{n,k(n)}$ we mean any minimizer of the weighted L^2-distance

$$\int (\hat{p}_n - \tilde{p}_{k(n)}(\theta))^2 \frac{1}{\hat{p}_n} d\lambda \tag{2}$$

over the set Θ, where λ denotes the Lebegue measure. The main result in Section 6 can be found in Theorem 31, which states that $\hat{\theta}_{n,k(n)}$ is asymptotically efficient (in the sense that it is asymptotically normal with variance-covariance matrix equal to the Cramér-Rao bound) if the number of simulations $k(n)$ is of order $n^{2+\alpha}$ for some $\alpha > 0$. The classical way to prove such a result would be to apply the mean-value theorem to the derivative of the indirect inference objective function given in (2). Since we do not know whether $\tilde{p}_{k(n)}(\theta)$ is differentiable in θ, we split the proof into two steps instead. In the first step, the problem of finding the asymptotic distribution of $\hat{\theta}_{n,k(n)}$ is reduced to finding the one of a related (though infeasible) estimator $\hat{\theta}_n$, which is defined as minimizer of the weighted L^2-distance

$$\int (\hat{p}_n - p_\theta)^2 \frac{1}{\hat{p}_n} d\lambda \tag{3}$$

over Θ. This is possible due to Lemma 30, which allows to take advantage of the closeness of the objective functions in (2) and (3) to conclude that $\sqrt{n}(\hat{\theta}_{n,k(n)} - \hat{\theta}_n)$ converges to 0 in outer probability. We point out that establishing closeness of the above objective functions requires knowledge about the asymptotic behaviour of the stochastic process $\sqrt{k(n)} \int_\Omega (\tilde{p}_{k(n)}(\theta) - p_\theta) f d\lambda$, $(\theta, f) \in \Theta \times \mathcal{F}$, for specific classes $\mathcal{F}$ of functions, and is by no means trivial. In the second step, we proceed along the lines of the classical proof. This time we can just require the parametrization $\theta \mapsto p_\theta$ to be sufficiently often differentiable in θ. We note that $\hat{\theta}_n$ is related to the minimum Hellinger distance estimator introduced in Beran (1977), which has been shown to be asymptotically efficient and at the same time robust against small perturbations of $\mathcal{P}_\Theta$.

The aim of Section 7 is to underline the statistical significance of the asymptotic efficiency result for indirect inference estimators: In point of fact, Theorem 39 states that $\hat{\theta}_{n,k(n)}$ is asymptotically efficient under convergent sequences of parameters.

Section 8 suggests how $\hat{p}_n$ can be computed in principle: Theorem 40 states that $\hat{p}_n$ (asymptotically) is the solution of a finite-dimensional system of equations in n variables.

In the appendix we gather properties of the relevant objective functions, high-level theorems, and auxiliary results that are used throughout the text.

Note that misspecification of $\mathcal{P}_\Theta$ is treated from the beginning. Misspecification in the context of indirect inference is the main issue in Dridi, Guay & Renault (2007). There the (misspecified) model is meant to capture some parametric aspect of the true data generating process, which can be estimated consistently if the auxiliary model is encompassed by the given model.

The main findings of this thesis are collected in the following theorems: Theorem 6 states the uniqueness of the non-parametric density estimator $\hat{p}_n$, and Theorem 40 captures its computability; Theorem 31 establishes that any indirect inference estimator is asymptotically normal and, under correct specification of the underlying parametric model, asymptotically efficient; Theorem 39 does so under convergent sequences of parameters.

2 Preliminaries and notation

We will use the set of natural numbers $\mathbb{N} = \{1, 2, \ldots\}$ and will denote by $\|\cdot\|$ the 2-norm on Euclidean space. For two real-valued functions f, g on $(0,\infty)$, we will write $f(\varepsilon) \lesssim g(\varepsilon)$ if there is a constant C, $0 < C < \infty$, such that $f(\varepsilon) \leq Cg(\varepsilon)$ holds true for all $\varepsilon > 0$.

For X a non-empty set and f a real-valued function on X, define the sup-norm of f by

$$\|f\|_X = \sup_{x\in X} |f(x)|$$

and $\ell^\infty(X)$ as the Banach space of all bounded functions on X, equipped with $\|\cdot\|_X$. For $(S,\mathcal{A})$ some measurable space, let $\mathcal{L}^0(S,\mathcal{A})$ denote the vector space of all $\mathcal{A}$-measurable real-valued functions on S. Let $\mathsf{L}^\infty(S,\mathcal{A})$ denote the space of those $f \in \mathcal{L}^0(S,\mathcal{A})$ that satisfy $\|f\|_S < \infty$. Note that any subset $\mathcal{F}$ of $\mathsf{L}^\infty(S,\mathcal{A})$ is a metric space with respect to the metric given by $(f,g) \mapsto \|f-g\|_S$, and as such will be denoted by $(\mathcal{F}, \|\cdot\|_S)$. For $f \in \mathcal{L}^0(S,\mathcal{A})$ and μ a non-negative measure on $(S,\mathcal{A})$, define

$$\|f\|_{2,\mu} = \left[\int_S f^2\, d\mu\right]^{1/2}$$

and $\mathcal{L}^2(S,\mathcal{A},\mu)$ as the space of those $f \in \mathcal{L}^0(S,\mathcal{A})$ that satisfy $\|f\|_{2,\mu} < \infty$. For any metric space (T,d), we denote by $\mathcal{B}(T,d)$ its Borel σ-algebra and by $\mathsf{C}(T,d)$ the Banach space of all bounded, d-continuous real-valued functions on T, equipped with the sup-norm.

For Ω some non-empty Borel set in $\mathbb{R}$, let $\mathcal{B}(\Omega)$ be the trace of the Borel σ-algebra of $\mathbb{R}$ on Ω, and let λ be the Lebesgue measure on $(\Omega,\mathcal{B}(\Omega))$. We will suppress $\mathcal{B}(\Omega)$ as well as λ in the notation introduced above, so that $\mathcal{L}^0(\Omega,\mathcal{B}(\Omega))$ is replaced by $\mathcal{L}^0(\Omega)$, $\mathcal{L}^2(\Omega,\mathcal{B}(\Omega),\lambda)$ by $\mathcal{L}^2(\Omega)$, $\mathsf{L}^\infty(\Omega,\mathcal{B}(\Omega))$ by $\mathsf{L}^\infty(\Omega)$, and $\|\cdot\|_{2,\lambda}$ by $\|\cdot\|_2$. Further, we will write a.e. instead of λ-a.e. For any non-empty subset Ω of $\mathbb{R}$, we denote by $\mathsf{C}(\Omega)$ the Banach space of all bounded, continuous real-valued functions on Ω, equipped with the sup-norm.

2.1 Hölder spaces

For Ω a non-empty open subset of $\mathbb{R}$, $f : \Omega \to \mathbb{R}$, and $s \geq 0$, define

$$\|f\|_{s,\Omega} = \begin{cases} \sum_{0\leq\alpha\leq\lfloor s\rfloor} \|D^\alpha f\|_\Omega + \sup\left\{\frac{|D^{\lfloor s\rfloor}f(x)-D^{\lfloor s\rfloor}f(y)|}{|x-y|^{s-\lfloor s\rfloor}} : x \neq y\right\} & \text{if } s \text{ is non-integer,}\\ \sum_{0\leq\alpha\leq s} \|D^\alpha f\|_\Omega & \text{otherwise.}\end{cases}$$

Here, $D^\alpha f$ denotes the classical derivative of f of order α, and $\lfloor s\rfloor$ denotes the integer part of s. For any non-integer $s > 0$, let $\mathsf{C}^s(\Omega)$ be the space of all $f : \Omega \to \mathbb{R}$ such that $\|f\|_{s,\Omega} < \infty$; for any integer $s \geq 0$, let $\mathsf{C}^s(\Omega)$ be the space of all $f : \Omega \to \mathbb{R}$ such that $\|f\|_{s,\Omega} < \infty$ and $D^s f$ is uniformly continuous.

2.2 Sobolev spaces

For Ω a non-empty open subset of $\mathbb{R}$, $s \geq 0$, and $f, g : \Omega \to \mathbb{R}$, let

$$\langle f|g\rangle_{s,2} = \begin{cases} \sum_{0\leq\alpha\leq\lfloor s\rfloor} \langle D_w^\alpha f | D_w^\alpha g\rangle_2 \\ \qquad + \int_\Omega \int_\Omega \frac{(D_w^{\lfloor s\rfloor} f(x) - D_w^{\lfloor s\rfloor} f(y))(D_w^{\lfloor s\rfloor} g(x) - D_w^{\lfloor s\rfloor} g(y))}{|x-y|^{1+2(s-\lfloor s\rfloor)}} d\lambda(x) d\lambda(y) \\ \hfill \text{if } s \text{ is non-integer,} \\ \sum_{0\leq\alpha\leq s} \langle D_w^\alpha f | D_w^\alpha g\rangle_2 \\ \hfill \text{otherwise,} \end{cases}$$

and $\|f\|_{s,2} = \sqrt{\langle f|f\rangle_{s,2}}$. Here, $D_w^\alpha f$ denotes the weak derivative of f of order α, $\lfloor s\rfloor$ denotes the integer part of s, and $\langle \cdot | \cdot \rangle_2$ is the usual (non-negative definite) inner product on $\mathcal{L}^2(\Omega)$. We define $\mathcal{W}_2^s(\Omega)$ as the space of all $f : \Omega \to \mathbb{R}$ such that $\|f\|_{s,2}$ is finite. For $s > 1/2$ and Ω a non-empty bounded, open interval in $\mathbb{R}$, each $f \in \mathcal{W}_2^s(\Omega)$ is a.e. equal to exactly one bounded, continuous function on Ω. For $s > 1/2$, we consequently define the Sobolev space $\mathsf{W}_2^s(\Omega) = \mathcal{W}_2^s(\Omega) \cap \mathsf{C}(\Omega)$ and note that it is a Hilbert space. The Sobolev balls $\{f \in \mathsf{W}_2^s(\Omega) : \|f\|_{s,2} \leq B\}$ of radius B, $0 < B < \infty$, will be denoted by $\mathcal{U}_{s,B}$. The next proposition collects some properties of Sobolev spaces.

Proposition 1. *Let Ω be a non-empty bounded, open interval in $\mathbb{R}$.*

1. *For $s > 1/2$, the Sobolev space $\mathsf{W}_2^s(\Omega)$ is a multiplication algebra, meaning that there is a constant $M_s > 0$ such that*

$$\|fg\|_{2,s} \leq M_s \|f\|_{2,s} \|g\|_{2,s}$$

   holds true for all $f, g \in \mathsf{W}_2^s(\Omega)$.

2. *For $s > 1/2$, the Sobolev space $\mathsf{W}_2^s(\Omega)$ is embedded in $\mathsf{C}^{s-1/2}(\Omega)$. Consequently, $\mathsf{W}_2^s(\Omega)$ is embedded in $\mathsf{C}(\Omega)$ with embedding constant $C_s > 0$, that is,*

$$\|f\|_\Omega \leq C_s \|f\|_{s,2}$$

   holds true for all $f \in \mathsf{W}_2^s(\Omega)$.

3. *If $0 \leq r < s$, then $\mathcal{W}_2^s(\Omega)$ is compactly embedded in $\mathcal{W}_2^r(\Omega)$; if $1/2 < r < s$, then $\mathsf{W}_2^s(\Omega)$ is compactly embedded in $\mathsf{W}_2^r(\Omega)$.*

4. *Let $\mathcal{F}$ be a set of real-valued functions on Ω that are uniformly bounded away from 0. If $\mathcal{F}$ is bounded in some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$, then so is $\{1/f : f \in \mathcal{F}\}$.*

Proof. Part 1: See, e.g., Part 6 of Proposition 2 in Nickl (2007).

Part 2: See Section 2.7.1, Formula (12), in Triebel (1983) and additionally note that (by Proposition 3.4.2 in Triebel, 1983) $\mathcal{W}_2^s(\Omega)$ is equal to the space $B_{2,2}^s(\Omega)$ (as defined there) up to equivalent semi-norms.

Part 3: See Theorem 1.16.1 in Lions & Magenes (1972).

Part 4: By Part 2, $\mathcal{F}$ is a bounded subset of $\mathsf{C}^{s-1/2}$. If $s - 1/2$ is integer, then every $f \in \mathcal{F}$ has classical derivatives up to order $s - 1/2$. If $s - 1/2$ is non-integer, then every

$f \in \mathcal{F}$ has classical derivatives up to order $\lfloor s - 1/2 \rfloor$. Furthermore, $D^{\lfloor s-1/2 \rfloor} f$ is Hölder of order $s - 1/2 - \lfloor s - 1/2 \rfloor$, and therefore is differentiable a.e. It follows that, for every α satisfying $0 \le \alpha \le \lfloor s \rfloor$, the weak derivatives $D_w^\alpha f$ of order α coincide with the ordinary derivatives a.e. With help of the chain rule, this allows to compute $D_w^\alpha(1/f)$ a.e. for α, $0 \le \alpha \le \lfloor s \rfloor$. Now, use the definition of the Sobolev norm $\| \cdot \|_{s,2}$ together with the hypotheses on $\mathcal{F}$ to see that $\{1/f : f \in \mathcal{F}\}$ is a bounded subset of $\mathsf{W}_2^s(\Omega)$. □

2.3 Extension of the logarithm

We extend the logarithm to the left in the usual way by setting $\log 0 = -\infty$, and denote by $\mathcal{B}([-\infty, \infty))$ the σ-algebra on $[-\infty, \infty)$ that is generated by $\mathcal{B}(\mathbb{R}) \cup \{\{-\infty\}\}$. The resulting mapping is $\mathcal{B}([0,\infty))$-$\mathcal{B}([-\infty,\infty))$-measurable and continuous. We also define $\log \infty = \infty$, but will not make explicit use of this right-sided extension.

2.4 Covering numbers and metric entropy

Let $(E, \| \cdot \|)$ be a normed space. Let $\varepsilon > 0$ and X be a totally bounded subset of E. Then we denote by $N(\varepsilon, X, E, \| \cdot \|)$ the minimal number of closed balls of radius ε needed to cover X. We call $N(\varepsilon, X, E, \| \cdot \|)$ the covering number of X and define the metric entropy of X as

$$H(\varepsilon, X, E, \| \cdot \|) = \log N(\varepsilon, X, E, \| \cdot \|).$$

Let $(S, \mathcal{A}, \mu)$ be a measure space. For any two elements $l, u \in \mathcal{L}^0(S, \mathcal{A})$, the set

$$[l, u] = \{f \in \mathcal{L}^0(S, \mathcal{A}) : l \le f \le u\}$$

is called a bracket and $\|u - l\|_{2,\mu}$ its $\mathcal{L}^2(\mu)$-bracketing size. For $\varepsilon > 0$ and $\mathcal{F}$ some subset of $\mathcal{L}^0(S, \mathcal{A})$, we define $N_{[\,]}(\varepsilon, \mathcal{F}, \| \cdot \|_{2,\mu})$ to be the minimal number of brackets of $\mathcal{L}^2(\mu)$-bracketing size ε needed to cover $\mathcal{F}$; if there is no finite number of such brackets, we set $N_{[\,]}(\varepsilon, \mathcal{F}, \| \cdot \|_{2,\mu}) = \infty$ for convenience. The $\mathcal{L}^2(\mu)$-bracketing metric entropy of $\mathcal{F}$ is defined as

$$H_{[\,]}(\varepsilon, \mathcal{F}, \| \cdot \|_{2,\mu}) = \log N_{[\,]}(\varepsilon, \mathcal{F}, \| \cdot \|_{2,\mu}).$$

Further, $I_{[\,]}(\eta, \mathcal{F}, \| \cdot \|_{2,\mu})$ denotes the $\mathcal{L}^2(\mu)$-bracketing metric integral of $\mathcal{F}$ given by

$$I_{[\,]}(\eta, \mathcal{F}, \| \cdot \|_{2,\mu}) = \int_{(0,\eta]} \sqrt{1 + H_{[\,]}(\varepsilon, \mathcal{F}, \| \cdot \|_{2,\mu})}\, d\varepsilon.$$

2.5 Weak convergence

Let $(\Omega_n, \mathcal{A}_n, P_n)$ be probability spaces, $Y_n : \Omega_n \to S$ (not necessary measurable) mappings with values in some metric space (S, d), and $s \in S$. We say that Y_n converges to s in outer P_n-probability if $P_n^*(d(Y_n, s) > \varepsilon)$ converges to 0 for all $\varepsilon > 0$.

Let $(\Omega, \mathcal{A}, P)$ be a probability space and $Y : \Omega \to S$ an $\mathcal{A}$-$\mathcal{B}(S,d)$-measurable mapping. We say that Y_n converges weakly to Y in (S, d), formally $Y_n \rightsquigarrow Y$, if the outer integrals $\int_{\Omega_n}^* g(Y_n) dP_n$ converge to $\int_\Omega g(Y) dP$ for every $g \in \mathsf{C}(S, d)$; Y_n converges weakly to a Borel probability measure L on $(S, \mathcal{B}(S, d))$, formally $Y_n \rightsquigarrow L$, if $\int_{\Omega_n}^* g(Y_n) dP_n$ converges to $\int_T g dL$ for every $g \in \mathsf{C}(S, d)$.

Let Y_n be real-valued and r_n a sequence of positive real numbers. We write $Y_n = o^*_{P_n}(r_n)$ if $r_n^{-1}Y_n$ converges to 0 in outer P_n-probability, and $Y_n = O^*_{P_n}(r_n)$ if

$$\lim_{M\to\infty} \limsup_{n\to\infty} P_n^*\left(r_n^{-1}Y_n > M\right) = 0.$$

In case the probability spaces $(\Omega_n, \mathcal{A}_n, P_n)$ are the n-fold products of a single probability space $(\Omega, \mathcal{A}, P)$, that is, $(\Omega_n, \mathcal{A}_n, P_n) = (\Omega^n, \mathcal{A}^n, P^n)$, we write $Y_n = o^*_P(r_n)$ instead of $Y_n = o^*_{P^n}(r_n)$ and $Y_n = O^*_P(r_n)$ for $Y_n = O^*_{P^n}(r_n)$.

3 The basic setup

Let Ω be a non-empty bounded, open interval in $\mathbb{R}$. We observe i.i.d. random variables $X_1, \ldots, X_n$ that take their values in $(\Omega, \mathcal{B}(\Omega))$ and each have law $\mathbb{P}$. Let $\mathcal{P}_\Theta = \{p_\theta : \theta \in \Theta\}$ be a parametric model of probability density functions p_θ on Ω, where Θ is a compact subset of $\mathbb{R}^m$. Further, let $(V, \mathcal{V}, \mu)$ be a probability space and $\rho : V \times \Theta \to \Omega$ a mapping that is $\mathcal{V}$-$\mathcal{B}(\Omega)$-measurable in the first argument. We assume that, for every $\theta \in \Theta$, the law of $\rho(\,\cdot\,, \theta)$ has density p_θ, so that ρ provides a way of simulating data according to the densities in the parametric model. More precisely, let $V_1, \ldots, V_k$ be independent random variables with values in $(V, \mathcal{V})$ and law μ, and assume that these variables are independent from the data. We then obtain simulated data $X_1(\theta), \ldots, X_k(\theta)$ via $X_i(\theta) = \rho(V_i, \theta)$, simultaneously so for all $\theta \in \Theta$.

Since the final statistical results in this thesis will only depend on the probability measures $\mathbb{P}$ and μ, we are free to choose the representation of the random variables $X_1, \ldots, X_n$ and $V_1, \ldots, V_k$, and will take them to be the respective coordinate projections of the measurable space $(\Omega^{\mathbb{N}} \times V^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}} \otimes \mathcal{V}^{\mathbb{N}})$. Furthermore, we consider the probability measure $\mathbb{P}^{\mathbb{N}} \otimes \mu^{\mathbb{N}}$ on $(\Omega^{\mathbb{N}} \times V^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}} \otimes \mathcal{V}^{\mathbb{N}})$ and will always write Pr for $\mathbb{P}^{\mathbb{N}} \otimes \mu^{\mathbb{N}}$ in the sequel. The empirical measures associated with $X_1, \ldots, X_n$ and $V_1, \ldots, V_k$ will be denoted by $\mathbb{P}_n$ and μ_k, respectively.

Finally, we define a class of non-parametric auxiliary models by

$$\mathcal{P}(t, \zeta, D) = \left\{ p \in \mathsf{W}_2^t(\Omega) : \int_\Omega p \, d\lambda = 1, \ \inf_{x \in \Omega} p(x) \geq \zeta, \ \|p\|_{t,2} \leq D \right\},$$

where $t > 1/2$, $\zeta \geq 0$, and $0 < D < \infty$. To ensure that $\mathcal{P}(t, \zeta, D)$ is not empty, we require that $\zeta \leq \lambda(\Omega)^{-1} \leq D^2$ (which is a necessary and sufficient condition for $\mathcal{P}(t, \zeta, D)$ to be non-empty). Throughout the thesis, we will always assume $t > 1/2$ and $\zeta \geq 0$.

Proposition 2. *Every auxiliary model $\mathcal{P}(t, \zeta, D)$ is a non-empty, convex set, which is compact in $\mathsf{C}(\Omega)$ as well as in $\mathsf{W}_2^s(\Omega)$ for every s satisfying $1/2 < s < t$.*

Proof. The constant density $\lambda(\Omega)^{-1}$ belongs to $\mathcal{P}(t, \zeta, D)$, and so the auxiliary model is non-empty. Since the defining conditions are convex, it is convex. That $\mathcal{P}(t, \zeta, D)$ is compact as claimed follows from Lemma 3 in Nickl (2007). [Note that the proof of this lemma does not use that $\zeta > 0$, as is implicit there, and therefore is also valid for $\zeta = 0$.] □

For later use we stress that any $p \in \mathcal{P}(t, \zeta, D)$ is continuous on Ω and satisfies $\|p\|_\Omega \leq C_t D$ in view of Part 2 of Proposition 1. We further note the fact that in any auxiliary model $\mathcal{P}(t, \zeta, D)$ pointwise convergence is equivalent to convergence in all Sobolev norms of smaller order as well as in the sup-norm. This is shown in the next proposition.

Proposition 3. *Let $p_n, p \in \mathcal{P}(t, \zeta, D)$. Then the following statements are equivalent:* (i) *$\|p_n - p\|_\Omega$ converges to 0;* (ii) *p_n converges pointwise to p;* (iii) *p_n converges to p a.e.;* (iv) *p_n converges to p on a dense subset of Ω;* (v) *$\|p_n - p\|_{r,2}$ converges to 0 for some r satisfying $0 \leq r < t$;* (vi) *$\|p_n - p\|_{r,2}$ converges to 0 for all r satisfying $0 \leq r < t$.*

Proof. Clearly, (vi) implies (v). To show that (v) implies (vi), it suffices, in light of Part 3 of Proposition 1, to show that $\|p_n - p\|_{s,2}$ converges to 0 for arbitrary $s \geq r$ satisfying

$1/2 < s < t$. Since $\mathcal{P}(t,\zeta,D)$ is a compact subset of $\mathsf{W}_2^s(\Omega)$ in view of Proposition 2, for any subsequence $p_{n'}$ of p_n there exists a further subsequence $p_{n''}$ of $p_{n'}$ and a $p^* \in \mathcal{P}(t,\zeta,D)$ such that $\|p_{n''} - p^*\|_{s,2}$ converges to 0. By Part 3 of Proposition 1, we then have that also $\|p_{n''} - p^*\|_{r,2}$ converges to 0 since $s \geq r$. Because also $\|p_{n''} - p\|_{r,2}$ converges to 0 as a consequence of (v), it follows that $p^* = p$. This shows that $\|p_n - p\|_{s,2}$ converges to 0.

Furthermore, (i) implies (ii), (ii) implies (iii), and (iii) implies (iv). That (vi) implies (i) is a direct consequence of Part 2 of Proposition 1. It remains to show that (iv) implies (v). Choose r such that $1/2 < r < t$. The same compactness argument as above shows that for any subsequence $p_{n'}$ of p_n there exists a further subsequence $p_{n''}$ of $p_{n'}$ and a $p^* \in \mathcal{P}(t,\zeta,D)$ such that $\|p_{n''} - p^*\|_{r,2}$ converges to 0. By Part 2 of Proposition 1, we have that $\|p_{n''} - p^*\|_{\Omega}$ converges to 0. Consequently, p and p^* coincide on a dense subset of Ω. Since p and p^* are continuous, they are identical. This shows that $\|p_{n''} - p\|_{r,2}$ converges to 0, and hence the same is true for the entire sequence p_n. $\square$

4 A catalogue of assumptions

Apart from the maintained assumptions layed out in the previous section, we will make use of the assumptions listed below. We will often write $p(x, \theta)$ for $p_\theta(x)$, and stress that $p(x, \theta)$ is a *function* from $\Omega \times \Theta$ to $\mathbb{R}$.

The following assumption will be crucial in proving the asymptotic efficiency result for the indirect inference estimators considered in this thesis.

Assumption P.1 *The parametric model is a subset of the auxiliary model:*

$$\mathcal{P}_\Theta \subseteq \mathcal{P}(t, \zeta, D).$$

Remark 4. If Assumption P.1 is satisfied, then the following statements are equivalent in view of Proposition 3: (i) Assumption P.4; (ii) $\theta \mapsto p_\theta$ is continuous as a mapping into the metric space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_{s,2})$ for every s satisfying $1/2 \leq s < t$; (iii) $\theta \mapsto p_\theta$ is continuous as a mapping into the metric space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$.

Assumption P.2 *The strict inequality*

$$\inf_{\Omega \times \Theta} p(x, \theta) > 0$$

holds true.

Assumption P.3 *The parametric model is a subset of the auxiliary model, and the strict inequalities*

$$\inf_{\Omega \times \Theta} p(x, \theta) > \zeta$$

as well as

$$\sup_{\theta \in \Theta} \|p_\theta\|_{t,2} < D$$

hold true.

Assumption P.4 *For every $x \in \Omega$, $\theta \mapsto p(x, \theta)$ is a continuous function on Θ.*

Assumption P.4 is typically used in maximum likelihood estimation to prove consistency of maximum likelihood estimators.

Assumption P.5 *The interior Θ° of Θ is non-empty. Further, for every $x \in \Omega$, $\theta \mapsto p(x, \theta)$ is twice continuously partially differentiable on Θ°, and the following domination conditions hold for all $i, j = 1, \ldots, m$:*

$$\int_\Omega \sup_{\theta \in \Theta^\circ} \left| \frac{\partial p}{\partial \theta_i}(x, \theta) \frac{\partial p}{\partial \theta_j}(x, \theta) \right| d\lambda(x) < \infty,$$

$$\int_\Omega \sup_{\theta \in \Theta^\circ} \left| \frac{\partial^2 p}{\partial \theta_i \partial \theta_j}(x, \theta) \right| d\lambda(x) < \infty.$$

In case of correct specification of the underlying model, Assumption P.5 is typically used to prove equality of the Cramér-Rao bound and the asymptotic variance-covariance matrix of maximum likelihood estimators.

Apart from the assumed measurability of the simulation mechanism $\rho(v, \theta)$ in the first argument, we will need assumptions to control its behaviour in the second argument.

Assumption R.1 *For every $v \in V$, the simulation mechanism $\rho(v, \theta)$ is continuous in θ.*

Assumption R.2 *For some constant γ, $0 < \gamma \leq 1$, and some measurable function $R : V \to (0, \infty)$ such that $\int_V R^2 d\mu < \infty$, the simulation mechanism $\rho : V \times \Theta \to \Omega$ satisfies*

$$|\rho(v, \theta') - \rho(v, \theta)| \leq R(v)\|\theta' - \theta\|^{\gamma}$$

for all $v \in V$ and all $\theta, \theta' \in \Theta$.

Assumption Z *In the definition of $\mathcal{P}(t, \zeta, D)$, ζ is required to be greater than* 0.

Assumption D *The probability measure $\mathbb{P}$ has a density p_0.*

In the following we treat the probability density p_0 as a *function* from Ω to $\mathbb{R}$, that is, we let p_0 denote a fixed representative of the Radon-Nikodym derivative of $\mathbb{P}$ with respect to λ.

Assumption D.1 *The density function p_0 belongs to $\mathcal{P}(t, \zeta, D)$.*

Assumption D.2 *The density function p_0 satisfies the strict inequality*

$$\inf_{x \in \Omega} p_0(x) > 0.$$

Assumption D.3 *The density function p_0 is an internal point of $\mathcal{P}(t, \zeta, D)$ in the sense of Nickl (2007), meaning that it belongs to $\mathcal{P}(t, \zeta, D)$ and satisfies the strict inequalities*

$$\inf_{x \in \Omega} p_0(x) > \zeta$$

as well as

$$\|p_0\|_{t,2} < D.$$

For later use we note the following: (i) If Assumption Z is satisfied, then Assumption P.1 implies Assumption P.2; (ii) If Assumptions Z and D are satisfied, then Assumption D.1 implies Assumption D.2.

4.1 On the interrelation of the simulation mechanism and the density functions in the parametric model

Assumptions on the density functions in the parametric model $\mathcal{P}_{\Theta}$ and on the simulation mechanism ρ are of course related to each other, but the interrelationship is somewhat intricate. The following proposition collects two important observations.

Proposition 5.

1. *If Assumption P.1 is satisfied, then Assumption R.1 implies Assumption P.4.*

2. *In general, Assumption R.1 does not imply Assumption P.4.*

Proof. Part 1: For every $z \in \Omega$ and every $\theta \in \Theta$, let $F(z, \theta) = \int_{\{x \in \Omega : x \leq z\}} p_{\theta} d\lambda$; and note that, for every $\theta \in \Theta$, $F(\,\cdot\,, \theta)$ is the distribution function on Ω that is associated with p_{θ}. Let $\theta_n, \theta \in \Theta$ be such that θ_n converges to θ. We show that $\|p_{\theta_n} - p_{\theta}\|_{\Omega}$ converges to 0, which we know to be equivalent to Assumption P.4 by Remark 4. Suppose, on the contrary, that $\|p_{\theta_n} - p_\theta\|_\Omega$ does not converge to 0. Then there is an $\varepsilon > 0$ and a subsequence $p_{\theta_{n'}}$ of p_{θ_n} such that $\|p_{\theta_{n'}} - p_\theta\|_\Omega \geq \varepsilon$. We make two observations: First, by sup-norm compactness of $\mathcal{P}(t, \zeta, D)$, there is a subsequence $p_{\theta_{n''}}$ of $p_{\theta_{n'}}$ and a $p^* \in \mathcal{P}(t, \zeta, D)$ such that $\|p_{\theta_{n''}} - p^*\|_\Omega$ converges to 0. Consequently, for every $z \in \Omega$, $F(z, \theta_{n''})$ converges to $\int_{\{x\in\Omega:\, x\leq z\}} p^* d\lambda$, which can be seen using the inequality $\|\cdot\|_1 \leq \lambda(\Omega)\|\cdot\|_\Omega$. Second, Assumption R.1 implies that $\rho(\,\cdot\,, \theta_n)$ converges in distribution to $\rho(\,\cdot\,, \theta)$. Hence, upon noting that the distribution of $\rho(\,\cdot\,, \theta_n)$ is $F(\,\cdot\,, \theta_n)$ and the one of $\rho(\,\cdot\,, \theta)$ is $F(\,\cdot\,, \theta)$, $F(z, \theta_n)$ converges to $F(z, \theta)$ for every $z \in \Omega$ by continuity of $F(z, \,\cdot\,)$.

It follows that $F(z, \theta) = \int_{\{x\in\Omega:\, x\leq z\}} p^* d\lambda$ for all $z \in \Omega$, and so p_θ must coincide with p^* a.e., hence everywhere on Ω by continuity of p_θ and p^*. We arrive at a contradiction, upon recalling that $\|p_{\theta_{n''}} - p^*\|_\Omega$ converges to 0 and $\|p_{\theta_{n''}} - p_\theta\|_\Omega \geq \varepsilon$ at the same time.

Part 2: We provide an example of a simulation mechanism that satisfies Assumption R.1, but where the family of density functions in the underlying parametric model does not satisfy Assumption P.4. To this end, let $\Omega = (0, 1)$, $\Theta = [0, 1/4]$, and $(V, \mathcal{V}, \mu)$ be the unit interval $(0, 1)$ together with its Borel σ-algebra and Lebesgue measure λ. Let the simulation mechanism $\rho(v, \theta)$ be given by

$$\rho(v, \theta) = \begin{cases} \frac{1-\theta}{1-2\theta} v & \text{if } 0 < v < \frac{1}{2} - \theta, \\ \frac{1}{2} v + \frac{1}{4} & \text{if } \frac{1}{2} - \theta \leq v \leq \frac{1}{2} + \theta, \\ \frac{1-\theta}{1-2\theta} v - \frac{\theta}{1-2\theta} & \text{if } \frac{1}{2} + \theta < v < 1, \end{cases}$$

and observe that it satisfies Assumption R.1. For each $\theta \in [0, 1/4]$, the inverse mapping $\rho^{-1}(x, \theta)$ can easily be obtained and looks as follows:

$$\rho^{-1}(x, \theta) = \begin{cases} \frac{1-2\theta}{1-\theta} x & \text{if } 0 < x < \frac{1-\theta}{2}; \\ 2x - \frac{1}{2} & \text{if } \frac{1-\theta}{2} \leq x \leq \frac{1+\theta}{2}; \\ \frac{1-2\theta}{1-\theta} x + \frac{\theta}{1-\theta} & \text{if } \frac{1+\theta}{2} < x < 1. \end{cases}$$

Apply the change of variables theorem for densities to conclude that, for every $\theta \in \Theta$, $p(\,\cdot\,, \theta)$ equals $\frac{\partial \rho^{-1}}{\partial x}(\,\cdot\,, \theta)$ a.e. on $\Omega = (0, 1)$, where

$$\frac{\partial \rho^{-1}}{\partial x}(x, \theta) = \begin{cases} 2 & \text{if } \frac{1-\theta}{2} < x < \frac{1+\theta}{2}, \\ \frac{1-2\theta}{1-\theta} & \text{if } 0 < x < \frac{1-\theta}{2} \text{ or } \frac{1+\theta}{2} < x < 1, \end{cases}$$

and does not exist for $x = (1 - \theta)/2$ and $x = (1 + \theta)/2$. We extend $\frac{\partial \rho^{-1}}{\partial x}(x, \theta)$ by setting it equal to 0 for all (x, θ) such that $x = (1 - \theta)/2$ or $x = (1 + \theta)/2$, and note that this extension is measurable on $\Omega \times \Theta$. For any fixed $x \in (0, 1)$, we may then look at $\frac{\partial \rho^{-1}}{\partial x}(x, \theta)$ as a function in θ, leading to

$$\frac{\partial \rho^{-1}}{\partial x}(x, \theta) = \begin{cases} 2 & \text{if } \theta > |2x - 1|, \\ 0 & \text{if } \theta = |2x - 1|, \\ \frac{1-2\theta}{1-\theta} & \text{if } \theta < |2x - 1|. \end{cases}$$

Observe that, for every $x \in (3/8, 1/2) \cup (1/2, 5/8)$, the left- and right-sided limits $\frac{\partial \rho^{-1}}{\partial x}(x, \theta-)$ and $\frac{\partial \rho^{-1}}{\partial x}(x, \theta+)$ at $\theta = |2x - 1|$ exist but are different. It follows that $\frac{\partial \rho^{-1}}{\partial x}(x, \theta)$, as a function in both variables, is discontinuous along the line segments

$$\{(x, |2x - 1|) : 3/8 < x < 1/2\}$$

and

$$\{(x, |2x-1|) : 1/2 < x < 5/8\}$$

in $\Omega \times \Theta$.

Suppose now that Assumption P.4 is satisfied. Then, by separability of Θ, $p(x,\theta)$ is jointly measurable, and hence the set

$$A := \left\{(x,\theta) \in \Omega \times \Theta : 1/2 < x < 5/8,\, p(x,\theta) \neq \frac{\partial \rho^{-1}}{\partial x}(x,\theta)\right\}$$

is a measurable subset of $\Omega \times \Theta$. We claim that there is some $\theta \in \Theta$ such that $\{x \in \Omega : (x,\theta) \in A\}$ has positive Lebesgue measure. This will contradict the fact that $p(\,\cdot\,,\theta)$ is equal to $\frac{\partial \rho^{-1}}{\partial x}(\,\cdot\,,\theta)$ a.e., and consequently also contradict Assumption P.4. To prove the claim, first observe the following: For every fixed x satisfying $1/2 < x < 5/8$, there is some (non-empty) open interval $U_x \subset \Theta$ such that $p(x,\theta)$ is different from $\frac{\partial \rho^{-1}}{\partial x}(x,\theta)$ for all $\theta \in U_x$: Otherwise there are $\theta_n^-, \theta_n^+ \in \Theta$ with the following properties: $\theta_n^- < 2x-1$, $\theta_n^+ > 2x-1$; θ_n^- and θ_n^+ converge to $2x-1$; $p(x,\theta_n^-) = \frac{\partial \rho^{-1}}{\partial x}(x,\theta_n^-)$ and $p(x,\theta_n^+) = \frac{\partial \rho^{-1}}{\partial x}(x,\theta_n^+)$. Using Assumption P.4, we infer that

$$\lim_{n\to\infty} \frac{\partial \rho^{-1}}{\partial x}(x,\theta_n^-) = \lim_{n\to\infty} \frac{\partial \rho^{-1}}{\partial x}(x,\theta_n^+),$$

which is in contradiction to what we have observed above.

Now,

$$\int_\Theta 1_A(x,\theta)\, d\lambda(\theta) \geq \int_\Theta 1_{U_x}(\theta) d\lambda(\theta) > 0$$

for all $x \in (1/2, 5/8)$, which implies that

$$\int_{(1/2,5/8)} \int_\Theta 1_A(x,\theta) d\lambda(\theta) d\lambda(x) > 0.$$

By Tonelli's theorem, the l.h.s. of the last display is equal to

$$\int_\Theta \int_{(1/2,5/8)} 1_A(x,\theta) d\lambda(x) d\lambda(\theta),$$

so that $\int_{(1/2,5/8)} 1_A(x,\theta) d\lambda(x)$ must be positive for some $\theta \in \Theta$. This verifies that $\{x \in \Omega : (x,\theta) \in A\}$ has positive Lebesgue measure. $\square$

5 Auxiliary maximum likelihood estimators

We now introduce auxiliary maximum likelihood estimators. Define the auxiliary log-likelihood function on $\mathcal{P}(t,\zeta,D)$ by

$$L_n(p) = \frac{1}{n}\sum_{i=1}^{n} \log p(X_i)$$

for the given data, and on $\Theta \times \mathcal{P}(t,\zeta,D)$ by

$$L_k(\theta,p) = \frac{1}{k}\sum_{i=1}^{k} \log p(X_i(\theta))$$

for the simulated data. Note that both functions $L_n(p)$ and $L_k(\theta,p)$ are well-defined in view of Section 2.3 and take their values in $[-\infty,\infty)$. In fact, $L_n(f)$ and $L_k(\theta,f)$ are well-defined and take their values in $[-\infty,\infty)$ for any non-negative real-valued function f on Ω.

An auxiliary maximum likelihood (AML-) estimator given $X_1,\dots,X_n$ is defined as an element $\hat{p}_n \in \mathcal{P}(t,\zeta,D)$ such that

$$L_n(\hat{p}_n) = \sup_{p\in\mathcal{P}(t,\zeta,D)} L_n(p)$$

holds true. An AML-estimator given $X_1(\theta),\dots,X_k(\theta)$ is an element $\tilde{p}_k(\theta) \in \mathcal{P}(t,\zeta,D)$ satisfying

$$L_k(\theta,\tilde{p}_k(\theta)) = \sup_{p\in\mathcal{P}(t,\zeta,D)} L_k(\theta,p).$$

In this section we investigate existence, uniqueness, consistency, rates of convergence, and uniform central limit theorems of AML-estimators. The results obtained here go beyond Nickl (2007) in three respects: First, we allow for auxiliary models $\mathcal{P}(t,\zeta,D)$ where the lower bound for the densities, ζ, can be equal to 0. Second, we show that the corresponding AML-estimators are uniquely defined, thus leading to *unique* measurable selections $\hat{p}_n$ and $\tilde{p}_k(\theta)$, $\theta\in\Theta$. Third, we prove that the results in Nickl (2007) on the consistency of $\hat{p}_n$ and its rates of convergence in different Sobolev norms hold uniformly over the parameter space for the AML-estimators $\tilde{p}_k(\theta)$, $\theta\in\Theta$. We further prove a uniform Donsker-type theorem, which extends Theorem 3 in Nickl (2007) in so far that the stochastic process $f \mapsto \int_\Omega (\tilde{p}_k(\theta) - p_\theta) f\, d\lambda$, when indexed by $\Theta\times\mathcal{F}$ for appropriate classes $\mathcal{F}$ of functions, converges weakly in $\ell^\infty(\Theta\times\mathcal{F})$ to a μ-Brownian bridge.

5.1 Existence of AML-estimators

In the following theorem we show that the AML-estimators defined above exist and are unique.

Theorem 6.

1. *There exists a unique $\hat{p}_n \in \mathcal{P}(t,\zeta,D)$ such that*

$$L_n(\hat{p}_n) = \sup_{p\in\mathcal{P}(t,\zeta,D)} L_n(p)$$

holds true. The resulting mapping $\hat{p}_n : \Omega^n \to \mathcal{P}(t,\zeta,D)$ is measurable with respect to the σ-algebras $\mathcal{B}(\Omega)^n$ and $\mathcal{B}(\mathcal{P}(t,\zeta,D), \|\cdot\|_\Omega)$.

2. *For each $\theta \in \Theta$, there exists a unique $\tilde{p}_k(\theta) \in \mathcal{P}(t, \zeta, D)$ such that*

$$L_k(\theta, \tilde{p}_k(\theta)) = \sup_{p \in \mathcal{P}(t,\zeta,D)} L_k(\theta, p)$$

holds true. The resulting mapping $\tilde{p}_k(\theta) : V^k \to \mathcal{P}(t, \zeta, D)$ is measurable with respect to the σ-algebras $\mathcal{V}^k$ and $\mathcal{B}(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$. Furthermore, if Assumption R.1 is satisfied, then, for any fixed values of the underlying simulated variables, $\theta \mapsto \tilde{p}_k(\theta)$ is continuous when viewed as a mapping from Θ into the metric space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$.

Proof. Part 1: Let $x_1, \ldots, x_n$ be fixed data points. The existence of a maximizer of L_n follows from the facts that L_n is a continuous mapping on the compact space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$: see Proposition 2 and use Part 2(a) of Proposition 43 with $\mathcal{F} = \mathcal{P}(t, \zeta, D)$. Denote by S the consequently non-empty set of all $p \in \mathcal{P}(t, \zeta, D)$ that maximize L_n. We have to show that S has only one element. Since L_n is a concave function on the convex set $\mathcal{P}(t, \zeta, D)$, we conclude from Lemma 59 that S is convex. If S is a subset of the Sobolev *sphere* of radius D, then S is a singleton since the Sobolev norm $\|\cdot\|_{t,2}$, being a Hilbert norm, is strictly convex. It remains to deal with the case when S is not contained in the Sobolev sphere of radius D. We claim that $S = \{\zeta\}$ then, and prove this claim in two steps.

Step 1: We show that any $p \in S$ with Sobolev norm less than D equals ζ. [In particular, it will follow that ζ belongs to S and has Sobolev norm less than D.] To this end, let $p \in S$ with $\|p\|_{t,2} < D$. Assume $p \neq \zeta$, so that there is some $z \in \Omega$ with $p(z) > \zeta$. By continuity of p we may assume that z is different from any of the finitely many data points $x_1, \ldots, x_n$. We claim that there is a $q \in \mathcal{P}(t, \zeta, D)$ such that $q(x_1) > p(x_1)$ and q coincides with p on the remaining (if any) observations, which will contradict the maximizing property of p. The existence of q can be seen as follows: Choose $\varepsilon > 0$ such that $I := [z - 2\varepsilon, z + 2\varepsilon]$, $U := (x_1 - 2\varepsilon, x_1 + 2\varepsilon)$, and $\{x_1, \ldots, x_n\} \setminus \{x_1\}$ are pairwise disjoint subsets of Ω and $\inf_{x \in I} p(x) > \zeta$. As $A := [x_1 - \varepsilon, x_1 + \varepsilon]$ is a closed set contained in the open set U, there is a C^∞-function $f : \Omega \to \mathbb{R}$ with values in $[0, 1]$ such that $f|_A = 1$ and $f|_{\Omega \setminus U} = 0$. For every $y \in \Omega$, let

$$\bar{f}(y) = \begin{cases} f(y + x_1 - z) & \text{if } y + x_1 - z \in \Omega, \\ 0 & \text{otherwise,} \end{cases}$$

so that $\bar{f}$ is the translation of f by $z - x_1$; and define $g : \Omega \to \mathbb{R}$ by $g = f - \bar{f}$. Then g has values in $[-1, 1]$, is contained in $\mathsf{W}_2^t(\Omega)$ since it is C^∞, and integrates to 0. Since $D - \|p\|_{t,2} > 0$ and $\inf_{x \in I} p(x) - \zeta > 0$, we can find a scalar $\beta > 0$ such that $\|\beta g\|_{t,2} \le D - \|p\|_{t,2}$ and $\beta \le \inf_{x \in I} p(x) - \zeta$. Let $q = p + \beta g$ and observe that $\|q\|_{t,2} \le \|p\|_{t,2} + \|\beta g\|_{t,2} \le D$. Further, $q(x) \ge \zeta$ for every $x \in \Omega$, which can be seen by distinguishing the two cases $x \in I$ and $x \in \Omega \setminus I$. For $x \in \Omega \setminus I$, we have that $g(x) \ge 0$, and so $q(x) \ge p(x) \ge \zeta$. If $x \in I$, then $q(x) \ge p(x) - \beta \ge p(x) - \inf_{x \in I} p(x) + \zeta \ge \zeta$, where the lower inequality holds true because $g(x) \ge -1$ for every $x \in \Omega$, the middle inequality holds true by the choice of β, and the upper one does so since $x \in I$ and therefore $p(x) - \inf_{x \in I} p(x) \ge 0$. It follows that $q \in \mathcal{P}(t, \zeta, D)$. Since $\beta > 0$ and $g(x_1) = 1$, $q(x_1) > p(x_1)$. Further, q coincides with p on the remaining (if any) data points because g is 0 there. The existence of q contradicts the maximizing property of p, and consequently p must be equal to ζ.

Step 2: We show that S cannot contain any element other than ζ. For any $q \in S$ and any α, $0 < \alpha < 1$, take the convex combination $s = \alpha\zeta + (1-\alpha)q$, which belongs to S by the convexity of S. Observe that $\|s\|_{t,2} < D$ since $\|\zeta\|_{t,2} < D$ by Step 1; hence $s = \zeta$ by Step 1, implying that $q = \zeta$.

To see that $\hat{p}_n : \Omega^n \to \mathcal{P}(t,\zeta,D)$ is measurable, we want to apply Lemma A3 in Pötscher & Prucha (1997). Because L_n can possibly attain the value $-\infty$, we cannot directly apply this lemma to L_n and so apply it to the *real-valued* function $h(L_n)$ instead. Here, h is defined by

$$h(y) = \begin{cases} \arctan(y) & \text{if } y \in \mathbb{R}, \\ -\frac{\pi}{2} & \text{if } y = -\infty, \\ \frac{\pi}{2} & \text{if } y = \infty, \end{cases}$$

is a monotonous real-valued function, and hence respects maxima.

Part 2: For any fixed $\theta \in \Theta$, the same arguments as in Part 1 give unique maximizers $\tilde{p}_k(\theta)$ of $L_k(\theta,\cdot)$, resulting in a $\mathcal{V}^k$-$\mathcal{B}(\mathcal{P}(t,\zeta,D),\|\cdot\|_\Omega)$-measurable mapping. To see that the mapping $\theta \mapsto \tilde{p}_k(\theta)$ is continuous as claimed, apply Lemma 60 with $X = \Theta$, $Y = (\mathcal{P}(t,\zeta,D),\|\cdot\|_\Omega)$, $u(x,y) = L_k(\theta,p)$, and $v(x) = \tilde{p}_k(\theta)$. Note that $(\mathcal{P}(t,\zeta,D),\|\cdot\|_\Omega)$ is a compact metric space by Proposition 2 and that, under Assumption R.1, $L_k(\theta,p)$ is continuous on $\Theta \times (\mathcal{P}(t,\zeta,D),\|\cdot\|_\Omega)$, as can be seen by applying Part 2(b) of Proposition 43 with $\mathcal{F} = \mathcal{P}(t,\zeta,D)$. $\square$

Remark 7. (i) The previous proof in particular shows that if $\zeta < \lambda(\Omega)^{-1}$, then $\hat{p}_n$ must lie on the Sobolev sphere of radius D. If also $\lambda(\Omega)^{-1} < D^2$, then $\hat{p}_n$ cannot be constant: Suppose indirectly that $\hat{p}_n$ is constant. It can only be equal to $\lambda(\Omega)^{-1}$ then, and so $\|\hat{p}_n\|_{t,2} = \lambda(\Omega)^{-1/2}$. Since $\lambda(\Omega)^{-1} < D^2$ by hypothesis, it follows that $\hat{p}_n$ does not belong to the Sobolev sphere of radius D, which is contradictory.

Similarly, if $\zeta < \lambda(\Omega)^{-1}$, then, for every $\theta \in \Theta$, $\tilde{p}_k(\theta)$ lies on the Sobolev sphere of radius D. If also $\lambda(\Omega)^{-1} < D^2$, then $\tilde{p}_k(\theta)$ cannot be constant.

(ii) The mapping $\hat{p}_n : \Omega^n \times \Omega \to \mathbb{R}$ is $\mathcal{B}(\Omega)^n$-measurable in the first and continuous in the second argument. Since Ω is separable, $\hat{p}_n$ is consequently jointly measurable. Similarly, the mappings $\tilde{p}_k(\theta) : V^k \times \Omega \to \mathbb{R}$ are jointly measurable for all $\theta \in \Theta$.

(iii) For any fixed data points $x_1,\dots,x_n$, we have that $\hat{p}_n(x_i) > 0$ for $i = 1,\dots,n$. This follows from the observation that for the constant density $\lambda(\Omega)^{-1}$, which belongs to $\mathcal{P}(t,\zeta,D)$, $L_n(\lambda(\Omega)^{-1}) = \log\lambda(\Omega)^{-1} > -\infty$, whereas $L_n(\hat{p}_n) = -\infty$ if $\hat{p}_n(x_i) = 0$ for some i, $1 \le i \le n$, which would contradict the maximizing property of $\hat{p}_n$. By an analogous argument we have that, for every $\theta \in \Theta$, $\tilde{p}_k(\theta)(x_i) > 0$ for $i = 1,\dots,n$.

5.2 Consistency of AML-estimators

By the law of large numbers, $L_n(f)$ and $L_k(\theta,f)$ converge a.s. to the respective expectations of $\log f$ and $\log f(\rho(\cdot,\theta))$. These expectations will be important in studying the limiting behaviour of $\hat{p}_n$ and $\tilde{p}_k(\theta)$. So let

$$L(f) = \int_\Omega \log f \, d\mathbb{P}$$

for any non-negative, measurable function f on Ω, provided the integral is defined; and let

$$L(\theta, f) = \int_V \log f(\rho(\,\cdot\,, \theta)) d\mu$$

for any $\theta \in \Theta$ and any non-negative, measurable function f on Ω, provided the integral is defined. If $f \in \mathsf{L}^\infty(\Omega)$, then both functions are well-defined and take their values in $[-\infty, \infty)$. The restrictions of $L(f)$ to $\mathcal{P}(t, \zeta, D)$ and of $L(\theta, f)$ to $\Theta \times \mathcal{P}(t, \zeta, D)$ are real-valued in case $\zeta > 0$. We will make use of the following facts.

Proposition 8. *$L(p_0)$ is well-defined and satisfies $L(p_0) > -\infty$, provided Assumption D holds. Similarly, for every $\theta \in \Theta$, $L(\theta, p_\theta)$ is well-defined and satisfies $L(\theta, p_\theta) > -\infty$.*

Proof. Observe that

$$L(p_0) = \int_{\{x\in\Omega:\, p_0(x)>0\}} \log p_0 \, p_0 d\lambda$$

and that

$$L(\theta, p_\theta) = \int_{\{x\in\Omega:\, p(x,\theta)>0\}} \log p_\theta \, p_\theta d\lambda$$

by the change of variable theorem. It is thus sufficient to show that

$$\int_{\{x\in\Omega:\, p_0(x)>0\}} \max(-p_0 \log p_0, 0) d\lambda < \infty$$

and

$$\int_{\{x\in\Omega:\, p(x,\theta)>0\}} \max(-p_\theta \log p_\theta, 0) d\lambda < \infty.$$

For every $x \in \Omega$, we have the identities

$$\max(-p_0(x) \log p_0(x), 0) \quad = \quad \begin{cases} h(p_0(x)) & \text{if } 0 < p_0(x) \leq 1, \\ 0 & \text{if } p_0(x) > 1, \end{cases}$$

and

$$\max(-p(x,\theta) \log p(x,\theta), 0) \quad = \quad \begin{cases} h(p(x,\theta)) & \text{if } 0 < p(x,\theta) \leq 1, \\ 0 & \text{if } p(x,\theta) > 1, \end{cases}$$

where $h(y)$ is defined by $h(y) = -y \log y$ for every $y \in (0, \infty)$. The function $h : (0, \infty) \to \mathbb{R}$ can be continuously extended to $[0, \infty)$ by setting $h(0) = 0$, and consequently is bounded on the compact interval $[0, 1]$. It follows that $\max(-p_0 \log p_0, 0)$ is bounded on $\{x \in \Omega : p_0(x) > 0\}$ and that $\max(-p_\theta \log p_\theta, 0)$ is bounded on $\{x \in \Omega : p(x, \theta) > 0\}$. This completes the proof. □

Lemma 9.

1. *If Assumptions D and D.1 are satisfied, then p_0 is the unique maximizer of the function $L(p)$ over $\mathcal{P}(t, \zeta, D)$.*

2. *If Assumption P.1 is satisfied, then p_θ is the unique element of $\mathcal{P}(t, \zeta, D)$ satisfying*

$$L(\theta, p_\theta) = \sup_{p \in \mathcal{P}(t,\zeta,D)} L(\theta, p).$$

Proof. Part 1: For any $p \in \mathcal{P}(t, \zeta, D)$ that is distinct from p_0, use that p and p_0 are continuous functions on Ω to verify that the set

$$\{x \in \Omega : p_0(x) > 0\} \cap \{x \in \Omega : p(x) \neq p_0(x)\}$$

has positive $\mathbb{P}$-probability. We may therefore apply the strict version of Jensen's inequality to see that

$$\begin{aligned} L(p) - L(p_0) &= \int_\Omega \log p \, d\mathbb{P} - \int_\Omega \log p_0 \, d\mathbb{P} \\ &= \int_{\{x \in \Omega :\, p_0(x) > 0\}} \log \frac{p}{p_0} d\mathbb{P} \\ &< \log \int_{\{x \in \Omega :\, p_0(x) > 0\}} \frac{p}{p_0} d\mathbb{P} \\ &\leq 0. \end{aligned}$$

Note that in view of Proposition 8, the expression $L(p) - L(p_0)$ is well-defined.

Part 2: Fix $\theta \in \Theta$. Note that

$$L(\theta, p) = \int_{\{x \in \Omega :\, p(x,\theta) > 0\}} \log p \, p_\theta d\lambda$$

by the change of variable theorem. Part 2 now follows by the same arguments as in Part 1, with p_0 replaced by p_θ. □

Remark 10. We note that maximizing $L(\,\cdot\,)$ over $\mathcal{P}(t, \zeta, D)$ is equivalent to minimizing the Kullback-Leibler distance

$$d_{\mathrm{KL}}(p_0, p) = \int_{\{x \in \Omega :\, p_0(x) > 0\}} \log \frac{p_0}{p} \, p_0 d\lambda$$

in $p \in \mathcal{P}(t, \zeta, D)$. Similarly, maximizing $L(\theta, \,\cdot\,)$ over $\mathcal{P}(t, \zeta, D)$ is equivalent to minimizing

$$d_{\mathrm{KL}}(p_\theta, p) = \int_{\{x \in \Omega :\, p(x,\theta) > 0\}} \log \frac{p_\theta}{p} \, p_\theta d\lambda$$

in $p \in \mathcal{P}(t, \zeta, D)$.

Theorem 11.

1. *Let Assumptions D and D.1 be satisfied. Then we have that*

$$\lim_{n \to \infty} \|\hat{p}_n - p_0\|_{s,2} = 0 \quad a.s.$$

for every s such that $0 \leq s < t$. In particular,

$$\lim_{n \to \infty} \|\hat{p}_n - p_0\|_\Omega = 0 \quad a.s.$$

2. *Let Assumption P.1 be satisfied. Then, for every $\theta \in \Theta$:*

$$\lim_{k \to \infty} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2} = 0 \quad a.s.$$

for every s such that $0 \leq s < t$; in particular,

$$\lim_{k \to \infty} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega = 0 \quad a.s.$$

3. Let Assumptions P.1, P.2, and R.1 be satisfied. Then we have that

$$\lim_{k\to\infty} \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2} = 0 \quad a.s.$$

for every s such that $0 \le s < t$. In particular,

$$\lim_{k\to\infty} \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{\Omega} = 0 \quad a.s.$$

Proof. Part 1: In view of Part 3 of Proposition 1, we may restrict ourselves to the case $1/2 < s < t$. By Kolomogorov's strong law of large numbers we have that

$$\lim_{n\to\infty} |L_n(p_0) - L(p_0)| = 0 \quad \text{a.s.} \tag{4}$$

[Note that $|L(p_0)| < \infty$ by Proposition 8 and Assumption D.1. Furthermore, this shows that the random variables $\log p_0(X_i)$ are a.s. real-valued.] Let ε_l be positive real numbers that converge monotonously to 0. Apply the uniform law of large numbers in Part 5(a) of Proposition 43 with $\mathcal{F} = \{p + \varepsilon_l : p \in \mathcal{P}(t, \zeta, D)\}$ to see that

$$\lim_{n\to\infty} \sup_{p\in\mathcal{P}(t,\zeta,D)} |L_n(p + \varepsilon_l) - L(p + \varepsilon_l)| = 0 \quad \text{a.s.} \tag{5}$$

for every $l \in \mathbb{N}$. The event where the statements in (4) and (5) hold true has probability 1. In the following arguments we fix an arbitrary element of this event.

We prove that $\|\hat{p}_n - p_0\|_{s,2}$ converges to 0 by showing that any subsequence of $\hat{p}_n$ has some subsequence converging to p_0 in the Sobolev norm $\|\cdot\|_{s,2}$. To this end, let $\hat{p}_{n'}$ be a subsequence of $\hat{p}_n$. Because $\mathcal{P}(t, \zeta, D)$ is compact in $\mathsf{W}_2^s(\Omega)$ by Proposition 2, there is a subsequence $\hat{p}_{n''}$ of $\hat{p}_{n'}$ and some $p^* \in \mathcal{P}(t, \zeta, D)$ such that $\|\hat{p}_{n''} - p^*\|_{s,2}$ converges to 0. We claim that $p^* = p_0$, which will complete the proof of Part 1:

Use Assumption D.1, the definition of $\hat{p}_{n''}$ as maximizer, and the monotonicity of the logarithm to obtain

$$\begin{aligned} L_{n''}(p_0) \le L_{n''}(\hat{p}_{n''}) &\le L_{n''}(\hat{p}_{n''} + \varepsilon_l) \\ &\le L(\hat{p}_{n''} + \varepsilon_l) + \sup_{p\in\mathcal{P}(t,\zeta,D)} |L_{n''}(p + \varepsilon_l) - L(p + \varepsilon_l)|. \end{aligned} \tag{6}$$

The first term on the r.h.s. of (6) converges to $L(p^* + \varepsilon_l)$ since $\|\hat{p}_{n''} - p^*\|_{s,2}$, hence also $\|\hat{p}_n - p^*\|_\Omega$, converges to 0 and $L(\cdot + \varepsilon_l)$ is sup-norm continuous on $\mathcal{P}(t, \zeta, D)$ by Part 4(a) of Proposition 43. Recall that the sup on the r.h.s. of (6) goes to 0, and that $L_{n''}(p_0)$ converges to $L(p_0)$. It follows that

$$L(p_0) \le L(p^* + \varepsilon_l). \tag{7}$$

Since ε_l converges monotonously to 0, the sequence of functions $\log(p^* + \varepsilon_l)$ is monotonously non-increasing with pointwise limit $\log p^*$, and is bounded above by the integrable function $\log(p^* + \varepsilon_1)$. Using the theorem of monotone convergence and (7), we conclude that $L(p_0) \le L(p^*)$. Hence, $p^* = p_0$ by Part 1 of Lemma 9.

Part 2: For fixed $\theta \in \Theta$, Part 2 follows by the same arguments as in Part 1, with p_0 replaced by p_θ.

Part 3: In view of Part 3 of Proposition 1, we may restrict ourselves to the case $1/2 < s < t$. Let $\xi = \inf_{\Omega\times\Theta} p(x, \theta)$. By hypothesis, $\xi > 0$, and so we may apply the uniform law of large numbers in Part 5(b) of Proposition 43 with $\mathcal{F} = \mathcal{P}(t, \xi, D)$ to get

$$\lim_{k\to\infty} \sup_{\Theta\times\mathcal{P}(t,\xi,D)} |L_k(\theta,p) - L(\theta,p)| = 0 \quad \text{a.s.}$$

Let ε_l be positive real numbers that converge monotonously to 0. For each $l \in \mathbb{N}$, the uniform law of large numbers stated in Part 5(b) of Proposition 43, with $\mathcal{F} = \{p + \varepsilon_l : p \in \mathcal{P}(t,\zeta,D)\}$, implies that

$$\lim_{k\to\infty} \sup_{\theta\in\Theta} \sup_{p\in\mathcal{P}(t,\zeta,D)} |L_k(\theta, p+\varepsilon_l) - L(\theta, p+\varepsilon_l)| = 0 \quad \text{a.s.}$$

The event where all statements in the last two displays hold true has probability 1. In the following arguments we fix an arbitrary element of this event.

To simplify notation, set $c_k = \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$. We claim that c_k converges to 0. Assume this is not true. Then there is some $\eta > 0$ such that for every $k \in \mathbb{N}$ there are $k' \in \mathbb{N}$, $k' \geq k$, and $\theta_{k'} \in \Theta$ that satisfy

$$\|\tilde{p}_{k'}(\theta_{k'}) - p_{\theta_{k'}}\|_{s,2} > \eta. \tag{8}$$

By compactness of Θ and compactness of $\mathcal{P}(t,\zeta,D)$ as a subset of $\mathsf{W}_2^s(\Omega)$, we find a subsequence $\tilde{p}_{k''}(\theta_{k''})$ of $\tilde{p}_{k'}(\theta_{k'})$ such that $\theta_{k''}$ converges to θ^* for some $\theta^* \in \Theta$, and $\|\tilde{p}_{k''}(\theta_{k''}) - p^*\|_{s,2}$ converges to 0 for some $p^* \in \mathcal{P}(t,\zeta,D)$. So, if p^* equals p_{θ^*} (which we verify below), then $\|\tilde{p}_{k''}(\theta_{k''}) - p_{\theta^*}\|_{s,2}$ converges to 0. Consequently, $\|\tilde{p}_{k''}(\theta_{k''}) - p_{\theta_{k''}}\|_{s,2}$ converges to 0 because $p_{\theta_{k''}}$ converges to p_{θ^*} in $(\mathcal{P}(t,\zeta,D), \|\cdot\|_{s,2})$ in view of Part 1 of Proposition 5 and Remark 4. This is in contradiction to (8) and therefore in contradiction to the assumption that c_k does not converge to 0.

It remains to show that p^* equals p_{θ^*}. Use Assumption P.1, the definition of $\tilde{p}_{k''}(\theta_{k''})$ as maximizer, and the monotonicity of the logarithm to obtain

$$\begin{aligned}
L_{k''}(\theta_{k''}, p_{\theta_{k''}}) &\leq L_{k''}(\theta_{k''}, \tilde{p}_{k''}(\theta_{k''})) \\
&\leq L_{k''}(\theta_{k''}, \tilde{p}_{k''}(\theta_{k''}) + \varepsilon_l) \\
&\leq L(\theta_{k''}, \tilde{p}_{k''}(\theta_{k''}) + \varepsilon_l) \\
&\quad + \sup_{\theta\in\Theta} \sup_{p\in\mathcal{P}(t,\zeta,D)} |L_{k''}(\theta, p+\varepsilon_l) - L(\theta, p+\varepsilon_l)|\,.
\end{aligned} \tag{9}$$

The first term on the r.h.s. of (9) converges to $L(\theta^*, p^* + \varepsilon_l)$ since $\theta_{k''}$ converges to θ^*, $\|\tilde{p}_{k''}(\theta_{k''}) - p^*\|_{s,2}$, hence also $\|\tilde{p}_{k''}(\theta_{k''}) - p^*\|_\Omega$, converges to 0, and $L(\,\cdot\,, \,\cdot\, + \varepsilon_l)$ is a continuous function on $\Theta \times (\mathcal{P}(t,\zeta,D), \|\cdot\|_\Omega)$ by Part 4(b) of Proposition 43. Recall that the sup on the r.h.s. of (9) goes to 0. Further, the sup on the r.h.s. of the inequality

$$\begin{aligned}
&|L_{k''}(\theta_{k''}, p_{\theta_{k''}}) - L(\theta^*, p_{\theta^*})| \\
&\quad \leq \sup_{\Theta\times\mathcal{P}(t,\xi,D)} |L_{k''}(\theta,p) - L(\theta,p)| + |L(\theta_{k''}, p_{\theta_{k''}}) - L(\theta^*, p_{\theta^*})|
\end{aligned}$$

converges to 0. The second term on the r.h.s. goes to 0 as $\theta_{k''}$ converges to θ^*, $\|p_{\theta_{k''}} - p_{\theta^*}\|_{s,2}$, hence also $\|p_{\theta_{k''}} - p_{\theta^*}\|_\Omega$, converges to 0, and $L(\theta,p)$ is a continuous function on $\Theta \times (\mathcal{P}(t,\xi,D), \|\cdot\|_\Omega)$ by Part 4(b) of Proposition 43. Hence, the l.h.s. of (9) goes to $L(\theta^*, p_{\theta^*})$. It follows that

$$L(\theta^*, p_{\theta^*}) \leq L(\theta^*, p^* + \varepsilon_l). \tag{10}$$

Since ε_l converges monotonously to 0, the sequence of functions $\log{(p^* + \varepsilon_l)(\rho(\,\cdot\,, \theta^*))}$ is monotonously non-increasing with pointwise limit $\log p^*(\rho(\,\cdot\,, \theta^*))$, and is bounded above by the integrable function $\log(p^* + \varepsilon_1)(\rho(\,\cdot\,, \theta^*))$. Using the theorem of monotone convergence and (10), we conclude that $L(\theta^*, p_{\theta^*}) \leq L(\theta^*, p^*)$. Hence, $p^* = p_{\theta^*}$ by Part 2 of Lemma 9. □

Remark 12. **(i)** Under the additional assumption Z, Part 1 of Theorem 11 is a direct consequence of Proposition 6 in Nickl (2007).

For later use we note the following:

(ii) Let Assumptions D and D.1 be satisfied, and $\inf_{x\in\Omega} p_0(x) > \chi$ for some $\chi \geq 0$. It follows from Part 1 of Theorem 11 that there are events $A_n \in \mathcal{B}(\Omega)^n$ that have probability tending to 1 as $n \to \infty$ on which $\inf_{x\in\Omega} \hat{p}_n(x) > \chi$ holds true.

(iii) Let Assumptions P.1 and R.1 be satisfied, and $\inf_{\Omega\times\Theta} p(x, \theta) > \chi$ for some $\chi \geq 0$. It follows from Part 3 of Theorem 11 that there are events $B_k \in \mathcal{V}^k$ that have probability tending to 1 as $k \to \infty$ on which $\inf_{\theta\in\Theta} \inf_{x\in\Omega} \tilde{p}_k(\theta)(x) > \chi$ holds true.

5.3 Rates of convergence for AML-estimators

The main result (Theorem 17) in this section is a uniform Donsker-type theorem for AML-estimators. It states that the stochastic process $f \mapsto \sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda$, when indexed by $\Theta \times \mathcal{F}$ for appropriate classes $\mathcal{F}$ of functions, converges weakly in $\ell^\infty(\Theta \times \mathcal{F})$ to a μ-Brownian bridge. An important step in proving this result is to establish rates of convergence for $\tilde{p}_k(\theta)$ in different Sobolev norms that hold uniformly over Θ.

5.3.1 Rates of convergence in Sobolev norms

Rates of convergence (in certain norms) for non-parametric M-estimators, such as AML-estimators, can be obtained in several distance measures by controlling the difference between the defining random objective function and its limiting counterpart. Following the idea of van de Geer (1993), this difference can be viewed as an empirical process over a certain class of functions, thus allowing to control its size by applying results from the theory of empirical processes. Along these lines, it was shown in Nickl (2007) that

$$\|\hat{p}_n - p_0\|_{s,2} = O^*_{\mathbb{P}}(n^{-(t-s)/(2t+1)})$$

for every $0 \leq s \leq t$, provided Assumptions Z, D, and D.1 hold. Clearly, for each $\theta \in \Theta$, an analogous result can immediately be obtained for $\|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ from the result of Nickl (2007), but we will need the same rates to hold uniformly over the parameter space Θ. This leads to the problem of bounding an expression of the form

$$\mathrm{E}^* \sup_{\Theta\times\mathcal{F}} \left|\sqrt{k}(L_k(\theta, f) - L(\theta, f))\right|, \tag{11}$$

where $\mathcal{F}$ is some specific class of functions on Ω and E^* is the outer expectation. Letting

$$\mathcal{F}^* = \{f(\rho(\,\cdot\,, \theta)) : \theta \in \Theta,\ f \in \mathcal{F}\},$$

(11) can be rewritten as

$$\mathrm{E}^* \left\|\sqrt{k}(\mu_k - \mu)\right\|_{\mathcal{F}^*}$$

and then can be bounded by the bracketing metric integral of $\mathcal{F}^*$ (cf. Theorem 58). Thus, we are led to investigate the $\mathcal{L}^2(\mu)$-bracketing metric entropy of $\mathcal{F}^*$. We start with a lemma.

Lemma 13. *Let Δ be a bounded, open interval in $\mathbb{R}$. Let $\mathcal{F}$ be a bounded subset of the Sobolev space $\mathsf{W}_2^s(\Delta)$ of order $s > 1/2$. Then the sup-norm metric entropy of $\mathcal{F}$ satisfies the inequality*

$$H(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Delta), \|\cdot\|_\Delta) \lesssim \varepsilon^{-1/s}$$

for all $\varepsilon > 0$. Furthermore, $\mathcal{F}$ is uniformly Donsker.

Proof. [See Nickl & Pötscher (2007) for a definition of the following spaces: $\mathsf{B}_{2,2}^s(\mathbb{R})$, $\mathsf{B}_{2,2}^s(\mathbb{R})|\Delta$, and $\mathsf{B}_{2,2}^s(\mathbb{R}, \langle x\rangle^\beta)$.] Since $\mathsf{B}_{2,2}^s(\mathbb{R})|\Delta$ coincides with $\mathsf{W}_2^s(\Delta)$ up to equivalent norms (cf. Proposition 3.4.2 in Triebel, 1983), and since the restriction mapping $\mathsf{B}_{2,2}^s(\mathbb{R}) \to \mathsf{B}_{2,2}^s(\mathbb{R})|\Delta$ is a retraction by Theorem 4.2.2 in Triebel (1995), we obtain a section $\sigma : \mathsf{W}_2^s(\Delta) \to \mathsf{B}_{2,2}^s(\mathbb{R})$. Consequently, $\sigma(\mathcal{F})$ is bounded in $\mathsf{B}_{2,2}^s(\mathbb{R})$. Now, let h be a C^∞-function on $\mathbb{R}$ with compact support such that $h(x) = 1$ for $x \in \Delta$, and define $\mathcal{F}' = \{h\sigma(f) : f \in \mathcal{F}\}$. [Note that such a function exists since Δ is bounded.] Making use of the fact that $\mathsf{B}_{2,2}^s(\mathbb{R})$ is a multiplication algebra, we conclude that $\mathcal{F}'$ is bounded in $\mathsf{B}_{2,2}^s(\mathbb{R}, \langle x\rangle^\beta)$ for any $\beta \in \mathbb{R}$, hence in particular for $\beta > s - 1/2$. By the choice of h, $\mathcal{F} = \mathcal{F}'|\Delta$, and so

$$\begin{aligned} H(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Delta), \|\cdot\|_\Delta) &= H(\varepsilon, \mathcal{F}'|\Delta, \mathsf{W}_2^s(\Delta), \|\cdot\|_\Delta) \\ &\leq H(\varepsilon, \mathcal{F}', \mathsf{B}_{2,2}^s(\mathbb{R}), \|\cdot\|_\mathbb{R}) \lesssim \varepsilon^{-1/s}, \end{aligned}$$

where the last inequality follows from Part 1 of Theorem 1 in Nickl & Pötscher (2007).

To prove the uniform Donsker property of $\mathcal{F}$, we verify the premises of Theorem 2.8.4 in van der Vaart & Wellner (1996): The envelope condition is trivially satisfied since $\mathcal{F}$ has a constant measurable envelope, whereas the condition on the convergence of the bracketing integral is a direct consequence of the first part of the lemma, upon noting that, by boundedness of $\mathcal{F}$ in $\mathsf{W}_2^s(\Omega)$, $H(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Omega), \|\cdot\|_\Delta)$ is 0 for ε large enough. □

Proposition 14. *Let $\mathcal{F}$ be a bounded subset of some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$.*

1. *Define the class of functions*

$$\mathcal{F}^* = \{f(\rho(\,\cdot\,, \theta)) : \theta \in \Theta,\ f \in \mathcal{F}\}.$$

If Assumption R.2 holds, then the $\mathcal{L}^2(\mu)$-bracketing metric entropy of $\mathcal{F}^*$ satisfies the inequality

$$H_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu}) \lesssim \varepsilon^{-1/s} \tag{12}$$

for all $\varepsilon > 0$. Furthermore, $\mathcal{F}^*$ is μ-Donsker.

2. *Let the elements of $\mathcal{F}$ be bounded below by some $\chi > 0$. Define the class of functions*

$$\log \mathcal{F}^* = \{\log f(\rho(\,\cdot\,, \theta)) : \theta \in \Theta,\ f \in \mathcal{F}\}.$$

If Assumption R.2 holds, then the $\mathcal{L}^2(\mu)$-bracketing metric entropy of $\log \mathcal{F}^*$ satisfies the inequality

$$H_{[\,]}(\varepsilon, \log \mathcal{F}^*, \|\cdot\|_{2,\mu}) \lesssim \varepsilon^{-1/s}$$

for all $\varepsilon > 0$.

Proof. Part 1: We have that

$$\begin{aligned}\sup_{f\in\mathcal{F}} |f(\rho(v,\theta')) - f(\rho(v,\theta))| &\leq L_s|\rho(v,\theta') - \rho(v,\theta)|^{\min(s-1/2,1)} \\ &\qquad \text{(for some } L_s > 0) \\ &\leq L_s R(v)\|\theta' - \theta\|^{\gamma\min(s-1/2,1)}\end{aligned}$$

for all $v \in V$ and all $\theta, \theta' \in \Theta$, where we made use of Assumption R.2 and the fact that $\mathcal{F}$ is a bounded subset of $\mathsf{C}^{\min(s-1/2,1)}(\Omega)$ (see Part 2 of Proposition 1). A cover of $\mathcal{F}^*$ is obtained from suitable covers of Θ and $\mathcal{F}$ in the following way: Fix $\varepsilon > 0$ and set $\delta(\varepsilon) = (\varepsilon/L_s)^{1/\nu}$, where $\nu := \gamma\min(s-1/2,1)$. To cover the space Θ, note that it is contained in an m-cube of edge length l and thus in the union of $\lceil l\sqrt{m}/\delta(\varepsilon)\rceil^m$-many closed balls $B(\theta_i, \delta(\varepsilon))$ with centers $\theta_i \in \Theta$ and Euclidean radius $\delta(\varepsilon)$. [Here, $\lceil l\sqrt{m}/\delta(\varepsilon)\rceil$ denotes the smallest integer not less than $l\sqrt{m}/\delta(\varepsilon)$.] To cover $\mathcal{F}$, we take $N(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Omega), \|\cdot\|_\Omega)$-many sup-norm closed balls $[f_j - 2\varepsilon, f_j + 2\varepsilon]$ of radius 2ε whose centers f_j already belong to $\mathcal{F}$. [Note that this can always be achieved.] We claim that the family of brackets

$$\begin{aligned}\Big\{&[f_j(\rho(\,\cdot\,,\theta_i)) - R(\,\cdot\,)\varepsilon - 2\varepsilon, f_j(\rho(\,\cdot\,,\theta_i)) + R(\,\cdot\,)\varepsilon + 2\varepsilon] : \\ &\qquad i = 1,\ldots,\lceil l\sqrt{m}/\delta(\varepsilon)\rceil^m,\, j = 1,\ldots,N(\varepsilon,\mathcal{F},\mathsf{W}_2^s(\Omega),\|\cdot\|_\Omega)\Big\}\end{aligned} \tag{13}$$

is a cover of $\mathcal{F}^*$. To prove this claim, let $h \in \mathcal{F}^*$, that is, $h = f(\rho(\,\cdot\,,\theta))$ for some $\theta \in \Theta$ and $f \in \mathcal{F}$, implying that there are indices i, j such that $\theta \in B(\theta_i, \delta(\varepsilon))$ and $f \in [f_j - 2\varepsilon, f_j + 2\varepsilon]$. Consequently,

$$h \in [f_j(\rho(\,\cdot\,,\theta)) - 2\varepsilon, f_j(\rho(\,\cdot\,,\theta)) + 2\varepsilon].$$

Now,

$$\begin{aligned}h(v) \leq f_j(\rho(v,\theta)) + 2\varepsilon &\leq f_j(\rho(v,\theta_i)) + |f_j(\rho(v,\theta)) - f_j(\rho(v,\theta_i))| + 2\varepsilon \\ &\leq f_j(\rho(v,\theta_i)) + R(v)\varepsilon + 2\varepsilon\end{aligned}$$

for all $v \in V$, where the last inequality follows from the first display in the proof and the choice of $\delta(\varepsilon)$. Similarly,

$$f_j(\rho(v,\theta_i)) - R(v)\varepsilon - 2\varepsilon \leq h(v).$$

Making use of Assumption R.2 and the norm inequality $\int_V |f|\, d\mu \leq \|f\|_{2,\mu}$, the $\mathcal{L}^2(\mu)$-bracketing size of any of the brackets in (13) can be bounded by ε times a constant that only depends on $R(\,\cdot\,)$ and μ. This leads to the relationship

$$N_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu}) \leq \max(1, 2^m l^m \sqrt{m}^{\,m} L_s^{m/\nu}\varepsilon^{-m/\nu})\, N(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Omega), \|\cdot\|_\Omega)$$

between the $\mathcal{L}^2(\mu)$-bracketing number of $\mathcal{F}^*$ and the sup-norm covering number of $\mathcal{F}$. Apply Lemma 13 with $\Delta = \Omega$ to get

$$H_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu}) \lesssim 1 + \max(0, \log\varepsilon^{-1}) + \varepsilon^{-1/s}.$$

It remains to show that $H_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu})$ can be bounded by $\varepsilon^{-1/s}$ up to some constant not depending on ε. To this end, we piece together the following observations: First, since $\mathcal{F}^*$ is bounded with respect to $\|\cdot\|_{2,\mu}$, it can be covered by a single bracket for $\varepsilon > \varepsilon_2$ for some suitable $\varepsilon_2 > 0$; hence $H_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu}) = 0$ for $\varepsilon > \varepsilon_2$. Second, there is an $\varepsilon_1 > 0$ such that $\varepsilon^{-1/s}$ dominates 1 and $\max(0, \log \varepsilon^{-1})$ on $(0, \varepsilon_1)$. Note finally that 1 and $\max(0, \log \varepsilon^{-1})$ can be dominated by some multiple of $\varepsilon^{-1/s}$ on the compact interval $[\varepsilon_1, \varepsilon_2]$ by monotonicity of the involved functions.

Since $\mathcal{F}^* \subseteq \mathcal{L}^2(V, \mathcal{V}, \mu)$, it follows from Ossiander's central limit theorem (see Theorem 7.2.1 in Dudley, 1999) that $\mathcal{F}^*$ is μ-Donsker.

Part 2: For any fixed $\varepsilon > 0$, we take for $\mathcal{F}^*$ the cover given in (13). Since the elements of $\mathcal{F}$ are bounded below by χ, the set of modified brackets

$$\Big\{[\max(\chi, f_j(\rho(\,\cdot\,,\theta_i)) - R(\,\cdot\,)\varepsilon - 2\varepsilon), f_j(\rho(\,\cdot\,,\theta_i)) + R(\,\cdot\,)\varepsilon + 2\varepsilon] :$$
$$i = 1, \ldots, \lceil l\sqrt{m}/\delta(\varepsilon)\rceil^m,\ j = 1, \ldots, N(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Omega), \|\cdot\|_\Omega)\Big\}$$

still covers $\mathcal{F}^*$. [Note that $\inf_{x\in\Omega} f_j(x) \geq \chi$ together with $R(v) > 0$ and $\varepsilon > 0$ implies that

$$f_j(\rho(v,\theta_i)) + R(v)\varepsilon + 2\varepsilon > \max(\chi, f_j(\rho(v,\theta_i)) - R(v)\varepsilon - 2\varepsilon)$$

for all $v \in V$.] Hence, by monotonicity of the logarithm and since $\chi > 0$, the brackets

$$[\log\max(\chi, f_j(\rho(\,\cdot\,,\theta_i)) - R(\,\cdot\,)\varepsilon - 2\varepsilon), \log(f_j(\rho(\,\cdot\,,\theta_i)) + R(\,\cdot\,)\varepsilon + 2\varepsilon)],$$

$i = 1, \ldots, \lfloor l\sqrt{m}/\delta(\varepsilon)\rfloor^m$, $j = 1, \ldots, N(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Omega), \|\cdot\|_\Omega)$, cover $\log \mathcal{F}^*$. Making use of the fact that the logarithm is Lipschitz on $[\chi, \infty)$ with Lipschitz constant χ^{-1}, the $\mathcal{L}^2(\mu)$-bracketing size of any bracket of the form

$$[\log\max(\chi, f_j(\rho(\,\cdot\,,\theta_i)) - R(\,\cdot\,)\varepsilon - 2\varepsilon), \log(f_j(\rho(\,\cdot\,,\theta_i)) + R(\,\cdot\,)\varepsilon + 2\varepsilon)]$$

can be bounded by χ^{-1} times the $\mathcal{L}^2(\mu)$-bracketing size of the corresponding bracket

$$[f_j(\rho(\,\cdot\,,\theta_i)) - R(\,\cdot\,)\varepsilon - 2\varepsilon, f_j(\rho(\,\cdot\,,\theta_i)) + R(\,\cdot\,)\varepsilon + 2\varepsilon].$$

Since these brackets are used to establish that

$$H_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu}) \lesssim \varepsilon^{-1/s}$$

in Part 1, the proof is complete. □

Part 1 of Proposition 14 provides a general condition under which the $\mathcal{L}^2(\mu)$-bracketing metric entropy of $\mathcal{F}^*$ can be bounded by some negative power of its $\mathcal{L}^2(\mu)$-bracketing size. In special situations, though, $\mathcal{F}^*$ itself can be identified as a bounded subset of some Sobolev space of order greater than 1/2. Lemma 13 can then be directly applied to $\mathcal{F}^*$ and (up to some constant not depending on the $\mathcal{L}^2(\mu)$-bracketing size) leads to the same bound for the $\mathcal{L}^2(\mu)$-bracketing metric entropy. This is exemplified in the following proposition.

Proposition 15. *Assume that* $\inf_{\Omega\times\Theta} p(x,\theta) > 0$ *and* $\sup_{\Omega\times\Theta} p(x,\theta) < \infty$. *Let* $(V, \mathcal{V}, \mu)$ *be the unit interval* $(0,1)$ *together with its Borel* σ*-algebra and Lebesgue measure* λ. *Take* $\rho(v,\theta)$ *to be the quantile function* $F^{-1}(v,\theta)$, *where* $F(v,\theta) := \int_{\{x\in\Omega:\, x\leq v\}} p(x,\theta)d\lambda(x)$.

Then, for any bounded subset $\mathcal{F}$ of some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order s such that $1/2 < s < 1$, the class of functions

$$\mathcal{F}^* := \{f(F^{-1}(\,\cdot\,,\theta)) : \theta \in \Theta,\ f \in \mathcal{F}\}$$

satisfies the inequality

$$H_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu}) \lesssim \varepsilon^{-1/s}$$

for all $\varepsilon > 0$. Furthermore, $\mathcal{F}^*$ is μ-Donsker.

Proof. Let $\xi = \inf_{\Omega\times\Theta} p(x,\theta)$ and $B = \sup_{\Omega\times\Theta} p(x,\theta)$. We first show that

$$|F^{-1}(u,\theta) - F^{-1}(v,\theta)| \le \xi^{-1}|u-v| \tag{14}$$

for all $u, v \in (0,1)$ and $\theta \in \Theta$. To this end, observe that the densities p_θ are everywhere positive, and so the c.d.f.s $F(\,\cdot\,,\theta)$ are invertible mappings from Ω to $(0,1)$. For $u, v \in (0,1)$ such that $u \neq v$ we then have

$$\frac{|F^{-1}(u,\theta) - F^{-1}(v,\theta)|}{|u-v|} = \left[\frac{|F(x,\theta) - F(y,\theta)|}{|x-y|}\right]^{-1}$$

for unique $x, y \in \Omega$. By the mean-value theorem, the r.h.s. of the previous display is equal to $1/p(z,\theta)$ for some z on the line segment joining x and y, and can thus be bounded by ξ^{-1}.

Since $1/2 < s < 1$, the squared Sobolev norm of any $f(F^{-1}(\,\cdot\,,\theta))$ in $\mathcal{F}^*$ can be bounded as follows:

$$\begin{aligned}
\int_{(0,1)} f(F^{-1}(v,\theta))^2 dv + \int_{(0,1)}\int_{(0,1)} \frac{|f(F^{-1}(u,\theta)) - f(F^{-1}(v,\theta))|^2}{|u-v|^{1+2s}}\,du\,dv \\
\le \int_{(0,1)} f(F^{-1}(v,\theta))^2 dv + \xi^{-1-2s} \\
\times \int_{(0,1)}\int_{(0,1)} \frac{|f(F^{-1}(u,\theta)) - f(F^{-1}(v,\theta))|^2}{|F^{-1}(u,\theta) - F^{-1}(v,\theta)|^{1+2s}}\,du\,dv \\
\text{(by (14))} \\
= \int_\Omega f(x)^2 p(x,\theta)dx \\
+\xi^{-1-2s}\int_\Omega\int_\Omega \frac{|f(x)-f(y)|^2}{|x-y|^{1+2s}} p(x,\theta)p(y,\theta)\,dx\,dy \\
\text{(by Tonelli's and the change of variables theorem)} \\
\le B\int_\Omega f(x)^2 dx + \xi^{-1-2s}B^2\int_\Omega\int_\Omega \frac{|f(x)-f(y)|^2}{|x-y|^{1+2s}}\,dx\,dy
\end{aligned}$$

Since $\mathcal{F}$ is a bounded subset of $\mathsf{W}_2^s(\Omega)$ by hypothesis, it follows that $\mathcal{F}^*$ is bounded in $\mathsf{W}_2^s(0,1)$. Apply Lemma 13 with $\Delta = (0,1)$ to $\mathcal{F}^*$ to complete the proof. $\square$

Theorem 16.

1. *Let Assumptions P.1, R.2, and Z be satisfied. Then*

$$\sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2} = O_\mu^*(k^{-(t-s)/(2t+1)}) \quad \text{as } k \to \infty$$

for every $0 \leq s \leq t$.

2. *Consider the auxiliary model* $\mathcal{P}(t,\zeta,D)$ *with* $\zeta = 0$. *Let Assumptions P.1, P.2, and R.2 be satisfied. Then*

$$\sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2} = O_\mu^*(k^{-(t-s)/(2t+1)}) \quad \text{as } k \to \infty$$

for every $0 \leq s \leq t$.

Proof. Part 1: We first prove the case when $s = 0$. To this end, we verify the conditions of Theorem 54 with $(\Lambda, \mathcal{A}, P) = (V^{\mathbb{N}}, \mathcal{V}^{\mathbb{N}}, \mu^{\mathbb{N}})$, $S = \Theta$, $T = \mathcal{P}(t,\zeta,D)$, $d(p,q) = \|p-q\|_2$, $H_k(s,t) = L_k(\theta,p)$, $H(s,t) = L(\theta,p)$, $\hat{t}_k(s) = \tilde{p}_k(\theta)$, and $t(s) = p_\theta$. Condition (54) is clear by definition of the AML-estimators $\tilde{p}_k(\theta)$. Condition (52) follows from the second-order Taylor expansion of $L(\theta,\,\cdot\,)$ around the density p_θ:

$$\begin{aligned}
L(\theta,p) - L(\theta,p_\theta) &= DL(\theta,p_\theta)(p-p_\theta) + \frac{1}{2}D^2L(\theta,\bar{p})(p-p_\theta,p-p_\theta) \\
&= \int_\Omega \frac{p-p_\theta}{p_\theta} p_\theta d\lambda - \frac{1}{2}\int_\Omega \frac{(p-p_\theta)^2}{\bar{p}^2} p_\theta d\lambda \\
&\qquad \text{(by Proposition 45)} \\
&\leq -\frac{1}{2}\zeta C_t^{-2} D^{-2} \|p-p_\theta\|_2^2,
\end{aligned}$$

where $\bar{p}$ is some density on the line segment joining p and p_θ; note that $\bar{p} \in \mathcal{P}(t,\zeta,D)$ by convexity of this set, and hence satisfies $\|\bar{p}\|_\Omega \leq C_t D$. This proves Condition (52) in Theorem 54 with $C = 2^{-1}\zeta C_t^{-2} D^{-2}$ and $\alpha = 2$, both constants being independent of θ and p.

Next is the verification of Condition (53). Letting

$$\mathcal{G}_\delta = \{\log p(\rho(\,\cdot\,,\theta)) - \log p_\theta(\rho(\,\cdot\,,\theta)) : \theta \in \Theta,\, p \in \mathcal{P}(t,\zeta,D),\, \|p-p_\theta\|_2 \leq \delta\}$$

for $\delta > 0$, the expression

$$\mathrm{E}^* \sup_{\theta\in\Theta} \sup_{\substack{p\in\mathcal{P}(t,\zeta,D),\\ \|p-p_\theta\|_2\leq\delta}} \left|\sqrt{k}(L_k - L)(\theta,p) - \sqrt{k}(L_k - L)(\theta,p_\theta)\right|$$

can be viewed as

$$\mathrm{E}^* \left\|\sqrt{k}(\mu_k - \mu)\right\|_{\mathcal{G}_\delta}.$$

So we are led to deriving an upper bound for the expected variation of the stochastic process $\sqrt{k}(\mu_k - \mu)$ over the class of functions $\mathcal{G}_\delta$. For this, we verify the premises of Theorem 58: the boundedness of $\mathcal{G}_\delta$ with respect to the sup-norm as well as the $\mathcal{L}^2(\mu)$-norm. By Assumption Z, the logarithm is Lipschitz on $[\zeta,\infty)$ with Lipschitz constant

ζ^{-1}. Use this to see that $\mathcal{G}_\delta$ is bounded by $B := 2\zeta^{-1}C_t D$ in the sup-norm and by $\eta(\delta) := \zeta^{-1}C_t^{1/2}D^{1/2}\delta$ in the $\mathcal{L}^2(\mu)$-norm, so that

$$\begin{aligned} &\mathrm{E}^* \left\| \sqrt{k}(\mu_k - \mu) \right\|_{\mathcal{G}_\delta} \\ &\quad \le (1696 + 64\sqrt{2})\, I_{[\,]}(\eta(\delta), \mathcal{G}_\delta, \|\cdot\|_{2,\mu}) \left[1 + \frac{B}{\eta(\delta)^2\sqrt{k}} I_{[\,]}(\eta(\delta), \mathcal{G}_\delta, \|\cdot\|_{2,\mu}) \right] \end{aligned}$$

in view of Theorem 58. Since

$$\begin{aligned} \mathcal{G}_\delta &\subseteq \{\log p(\rho(\,\cdot\,,\theta)) - \log p_\theta(\rho(\,\cdot\,,\theta)) : \theta \in \Theta,\, p \in \mathcal{P}(t,\zeta,D)\} \\ &\subseteq \{\log p(\rho(\,\cdot\,,\theta)) : \theta \in \Theta,\, p \in \mathcal{P}(t,\zeta,D)\} \\ &\qquad - \{\log p(\rho(\,\cdot\,,\theta)) : \theta \in \Theta,\, p \in \mathcal{P}(t,\zeta,D)\}, \end{aligned}$$

we have that

$$N_{[\,]}(\varepsilon, \mathcal{G}_\delta, \|\cdot\|_{2,\mu}) \le N_{[\,]}(\varepsilon, \{\log p(\rho(\,\cdot\,,\theta)) : \theta \in \Theta,\, p \in \mathcal{P}(t,\zeta,D)\}, \|\cdot\|_{2,\mu})^2.$$

Use this inequality and apply Part 2 of Proposition 14 with $s = t$ and $\mathcal{F} = \mathcal{P}(t,\zeta,D)$ to get

$$\begin{aligned} I_{[\,]}(\eta(\delta), \mathcal{G}_\delta, \|\cdot\|_{2,\mu}) &\lesssim I_{[\,]}(\eta(\delta), \log \mathcal{F}^*, \|\cdot\|_{2,\mu}) \\ &\lesssim \int_{(0,\eta(\delta)]} \sqrt{1 + \varepsilon^{-1/t}}\, d\varepsilon \\ &\lesssim \max(\eta(\delta), \int_{(0,\eta(\delta)]} \varepsilon^{-1/2t}\, d\varepsilon) \\ &\lesssim \max(\delta, \delta^{1-1/2t}). \end{aligned}$$

Hence there is some constant L, $0 < L < \infty$, such that

$$\mathrm{E}^* \left\| \sqrt{k}(\mu_k - \mu) \right\|_{\mathcal{G}_\delta} \le L \max(\delta, \delta^{1-1/2t}) \left[1 + \frac{\max(\delta, \delta^{1-1/2t})}{\delta^2\sqrt{k}} \right]$$

holds for all $\delta > 0$. Write $\varphi_k(\delta)$ for the r.h.s. of the last display and note that $\delta \mapsto \delta^{-\beta}\varphi_k(\delta)$ is non-increasing for $\beta = 1$. This establishes Condition (53) in Theorem 54.

Condition (55), with $\alpha = 2$, is satisfied for $r_k = k^{t/(2t+1)}$. This gives the desired rate and completes the proof in case $s = 0$.

In case $0 < s \le t$, we make use of the fact that $\sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{t,2} \le 2D$, and of the interpolation inequality

$$\|f\|_{s,2} \le C_{s,t}\, \|f\|_{t,2}^{s/t}\, \|f\|_2^{(t-s)/t}$$

for $f \in \mathsf{W}_2^t(\Omega)$, where $C_{s,t} > 0$; see Theorem 1.9.6 and Remark 1.9.1 in Lions & Magenes (1972).

Part 2: The case $s = t$ is clear in view of Assumption P.1 and the definition of $\tilde{p}_k(\theta)$, so assume $0 \le s < t$. Choose $\chi > 0$ such that $\inf_{\Omega\times\Theta} p(x,\theta) > \chi$. Then, by Remark 12(ii), there are events that have probability tending to 1 on which $\inf_{\theta\in\Theta} \inf_{x\in\Omega} \tilde{p}_k(\theta)(x) > \chi$ holds true. Since $\mathcal{P}(t,\chi,D) \subset \mathcal{P}(t,\zeta,D)$, we have that on these events $\tilde{p}_k(\theta)$ coincides with the AML-estimators over the auxiliary model $\mathcal{P}(t,\chi,D)$. Part 2 now follows from Part 1. $\square$

5.3.2 A uniform Donsker-type theorem for AML-estimators

It was shown in Nickl (2007) that, under Assumptions Z, D, and D.3, the stochastic process $f \mapsto \sqrt{n}\int_\Omega (\hat{p}_n - p_0) f d\lambda$, when indexed by a bounded subset $\mathcal{F}$ of some Sobolev space of order $s > 1/2$, converges weakly in $\ell^\infty(\mathcal{F})$ to a centered Gaussian process. Of course, this result immediately gives that, for every *fixed* $\theta \in \Theta$, $f \mapsto \sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda$ converges weakly in $\ell^\infty(\mathcal{F})$. It is therefore natural to ask whether there can even be obtained a weak limit theorem for the stochastic process $(\theta, f) \mapsto \sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda$ in the space $\ell^\infty(\Theta \times \mathcal{F})$. This question will be answered in the positive by the following theorem.

Theorem 17. *Let Assumptions P.3, R.2, and Z be satisfied. Let $\mathcal{F}$ be a non-empty bounded subset of some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$. Then the following statements hold true:*

1. *For all $j > 1/2$,*

$$\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda - \sqrt{k}(\mu_k - \mu) f(\rho(\,\cdot\,,\theta))\right| = o_\mu^*(k^{-(\min(s,t)-j)/(2t+1)}) \quad \text{as } k \to \infty;$$

   in particular,

$$\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda - \sqrt{k}(\mu_k - \mu) f(\rho(\,\cdot\,,\theta))\right| = o_\mu^*(1) \quad \text{as } k \to \infty.$$

2. *Let*

$$\mathcal{F}^* = \{f(\rho(\,\cdot\,,\theta)) : \theta \in \Theta,\ f \in \mathcal{F}\}\,.$$

   Then there exists a μ-Brownian bridge $\mathbb{G}$ indexed by $\mathcal{F}^*$, that is, $\mathbb{G}$ is a centered Gaussian process indexed by $\mathcal{F}^*$, which is measurable as a mapping with values in $\ell^\infty(\mathcal{F})$, and has covariance function

$$\begin{aligned}&\mathrm{Cov}[\mathbb{G}(f(\rho(\,\cdot\,,\theta))), \mathbb{G}(g(\rho(\,\cdot\,,\theta')))]\\ &\quad = \int_V \left(f(\rho(\,\cdot\,,\theta)) - \int_V f(\rho(\,\cdot\,,\theta))d\mu\right)\left(g(\rho(\,\cdot\,,\theta')) - \int_V g(\rho(\,\cdot\,,\theta'))d\mu\right)d\mu\end{aligned}$$

   and bounded sample paths that are uniformly continuous with respect to the pseudo-metric

$$\begin{aligned}&d(f(\rho(\,\cdot\,,\theta)), g(\rho(\,\cdot\,,\theta')))\\ &\quad = \left[\int_V \left(f(\rho(\,\cdot\,,\theta)) - g(\rho(\,\cdot\,,\theta')) - \int_V (f(\rho(\,\cdot\,,\theta)) - g(\rho(\,\cdot\,,\theta')))d\mu\right)^2 d\mu\right]^{1/2},\end{aligned}$$

   such that the stochastic process $(\theta, f) \mapsto \sqrt{k}\int_V (\tilde{p}_k(\theta) - p_\theta) f d\lambda$ converges weakly to $\mathbb{G}(\theta, f)$ in $\ell^\infty(\Theta \times \mathcal{F})$, where $\mathbb{G}(\theta, f) := \mathbb{G}(f(\rho(\,\cdot\,,\theta)))$ for every $\theta \in \Theta$ and every $f \in \mathcal{F}$, and $\mathbb{G}(\,\bullet\,,\,\cdot\,)$ is measurable as a mapping with values in $\ell^\infty(\Theta \times \mathcal{F})$.

3. *We have that*

$$\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda\right| = O_\mu^*(1) \quad \text{as } k \to \infty.$$

Proof. Part 1: The proof strategy for Part 1 is taken from Nickl (2007).

Step 1: We first consider the case $s \geq t$. By Part 3 of Proposition 1, $\mathcal{F}$ is then a bounded subset of $\mathsf{W}_2^t(\Omega)$ and therefore contained in $\mathcal{U}_{t,B}$ for some B, $0 < B < \infty$. W.l.o.g. we may therefore assume that $\mathcal{F}$ equals $\mathcal{U}_{t,B}$.

We define $\pi_\theta(f) = (f - \int_\Omega f p_\theta d\lambda)p_\theta$ for any function $f \in \mathsf{W}_2^t(\Omega)$. Then, for every $\theta \in \Theta$ and $f \in \mathcal{F}$,

$$\begin{aligned}
\|\pi_\theta(f)\|_{t,2} &\leq M_t \left\| f - \int_\Omega f p_\theta \, d\lambda \right\|_{t,2} \|p_\theta\|_{t,2} \\
&\qquad \text{(by Part 1 of Proposition 1)} \\
&\leq M_t D \left[\|f\|_{t,2} + \left\| \int_\Omega f p_\theta \, d\lambda \right\|_{t,2} \right] \\
&\qquad \text{(since } p_\theta \in \mathcal{P}(t, \zeta, D)\text{)} \\
&\leq M_t D B (1 + C_t \|1\|_{t,2}) \\
&\qquad \text{(since } \mathcal{F} = \mathcal{U}_{t,B}\text{)} \\
&\leq M_t D B (1 + C_t \lambda(\Omega)^{1/2}) < \infty.
\end{aligned}$$

Hence,

$$\sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \|\pi_\theta(f)\|_{t,2} \leq M_t D B (1 + C_t \lambda(\Omega)^{1/2}) < \infty. \tag{15}$$

From Proposition 45 we obtain the relations

$$\begin{aligned}
\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda &= -D^2 L(\theta, p_\theta)(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) \\
(\mu_k - \mu) f(\rho(\,\cdot\,, \theta)) &= D L_k(\theta, p_\theta)(\pi_\theta(f)),
\end{aligned}$$

so that

$$\begin{aligned}
&\int (\tilde{p}_k(\theta) - p_\theta) f d\lambda - (\mu_k - \mu) f(\rho(\,\cdot\,, \theta)) \\
&\quad = -D^2 L(\theta, p_\theta)(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) - D L_k(\theta, p_\theta)(\pi_\theta(f)).
\end{aligned} \tag{16}$$

Applying the pathwise mean-value theorem to the function $DL_k(\theta, \cdot\,)(\pi_\theta(f))$ gives

$$DL_k(\theta, \tilde{p}_k(\theta))(\pi_\theta(f)) = DL_k(\theta, p_\theta)(\pi_\theta(f)) + D^2 L_k(\theta, \bar{p}_k(\theta))(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)),$$

where $\bar{p}_k(\theta) = \xi(\theta, f)\tilde{p}_k(\theta) + (1 - \xi(\theta, f))p_\theta$ for some $\xi(\theta, f) \in (0, 1)$. Hence

$$\begin{aligned}
&D^2 L(\theta, p_\theta)(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) \\
&\quad = (D^2 L(\theta, p_\theta) - D^2 L_k(\theta, \bar{p}_k(\theta)))(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) \\
&\qquad + D^2 L_k(\theta, \bar{p}_k(\theta))(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) \\
&\quad = (D^2 L(\theta, p_\theta) - D^2 L_k(\theta, \bar{p}_k(\theta)))(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) \\
&\qquad + DL_k(\theta, \tilde{p}_k(\theta))(\pi_\theta(f)) - DL_k(\theta, p_\theta)(\pi_\theta(f)) \\
&\quad = (D^2 L(\theta, p_\theta) - D^2 L(\theta, \bar{p}_k(\theta)))(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) \\
&\qquad + (D^2 L(\theta, \bar{p}_k(\theta)) - D^2 L_k(\theta, \bar{p}_k(\theta)))(\tilde{p}_k(\theta) - p_\theta, \pi_\theta(f)) \\
&\qquad + DL_k(\theta, \tilde{p}_k(\theta))(\pi_\theta(f)) - DL_k(\theta, p_\theta)(\pi_\theta(f)),
\end{aligned}$$

which together with (16) leads to the following inequality:

$$\begin{aligned}
&\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\int_\Omega(\tilde{p}_k(\theta)-p_\theta)f\,d\lambda-(\mu_k-\mu)f(\rho(\,\cdot\,,\theta))\right|\\
&\quad\le\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|(D^2L(\theta,p_\theta)-D^2L(\theta,\bar{p}_k(\theta)))(\tilde{p}_k(\theta)-p_\theta,\pi_\theta(f))\right|\\
&\qquad+\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|(D^2L(\theta,\bar{p}_k(\theta))-D^2L_k(\theta,\bar{p}_k(\theta)))(\tilde{p}_k(\theta)-p_\theta,\pi_\theta(f))\right|\\
&\qquad+\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|DL_k(\theta,\tilde{p}_k(\theta))(\pi_\theta(f))\right|\\
&\quad=\mathrm{I}+\mathrm{II}+\mathrm{III}.
\end{aligned}$$

To complete the proof of Part 1 when $s\ge t$, we first consider only j such that $1/2<j<t$, and derive bounds for each of the above expressions.

Bound for I: Use Proposition 45 to compute Expression I and then bound it by

$$\begin{aligned}
&2\zeta^{-3}C_tD\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}[\|\pi_\theta(f)\|_2\,\|(\bar{p}_k(\theta)-p_\theta)(\tilde{p}_k(\theta)-p_\theta)\|_2]\\
&\qquad\text{(by the Cauchy-Schwarz inequality)}\\
&=2\zeta^{-3}C_tD\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left[\|\pi_\theta(f)\|_2\,\xi(\theta,f)\|\tilde{p}_k(\theta)-p_\theta\|_2^2\right]\\
&\qquad\text{(since }\bar{p}_k(\theta)=\xi(\theta,f)(\tilde{p}_k(\theta)-p_\theta))\\
&\le2\zeta^{-3}C_tM_tD^2B(1+C_t\lambda(\Omega)^{1/2})\sup_{\theta\in\Theta}\|\tilde{p}_k(\theta)-p_\theta\|_2^2\\
&\qquad\text{(by (15))}.
\end{aligned}$$

Part 1 of Theorem 16 implies that the r.h.s. of the last display, and hence Expression I, is $O^*_\mu(k^{-2t/(2t+1)})$.

Bound for II: Clearly,

$$\mathrm{II}\le\sup_{\theta\in\Theta}\|\tilde{p}_k(\theta)-p_\theta\|_{j,2}\sup_{\Theta\times\mathcal{P}(t,\zeta,D)}\left\|D^2L(\theta,p)-D^2L_k(\theta,p)\right\|_{\mathcal{U}_{j,1}\times\{\pi_\theta(f):\theta\in\Theta,\,f\in\mathcal{F}\}}.$$

The first supremum in the last display is $O^*_\mu(k^{-(t-j)/(2t+1)})$ by Part 1 of Theorem 16. Since the set $\{\pi_\theta(f):\theta\in\Theta,\,f\in\mathcal{F}\}$ is bounded in $\mathsf{W}^j_2(\Omega)$ by (15) and Part 3 of Proposition 1, and $\mathcal{U}_{j,1}$ is so too, the second supremum in the last display is $O_\mu(k^{-1/2})$ by Proposition 46, when applied with $\alpha=2$, $\mathcal{H}_1=\mathcal{U}_{j,1}$, and $\mathcal{H}_2=\{\pi_\theta(f):\theta\in\Theta,\,f\in\mathcal{F}\}$. Hence, Expression II is $O^*_\mu(k^{-(t-j)/(2t+1)-1/2})$.

Bound for III: Let

$$\eta\;=\;D-\sup_{\theta\in\Theta}\|p_\theta\|_{t,2},$$

which is positive by Assumption P.3,

$$\mathcal{U}_{t,\eta,0}=\left\{w\in\mathsf{W}^t_2(\Omega):\|w\|_{t,2}\le\eta,\;\int_\Omega w\,d\lambda=0\right\},$$

and

$$\tilde{h}_k(w,\theta,\varepsilon)=\varepsilon(p_\theta+w)+(1-\varepsilon)\tilde{p}_k(\theta)\quad\text{for }w\in\mathsf{W}^t_2(\Omega)\text{ and }0<\varepsilon\le1.$$

We show that, on events that have probability tending to 1, the inclusion

$$\{\tilde{h}_k(w,\theta,\varepsilon):\theta\in\Theta,\,w\in\mathcal{U}_{t,\eta,0}\}\subseteq\mathcal{P}(t,\zeta,D)$$

holds for every ε, $0 < \varepsilon \le \varepsilon^*$, where ε^* is non-random and does not depend on k. First,

$$\left\|\tilde{h}_k(w,\theta,\varepsilon)\right\|_{t,2} \le D$$

by the triangle inequality. Second, by Assumption P.3, there is a $\beta > 0$ such that $\inf_{\Omega\times\Theta} p(x,\theta) > \zeta + \beta$ holds. By Remark 12(iii), $\inf_{\Omega\times\Theta} \tilde{p}_k(x,\theta) > \zeta + \beta$ on events that have probability going to 1. On these events, $\tilde{h}_k(w,\theta,\varepsilon) \ge \zeta + \beta - \varepsilon C_t \eta$ (use that $w \ge -\|w\|_\Omega \ge -C_t\|w\|_{t,2} \ge -C_t\eta$), and therefore $\tilde{h}_k(w,\theta,\varepsilon) \ge \zeta$ if only $\varepsilon > 0$ satisfies $\varepsilon \le \varepsilon^* := \beta/C_t\eta$. Note that $\tilde{h}_k(w,\theta,\varepsilon)$ integrates to 1 since $\int_\Omega w\, d\lambda = 0$.

Next, define $s : \mathsf{W}_2^t(\Omega) \to \mathsf{W}_2^t(\Omega)$ by

$$s(f) = \begin{cases} \eta f \|f\|_{t,2}^{-1} & \text{if } f \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that $\{s(\pi_\theta(f)) : \theta \in \Theta,\ f \in \mathcal{F}\} \subseteq \mathcal{U}_{t,\eta,0}$. Since $\tilde{p}_k(\theta)$ maximizes $L_k(\theta, \cdot)$ over the auxiliary model $\mathcal{P}(t,\zeta,D)$ and, for all ε such that $0 < \varepsilon \le \varepsilon^*$, $\tilde{h}_k(s(\pi_\theta(f)),\theta,\varepsilon)$ belongs to $\mathcal{P}(t,\zeta,D)$ on events that have probability tending to 1, we conclude that on these events

$$DL_k(\theta,\tilde{p}_k(\theta))(p_\theta - \tilde{p}_k(\theta) + s(\pi_\theta(f))) \le 0$$

holds for all $\theta \in \Theta$ and $f \in \mathcal{F}$ (see the erratum to Nickl, 2007). This implies

$$\begin{aligned} DL_k(\theta,\tilde{p}_k(\theta))(s(\pi_\theta(f))) &\le DL_k(\theta,\tilde{p}_k(\theta))(\tilde{p}_k(\theta) - p_\theta) \\ &= (DL_k(\theta,\tilde{p}_k(\theta)) - DL(\theta,p_\theta))(\tilde{p}_k(\theta) - p_\theta) \end{aligned}$$

for all $\theta \in \Theta$ and $f \in \mathcal{F}$, where the equality is true since $DL(\theta,p_\theta)(\tilde{p}_k(\theta) - p_\theta) = 0$, as can be seen using Proposition 45. Since $\mathcal{U}_{t,\eta,0}$ also contains $-s(\pi_\theta(f))$, we can replace the l.h.s. in the last display by its absolute value. Hence

$$\begin{aligned} &\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}} |DL_k(\theta,\tilde{p}_k(\theta))(s(\pi_\theta(f)))| \\ &\quad \le \sup_{\theta\in\Theta} |(DL_k(\theta,\tilde{p}_k(\theta)) - DL(\theta,\tilde{p}_k(\theta)))(\tilde{p}_k(\theta) - p_\theta)| \\ &\qquad + \sup_{\theta\in\Theta} |(DL(\theta,\tilde{p}_k(\theta)) - DL(\theta,p_\theta))(\tilde{p}_k(\theta) - p_\theta)| \\ &\quad \le \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{j,2} \sup_{\Theta\times\mathcal{P}(t,\zeta,D)} \|DL_k(\theta,p) - DL(\theta,p)\|_{\mathcal{U}_{j,1}} \\ &\qquad + \zeta^{-1} \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_2^2, \end{aligned}$$

where we used Proposition 45. Use Part 1 of Theorem 16 and Proposition 46 with $\alpha = 1$ and $\mathcal{H}_1 = \mathcal{U}_{j,1}$ to conclude that the r.h.s. of the last display is

$$O_\mu^*(k^{-(t-j)/(2t+1)-1/2}) + O_\mu^*(k^{-2t/(2t+1)}),$$

hence $O_\mu^*(k^{-(t-j)/(2t+1)-1/2})$ as $j > 1/2$. Together with (15) and the definition of $s(\cdot)$ this results in

$$\mathrm{III} = O_\mu^*(k^{-(t-j)/(2t+1)-1/2}).$$

The above bounds imply that

$$\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}} \left| \sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda - \sqrt{k}(\mu_k - \mu) f(\rho(\cdot,\theta)) \right| = O_\mu^*(k^{-(t-j)/(2t+1)})$$

for every j, $1/2 < j < t$, and hence a fortiori for all $j > 1/2$. Consequently, the expression on the l.h.s. in the last display is $o^*_\mu(k^{-(t-j)/(2t+1)})$ for all $j > 1/2$, hence in particular $o^*_\mu(1)$ as $k \to \infty$. This completes the proof of Part 1 when $s \geq t$.

Step 2: We now consider the case when $1/2 < s < t$, and assume w.lo.g. that $\mathcal{F} = \mathcal{U}_{s,B}$ for some B, $0 < B < \infty$. In the proof of Proposition 1 in Nickl (2007) there are defined approximating sequences $u_k(f) \in \mathsf{W}^t_2(\Omega)$ for all $f \in \mathcal{F}$, with the following properties:

$$\sup_{f\in\mathcal{F}} \|u_k(f)\|_{t,2} = O(k^{(t-s)/(2t+1)}) \quad \text{as } k \to \infty, \tag{17}$$

where $\sup_{f\in\mathcal{F}} \|u_k(f)\|_{t,2}$ is real for every $k \in \mathbb{N}$; and, for every r, $0 \leq r < s$,

$$\sup_{f\in\mathcal{F}} \|f - u_k(f)\|_{r,2} = O(k^{-(s-r)/(2t+1)}) \quad \text{as } k \to \infty. \tag{18}$$

We have that

$$\begin{aligned}
&\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}\int_\Omega(\tilde{p}_k(\theta)-p_\theta)fd\lambda - \sqrt{k}(\mu_k-\mu)f(\rho(\,\cdot\,,\theta))\right| \\
&\quad\leq \sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}\int_\Omega(\tilde{p}_k(\theta)-p_\theta)(f-u_k(f))d\lambda\right| \\
&\qquad+\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}(\mu_k-\mu)(f(\rho(\,\cdot\,,\theta))-u_k(f)(\rho(\,\cdot\,,\theta)))\right| \\
&\qquad+\sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}\int_\Omega(\tilde{p}_k(\theta)-p_\theta)u_k(f)d\lambda - \sqrt{k}(\mu_k-\mu)u_k(f)(\rho(\,\cdot\,,\theta))\right| \\
&\quad= \mathrm{IV}+\mathrm{V}+\mathrm{VI}.
\end{aligned}$$

To complete the proof of Part 1 when $1/2 < s < t$, we choose an arbitrary j' such that $1/2 < j' < \min(j, s)$, where j is as in the theorem, and derive bounds for each of the above expressions.

Bound for IV: By using (18), with $r = 0$, and Part 1 of Theorem 16,

$$\begin{aligned}
\mathrm{IV} &\leq \sqrt{k}\sup_{f\in\mathcal{F}}\|f-u_k(f)\|_2 \sup_{\theta\in\Theta}\|\tilde{p}_k(\theta)-p_\theta\|_2 \\
&\qquad\text{(by the Cauchy-Schwarz inequality)} \\
&= O^*_\mu(k^{-(s-1/2)/(2t+1)}).
\end{aligned}$$

Bound for V: Observe that

$$\begin{aligned}
\mathrm{V} &= \sup_{\theta\in\Theta}\sup_{f\in\mathcal{F}}\left|\sqrt{k}(\mu_k-\mu)(f-u_k(f))(\rho(\,\cdot\,,\theta))\right| \\
&\leq \sup_{\theta\in\Theta}\sup_{h\in\mathcal{U}_{j',1}}\left|\sqrt{k}(\mu_k-\mu)h(\rho(\,\cdot\,,\theta))\right| \times \sup_{f\in\mathcal{F}}\|f-u_k(f)\|_{j',2} \\
&= \|\sqrt{k}(\mu_k-\mu)\|_{\mathcal{U}^*_{j',1}} \sup_{f\in\mathcal{F}}\|f-u_k(f)\|_{j',2},
\end{aligned} \tag{19}$$

where

$$\mathcal{U}^*_{j',1} = \left\{h(\rho(\,\cdot\,,\theta)) : \theta\in\Theta,\, h\in\mathcal{U}_{j',1}\right\}.$$

Since $j' > 1/2$, the class of functions $\mathcal{U}^*_{j',1}$ is μ-Donsker by Part 1 of Proposition 14, hence

$$\left\| \sqrt{k}(\mu_k - \mu) \right\|_{\mathcal{U}^*_{j',1}} = O_\mu(1)$$

in view of Prohorov's theorem. Making use of (18), it follows that the r.h.s. of (19), and hence Expression V, is $O_\mu(k^{-(s-j')/(2t+1)})$.

Bound for VI: Note that Expression VI is bounded by

$$\begin{aligned}
&\sup_{\theta\in\Theta} \sup_{\substack{f\in\mathcal{F},\\ \|u_k(f)\|_{t,2}\neq 0}} \Bigg[\left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) \frac{u_k(f)}{\|u_k(f)\|_{t,2}} d\lambda - \sqrt{k}(\mu_k - \mu) \frac{u_k(f)}{\|u_k(f)\|_{t,2}} (\rho(\,\cdot\,,\theta)) \right| \\
&\qquad \times \|u_k(f)\|_{t,2} \Bigg] \\
&\quad \leq \sup_{\theta\in\Theta} \sup_{h\in\mathcal{U}_{t,1}} \left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) h d\lambda - \sqrt{k}(\mu_k - \mu) h(\rho(\,\cdot\,,\theta)) \right| \times \sup_{f\in\mathcal{F}} \|u_k(f)\|_{t,2}.
\end{aligned}$$

Applying Step 1 to the first term on the r.h.s. of the last display, and using (17), we conclude that

$$\mathrm{VI} = O^*_\mu(k^{-(s-j)/(2t+1)})$$

for all $j > 1/2$.

The above bounds imply that

$$\sup_{\theta\in\Theta} \sup_{f\in\mathcal{F}} \left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda - \sqrt{k}(\mu_k - \mu) f(\rho(\,\cdot\,,\theta)) \right| = O^*_\mu(k^{-(s-j)/(2t+1)})$$

for all $j > 1/2$. Consequently, the expression on the l.h.s. in the last display is

$$o^*_\mu(k^{-(s-j)/(2t+1)})$$

for all $j > 1/2$, hence in particular $o^*_\mu(1)$ as $k \to \infty$.

Part 2: In view of Part 1 it is sufficient to show that $(\theta, f) \mapsto \sqrt{k}(\mu_k - \mu) f(\rho(\,\cdot\,,\theta))$ converges weakly in $\ell^\infty(\Theta \times \mathcal{F})$ to $\mathbb{G}(\theta, f)$. To this end, let $H(\varphi)(\theta, f) = \varphi(f(\rho(\,\cdot\,,\theta)))$ for every $\varphi \in \ell^\infty(\mathcal{F}^*)$, $\theta \in \Theta$, and $f \in \mathcal{F}$; and note that the resulting mapping $H : \ell^\infty(\mathcal{F}^*) \to \ell^\infty(\Theta \times \mathcal{F})$ is continuous since H is linear and

$$\begin{aligned}
\|H(\varphi)\|_{\Theta\times\mathcal{F}} &= \sup_{\theta\in\Theta} \sup_{f\in\mathcal{F}} |H(\varphi)(\theta, f)| \\
&= \sup_{\theta\in\Theta} \sup_{f\in\mathcal{F}} |\varphi(f(\rho(\,\cdot\,,\theta)))| \\
&= \|\varphi\|_{\mathcal{F}^*}
\end{aligned}$$

for all $\varphi \in \ell^\infty(\mathcal{F}^*)$. Since the class of functions $\mathcal{F}^*$ is μ-Donsker by Part 1 of Proposition 14, $\sqrt{k}(\mu_k - \mu)$ converges weakly in $\ell^\infty(\mathcal{F}^*)$ to $\mathbb{G}$, and therefore, by the continuous mapping theorem, in $\ell^\infty(\Theta\times\mathcal{F})$ to $\mathbb{G}(\theta, f)$, noting that $\mathbb{G}(\theta, f) = H(\mathbb{G})$. Since $\mathbb{G}$ is measurable mapping with values in $\ell^\infty(\mathcal{F}^*)$, this also shows that $\mathbb{G}(\bullet, \cdot)$ is a measurable mapping with values in $\ell^\infty(\Theta \times \mathcal{F})$.

Part 3: In view of Prohorov's theorem, Part 3 follows directly from Part 2. □

We will show next that $\int_\Omega(\tilde{p}_k(\theta) - p_\theta)(\,\cdot\,)d\lambda$ converges in $\ell^\infty(\mathcal{F})$ to $\mathbb{G}(\theta)$ *uniformly* over Θ, where $\mathbb{G}(\theta)(f) := \mathbb{G}(f(\rho(\,\cdot\,,\theta)))$ for all $f \in \mathcal{F}$, and $\mathbb{G}$ is the Gaussian process in Part 2 of Theorem 17. For this we need the following definitions: (i) For (S, d) some metric space and h a real-valued function on S, let

$$\|h\|_{BL(S,d)} = \|h\|_S + \sup\left\{\frac{|h(x) - h(y)|}{d(x, y)} : x \neq y\right\};$$

(ii) Let $(\Omega_1, \mathcal{A}_1, P_1)$ and $(\Omega_2, \mathcal{A}_2, P_2)$ be probability spaces, and $Y_1 : \Omega_1 \to S$ and $Y_2 : \Omega_2 \to S$ be mappings such that Y_2 is $\mathcal{A}_2$-$\mathcal{B}(S, d)$-measurable and has separable range. Then, an analogue of the dual bounded Lipschitz metric is defined by

$$\beta_{(S,d)}(Y_1, Y_2) = \sup\left\{\left|\int_{\Omega_1}^{*} h(Y_1)dP_1 - \int_{\Omega_2} h(Y_2)dP_2\right| : h : S \to \mathbb{R},\ \|h\|_{BL(S,d)} \leq 1\right\},$$

where $\int_{\Omega_1}^{*}(\,\cdot\,)dP_1$ is the outer integral with respect to P_1; cf. the definition on p. 115 in Dudley (1999). We note that, by Theorem 3.6.4 in Dudley (1999), $Y_n \rightsquigarrow Y$ if and only if

$$\lim_{n\to\infty} \beta_{(S,d)}(Y_n, Y) = 0.$$

Corollary 18. *Let the hypotheses of Theorem 17 be satisfied. Then, for every $\theta \in \Theta$, $\mathbb{G}(\theta, \,\cdot\,)$ is a measurable mapping with values in $\ell^\infty(\mathcal{F})$, where $\mathbb{G}$ is the stochastic process from Part 2 of Theorem 17. Furthermore,*

$$\lim_{k\to\infty} \sup_{\theta\in\Theta} \beta_{\ell^\infty(\mathcal{F})}(\sqrt{k}\int_\Omega(\tilde{p}_k(\theta) - p_\theta)(\,\cdot\,)d\lambda, \mathbb{G}(\theta, \,\cdot\,)) = 0.$$

Proof. Let $\theta \in \Theta$ be fixed, and define $H_\theta(\varphi)(f) = \varphi(\theta, f)$ for every $\varphi \in \ell^\infty(\Theta \times \mathcal{F})$ and $f \in \mathcal{F}$. Note that this gives a Lipschitz mapping $H_\theta : \ell^\infty(\Theta \times \mathcal{F}) \to \ell^\infty(\mathcal{F})$ whose Lipschitz constant is 1 and hence is independent of θ. Since $\mathbb{G}(\,\bullet\,, \,\cdot\,)$ is a measurable mapping with values in $\ell^\infty(\Theta \times \mathcal{F})$ by Part 2 of Theorem 17, this shows that, for every $\theta \in \Theta$, $\mathbb{G}(\theta, \,\cdot\,)$ is measurable in $\ell^\infty(\mathcal{F})$. Further, since the composition of Lipschitz mappings with Lipschitz constant at most 1 is again Lipschitz with Lipschitz constant at most 1, it follows that

$$\begin{aligned}
&\sup_{\theta\in\Theta} \beta_{\ell^\infty(\mathcal{F})}(\sqrt{k}\int_\Omega(\tilde{p}_k(\theta) - p_\theta)(\,\cdot\,)d\lambda, \mathbb{G}(\theta, \,\cdot\,)) \\
&\quad= \sup_{\theta\in\Theta} \beta_{\ell^\infty(\mathcal{F})}(H_\theta(\sqrt{k}\int_\Omega(\tilde{p}_k(\,\bullet\,) - p_\bullet)(\,\cdot\,)d\lambda), H_\theta(\mathbb{G}(\,\bullet\,, \,\cdot\,))) \\
&\quad\leq \beta_{\ell^\infty(\Theta\times\mathcal{F})}(\sqrt{k}\int_\Omega(\tilde{p}_k(\,\bullet\,) - p_\bullet)(\,\cdot\,)d\lambda, \mathbb{G}(\,\bullet\,, \,\cdot\,)).
\end{aligned}$$

The r.h.s., and therefore the l.h.s., of the previous display converges to 0 by Part 2 of Theorem 17. □

The statement in Corollary 18 is in fact independent of any distance describing the concept of weak convergence in $\ell^\infty(\mathcal{F})$. More precisely, we will show below that, under the same assumptions as in Corollary 18, the following statements are equivalent: (i)

$$\lim_{k\to\infty} \sup_{\theta\in\Theta} \beta_{\ell^\infty(\mathcal{F})}(\sqrt{k}\int_\Omega(\tilde{p}_k(\theta) - p_\theta)(\,\cdot\,)d\lambda, \mathbb{G}(\theta, \,\cdot\,)) = 0;$$

(ii) For any $\theta_k, \theta \in \Theta$ such that θ_k converges to θ,

$$\sqrt{k}\int_{\Omega}(\tilde{p}_k(\theta_k) - p_{\theta_k})(\,\cdot\,)d\lambda \rightsquigarrow \mathbb{G}(\theta,\,\cdot\,) \quad \text{as } k \to \infty.$$

We first prove an abstract result:

Lemma 19. *Let T be a non-empty compact metric space, and (S,d) a metric space. For every $t \in T$, let $Y_n(t)$ be a sequence of (not necessary measurable) mappings with values in S; furthermore, let Y be a measurable mapping with separable range in S. If $Y(t_n) \rightsquigarrow Y(t)$ whenever $t_n, t \in T$ are such that t_n converges to t, then the following statements are equivalent:* (i)

$$\lim_{n\to\infty} \sup_{t\in T} \beta_{(S,d)}(Y_n(t), Y(t)) = 0;$$

(ii) *For any $t_n, t \in T$ such that t_n converges to t, $Y_n(t_n) \rightsquigarrow Y(t)$.*

Proof. We first prove that (i) implies (ii). Let $t_n, t \in T$ such that t_n converges to t. Observe that

$$\begin{aligned}\beta_{(S,d)}(Y_n(t_n), Y(t)) &\leq \beta_{(S,d)}(Y_n(t_n), Y(t_n)) + \beta_{(S,d)}(Y(t_n), Y(t)) \\ &\leq \sup_{t\in T} \beta_{(S,d)}(Y_n(t), Y(t)) + \beta_{(S,d)}(Y(t_n), Y(t)).\end{aligned}$$

The first term on the r.h.s. of the last display converges to 0 by (i), whereas the second term converges to 0 because $Y(t_n) \rightsquigarrow Y(t)$. Hence, the l.h.s. converges to 0, which just means that $Y_n(t_n) \rightsquigarrow Y(t)$.

That (ii) implies (i) can be seen as follows: Assume that $\sup_{t\in T} \beta_{(S,d)}(Y_n(t), Y(t))$ does not converge to 0. Then there is some $\eta > 0$ such that for every n there are n', $n' \geq n$, and $t_{n'} \in T$ such that

$$\beta_{(S,d)}(Y_{n'}(t_{n'}), Y(t_{n'})) > \eta. \tag{20}$$

Since T is compact, there is a subsequence $t_{n''}$ of $t_{n'}$ and some $t \in T$ such that $t_{n''}$ converges to t. It follows that the l.h.s. of the inequality

$$\beta_{(S,d)}(Y_{n''}(t_{n''}), Y(t_{n''})) \leq \beta_{(S,d)}(Y_{n''}(t_{n''}), Y(t)) + \beta_{(S,d)}(Y(t_{n''}), Y(t))$$

converges to 0 because the r.h.s. does so in view of (ii) and since $Y(t_n) \rightsquigarrow Y(t)$ by hypothesis. This is then in contradiction to (20). □

We are now able to prove the equivalence of the statements preceding Lemma 19. In order to do so, we apply this lemma to $T = \Theta$, $(S,d) = \ell^\infty(\mathcal{F})$, $Y_k(\theta) = \sqrt{k}\int_\Omega(\tilde{p}_k(\theta) - p_\theta)(\,\cdot\,)d\lambda$, $Y(\theta) = \mathbb{G}(\theta,\,\cdot\,)$, and therefore have to verify that $\mathbb{G}(\theta,\,\cdot\,)$ is measurable in $\ell^\infty(\mathcal{F})$ and $\mathbb{G}(\theta_k,\,\cdot\,) \rightsquigarrow \mathbb{G}(\theta,\,\cdot\,)$ for any $\theta_k, \theta \in \Theta$ such that θ_k converges to θ. Since $\mathbb{G}(\theta,\,\cdot\,)$ is measurable in $\ell^\infty(\mathcal{F})$ by Corollary 18, it is sufficient to show that $\mathbb{G}(\theta_k,\,\cdot\,)$ converges a.s. to $\mathbb{G}(\theta,\,\cdot\,)$ in $\ell^\infty(\mathcal{F})$, that is,

$$\lim_{k\to\infty} \sup_{f\in\mathcal{F}} |\mathbb{G}(f(\rho(\,\cdot\,,\theta_k))) - \mathbb{G}(f(\rho(\,\cdot\,,\theta)))| = 0. \tag{21}$$

So fix an arbitrary $\varepsilon > 0$. Because the sample paths of $\mathbb{G}$ are uniformly continuous with respect to the pseudo-metric d, there is a $\delta(\varepsilon) > 0$ such that $|\mathbb{G}(f(\rho(\,\cdot\,,\theta))) - \mathbb{G}(g(\rho(\,\cdot\,,\theta')))| < \varepsilon$ whenever $d(f(\rho(\,\cdot\,,\theta)), g(\rho(\,\cdot\,,\theta'))) < \delta(\varepsilon)$. Now, observe that

$$\begin{aligned} d(f(\rho(\,\cdot\,,\theta_k)), f(\rho(\,\cdot\,,\theta))) &\leq \left[\int_V |f(\rho(\,\cdot\,,\theta_k)) - f(\rho(\,\cdot\,,\theta))|^2 d\mu\right]^{1/2} \\ &\leq L_s \left[\int_V |\rho(\,\cdot\,,\theta_k) - \rho(\,\cdot\,,\theta)|^{2\min(s-1/2,1)} d\mu\right]^{1/2} \end{aligned}$$

for some constant $L_s > 0$ not depending on $f \in \mathcal{F}$. Here, we made use of the fact that, by Part 2 of Proposition 1, $\mathcal{F}$ is a bounded subset of $\mathsf{C}^{\min(s-1/2,1)}$.

The r.h.s. of the last display converges to 0 in view of the theorem of dominated convergence (use Assumption R.1 together with the fact that ρ takes its values in the bounded set Ω). Consequently, there is an index $n(\delta(\varepsilon))$ such that it is less $\delta(\varepsilon)$, which in turn implies that

$$\sup_{f\in\mathcal{F}} |\mathbb{G}(f(\rho(\,\cdot\,,\theta_k))) - \mathbb{G}(f(\rho(\,\cdot\,,\theta)))| \leq \varepsilon$$

for all $n \geq n(\delta(\varepsilon))$. This establishes (21).

6 Indirect inference estimators

Based on a given auxiliary model $\mathcal{P}(t, \zeta, D)$ and resulting AML-estimators $\hat{p}_n$ and $\tilde{p}_k(\theta)$, we define

$$\mathbb{Q}_{n,k}(\theta) = \begin{cases} \int_\Omega (\hat{p}_n - \tilde{p}_k(\theta))^2 \frac{1}{\hat{p}_n} d\lambda & \text{if } \hat{p}_n(x) > 0 \text{ for all } x \in \Omega, \\ 0 & \text{otherwise.} \end{cases} \tag{22}$$

Then $\mathbb{Q}_{n,k}$ takes its values in $[0, \infty]$. By separability of Ω and continuity of $\hat{p}_n$, the event $\{\hat{p}_n(x) > 0 \text{ for all } x \in \Omega\}$ can actually be defined by countably many measurable conditions, and therefore belongs to the σ-algebra $\mathcal{B}(\Omega)^{\mathbb{N}} \otimes \mathcal{V}^{\mathbb{N}}$. Since $\hat{p}_n$ and $\tilde{p}_k(\theta)$, respectively, are jointly measurable by Remark 7(ii), it follows from Tonelli's theorem that $\mathbb{Q}_{n,k}(\theta)$ is measurable. That we assign the value 0 on the complement of $\{\hat{p}_n(x) > 0 \text{ for all } x \in \Omega\}$ is arbitrary and irrelevant for the asymptotic considerations to follow.

Replacing the AML-estimators $\tilde{p}_k(\theta)$ in (22) by the densities p_θ, we define

$$\mathbb{Q}_n(\theta) = \begin{cases} \int_\Omega (\hat{p}_n - p_\theta)^2 \frac{1}{\hat{p}_n} d\lambda & \text{if } \hat{p}_n(x) > 0 \text{ for all } x \in \Omega, \\ 0 & \text{otherwise.} \end{cases}$$

The objective function $\mathbb{Q}_n$ corresponds to the hypothetical case in which no simulations are necessary, and serves as an auxiliary device in proving asymptotic efficiency of simulation-based indirect inference estimators. Note that $\mathbb{Q}_n$ takes its values in $[0, \infty]$. By separability of Ω and continuity of $\hat{p}_n$, we have that the event $\{\hat{p}_n(x) > 0 \text{ for all } x \in \Omega\}$ belongs to the σ-algebra $\mathcal{B}(\Omega)^{\mathbb{N}}$. Since $\hat{p}_n$ is jointly measurable by Remark 7(ii), it follows from Tonelli's theorem that $\mathbb{Q}_n(\theta)$ is measurable.

A simulation-based indirect inference (II-) estimator $\hat{\theta}_{n,k}$ is defined as *any* mapping from $\Omega^n \times V^k$ to Θ that satisfies the condition

$$\mathbb{Q}_{n,k}(\hat{\theta}_{n,k}) = \inf_{\theta \in \Theta} \mathbb{Q}_{n,k}(\theta)$$

whenever $\mathbb{Q}_{n,k}(\theta)$ attains a minimum on Θ. By an II-estimator $\hat{\theta}_n$ we mean *any* mapping from Ω^n to Θ that satisfies the condition

$$\mathbb{Q}_n(\hat{\theta}_n) = \inf_{\theta \in \Theta} \mathbb{Q}_n(\theta)$$

whenever $\mathbb{Q}_n(\theta)$ attains a minimum on Θ.

Finally, let Assumptions D and D.2 be satisfied. Define

$$Q(\theta) = \int_\Omega (p_0 - p_\theta)^2 \frac{1}{p_0} d\lambda$$

and note that Q takes its values in $[0, \infty]$. Since, under the additional assumptions P.1 and D.1, $\hat{p}_n$ converges in probability to p_0 and $\tilde{p}_k(\theta)$ to p_θ, Q can be considered as the limiting counterpart of both $\mathbb{Q}_{n,k}$ as well as $\mathbb{Q}_n$.

6.1 Consistency of II-estimators

It is clear from the definition above that there always exist (simulation-based) II-estimators, but note that they do not necessarily minimize the respective objective function for *all* realizations of the data. The existence of (simulation-based) II-estimators that do so, or at least do so on events that have probability tending to 1, is secured in Propositions 20 and 25.

Proposition 20. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$.*

1. *Let Assumption Z be satisfied. Then any II-estimator $\hat{\theta}_n$ minimizes $\mathbb{Q}_n$ for every $(x_1, \ldots, x_n) \in \Omega^n$. Furthermore, there exists an II-estimator $\hat{\theta}_n$ that is $\mathcal{B}(\Omega)^n$-$\mathcal{B}(\Theta)$-measurable.*

2. *Consider the auxiliary model $\mathcal{P}(t, \zeta, D)$ with $\zeta = 0$. Let Assumptions D, D.1, and D.2 be satisfied. Then there are events $A_n \in \mathcal{B}(\Omega)^n$ that have probability tending to 1 as $n \to \infty$ such that any II-estimator $\hat{\theta}_n$ minimizes $\mathbb{Q}_n$ on these events. [In fact, more is true: If $\chi > 0$ satisfies $\inf_{x\in\Omega} p_0(x) > \chi$, then, on A_n, any II-estimator $\hat{\theta}_n$ coincides with an II-estimator that is obtained by using $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$ as the underlying auxiliary model.]*

Proof. Part 1: By Part 1(b) of Proposition 48, $\mathbb{Q}_n$ is continuous and real-valued on Θ for each $(x_1, \ldots, x_n) \in \Omega^n$. Since Θ is compact, any $\hat{\theta}_n$ is a minimizer for each $(x_1, \ldots, x_n) \in \Omega^n$. Since $\mathbb{Q}_n$ is also a measurable function in $(x_1, \ldots, x_n)$ for each fixed $\theta \in \Theta$, as shown earlier, the existence of a measurable selection follows from Lemma A3 in Pötscher & Prucha (1997).

Part 2: Let χ be as in the proposition. By Remark 12(ii) there are events $A_n \in \mathcal{B}(\Omega)^n$ that have probability tending to 1 as $n \to \infty$ on which $\inf_{x\in\Omega} \hat{p}_n(x) > \chi$ holds. Since $\mathcal{P}(t, \chi, D) \subseteq \mathcal{P}(t, \zeta, D)$, it follows that on these events $\hat{p}_n$ coincides with the corresponding AML-estimator derived from $\mathcal{P}(t, \chi, D)$. Therefore, on these events, $\mathbb{Q}_n$ coincides with the corresponding objective function, and thus the claim follows from the already established Part 1 applied to the II-estimator based on the auxiliary model $\mathcal{P}(t, \chi, D)$. □

Since Θ is compact, the following lemma is an immediate consequence of Lemma 47.

Lemma 21. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous mapping from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let Assumptions D and D.2 be satisfied. Then Q attains its minimum on Θ.*

Remark 22. Assumption P.4 together with a uniform integrability condition on $\{p_\theta^2 : \theta \in \Theta\}$ clearly implies that $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous mapping from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. In particular, Assumptions P.1 and P.4 together are sufficient.

Proposition 23. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let Assumptions D, D.1, and D.2 be satisfied. If Q has a unique minimizer θ_0^* over Θ, then any II-estimator $\hat{\theta}_n$ converges to θ_0^* in outer probability as $n \to \infty$.*

Proof. For $\varepsilon > 0$, let

$$c(\varepsilon) = \inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [Q(\theta) - Q(\theta_0^*)].$$

Since Q is a continuous function by Part 3 of Proposition 48, since the set $\{\theta \in \Theta : \|\theta - \theta_0^*\| \geq \varepsilon\}$ is compact, and since $Q(\theta) > Q(\theta_0^*)$ for any $\theta \neq \theta_0^*$, we see that $c(\varepsilon) > 0$. It follows from Part 1 of Proposition 49 that, for any $\delta > 0$,

$$\sup_{\theta\in\Theta} |\mathbb{Q}_n(\theta) - Q(\theta)| \leq \delta$$

on events having inner probability tending to 1. Choose $\delta > 0$ such that $c(\varepsilon) - 2\delta > 0$. With help of the inequality

$$\inf_{\|\theta-\theta_0^*\|\geq\varepsilon}[\mathbb{Q}_n(\theta) - \mathbb{Q}_n(\theta_0^*)] \geq \inf_{\|\theta-\theta_0^*\|\geq\varepsilon}[Q(\theta) - Q(\theta_0^*)] - 2\sup_{\theta\in\Theta}|\mathbb{Q}_n(\theta) - Q(\theta)|$$

it then follows that

$$\inf_{\|\theta-\theta_0^*\|\geq\varepsilon}[\mathbb{Q}_n(\theta) - \mathbb{Q}_n(\theta_0^*)] \geq c(\varepsilon) - 2\delta > 0$$

on events having inner probability going to 1. Together with Proposition 20 this implies that $\hat{\theta}_n$ does not belong to the set $\{\theta \in \Theta : \|\theta - \theta_0^*\| \geq \varepsilon\}$, at least on events having inner probability tending to 1. □

Remark 24. **(i)** We do not strive for utmost generality here: Possible relaxations lie in weakening the assumptions that $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and that θ_0^* is unique.

(ii) If, in addition to the assumptions in the preceding proposition, Assumption Z is satisfied and $\hat{\theta}_n$ is $\mathcal{B}(\Omega)^n$-$\mathcal{B}(\Theta)$-measurable, then $\hat{\theta}_n$ converges in probability to θ_0^* by Lemma 3.1 in Pötscher & Prucha (1997). More precisely, we apply this lemma to $R_n = \mathbb{Q}_n$, $\bar{R}_n = Q$, $\hat{\beta}_n = \hat{\theta}_n$, and $\bar{\beta}_n = \theta_0^*$, and thus check its premises: Condition (3.2) holds by Part 1 of Proposition 49; θ_0^* is identifiably unique since it is the unique minimizer of Q, and Q is a continuous function on the compact set Θ; finally, Condition (3.3) holds by Part 1 of Proposition 20.

Note that in the following proposition the statement of Part 3 is stronger than the one of Part 2, but also requires additional assumptions.

Proposition 25. *Let Assumption R.1 be satisfied.*

1. *If Assumption Z is satisfied, then any simulation-based II-estimator $\hat{\theta}_{n,k}$ minimizes $\mathbb{Q}_{n,k}$ for every $(x_1, \ldots, x_n, v_1, \ldots, v_k) \in \Omega^n \times V^k$. Furthermore, there exists a simulation-based II-estimator that is $\mathcal{B}(\Omega)^n \otimes \mathcal{V}^k$-$\mathcal{B}(\Theta)$-measurable.*

2. *Consider the auxiliary model $\mathcal{P}(t, \zeta, D)$ with $\zeta = 0$. Let Assumptions D, D.1, and D.2 be satisfied. Then there are events $A_n \in \mathcal{B}(\Omega)^n$ having probability converging to 1 as $n \to \infty$ such that, for every $k \in \mathbb{N}$, any simulation-based II-estimator $\hat{\theta}_{n,k}$ minimizes $\mathbb{Q}_{n,k}$ on $A_n \times V^k$.*

3. *Consider the auxiliary model $\mathcal{P}(t, \zeta, D)$ with $\zeta = 0$. Let Assumptions P.1, P.2, D, D.1, and D.2 be satisfied. Then, for every constant $\chi > 0$ satisfying $\inf_{x\in\Omega} p_0(x) > \chi$ and $\inf_{\Omega\times\Theta} p(x, \theta) > \chi$, there are events $C_{n,k} \in \mathcal{B}(\Omega)^n \times \mathcal{V}^k$ that have probability tending to 1 as $\min(n, k) \to \infty$ such that on $C_{n,k}$ any simulation-based II-estimator $\hat{\theta}_{n,k}$ coincides with a simulation-based II-estimator that is obtained from using $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$ as the underlying auxiliary model.*

Proof. Part 1: By Part 2(b) of Proposition 48, $\mathbb{Q}_{n,k}$ is continuous and real-valued on Θ for each $(x_1, \ldots, x_n, v_1, \ldots, v_k) \in \Omega^n \times V^k$. Since Θ is compact, any $\hat{\theta}_{n,k}$ is a minimizer for each $(x_1, \ldots, x_n, v_1, \ldots, v_k) \in \Omega^n \times V^k$. Since $\mathbb{Q}_{n,k}$ is also a measurable function in $(x_1, \ldots, x_n, v_1, \ldots, v_k)$ for each fixed $\theta \in \Theta$, as shown earlier, the existence of a measurable selection follows from Lemma A3 in Pötscher & Prucha (1997).

Part 2: By Remark 12(ii) there are events $A_n \in \mathcal{B}(\Omega)^n$ that have probability tending to 1 as $n \to \infty$ on which $\inf_{x\in\Omega} \hat{p}_n(x) > 2^{-1} \inf_{x\in\Omega} p_0(x) > 0$. From Part 2(a) of Proposition 48 it follows that $\mathbb{Q}_{n,k}$ is continuous and real-valued on Θ for each $(x_1, \dots, x_n, v_1, \dots, v_k) \in A_n \times V^k$. Compactness of Θ completes the proof.

Part 3: Let χ be as in the proposition. Set $C_{n,k} = A_n \cap B_k$, where A_n and B_k are as in Remarks 12(ii) and (iii), and observe that $C_{n,k}$ has probability tending to 1 as $\min(n, k) \to \infty$. By Remark 12, we have on $C_{n,k}$ that $\inf_{x \in \Omega} \hat{p}_n(x) > \chi$ and $\inf_{\Omega \times \Theta} \tilde{p}_k(\theta)(x) > \chi$. Since $\mathcal{P}(t, \chi, D) \subseteq \mathcal{P}(t, \zeta, D)$, it follows that on $C_{n,k}$ the AML-estimators $\hat{p}_n$ and $\tilde{p}_k(\theta)$ coincide with the corresponding AML-estimators based on $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$ as auxiliary model. Therefore, on $C_{n,k}$, the objective function $\mathbb{Q}_{n,k}$ coincides with the corresponding objective function, and thus $\hat{\theta}_{n,k}$ coincides with the corresponding simulation-based II-estimator based on the auxiliary model $\mathcal{P}(t, \chi, D)$. □

Proposition 26. *Let Assumptions P.1, P.2, R.1, D, D.1, and D.2 be satisfied. If Q has a unique minimizer θ_0^* over Θ, then any simulation-based II-estimator $\hat{\theta}_{n,k}$ converges to θ_0^* in outer probability as $\min(n, k) \to \infty$.*

Proof. The proof is completely analogous to the proof of Proposition 23: For $\varepsilon > 0$, let

$$c(\varepsilon) = \inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [Q(\theta) - Q(\theta_0^*)].$$

Since Q is a continuous function by Part 3 of Proposition 48, since the set $\{\theta \in \Theta : \|\theta - \theta_0^*\| \geq \varepsilon\}$ is compact, and since $Q(\theta) > Q(\theta_0^*)$ for any $\theta \neq \theta_0^*$, we conclude that $c(\varepsilon) > 0$. It follows from Part 2 of Proposition 49 that, for any $\delta > 0$,

$$\sup_{\theta \in \Theta} |\mathbb{Q}_{n,k}(\theta) - Q(\theta)| \leq \delta$$

on events having inner probability tending to 1 as $\min(n, k) \to \infty$. Choose $\delta > 0$ such that $c(\varepsilon) - 2\delta > 0$. With help of the inequality

$$\inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_{n,k}(\theta_0^*)] \geq \inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [Q(\theta) - Q(\theta_0^*)] - 2 \sup_{\theta \in \Theta} |\mathbb{Q}_{n,k}(\theta) - Q(\theta)|$$

it then follows that

$$\inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_{n,k}(\theta_0^*)] \geq c(\varepsilon) - 2\delta > 0$$

on events having inner probability going to 1 as $\min(n, k) \to \infty$. Together with Proposition 25 this implies that $\hat{\theta}_{n,k}$ does not belong to the set $\{\theta \in \Theta : \|\theta - \theta_0^*\| \geq \varepsilon\}$, at least on events having inner probability tending to 1 as $\min(n, k) \to \infty$. □

Remark 27. It follows from Part 1 of Proposition 5 together with Remark 22 that the assumptions of Lemma 21 for Q are satisfied. Hence, Q has a minimizer over Θ, and the assumption in Proposition 26 that Q has a unique minimizer is in fact a uniqueness assumption.

6.2 Asymptotic efficiency of II-estimators

Note that under the assumptions of the subsequent theorem, as well as under the assumptions of Theorem 31, the function Q possesses a minimizer (cf. Lemma 21) and the matrix $\frac{\partial^2 Q}{\partial \theta \partial \theta'}(\theta)$ exists for every θ in the (non-empty) interior of Θ (see Part 2 of Lemma 51).

Theorem 28. *Let Assumptions P.1, P.4, P.5, D, D.3, and the following conditions be satisfied:*

1. *The minimizer θ_0^* of Q over Θ is unique and lies in the interior of Θ.*
2. *The $m \times m$ matrix with (i,j)-th entry*
$$\frac{\partial^2 Q}{\partial \theta_i \partial \theta_j}(\theta_0^*)$$
is positive definite.
3. *For $i = 1, \dots, m$, the first-order partial derivatives $\frac{\partial p}{\partial \theta_i}(\,\cdot\,, \theta_0^*)$ belong to some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$.*

Then
$$\sqrt{n}(\hat{\theta}_n - \theta_0^*) \rightsquigarrow N(0, \Sigma(\theta_0^*)) \quad \textit{as } n \to \infty,$$
where $\Sigma(\theta_0^*)$ is positive definite and is given by

$$\begin{aligned}
&\Sigma(\theta_0^*) \\
&= \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,, \theta_0^*) \frac{\partial p}{\partial \theta'}(\,\cdot\,, \theta_0^*) \frac{1}{p_0} d\lambda - \int_\Omega (p_0 - p_{\theta_0^*}) \frac{\partial^2 p}{\partial \theta \partial \theta'}(\,\cdot\,, \theta_0^*) \frac{1}{p_0} d\lambda\right]^{-1} \\
&\quad \times \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,, \theta_0^*) \frac{\partial p}{\partial \theta'}(\,\cdot\,, \theta_0^*) \frac{p_{\theta_0^*}^2}{p_0^3} d\lambda - \int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,, \theta_0^*) \frac{p_{\theta_0^*}}{p_0} d\lambda \int_\Omega \frac{\partial p}{\partial \theta'}(\,\cdot\,, \theta_0^*) \frac{p_{\theta_0^*}}{p_0} d\lambda\right] \\
&\quad \times \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,, \theta_0^*) \frac{\partial p}{\partial \theta'}(\,\cdot\,, \theta_0^*) \frac{1}{p_0} d\lambda - \int_\Omega (p_0 - p_{\theta_0^*}) \frac{\partial^2 p}{\partial \theta \partial \theta'}(\,\cdot\,, \theta_0^*) \frac{1}{p_0} d\lambda\right]^{-1}.
\end{aligned}$$

If $\mathcal{P}_\Theta$ is correctly specified in the sense that there is a parameter $\theta_0 \in \Theta$ such that the density p_0 is a.e. equal to p_{θ_0}, then $\theta_0^* = \theta_0$ and
$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, I^{-1}(\theta_0)) \quad \textit{as } n \to \infty,$$
where $I^{-1}(\theta_0)$ denotes the inverse of the Fisher information matrix.

Proof. Step 1: We first establish the theorem under the additional assumption Z. By Proposition 23, $\hat{\theta}_n$ comes to lie in Θ° on events whose inner probability goes to 1. We have that
$$\frac{\partial \mathbb{Q}_n}{\partial \theta}(\hat{\theta}_n) = 0$$
on these events. Applying the mean-value theorem to each component of $\frac{\partial \mathbb{Q}_n}{\partial \theta}$ then yields
$$\frac{\partial \mathbb{Q}_n}{\partial \theta}(\theta_0^*) + H_n(\hat{\theta}_n - \theta_0^*) = 0$$
on these events, where H_n is the Hessian matrix of $\mathbb{Q}_n$ with i-th row evaluated at some mean value $\bar{\theta}_{n,i}$ lying on the line segment that joins θ_0^* and $\hat{\theta}_n$. Since $\hat{\theta}_n$ converges to θ_0^* in outer probability by Proposition 23, $\bar{\theta}_{n,i}$ does so too. Hence, H_n converges to $\frac{\partial^2 Q}{\partial \theta \partial \theta'}(\theta_0^*)$ in outer probability by Proposition 52 and continuity of $\frac{\partial^2 Q}{\partial \theta \partial \theta'}$ on Θ° (see Lemma 51).

Since the latter matrix is invertible by Condition 2, it follows that H_n is invertible on events that have inner probability tending to 1, which leads to the relation

$$\sqrt{n}(\hat{\theta}_n - \theta_0^*) = -H_n^{-1}\sqrt{n}\frac{\partial \mathbb{Q}_n}{\partial \theta}(\theta_0^*) \tag{23}$$

on events that have inner probability tending to 1. Since inverting matrices is continuous, we infer from the continuous mapping theorem that H_n^{-1} converges in outer probability to the inverse of $\frac{\partial^2 Q}{\partial\theta\partial\theta'}(\theta_0^*)$, hence to

$$\left[2\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda - 2\int_\Omega (p_0 - p_{\theta_0^*})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda\right]^{-1} \tag{24}$$

by Part 2 of Lemma 51. To see that $\sqrt{n}\frac{\partial \mathbb{Q}_n}{\partial \theta}(\theta_0^*)$ satisfies a central limit theorem, we proceed as follows: Let $v \in \mathbb{R}^m$ be arbitrary, and use Part 1 of Lemma 51 to obtain

$$\begin{aligned}
v'\sqrt{n}\frac{\partial \mathbb{Q}_n}{\partial \theta}(\theta_0^*) &= 2\sqrt{n}\int_\Omega (\hat{p}_n - p_0)^2 v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{\hat{p}_n p_0^2}d\lambda \\
&\quad -2\sqrt{n}\int_\Omega (\hat{p}_n - p_0)v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0^2}d\lambda \\
&\quad -2\sqrt{n}\int_\Omega (p_0 - p_{\theta_0^*})v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda \\
&= \mathrm{I} + \mathrm{II} + \mathrm{III}.
\end{aligned}$$

Convergence of I: By Condition 3, the expression $v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)$ belongs to some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$ and is thus bounded by $C_s\|v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\|_{s,2}$ in the sup-norm. Together with the inequality

$$\left\|\frac{p_{\theta_0^*}}{\hat{p}_n p_0^2}\right\|_\Omega \le \zeta^{-3}C_t D$$

we get that

$$\mathrm{I} \le 2C_s\left\|v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\right\|_{s,2}\zeta^{-3}C_t D\,\sqrt{n}\|\hat{p}_n - p_0\|_2^2,$$

Now, Expression I converges to 0 in outer probability since $\sqrt{n}\|\hat{p}_n - p_0\|_2^2$ does so by Proposition 6 in Nickl (2007).

Convergence of II: Let $r = \min(s,t)$. We show that the function

$$f := -2v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0^2}$$

belongs to $\mathsf{W}_2^r(\Omega)$: Since $\mathsf{W}_2^r(\Omega)$ is a multiplication algebra by Part 1 of Proposition 1, it is sufficient to show that each factor of f belongs to $\mathsf{W}_2^r(\Omega)$: $-2v'\frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*) \in \mathsf{W}_2^s(\Omega)$ by Condition 3, and $p_{\theta_0^*} \in \mathsf{W}_2^t(\Omega)$ by Assumption P.1; hence these factors belong to $\mathsf{W}_2^r(\Omega)$ by Part 3 of Proposition 1. By Assumption D.1 and Part 3 of Proposition 1, $p_0 \in \mathsf{W}_2^r(\Omega)$; hence $1/p_0$ lies in $\mathsf{W}_2^r(\Omega)$ by Part 4 of Proposition 1, where we have used Assumption D.2.

We can now apply Theorem 3 in Nickl (2007) to $\mathcal{F} = \{f\}$ and obtain that

$$\sqrt{n}\int_\Omega (\hat{p}_n - p_0)f\,d\lambda \rightsquigarrow N(0, \mathbb{P}(f - \mathbb{P}f)^2).$$

Convergence of III: Note that, up to the factor $\sqrt{n}$, Expression *III* is equal to $v'\frac{\partial Q}{\partial \theta}(\theta_0^*)$ by Part 2 of Lemma 51. Since θ_0^* minimizes Q and lies in the interior of Θ by Condition 1, we have that $\frac{\partial Q}{\partial \theta}(\theta_0^*) = 0$. Consequently, Expression *III* is 0.

By the Cramér-Wold device, $\sqrt{n}\frac{\partial \mathbb{Q}_n}{\partial \theta}(\theta_0^*)$ asymptotically follows a central normal distribution with variance-covariance matrix given by

$$4\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}^2}{p_0^3}d\lambda - 4\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0}\,d\lambda\int_\Omega \frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0}d\lambda. \tag{25}$$

Combine that $-H_n^{-1}$ converges in outer probability to (24) with (25) to obtain

$$\sqrt{n}(\hat{\theta}_n - \theta_0^*) \rightsquigarrow N(0, \Sigma(\theta_0^*)),$$

where

$$\begin{aligned}
&\Sigma(\theta_0^*)\\
&\quad= \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda - \int_\Omega (p_0 - p_{\theta_0^*})\frac{\partial^2 p}{\partial \theta \partial \theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda\right]^{-1}\\
&\qquad\times\left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}^2}{p_0^3}d\lambda - \int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0}d\lambda\int_\Omega \frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0}d\lambda\right]\\
&\qquad\times\left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda - \int_\Omega (p_0 - p_{\theta_0^*})\frac{\partial^2 p}{\partial \theta \partial \theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda\right]^{-1}.
\end{aligned}$$

If the given model is correctly specified, then θ_0^* coincides with θ_0 in view of Condition 1, and $\Sigma(\theta_0^*)$ simplifies to

$$\Sigma(\theta_0) = \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0)\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_0)\frac{1}{p_0}d\lambda\right]^{-1}.$$

[Here, we have used that $\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_0)d\lambda = 0$, which follows from Assumption P.5 and the inequality $\int_\Omega |f|d\lambda \leq \lambda(\Omega)^{1/2}\|f\|_2$.] We conclude that in this case

$$\sqrt{n}(\hat{\theta}_n - \theta_0) \rightsquigarrow N(0, I^{-1}(\theta)) \quad \text{as } n \to \infty.$$

Step 2: We now turn to the case where $\zeta = 0$. Note that $\inf_{x\in\Omega} p_0(x) > 0$ because of Assumption D.3. Let $\chi > 0$ be such that $\inf_{x\in\Omega} p_0(x) > \chi$. Then it follows from Part 2 of Proposition 20 that there are events that have probability tending to 1 such that on these events $\hat{\theta}_n$ coincides with an II-estimator $\breve{\theta}_n$ that is based on $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$. Applying what has already been established in Step 1 to $\breve{\theta}_n$ completes the proof. □

Remark 29. **(i)** Beran (1977) defines a minimum Hellinger distance estimator $\hat{\hat{\theta}}_n$ as minimizer of the function

$$\mathbb{H}_n(\theta) = \frac{1}{2}\int_\Omega \left(\sqrt{\hat{p}_n} - \sqrt{p_\theta}\right)^2 d\lambda$$

and shows that $\hat{\hat{\theta}}_n$ is both asymptotically efficient *and* minimax robust in a Hellinger neighbourhood of the given model. This is of interest here as $\hat{\theta}_n$ can be viewed as a minimum distance estimator and as such is related to $\hat{\hat{\theta}}_n$. To see this, rewrite $\mathbb{H}_n$ as

$$\mathbb{H}_n(\theta) = \frac{1}{2}\int_\Omega (\hat{p}_n - p_\theta)^2 \frac{1}{\hat{p}_n} \frac{\hat{p}_n}{\left(\sqrt{\hat{p}_n} + \sqrt{p_\theta}\right)^2}\, d\lambda$$

and suppose the given model is correctly specified. Then $\hat{p}_n/(\sqrt{\hat{p}_n} + \sqrt{p_\theta})^2$ approaches $1/4$ as $n \to \infty$ and $\|\theta - \theta_0\|$ converges to 0, and so $\mathbb{H}_n$ gets close to $\mathbb{Q}_n/8$.

(ii) A crucial point in obtaining the rate $n^{-1/2}$ for II-estimators is that the density estimator $\hat{p}_n$ has precisely this rate in the weak topology; see Theorem 3 in Nickl (2007).

We now turn to the asymptotic theory for *simulation-based* II-estimators. Before stating the main result of this thesis, we prove an important lemma, which will allow us to reduce the problem of deriving the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{n,k(n)} - \theta_0^*)$ to deriving the one of $\sqrt{n}(\hat{\theta}_n - \theta_0^*)$.

Lemma 30. *Let $U \subseteq \mathbb{R}^m$ be a non-empty open, convex set. Let $f, g : U \to \mathbb{R}$, where g has partial derivatives up to order 2 and satisfies the following property: There is some $K > 0$ such that*

$$\inf_{x\in U}\left[y' \frac{\partial^2 g}{\partial x \partial x'}(x) y\right] \geq K\|y\|^2 \tag{26}$$

for all $y \in \mathbb{R}^m$. If u is a minimizer of f and v is a minimizer of g, then

$$\|u - v\| \leq 2K^{-1/2}\sqrt{\|f - g\|_U}.$$

Proof. Let u be a minimizer of f and v be a minimizer of g. As v is an inner point of U, we have that

$$\frac{\partial g}{\partial x}(v) = 0.$$

A Taylor expansion of g around v yields

$$g(x) = g(v) + \frac{1}{2}(x - v)' \frac{\partial^2 g}{\partial x \partial x'}(\bar{v})(x - v), \tag{27}$$

where $\bar{v}$ lies on the line segment joining x and v; hence $\bar{v} \in U$ by convexity of U. Using (26), (27) gives

$$\|x - v\| \leq \sqrt{2}K^{-1/2}\sqrt{|g(x) - g(v)|}. \tag{28}$$

Next, note that the inequalities

$$f(u) - g(v) \leq f(v) - g(v) \leq \|f - g\|_U$$

and

$$f(u) - g(v) \geq f(u) - g(u) \geq -\|f - g\|_U$$

imply that

$$|f(u) - g(v)| \leq \|f - g\|_U,$$

which in turn yields

$$\begin{aligned} |g(u) - g(v)| &\leq |g(u) - f(u)| + |f(u) - g(v)| \\ &\leq 2\|f - g\|_U. \end{aligned}$$

By (28) this results in

$$\|u-v\| \leq 2K^{-1/2}\sqrt{\|f-g\|_U}.$$

□

We are now ready to state the main result of this thesis. It provides the asymptotic distribution of simulation-based II-estimators for a proper choice of $k = k(n)$. Furthermore, it shows that under correct specification of the underlying model simulation-based II-estimators are efficient in the sense that they asymptotically attain the Cramér-Rao bound.

Theorem 31. *Let Assumptions P.1, P.5, R.2, D, D.3, and the following conditions be satisfied:*

1. *The minimizer θ_0^* of Q over Θ is unique and lies in the interior of Θ.*
2. *The $m \times m$ matrix with (i,j)-th entry*
$$\frac{\partial^2 Q}{\partial\theta_i \partial\theta_j}(\theta_0^*)$$
is positive definite.
3. *For $i = 1, \ldots, m$, the first-order partial derivatives $\frac{\partial p}{\partial\theta_i}(\,\cdot\,, \theta_0^*)$ belong to some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$.*

Let, in addition, either (i) *Assumption P.2 be satisfied, and choose $k(n)$ such that $\liminf_{n\to\infty} k(n)/n^{2+1/t} > 0$; or* (ii) *Assumption P.3 be satisfied, and choose $k(n)$ such that $\liminf_{n\to\infty} k(n)/n^{2+\alpha} > 0$ for some (arbitrarily small) $\alpha > 0$. Then*

$$\sqrt{n}(\hat{\theta}_{n,k(n)} - \theta_0^*) \rightsquigarrow N(0, \Sigma(\theta_0^*)) \quad \text{as } n \to \infty,$$

where $\Sigma(\theta_0^*)$ is positive definite and is given by

$$\begin{aligned}
&\Sigma(\theta_0^*) \\
&= \left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda - \int_\Omega (p_0 - p_{\theta_0^*})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda\right]^{-1} \\
&\quad \times \left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}^2}{p_0^3}d\lambda - \int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0}d\lambda \int_\Omega \frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta_0^*)\frac{p_{\theta_0^*}}{p_0}d\lambda\right] \\
&\quad \times \left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta_0^*)\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda - \int_\Omega (p_0 - p_{\theta_0^*})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta_0^*)\frac{1}{p_0}d\lambda\right]^{-1}.
\end{aligned}$$

If $\mathcal{P}_\Theta$ is correctly specified in the sense that there is a parameter $\theta_0 \in \Theta$ such that the density p_0 is a.e. equal to p_{θ_0}, then $\theta_0^* = \theta_0$ and

$$\sqrt{n}(\hat{\theta}_{n,k(n)} - \theta_0) \rightsquigarrow N(0, I^{-1}(\theta_0)) \quad \text{as } n \to \infty,$$

where $I^{-1}(\theta_0)$ denotes the inverse of the Fisher information matrix.

Proof. Step 1: We first establish the theorem under the additional assumption Z. The key step lies in proving that

$$\sqrt{n}(\hat{\theta}_{n,k(n)} - \hat{\theta}_n) = o^*_{\mathrm{Pr}}(1) \quad \text{as } n \to \infty.$$

With help of Lemma 30 this can be achieved by exploiting the closeness of the objective functions $\mathbb{Q}_{n,k}$ and $\mathbb{Q}_n$: The matrix $\frac{\partial^2 Q}{\partial\theta\partial\theta'}(\theta_0^*)$ exists by Part 2 of Lemma 51 and is positive definite by Condition 2. Use this together with the continuity of $\frac{\partial^2 Q}{\partial\theta\partial\theta'}$ on the interior Θ° of Θ (see Part 2 of Lemma 51) and the fact that, for all $i, j = 1, \ldots, m$,

$$\sup_{\theta\in\Theta^\circ} \left| \frac{\partial^2 \mathbb{Q}_n}{\partial\theta_i\partial\theta_j}(\theta) - \frac{\partial^2 Q}{\partial\theta_i\partial\theta_j}(\theta) \right| = o_{\mathbb{P}}(1) \quad \text{as } n \to \infty$$

(see Proposition 52) to conclude that there are events E_n having probability tending to 1 as $n \to \infty$ such that

$$\inf_{\theta\in U} \left[y' \frac{\partial^2 \mathbb{Q}_n}{\partial\theta\partial\theta'}(\theta) y \right] \geq K\|y\|^2 \quad \text{for all } y \in \mathbb{R}^m$$

holds true for some open, convex neighbourhood U of θ_0^* and some constant $K > 0$, where neither U nor K depends on n or the data. By Propositions 23 and 26, $\hat{\theta}_n$ and $\hat{\theta}_{n,k(n)}$ both come to lie in U on events E'_n whose inner probability goes to 1 as $n \to \infty$. For the rest of the proof of Step 1 we restrict our reasoning to the events $E_n \cap E'_n$, and note that they have inner probability tending to 1 as $n \to \infty$. Since, by Part 1 of Proposition 20 and Part 1 of Proposition 25, $\hat{\theta}_n$ and $\hat{\theta}_{n,k(n)}$, respectively, minimize the objective functions $\mathbb{Q}_n$ and $\mathbb{Q}_{n,k(n)}$, we may apply Lemma 30 with $f = \mathbb{Q}_{n,k(n)}|U$, $g = \mathbb{Q}_n|U$, $u = \hat{\theta}_{n,k(n)}$, and $v = \hat{\theta}_n$ to get

$$\|\hat{\theta}_{n,k(n)} - \hat{\theta}_n\| \leq 2K^{-1/2}\sqrt{\|\mathbb{Q}_{n,k(n)} - \mathbb{Q}_n\|_U}.$$

It follows from Part 3 of Proposition 49 and the choice of $k(n)$ that

$$\sqrt{n}(\hat{\theta}_{n,k(n)} - \hat{\theta}_n) = o^*_{\mathrm{Pr}}(1) \quad \text{as } n \to \infty$$

holds under (i) as well as under (ii). The proof of Step 1 now follows from Theorem 28, noting that Assumption P.4 follows from Assumption R.2; cf. Part 1 of Proposition 5.

Step 2: We now prove the theorem in case $\zeta = 0$. Note that $\inf_{x\in\Omega} p_0(x) > 0$ and $\inf_{\Omega\times\Theta} p(x,\theta) > 0$ by Assumptions D.2 and P.2. Let $\chi > 0$ be such that $\inf_{x\in\Omega} p_0(x) > \chi$ and $\inf_{\Omega\times\Theta} p(x,\theta) > \chi$. Then it follows from Part 2 of Proposition 20 and Part 3 of Proposition 25 that there are events $C_{n,k(n)}$ having probability tending to 1 as $n \to \infty$ such that on these events $\hat{\theta}_{n,k(n)}$ coincides with a simulation-based II-estimator $\check{\theta}_{n,k(n)}$ based on $\mathcal{P}(t,\chi,D)$ instead of $\mathcal{P}(t,\zeta,D)$. Applying what has already been established in Step 1 to $\check{\theta}_{n,k(n)}$ gives the result. □

Remark 32. **(i)** The classical way to derive the asymptotic distribution of $\sqrt{n}(\hat{\theta}_{n,k(n)} - \theta_0^*)$ would be to apply the mean-value theorem directly to the 'score' $\frac{\partial\mathbb{Q}_{n,k(n)}}{\partial\theta}$, but this requires knowledge about the differentiability of the mapping $\theta \mapsto \tilde{p}_{k(n)}(\theta)$ that we do not have. The usual way to show that $\theta \mapsto \tilde{p}_{k(n)}(\theta)$ is differentiable (or higher-order differentiable) is to make use of the fact that $\tilde{p}_{k(n)}(\theta)$ is implicitly defined as the maximizer of $L_{k(n)}(\theta, \cdot)$ over the auxiliary model $\mathcal{P}(t,\zeta,D)$, and to apply the implicit function theorem based on the equation

$$DL_{k(n)}(\theta, \tilde{p}_{k(n)}(\theta)) = 0. \tag{29}$$

This approach is not feasible here since $\tilde{p}_{k(n)}(\theta)$ lies on the 'boundary' of $\mathcal{P}(t, \zeta, D)$ by Remark 7(i), so that we only know that

$$DL_{k(n)}(\theta, \tilde{p}_{k(n)}(\theta)) \leq 0$$

is satisfied. To circumvent this problem, we make use of Lemma 30 to conclude that $\sqrt{n}(\hat{\theta}_{n,k(n)} - \hat{\theta}_n) = o^*_{\Pr}(1)$, and only then apply the mean-value theorem to $\frac{\partial \mathbb{Q}_n}{\partial \theta}$ to get an explicit expression for $\sqrt{n}(\hat{\theta}_n - \theta^*_0)$. Now, we need to differentiate $\theta \mapsto p_\theta$ instead of $\theta \mapsto \tilde{p}_{k(n)}(\theta)$, which is possible simply by assumption.

(ii) The use of Lemma 30 in the proof of the preceding theorem entails that $k(n)$ has to be chosen such that at least $\liminf_{n\to\infty} k(n)/n^{2+\alpha} > 0$ for some $\alpha > 0$. Whether this can be weakened to requiring that $\liminf_{n\to\infty} k(n)/n^{1+\alpha} > 0$ for some $\alpha > 0$ by another method of proof is unclear for the reasons outlined above. For indirect inference estimators based on other non-parametric auxiliary estimators this can be established; see Nickl & Pötscher (2009).

(iii) When using *parametric* auxiliary models, the resulting simulation-based II-estimators are only asymptotically efficient under the somewhat restrictive assumption that the auxiliary model is correctly specified. This is in stark contrast to the situation here: The requirement that p_0 and all p_θ belong to the non-parametric auxiliary model $\mathcal{P}(t, \zeta, D)$ amounts to imposing a certain common degree of smoothness (slightly more than continuity) on these density functions and is already sufficient to establish asymptotic efficiency.

7 Moving-parameter asymptotics for indirect inference estimators

In this section we investigate the asymptotic behaviour of $\hat{\theta}_{n,k(n)}$ under convergent sequences of parameters. More precisely, we will prove that $\hat{\theta}_{n,k(n)}$ is asymptotically normal with variance-covariance matrix equal to the Cramér-Rao bound whenever the parameters $\theta_{0,n}$ converge to $\theta_{0,\infty}$ and $\theta_{0,\infty}$ belongs to the interior of the parameter space Θ. This implies that the estimator is regular and asymptotically efficient. While so far the data $X_1, \ldots, X_n$ have been assumed to be i.i.d. with *fixed* law $\mathbb{P}$, we now let $\mathbb{P}$ vary in certain subsets of the class of all probability measures on $(\Omega, \mathcal{B}(\Omega))$. For any probability density function q on Ω we will write $\mathbb{P}(q)$ for the associated probability measure on $(\Omega, \mathcal{B}(\Omega))$, and $\Pr(q)$ for the probability measure $\mathbb{P}(q)^{\mathbb{N}} \otimes \mu^{\mathbb{N}}$ on $(\Omega^{\mathbb{N}} \times V^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}} \otimes \mathcal{V}^{\mathbb{N}})$. We only consider auxiliary models that satisfy Assumption Z in this section.

7.1 AML-estimators: uniformity in the underlying probability measure

Recall from Section 5 that

$$L_n(f) = \frac{1}{n} \sum_{i=1}^{n} \log f(X_i)$$

for any non-negative real-valued function f on Ω. If $X_1, \ldots, X_n$ are i.i.d. with law $\mathbb{P}$ and $f \in \mathsf{L}^\infty(\Omega)$, then $L_n(f)$ converges a.s. to $\int_\Omega \log f d\mathbb{P}$ by the strong law of large numbers. We thus define

$$L_{\mathbb{P}}(f) = \int_\Omega \log f d\mathbb{P}$$

for any non-negative $f \in \mathsf{L}^\infty(\Omega)$ and any probability measure $\mathbb{P}$ on $(\Omega, \mathcal{B}(\Omega))$. The functions $L_{\mathbb{P}}$ take their values in $[-\infty, \infty)$.

Theorem 33. *Under Assumption Z,*

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{q\in\mathcal{P}(t,\zeta,D)} \Pr(q)^* \left(n^{(t-r)/(2t+1)} \|\hat{p}_n - q\|_{r,2} > M \right) = 0$$

for every r, $0 \le r \le t$. In particular,

$$\lim_{n\to\infty} \sup_{q\in\mathcal{P}(t,\zeta,D)} \Pr(q)^* \left(\|\hat{p}_n - q\|_\Omega > \eta \right) = 0$$

for every $\eta > 0$.

Proof. Throughout the proof we denote by $\mathrm{E}^*_{\mathbb{P}(q)}$ the outer expectation with respect to the probability measure $\mathbb{P}(q)^{\mathbb{N}}$ on $(\Omega^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}})$.

We first prove the case when $r = 0$. To this end, we verify the conditions of Theorem 55 with $(\Lambda, \mathcal{A}) = (\Omega^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}})$, $\mathfrak{P} = \{\mathbb{P}(q)^{\mathbb{N}} : q \in \mathcal{P}(t, \zeta, D)\}$, $T = \mathcal{P}(t, \zeta, D)$, $d(p, q) = \|p - q\|_2$, $H_n = L_n$, $H_P = L_{\mathbb{P}(q)}$, $\hat{t}_n = \hat{p}_n$, and $t_P = q$. Condition (58) is clear by definition of $\hat{p}_n$. Condition (56) follows from a second-order Taylor expansion of $L_{\mathbb{P}(q)}$ around the density q:

$$\begin{aligned} L_{\mathbb{P}(q)}(p) - L_{\mathbb{P}(q)}(q) &= DL_{\mathbb{P}(q)}(q)(p-q) + \frac{1}{2}D^2 L_{\mathbb{P}(q)}(\bar{p})(p-q,p-q) \\ &= \int_\Omega \frac{p-q}{q} q d\lambda - \frac{1}{2}\int_\Omega \frac{(p-q)^2}{\bar{p}^2} q d\lambda \\ &\qquad \text{(by Proposition 61)} \\ &\leq -\frac{1}{2}\zeta C_t^{-2} D^{-2} \|p-q\|_2^2, \end{aligned}$$

where $\bar{p}$ is some density on the line segment joining p and q; note that $\bar{p} \in \mathcal{P}(t,\zeta,D)$ by convexity of this set, and hence satisfies $\|\bar{p}\|_\Omega \leq C_t D$. This proves Condition (56) in Theorem 55 with constants $C = 2^{-1}\zeta C_t^{-2} D^{-2}$ and $\alpha = 2$, which are both independent of p_0.

Next is the verification of Condition (57). Letting

$$\mathcal{F}_{q,\delta} = \{\log p - \log q : p \in \mathcal{P}(t,\zeta,D),\ \|p-q\|_2 \leq \delta\}$$

for $\delta > 0$, the expression

$$\sup_{q\in\mathcal{P}(t,\zeta,D)} \mathrm{E}^*_{\mathbb{P}(q)} \sup_{\substack{p\in\mathcal{P}(t,\zeta,D),\\ \|p-q\|_2\leq\delta}} \left|\sqrt{n}(L_n - L_{\mathbb{P}(q)})(p) - \sqrt{n}(L_n - L_{\mathbb{P}(q)})(q)\right|$$

can be viewed as

$$\sup_{q\in\mathcal{P}(t,\zeta,D)} \mathrm{E}^*_{\mathbb{P}(q)} \left\|\sqrt{n}(\mathbb{P}_n - \mathbb{P}(q))\right\|_{\mathcal{F}_{q,\delta}}$$

and can be bounded above by

$$\sup_{q\in\mathcal{P}(t,\zeta,D)} \mathrm{E}^*_{\mathbb{P}(q)} \left\|\sqrt{n}(\mathbb{P}_n - \mathbb{P}(q))\right\|_{\mathcal{F}_\delta},$$

where

$$\mathcal{F}_\delta = \{\log p - \log q : p, q \in \mathcal{P}(t,\zeta,D),\ \|p-q\|_2 \leq \delta\}.$$

So we are led to deriving an upper bound for the expected variation of the empirical process $\sqrt{n}(\mathbb{P}_n - \mathbb{P}(q))$ over the class of functions $\mathcal{F}_\delta$. To this end, we verify the premises of Theorem 58: the boundedness of $\mathcal{F}_\delta$ with respect to the sup-norm as well as the $\mathcal{L}^2(\mathbb{P}(q))$-norm. By Assumption Z, the logarithm is Lipschitz on $[\zeta,\infty)$ with Lipschitz constant ζ^{-1}. Use this to see that $\mathcal{F}_\delta$ is bounded by $B := 2\zeta^{-1}C_t D$ in the sup-norm and by $\eta(\delta) := \zeta^{-1} C_t^{1/2} D^{1/2}\delta$ in the $\mathcal{L}^2(\mathbb{P}(q))$-norm. We conclude from Theorem 58 that

$$\begin{aligned} &\sup_{q\in\mathcal{P}(t,\zeta,D)} \mathrm{E}^*_{\mathbb{P}(q)} \left\|\sqrt{n}(\mathbb{P}_n - \mathbb{P}(q))\right\|_{\mathcal{F}_\delta} \\ &\quad\leq (1696 + 64\sqrt{2}) \sup_{q\in\mathcal{P}(t,\zeta,D)} I_{[\,]}(\eta(\delta), \mathcal{F}_\delta, \|\cdot\|_{2,\mathbb{P}(q)}) \\ &\qquad\times\left[1 + \frac{B}{\eta(\delta)^2\sqrt{n}} \sup_{q\in\mathcal{P}(t,\zeta,D)} I_{[\,]}(\eta(\delta), \mathcal{F}_\delta, \|\cdot\|_{2,\mathbb{P}(q)})\right] \end{aligned} \tag{30}$$

for all $\delta > 0$. Letting $\mathcal{F} = \{\log p - \log q : p, q \in \mathcal{P}(t,\zeta,D)\}$,

$$\begin{aligned} \sup_{q\in\mathcal{P}(t,\zeta,D)} N_{[\,]}(\varepsilon, \mathcal{F}_\delta, \|\cdot\|_{2,\mathbb{P}(q)}) &\leq \sup_{q\in\mathcal{P}(t,\zeta,D)} N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,\mathbb{P}(q)}) \\ &\leq N(\varepsilon, \mathcal{F}, \mathsf{L}^\infty(\Omega), \|\cdot\|_\Omega). \end{aligned} \tag{31}$$

In order to bound $N(\varepsilon, \mathcal{F}, \mathsf{L}^\infty(\Omega), \|\cdot\|_\Omega)$ above, observe that $\mathcal{F} = \log \mathcal{P}(t,\zeta,D) - \log \mathcal{P}(t,\zeta,D)$, so that it is sufficient to control the covering number

$$N(\varepsilon, \log \mathcal{P}(t,\zeta,D), \mathsf{L}^\infty(\Omega), \|\cdot\|_\Omega).$$

To cover $\log \mathcal{P}(t,\zeta,D)$, first take $N(\varepsilon, \mathcal{P}(t,\zeta,D), \mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)$-many sup-norm closed balls $[p_i - 2\varepsilon, p_i + 2\varepsilon]$ of radius 2ε whose centers p_i already belong to $\mathcal{P}(t,\zeta,D)$. [This can always be achieved.] Since the elements of $\mathcal{P}(t,\zeta,D)$ are bounded below by ζ,

$$\left\{[\max(\zeta, p_i - 2\varepsilon), p_i + 2\varepsilon] : i = 1, \ldots, N(\varepsilon, \mathcal{P}(t,\zeta,D), \mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)\right\}$$

forms a cover of $\mathcal{P}(t,\zeta,D)$. Hence, by monotonicity of the logarithm and since $\zeta > 0$, the brackets

$$[\log \max(\zeta, p_i - 2\varepsilon), \log(p_i + 2\varepsilon)],$$

$i = 1, \ldots, N(\varepsilon, \mathcal{P}(t,\zeta,D), \mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)$, cover $\log \mathcal{P}(t,\zeta,D)$. Making use of the fact that the logarithm is Lipschitz on $[\zeta, \infty)$ with Lipschitz constant ζ^{-1}, these brackets are bounded by $4\zeta^{-1}\varepsilon$ in the sup-norm. It follows that

$$N(\varepsilon, \mathcal{F}, \mathsf{L}^\infty(\Omega), \|\cdot\|_\Omega) \lesssim N(\varepsilon, \mathcal{P}(t,\zeta,D), \mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)^2.$$

Use this inequality and apply Lemma 13 with $\Delta = \Omega$ to $\mathcal{P}(t,\zeta,D)$ to see that

$$\begin{aligned}
\int_{(0,\eta(\delta)]} \sqrt{1 + H(\varepsilon, \mathcal{F}, \mathsf{L}^\infty(\Omega), \|\cdot\|_\Omega)}\, d\varepsilon &\lesssim \int_{(0,\eta(\delta)]} \sqrt{1 + \varepsilon^{-1/t}}\, d\varepsilon \\
&\lesssim \max(\eta(\delta), \int_{(0,\eta(\delta)]} \varepsilon^{-1/2t} d\varepsilon) \\
&\lesssim \max(\delta, \delta^{1-1/2t}).
\end{aligned}$$

By (31) this bound is also valid for

$$\sup_{q \in \mathcal{P}(t,\zeta,D)} I_{[\,]}(\eta(\delta), \mathcal{F}_\delta, \|\cdot\|_{2,\mathbb{P}(q)})$$

and can be used to bound the r.h.s. of (30), so that there is some constant L, $0 < L < \infty$, such that

$$\sup_{q \in \mathcal{P}(t,\zeta,D)} \mathrm{E}^*_{\mathbb{P}(q)} \left\|\sqrt{n}(\mathbb{P}_n - \mathbb{P}(q))\right\|_{\mathcal{F}_{q,\delta}} \le L \max(\delta, \delta^{1-1/2t}) \left[1 + \frac{\max(\delta, \delta^{1-1/2t})}{\delta^2 \sqrt{n}}\right]$$

holds for all $\delta > 0$. Write $\varphi_n(\delta)$ for the r.h.s. in the last display and note that $\delta \mapsto \delta^{-\beta}\varphi_n(\delta)$ is non-increasing for $\beta = 1$. This shows Condition (57) in Theorem 55.

Condition (59), with $\alpha = 2$, is satisfied for $r_n = n^{t/(2t+1)}$. This gives the desired rate and completes the proof in case $r = 0$.

In case $0 < r \le t$, we make use of the fact that $\sup_{q \in \mathcal{P}(t,\zeta,D)} \|\hat{p}_n - q\|_{t,2} \le 2D$ and of the interpolation inequality

$$\|f\|_{r,2} \le C_{r,t} \|f\|_{t,2}^{r/t} \|f\|_2^{(t-r)/t}$$

for $f \in \mathsf{W}_2^t(\Omega)$, where $C_{r,t} > 0$; see Theorem 1.9.6 and Remark 1.9.1 in Lions & Magenes (1972). □

Part 2 of the next theorem is a modification of Theorem 3 in Nickl (2007) allowing for moving true densities $p_{0,n}$. It states that the stochastic process

$$f \to \sqrt{n} \int_{\Omega} (\hat{p}_n - p_{0,n}) f d\lambda$$

converges weakly to a Gaussian process that is determined by the pointwise limit of the densities $p_{0,n}$. The proof of this fact needs a non-trivial result from Giné & Zinn (1991); see Corollary 2.7 there.

Theorem 34. *Let Assumption Z be satisfied. Let $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t, \zeta, D)$, $p_{0,n}$ converge pointwise to $p_{0,\infty}$, and suppose that the strict inequalities*

$$\inf_{x\in\Omega} p_{0,\infty}(x) > \zeta \quad \textit{and} \quad \|p_{0,\infty}\|_{t,2} < D$$

as well as

$$\inf_{n\in\mathbb{N}} \inf_{x\in\Omega} p_{0,n}(x) > \zeta \qquad \textit{and} \qquad \sup_{n\in\mathbb{N}} \|p_{0,n}\|_{t,2} < D$$

hold true. For each $n \in \mathbb{N}$, consider the stochastic processes

$$f \mapsto \sqrt{n} \int_{\Omega} (\hat{p}_n - p_{0,n}) f d\lambda \quad \textit{and} \quad f \mapsto \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))f$$

on $(\Omega^n, \mathcal{B}(\Omega)^n, \mathbb{P}(p_{0,n})^n)$. Then, for any non-empty bounded subset $\mathcal{F}$ of some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$, the following statements hold true:

1. *For all $j > 1/2$,*

$$\sup_{f\in\mathcal{F}} \left| \sqrt{n} \int_{\Omega} (\hat{p}_n - p_{0,n}) f d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))f \right|$$
$$= o^*_{\mathbb{P}(p_{0,n})^n}(n^{-(\min(s,t)-j)/(2t+1)}) \quad \textit{as } n \to \infty;$$

in particular,

$$\sup_{f\in\mathcal{F}} \left| \sqrt{n} \int_{\Omega} (\hat{p}_n - p_{0,n}) f d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))f \right| = o^*_{\mathbb{P}(p_{0,n})^n}(1) \quad \textit{as } n \to \infty.$$

2. *There exists a $\mathbb{P}(p_{0,\infty})$-Brownian bridge $\mathbb{G}$ indexed by $\mathcal{F}$, that is, $\mathbb{G}$ is a centered Gaussian process indexed by $\mathcal{F}$, which is measurable as a mapping with values in $\ell^\infty(\mathcal{F})$ and has covariance function*

$$\mathrm{Cov}[\mathbb{G}(f), \mathbb{G}(g)] = \int_{\Omega} \left(f - \int_{\Omega} f p_{0,\infty} d\lambda \right) \left(g - \int_{\Omega} g p_{0,\infty} d\lambda \right) p_{0,\infty} d\lambda$$

and bounded sample paths that are uniformly continuous with respect to the pseudo-metric

$$d(f,g) = \left[\int_{\Omega} \left(f - g - \int_{\Omega} (f-g) p_{0,\infty} d\lambda \right)^2 p_{0,\infty} d\lambda \right]^{1/2},$$

such that $\sqrt{n}\int_{\Omega}(\hat{p}_n - p_{0,n})(\,\cdot\,)d\lambda$ converges weakly to $\mathbb{G}$ in $\ell^\infty(\mathcal{F})$. Up to the probability space on which $\mathbb{G}$ is defined, weak convergence refers to the underlying probability spaces $(\Omega^n, \mathcal{B}(\Omega)^n, \mathbb{P}(p_{0,n})^n)$.

3. Let $\mathbb{G}$ be the Gaussian process from Part 2. Then

$$\sqrt{n}\int_{\Omega}(\hat{p}_n - p_{0,n})f_n d\lambda \rightsquigarrow \mathbb{G}(f_0)$$

whenever f_n, f_0 belong to $\mathcal{F}$ and f_n converges pointwise to f_0. Up to the probability space on which $\mathbb{G}$ is defined, weak convergence refers to the underlying probability spaces $(\Omega^n, \mathcal{B}(\Omega)^n, \mathbb{P}(p_{0,n})^n)$.

Proof. Throughout the proof, convergence in probability (as n tends to infinity) always refers to the underlying probability spaces $(\Omega^n, \mathcal{B}(\Omega)^n, \mathbb{P}(p_{0,n})^n)$.

Part 1: The proof strategy for Part 1 is as in Nickl (2007).

Step 1: We first consider the case $s \geq t$. By Part 3 of Proposition 1, $\mathcal{F}$ is then bounded in $\mathsf{W}_2^t(\Omega)$ and hence is contained in $\mathcal{U}_{t,B}$ for some B, $0 < B < \infty$. W.l.o.g. we may therefore assume that $\mathcal{F}$ equals $\mathcal{U}_{t,B}$. We define

$$\pi_n(f) = \left(f - \int_{\Omega} f p_{0,n} d\lambda\right) p_{0,n}$$

for any function $f \in \mathsf{W}_2^t(\Omega)$. Then, for every $f \in \mathcal{F}$,

$$\begin{aligned}
\|\pi_n(f)\|_{t,2} &\leq M_t \left\| f - \int_{\Omega} f p_{0,n} d\lambda \right\|_{t,2} \|p_{0,n}\|_{t,2} \\
&\qquad \text{(by Part 1 of Proposition 1)} \\
&\leq M_t D \left[\|f\|_{t,2} + \left\| \int_{\Omega} f p_{0,n}\, d\lambda \right\|_{t,2} \right] \\
&\qquad \text{(since } p_{0,n} \in \mathcal{P}(t, \zeta, D)) \\
&\leq M_t D B(1 + C_t \|1\|_{t,2}) \\
&\qquad \text{(since } \mathcal{F} = \mathcal{U}_{t,B}) \\
&\leq M_t D B(1 + C_t \lambda(\Omega)^{1/2}).
\end{aligned}$$

Hence,

$$\sup_{n\in\mathbb{N}} \sup_{f\in\mathcal{F}} \|\pi_n(f)\|_{t,2} \leq M_t D B(1 + C_t \lambda(\Omega)^{1/2}) < \infty. \tag{32}$$

Use Proposition 61 with $P = \mathbb{P}(p_{0,n})$ and the first part of Proposition 44 to infer the relations

$$\int_{\Omega}(\hat{p}_n - p_{0,n}) f d\lambda = -D^2 L_{\mathbb{P}(p_{0,n})}(p_{0,n})(\hat{p}_n - p_{0,n}, \pi_n(f))$$
$$(\mathbb{P}_n - \mathbb{P}(p_{0,n}))f = DL_n(p_{0,n})(\pi_n(f)).$$

Therefore,

$$\begin{aligned}
&\int_{\Omega}(\hat{p}_n - p_{0,n}) f d\lambda - (\mathbb{P}_n - \mathbb{P}(p_{0,n}))f \\
&\quad = -D^2 L_{\mathbb{P}(p_{0,n})}(p_{0,n})(\hat{p}_n - p_{0,n}, \pi_n(f)) - DL_n(p_{0,n})(\pi_n(f)).
\end{aligned} \tag{33}$$

Applying the pathwise mean-value theorem to the function $DL_n(\,\cdot\,)(\pi_n(f))$ gives

$$DL_n(\hat{p}_n)(\pi_n(f)) = DL_n(p_{0,n})(\pi_n(f)) + D^2 L_n(\bar{p}_n)(\hat{p}_n - p_{0,n}, \pi_n(f)),$$

where $\bar{p}_n = \xi(n,f)\hat{p}_n + (1-\xi(n,f))p_{0,n}$ for some $\xi(n,f) \in (0,1)$. Hence

$$\begin{aligned}
&D^2 L_{\mathbb{P}(p_{0,n})}(p_{0,n})(\hat{p}_n - p_{0,n}, \pi_n(f)) \\
&\quad = (D^2 L_{\mathbb{P}(p_{0,n})}(p_{0,n}) - D^2 L_n(\bar{p}_n))(\hat{p}_n - p_{0,n}, \pi_n(f)) + D^2 L_n(\bar{p}_n)(\hat{p}_n - p_{0,n}, \pi_n(f)) \\
&\quad = (D^2 L_{\mathbb{P}(p_{0,n})}(p_{0,n}) - D^2 L_n(\bar{p}_n))(\hat{p}_n - p_{0,n}, \pi_n(f)) \\
&\qquad + DL_n(\hat{p}_n)(\pi_n(f)) - DL_n(p_{0,n})(\pi_n(f)) \\
&\quad = (D^2 L_{\mathbb{P}(p_{0,n})}(p_{0,n}) - D^2 L_{\mathbb{P}(p_{0,n})}(\bar{p}_n))(\hat{p}_n - p_{0,n}, \pi_n(f)) \\
&\qquad + (D^2 L_{\mathbb{P}(p_{0,n})}(\bar{p}_n) - D^2 L_n(\bar{p}_n))(\hat{p}_n - p_{0,n}, \pi_n(f)) \\
&\qquad + DL_n(\hat{p}_n)(\pi_n(f)) - DL_n(p_{0,n})(\pi_n(f)),
\end{aligned}$$

which together with (33) leads to the following inequality:

$$\begin{aligned}
&\sup_{f\in\mathcal{F}} \left| \int_\Omega (\hat{p}_n - p_{0,n}) f d\lambda - (\mathbb{P}_n - \mathbb{P}(p_{0,n}))f \right| \\
&\qquad \le \sup_{f\in\mathcal{F}} \left| (D^2 L_{\mathbb{P}(p_{0,n})}(p_{0,n}) - D^2 L_{\mathbb{P}(p_{0,n})}(\bar{p}_n))(\hat{p}_n - p_{0,n}, \pi_n(f)) \right| \\
&\qquad\quad + \sup_{f\in\mathcal{F}} \left| (D^2 L_{\mathbb{P}(p_{0,n})}(\bar{p}_n) - D^2 L_n(\bar{p}_n))(\hat{p}_n - p_{0,n}, \pi_n(f)) \right| \\
&\qquad\quad + \sup_{f\in\mathcal{F}} |DL_n(\hat{p}_n)(\pi_n(f))| \\
&\quad = \mathrm{I} + \mathrm{II} + \mathrm{III}.
\end{aligned}$$

To complete the proof of Part 1 when $s \ge t$, we first consider only j such that $1/2 < j < t$, and derive bounds for each of the above expressions.

Bound for I: Compute Expression I by making use of Proposition 61 with $P = \mathbb{P}(p_{0,n})$ and then bound it by

$$\begin{aligned}
&2\zeta^{-3} C_t D \sup_{f\in\mathcal{F}} [\|\pi_n(f)\|_2 \, \|(\bar{p}_n - p_{0,n})(\hat{p}_n - p_{0,n})\|_2] \\
&\qquad \text{(by the Cauchy-Schwarz inequality)} \\
&= \; 2\zeta^{-3} C_t D \sup_{f\in\mathcal{F}} \left[\|\pi_n(f)\|_2 \xi(n,f) \|\hat{p}_n - p_{0,n}\|_2^2\right] \\
&\qquad \text{(since } \bar{p}_n = \xi(n,f)(\hat{p}_n - p_{0,n})\text{)} \\
&\le \; 2\zeta^{-3} C_t M_t D^2 B (1 + C_t \lambda(\Omega)^{1/2}) \, \|\hat{p}_n - p_{0,n}\|_2^2 \\
&\qquad \text{(by (32))}.
\end{aligned}$$

We conclude from Theorem 33 that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}(p_{0,n})^{n*} \left(n^{2t/(2t+1)} \|\hat{p}_n - p_{0,n}\|_2^2 > M \right) = 0;$$

hence $\mathrm{I} = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-2t/(2t+1)})$.

Bound for II: We have

$$\mathrm{II} \le \|\hat{p}_n - p_{0,n}\|_{j,2} \sup_{p\in\mathcal{P}(t,\zeta,D)} \left\| D^2 L_{\mathbb{P}(p_{0,n})}(p) - D^2 L_n(p) \right\|_{\mathcal{U}_{j,1}\times\{\pi_n(f):n\in\mathbb{N},\, f\in\mathcal{F}\}}.$$

Now, $\|\hat{p}_n - p_{0,n}\|_{j,2}$ is $O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(t-j)/(2t+1)})$ by Theorem 33. Since $\{\pi_n(f) : n \in \mathbb{N},\, f \in \mathcal{F}\}$ is bounded in $\mathsf{W}_2^j(\Omega)$ by (32) and Part 3 of Proposition 1, and $\mathcal{U}_{j,1}$ is so too, it follows from Proposition 62, with $\alpha = 2$, $\mathcal{H}_1 = \mathcal{U}_{j,1}$, and $\mathcal{H}_2 = \{\pi_n(f) : n \in \mathbb{N},\, f \in \mathcal{F}\}$, that

$$\sup_{p\in\mathcal{P}(t,\zeta,D)} \left\| D^2 L_n(p) - D^2 L_{\mathbb{P}(p_{0,n})}(p) \right\|_{\mathcal{U}_{j,1}\times\{\pi_n(f):n\in\mathbb{N},\, f\in\mathcal{F}\}} = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-1/2}).$$

Hence, $\mathrm{II} = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(t-j)/(2t+1)-1/2})$.

Bound for III: Let

$$\eta \quad = \quad D - \sup_{n\in\mathbb{N}} \|p_{0,n}\|_{t,2},$$

which is positive by hypothesis,

$$\mathcal{U}_{t,\eta,0} = \left\{ w \in \mathsf{W}_2^t(\Omega) : \|w\|_{t,2} \le \eta, \int_\Omega w\, d\lambda = 0 \right\},$$

and

$$\hat{h}_n(w,\varepsilon) = \varepsilon(p_{0,n} + w) + (1-\varepsilon)\hat{p}_n \quad \text{for } w \in \mathsf{W}_2^t(\Omega) \text{ and } 0 < \varepsilon \le 1.$$

We show that, on events that have probability tending to 1, the inclusion

$$\left\{ \hat{h}_n(w,\varepsilon) : w \in \mathcal{U}_{t,\eta,0} \right\} \subseteq \mathcal{P}(t,\zeta,D)$$

holds for every ε, $0 < \varepsilon \le \varepsilon^*$, where ε^* is non-random and does not depend on n. First, $\|\hat{h}_n(w,\varepsilon)\|_{t,2} \le D$ by the triangle inequality. Second, by the hypothesis on the lower bound for the densities $p_{0,n}$, there is a $\beta > 0$ such that $\inf_{n\in\mathbb{N}} \inf_{x\in\Omega} p_{0,n}(x) > \zeta + \beta$ holds. Using the second statement in Theorem 33, we see that $\inf_{x\in\Omega} \hat{p}_n(x) > \zeta + \beta$ on events with inner probability going to 1. On these events, $\hat{h}_n(w,\varepsilon) \ge \zeta + \beta - \varepsilon C_t \eta$ (use that $w \ge -\|w\|_\Omega \ge -C_t\|w\|_{t,2} \ge -C_t\eta$), and therefore $\hat{h}_n(w,\varepsilon) \ge \zeta$ if only $\varepsilon > 0$ satisfies $\varepsilon \le \varepsilon^* := \beta/C_t\eta$. Note that $\hat{h}_n(w,\varepsilon)$ integrates to 1 since $\int_\Omega w\, d\lambda = 0$.

Next, define $s : \mathsf{W}_2^t(\Omega) \to \mathsf{W}_2^t(\Omega)$ by

$$s(f) = \begin{cases} \eta f \|f\|_{t,2}^{-1} & \text{if } f \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

Note that

$$\{ s(\pi_n(f)) : n \in \mathbb{N},\ f \in \mathcal{F} \} \subseteq \mathcal{U}_{t,\eta,0}.$$

Since $\hat{p}_n$ maximizes L_n over $\mathcal{P}(t,\zeta,D)$ and, for all ε such that $0 < \varepsilon \le \varepsilon^*$,

$$\hat{h}_n(s(\pi_n(f)), \varepsilon)$$

belongs to $\mathcal{P}(t,\zeta,D)$ on events with inner probability tending to 1, we conclude that on these events

$$DL_n(\hat{p}_n)(p_{0,n} - \hat{p}_n + s(\pi_n(f))) \le 0$$

holds for all $f \in \mathcal{F}$ (see the erratum to Nickl, 2007). This implies

$$\begin{aligned} DL_n(\hat{p}_n)(s(\pi_n(f))) &\le DL_n(\hat{p}_n)(\hat{p}_n - p_{0,n}) \\ &= (DL_n(\hat{p}_n) - DL_{\mathbb{P}(p_{0,n})}(p_{0,n}))(\hat{p}_n - p_{0,n}) \end{aligned}$$

for all $f \in \mathcal{F}$, where the equality in the last display is true since $DL_{\mathbb{P}(p_{0,n})}(p_{0,n})(\hat{p}_n - p_{0,n}) = 0$, as can be seen using Proposition 61 with $P = \mathbb{P}(p_{0,n})$. Since $\mathcal{U}_{t,\eta,0}$ also contains $-s(\pi_n(f))$, we can replace the l.h.s. in the last display by its absolute value. Hence

$$\begin{aligned} \sup_{f\in\mathcal{F}} |DL_n(\hat{p}_n)(s(\pi_n(f)))| &\le \left|(DL_n(\hat{p}_n) - DL_{\mathbb{P}(p_{0,n})}(\hat{p}_n))(\hat{p}_n - p_{0,n})\right| \\ &\quad + \left|(DL_{\mathbb{P}(p_{0,n})}(\hat{p}_n) - DL_{\mathbb{P}(p_{0,n})}(p_{0,n}))(\hat{p}_n - p_{0,n})\right| \\ &\le \|\hat{p}_n - p_{0,n}\|_{j,2} \sup_{p\in\mathcal{P}(t,\zeta,D)} \left\|DL_n(p) - DL_{\mathbb{P}(p_{0,n})}(p)\right\|_{\mathcal{U}_{j,1}} \\ &\quad + \zeta^{-1}\|\hat{p}_n - p_{0,n}\|_2^2, \end{aligned}$$

where we used Proposition 61 with $P = \mathbb{P}(p_{0,n})$. Use Theorem 33 and Proposition 62 with $\alpha = 1$ and $\mathcal{H}_1 = \mathcal{U}_{j,1}$ to conclude that the r.h.s. of the last display is

$$O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(t-j)/(2t+1)-1/2}) + O^*_{\mathbb{P}(p_{0,n})^n}(n^{-2t/(2t+1)}),$$

hence $O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(t-j)/(2t+1)-1/2})$ because $j > 1/2$. Together with (32) and the definition of $s(\,\cdot\,)$ this results in

$$\mathrm{III} = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(t-j)/(2t+1)}).$$

The above bounds imply that

$$\sup_{f\in\mathcal{F}}\left|\sqrt{n}\int_\Omega(\hat{p}_n - p_{0,n})fd\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))f\right| = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(t-j)/(2t+1)})$$

for every j satisfying $1/2 < j < t$, and hence a fortiori for all $j > 1/2$. Consequently, the l.h.s. of the last display is $o^*_{\mathbb{P}(p_{0,n})^n}(n^{-(t-j)/(2t+1)})$ for all $j > 1/2$, hence in particular $o^*_{\mathbb{P}(p_{0,n})^n}(1)$ as $n \to \infty$. This completes the proof of Part 1 when $s \geq t$.

Step 2: We now consider the case when $1/2 < s < t$, and assume w.l.o.g. that $\mathcal{F} = \mathcal{U}_{t,B}$ for some B, $0 < B < \infty$. In the proof of Proposition 1 in Nickl (2007) there are defined approximating sequences $u_n(f) \in \mathsf{W}_2^t(\Omega)$ for all $f \in \mathcal{F}$, with the following properties:

$$\sup_{f\in\mathcal{F}}\|u_n(f)\|_{t,2} = O(n^{(t-s)/(2t+1)}) \qquad \text{as } n\to\infty, \tag{34}$$

where $\sup_{f\in\mathcal{F}}\|u_n(f)\|_{t,2}$ is real for every $n \in \mathbb{N}$; and, for every r, $0 \leq r < s$,

$$\sup_{f\in\mathcal{F}}\|f - u_n(f)\|_{r,2} = O(n^{-(s-r)/(2t+1)}) \qquad \text{as } n\to\infty. \tag{35}$$

We have that

$$\begin{aligned}
&\sup_{f\in\mathcal{F}}\left|\sqrt{n}\int_\Omega(\hat{p}_n - p_{0,n})fd\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))f\right| \\
&\quad\leq \sup_{f\in\mathcal{F}}\left|\sqrt{n}\int_\Omega(\hat{p}_n - p_{0,n})(f - u_n(f))d\lambda\right| \\
&\qquad + \sup_{f\in\mathcal{F}}\left|\sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))(f - u_n(f))\right| \\
&\qquad + \sup_{f\in\mathcal{F}}\left|\sqrt{n}\int_\Omega(\hat{p}_n - p_{0,n})u_n(f)d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))u_n(f)\right| \\
&\quad= \mathrm{IV} + \mathrm{V} + \mathrm{VI}.
\end{aligned}$$

To complete the proof of Part 1 when $1/2 < s < t$, we choose an arbitrary j' such that $1/2 < j' < \min(j, s)$, where j is as in the theorem, and derive bounds for each of the above expressions.

Bound for IV: By (34), with $r = 0$, and by Theorem 33,

$$\begin{aligned}
\mathrm{IV} &\leq \sqrt{n}\|\hat{p}_n - p_{0,n}\|_2 \sup_{f\in\mathcal{F}}\|f - u_n(f)\|_2 \\
&\qquad\qquad \text{(by the Cauchy-Schwarz inequality)} \\
&= O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(s-1/2)/(2t+1)}).
\end{aligned}$$

Bound for V: Expression V can be bounded by

$$\|\sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))\|_{\mathcal{U}_{j',1}} \sup_{f\in\mathcal{F}} \|f - u_n(f)\|_{j',2},$$

so that, by Proposition 62, with $\alpha = 1$ and $\mathcal{H}_1 = \mathcal{U}_{j',1}$, and by (35), with $r = j'$,

$$\mathrm{V} = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(s-j')/(2t+1)}).$$

Bound for VI: Note that

$$\begin{aligned}
\mathrm{VI} &= \sup_{\substack{f\in\mathcal{F},\\ \|u_n(f)\|_{t,2}\neq 0}} \Bigg[\bigg|\sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n})\frac{u_n(f)}{\|u_n(f)\|_{t,2}}d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))\frac{u_n(f)}{\|u_n(f)\|_{t,2}}\bigg| \\
&\qquad \times \|u_n(f)\|_{t,2}\Bigg] \\
&\leq \sup_{h\in\mathcal{U}_{t,1}} \left|\sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n})h d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))h\right| \times \sup_{f\in\mathcal{F}}\|u_n(f)\|_{t,2}.
\end{aligned}$$

Applying Step 1 to the first term on the r.h.s. of the last display, and using (34), we conclude that

$$\mathrm{VI} = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(s-j)/(2t+1)})$$

for all $j > 1/2$.

The above bounds imply that

$$\sup_{f\in\mathcal{F}}\left|\sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n})f d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))f\right| = O^*_{\mathbb{P}(p_{0,n})^n}(n^{-(s-j)/(2t+1)})$$

for all $j > 1/2$. Consequently, the l.h.s. of the last display is $o^*_{\mathbb{P}(p_{0,n})^n}(n^{-(s-j)/(2t+1)})$ for all $j > 1/2$, hence in particular $o^*_{\mathbb{P}(p_{0,n})^n}(1)$ as $n\to\infty$.

Part 2: Note first that the empirical process $\sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))(\,\cdot\,)$ converges to $\mathbb{G}$ in $\ell^\infty(\mathcal{F})$, where $\mathbb{G}$ is a centered Gaussian process indexed by $\mathcal{F}$ with covariance function

$$\mathrm{Cov}[\mathbb{G}(f), \mathbb{G}(g)] = \int_\Omega \left(f - \int_\Omega f p_{0,\infty} d\lambda\right)\left(g - \int_\Omega g p_{0,\infty} d\lambda\right) p_{0,\infty} d\lambda$$

and bounded sample paths that are uniformly continuous with respect to d. This follows from Corollary 2.7 in Giné & Zinn (1991) since $\mathcal{F}$ is uniformly Donsker by Lemma 13 and (in view of Proposition 3)

$$\begin{aligned}
\lim_{n\to\infty}\|\mathbb{P}(p_{0,n}) - \mathbb{P}(p_{0,\infty})\|_{\mathcal{G}} &= \lim_{n\to\infty}\sup_{g\in\mathcal{G}}\left|\int_\Omega g(p_{0,n} - p_{0,\infty})d\lambda\right| \\
&\leq \lambda(\Omega)\sup_{g\in\mathcal{G}}\|g\|_\Omega \lim_{n\to\infty}\|p_{0,n} - p_{0,\infty}\|_\Omega \\
&= 0.
\end{aligned}$$

for any class $\mathcal{G}$ of functions that is bounded in the sup-norm. Since the difference

$$\sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n})(\,\cdot\,)d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))(\,\cdot\,)$$

converges in outer probability to 0 in the space $\ell^\infty(\mathcal{F})$, as shown in Part 1, we have verified Part 2.

Part 3: For any $f \in \mathcal{F}$, define $\delta_f : \ell^\infty(\mathcal{F}) \to \mathbb{R}$ by $\delta_f(\varphi) = \varphi(f)$. We apply Theorem 1.11.1 in van der Vaart & Wellner (1996) with $\mathbb{D}_n = \mathbb{D} = \ell^\infty(\mathcal{F})$, $\mathbb{D}_0 = \mathsf{C}(\mathcal{F}, d)$, $\mathbb{E} = \mathbb{R}$, $g_n = \delta_{f_n}$, $g = \delta_{f_0}$, $X_n = \sqrt{n}\int_\Omega(\hat{p}_n - p_{0,n})(\,\cdot\,)d\lambda$, and $X = \mathbb{G}$. To verify its premises, note that $\mathbb{G}$ takes its values in $\mathsf{C}(\mathcal{F}, d)$ and is measurable. Further, for $\varphi_n \in \ell^\infty(\mathcal{F})$ and $\varphi \in \mathsf{C}(\mathcal{F}, \rho)$ such that $\|\varphi_n - \varphi\|_{\mathcal{F}}$ converges to 0, we conclude from

$$|\varphi_n(f_n) - \varphi(f_0)| \le \sup_{f\in\mathcal{F}} |\varphi_n(f) - \varphi(f)| + |\varphi(f_n) - \varphi(f_0)|$$

that $\delta_{f_n}(\varphi_n)$ converges to $\delta_{f_0}(\varphi)$. [Clearly, $\varphi(f_n)$ converges to $\varphi(f_0)$ since pointwise convergence of f_n to f_0 implies that $d(f_n, f_0)$ goes to 0 by the theorem of dominated convergence.] Since $X_n \rightsquigarrow X$ by Part 2, we conclude that $\delta_{f_n}(X_n) \rightsquigarrow \delta_{f_0}(X)$, which just means that

$$\sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n}) f_n d\lambda \rightsquigarrow \mathbb{G}(f_0).$$

□

7.2 II-estimators: regularity and asymptotic efficiency

Maintaining Assumption Z, we define

$$Q_q(\theta) = \int_\Omega (q - p_\theta)^2 \frac{1}{q} d\lambda$$

for every $q \in \mathcal{P}(t, \zeta, D)$.

Proposition 35. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let Assumption Z be satisfied. Let $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t, \zeta, D)$ be such that $p_{0,n}$ converges pointwise to $p_{0,\infty}$, and let $\theta^*_{0,\infty} \in \Theta$ be the unique minimizer of $Q_{p_{0,\infty}}$. Then any II-estimator $\hat{\theta}_n$ converges to $\theta^*_{0,\infty}$ in outer probability as $n \to \infty$, where the underlying probability spaces are $(\Omega^n, \mathcal{B}(\Omega)^n, \mathbb{P}(p_{0,n})^n)$.*

Proof. For any fixed $\varepsilon > 0$, let

$$c(\varepsilon) = \inf_{\|\theta - \theta^*_{0,\infty}\| \ge \varepsilon} [Q_{p_{0,\infty}}(\theta) - Q_{p_{0,\infty}}(\theta^*_{0,\infty})].$$

Since $Q_{p_{0,\infty}}$ is a continuous function by Part 1 of Proposition 64, since the set $\{\theta \in \Theta : \|\theta - \theta^*_{0,\infty}\| \ge \varepsilon\}$ is compact, and since $Q_{p_{0,\infty}}(\theta) > Q_{p_{0,\infty}}(\theta^*)$ for any $\theta \ne \theta^*_{0,\infty}$, we see that $c(\varepsilon) > 0$.

Observe that

$$\begin{aligned}
\sup_{\theta\in\Theta} |Q_{p_{0,n}}(\theta) - Q_{p_{0,\infty}}(\theta)| &= \sup_{\theta\in\Theta} \left| \int_\Omega (p_{0,n} - p_{0,\infty}) \frac{-p_\theta^2}{p_{0,n} p_{0,\infty}} d\lambda \right| \\
&\le \zeta^{-2} \|p_{0,n} - p_{0,\infty}\|_\Omega \sup_{\theta\in\Theta} \|p_\theta\|_2^2.
\end{aligned}$$

Since, by hypothesis, Θ is compact and $\theta \mapsto p_\theta$ is a continuous mapping from Θ to $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$, we have that $\sup_{\theta\in\Theta} \|p_\theta\|_2 < \infty$. Since $\|p_{0,n} - p_{0,\infty}\|_\Omega$ converges to 0 in view of Proposition 3, we see that the r.h.s. of the previous display, and therefore $\sup_{\theta\in\Theta}|Q_{p_{0,n}}(\theta) - Q_{p_{0,\infty}}(\theta)|$, converges to 0 as $n \to \infty$. This together with Part 2 of Proposition 64 shows that, for any $\delta > 0$,

$$\sup_{\theta\in\Theta}|\mathbb{Q}_n(\theta) - Q_{p_{0,n}}(\theta)| \le \delta$$

on events whose inner $\mathbb{P}(p_{0,n})^n$-probability tends to 1 as $n \to \infty$. Choose $\delta > 0$ such that $c(\varepsilon) - 2\delta > 0$. With the help of the inequality

$$\begin{aligned}&\inf_{\|\theta-\theta^*_{0,\infty}\|\ge\varepsilon}[\mathbb{Q}_n(\theta) - \mathbb{Q}_n(\theta^*_{0,\infty})] \\ &\quad\ge \inf_{\|\theta-\theta^*_{0,\infty}\|\ge\varepsilon}[Q_{p_{0,\infty}}(\theta) - Q_{p_{0,\infty}}(\theta^*_{0,\infty})] - 2\sup_{\theta\in\Theta}|\mathbb{Q}_n(\theta) - Q_{p_{0,\infty}}(\theta)|\end{aligned}$$

it then follows that

$$\inf_{\|\theta-\theta^*_{0,\infty}\|\ge\varepsilon}[\mathbb{Q}_n(\theta) - \mathbb{Q}_n(\theta^*_{0,\infty})] \ge c(\varepsilon) - 2\delta > 0$$

on events having inner $\mathbb{P}(p_{0,n})^n$-probability going to 1. This implies that, on these events, $\hat{\theta}_n$ does not belong to the set $\{\theta \in \Theta : \|\theta - \theta^*_{0,\infty}\| \ge \varepsilon\}$ since $\hat{\theta}_n$ minimizes $\mathbb{Q}_n$ by Part 1 of Proposition 20. $\square$

Proposition 36. *Let Assumptions P.1, P.2, R.1, and Z be satisfied. Let $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t,\zeta,D)$ be such that $p_{0,n}$ converges pointwise to $p_{0,\infty}$, and let $\theta^*_{0,\infty}$ be the unique minimizer of $Q_{p_{0,\infty}}$. Then any simulation-based II-estimator $\hat{\theta}_{n,k}$ converges to $\theta^*_{0,\infty}$ in outer probability as $\min(n,k) \to \infty$, where the underlying probability spaces are $(\Omega^n \times V^k, \mathcal{B}(\Omega)^n \otimes \mathcal{V}^k, \mathbb{P}(p_{0,n})^n \otimes \mu^k)$.*

Proof. The proof is completely analogous to the proof of Proposition 35: For any fixed $\varepsilon > 0$, let

$$c(\varepsilon) = \inf_{\|\theta-\theta^*_{0,\infty}\|\ge\varepsilon}[Q_{p_{0,\infty}}(\theta) - Q_{p_{0,\infty}}(\theta^*_{0,\infty})].$$

Since $Q_{p_{0,\infty}}$ is a continuous function by Part 1 of Proposition 64, since the set $\{\theta \in \Theta : \|\theta - \theta^*_{0,\infty}\| \ge \varepsilon\}$ is compact, and since $Q_{p_{0,\infty}}(\theta) > Q_{p_{0,\infty}}(\theta^*)$ for any $\theta \ne \theta^*_{0,\infty}$, we see that $c(\varepsilon) > 0$.

Observe that

$$\begin{aligned}\sup_{\theta\in\Theta}|Q_{p_{0,n}}(\theta) - Q_{p_{0,\infty}}(\theta)| &= \sup_{\theta\in\Theta}\left|\int_\Omega (p_{0,n} - p_{0,\infty})\frac{-p_\theta^2}{p_{0,n}p_{0,\infty}}d\lambda\right| \\ &\le \zeta^{-2}\|p_{0,n} - p_{0,\infty}\|_\Omega \sup_{\theta\in\Theta}\|p_\theta\|_2^2.\end{aligned}$$

Since, by hypothesis, Θ is compact and $\theta \mapsto p_\theta$ is a continuous mapping from Θ to $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$, we have that $\sup_{\theta\in\Theta}\|p_\theta\|_2 < \infty$. Since $\|p_{0,n} - p_{0,\infty}\|_\Omega$ converges to 0 in view of Proposition 3, we see that the r.h.s. of the previous display, and therefore $\sup_{\theta\in\Theta}|Q_{p_{0,n}}(\theta) - Q_{p_{0,\infty}}(\theta)|$, converges to 0 as $n \to \infty$. This together with Part 3 of Proposition 64 shows that, for any $\delta > 0$,

$$\sup_{\theta\in\Theta}|\mathbb{Q}_{n,k}(\theta) - Q_{p_{0,\infty}}(\theta)| \le \delta$$

on events whose inner $\mathbb{P}(p_{0,n})^n \otimes \mu^{k(n)}$-probability tends to 1 as $\min(n,k) \to \infty$. Choose $\delta > 0$ such that $c(\varepsilon) - 2\delta > 0$. With help of the inequality

$$\begin{aligned}&\inf_{\|\theta-\theta^*_{0,\infty}\|\geq\varepsilon}[\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_{n,k}(\theta^*_{0,\infty})] \\ &\quad\geq \inf_{\|\theta-\theta^*_{0,\infty}\|\geq\varepsilon}[Q_{p_{0,\infty}}(\theta) - Q_{p_{0,\infty}}(\theta^*_{0,\infty})] - 2\sup_{\theta\in\Theta}|\mathbb{Q}_{n,k}(\theta) - Q_{p_{0,\infty}}(\theta)|\end{aligned}$$

it then follows that

$$\inf_{\|\theta-\theta^*_{0,\infty}\|\geq\varepsilon}[\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_{n,k}(\theta^*_{0,\infty})] \geq c(\varepsilon) - 2\delta > 0$$

on events having inner $\mathbb{P}(p_{0,n})^n \otimes \mu^{k(n)}$-probability going to 1 as $\min(n,k) \to \infty$. This implies that, on these events, $\hat{\theta}_{n,k}$ does not belong to the set $\{\theta \in \Theta : \|\theta - \theta^*_{0,\infty}\| \geq \varepsilon\}$ since $\hat{\theta}_{n,k}$ minimizes $\mathbb{Q}_{n,k}$ by Part 1 of Proposition 25. $\square$

Note that under the assumptions of the following theorem, as well as Theorem 39, the functions $Q_{p_{0,n}}$ and $Q_{p_{0,\infty}}$ have minimizers (cf. Lemma 47) and the matrix $\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta\partial\theta'}(\theta)$ exists for every θ in the (non-empty) interior of Θ (see Lemma 66).

Theorem 37. *Let Assumptions P.1, P.4, P.5, and Z be satisfied. Let $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t,\zeta,D)$ and $p_{0,n}$ converge pointwise to $p_{0,\infty}$. For each $n \in \mathbb{N}$, let $\theta^*_{0,n} \in \Theta$ be a minimizer of $Q_{p_{0,n}}$; let $\theta^*_{0,\infty} \in \Theta$ be the unique minimizer of $Q_{p_{0,\infty}}$, and suppose it belongs to the interior Θ° of Θ. Furthermore, let the following conditions be satisfied:*

1. *The density function $p_{0,\infty}$ satisfies the strict inequalities*

$$\inf_{x\in\Omega} p_{0,\infty}(x) > \zeta \qquad \text{and} \qquad \|p_{0,\infty}\|_{t,2} < D.$$

2. *The $m \times m$ matrix with (i,j)-th entry*

$$\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta_i\partial\theta_j}(\theta^*_{0,\infty})$$

is positive definite.

3. *The set*

$$\left\{\frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta^*_{0,\infty}) : i = 1,\dots,m\right\} \cup \left\{\frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta^*_{0,n}) : \theta^*_{0,n} \in \Theta^\circ,\, i = 1,\dots,m\right\}$$

is bounded in some Sobolev space $\mathsf{W}^s_2(\Omega)$ of order $s > 1/2$.

Then

$$\sqrt{n}(\hat{\theta}_n - \theta^*_{0,n}) \rightsquigarrow N(0, \Sigma(\theta^*_{0,\infty})) \quad \text{as } n \to \infty,$$

where $\Sigma(\theta^*_{0,\infty})$ is positive definite and is given by

$$\begin{aligned}
&\Sigma(\theta^*_{0,\infty})\\
&\quad= \left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda - \int_\Omega (p_{0,\infty}-p_{\theta^*_{0,\infty}})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda\right]^{-1}\\
&\qquad\times\left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p^2_{\theta^*_{0,\infty}}}{p^3_{0,\infty}}d\lambda\right.\\
&\qquad\qquad\left.-\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}}d\lambda\int_\Omega \frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}}d\lambda\right]\\
&\qquad\times\left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda\right.\\
&\qquad\qquad\left.-\int_\Omega (p_{0,\infty}-p_{\theta^*_{0,\infty}})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda\right]^{-1}.
\end{aligned}$$

Here, weak convergence refers to the underlying probability spaces $(\Omega^n,\mathcal{B}(\Omega)^n,\mathbb{P}(p_{0,n})^n)$. If there is a parameter $\theta_{0,\infty}\in\Theta$ such that $p_{0,\infty}$ equals $p_{\theta_{0,\infty}}$ a.e., then $\theta^*_{0,\infty}=\theta_{0,\infty}$ and

$$\sqrt{n}(\hat\theta_n-\theta^*_{0,n}) \rightsquigarrow N(0,I^{-1}(\theta_{0,\infty})) \quad \textit{as } n\to\infty,$$

where $I^{-1}(\theta_{0,\infty})$ denotes the inverse of the Fisher information matrix.

Proof. By Proposition 35, $\hat\theta_n$ comes to lie in Θ° on events that have inner $\mathbb{P}(p_{0,n})^n$-probability going to 1. We have that

$$\frac{\partial\mathbb{Q}_n}{\partial\theta}(\hat\theta_n)=0$$

on these events. Applying the mean-value theorem to each component of $\frac{\partial\mathbb{Q}_n}{\partial\theta}$ gives

$$\frac{\partial\mathbb{Q}_n}{\partial\theta}(\theta^*_{0,n})+H_n(\hat\theta_n-\theta^*_{0,n})=0$$

on these events, where H_n is the Hessian matrix of $\mathbb{Q}_n$ with i-th row evaluated at some mean value $\bar\theta_{n,i}$ lying on the line segment that joins $\theta^*_{0,n}$ and $\hat\theta_n$. Since $\theta^*_{0,n}$ converges to $\theta^*_{0,\infty}$ by Lemma 65 and $\hat\theta_n$ converges to $\theta^*_{0,\infty}$ in outer $\mathbb{P}(p_{0,n})^n$-probability by Proposition 35, we see that $\bar\theta_{n,i}$ converges to $\theta^*_{0,\infty}$ in outer $\mathbb{P}(p_{0,n})^n$-probability. Hence, H_n converges to $\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta\partial\theta'}(\theta^*_{0,\infty})$ in outer $\mathbb{P}(p_{0,n})^n$-probability by Proposition 67 and continuity of $\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta\partial\theta'}$ on Θ°. Since the latter matrix is invertible by Condition 2, it follows that H_n is invertible on events that have inner $\mathbb{P}(p_{0,n})^n$-probability tending to 1, which leads to the relation

$$\sqrt{n}(\hat\theta_n-\theta^*_{0,n}) = -H_n^{-1}\sqrt{n}\frac{\partial\mathbb{Q}_n}{\partial\theta}(\theta^*_{0,n}) \tag{36}$$

on these events. Since inverting matrices is continuous, we infer from the continuous mapping theorem that H_n^{-1} converges in outer probability to the inverse of $\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta\partial\theta'}(\theta^*_{0,\infty})$, hence to

$$\left[2\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda - 2\int_\Omega (p_{0,\infty}-p_{\theta^*_{0,\infty}})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda\right]^{-1} \tag{37}$$

by Lemma 66. To see that $\sqrt{n}\frac{\partial \mathbb{Q}_n}{\partial\theta}(\theta^*_{0,n})$ satisfies a central limit theorem, we proceed as follows: Let $v \in \mathbb{R}^m$ be arbitrary, and use Part 1 of Lemma 51 to obtain

$$\begin{aligned} v'\sqrt{n}\frac{\partial \mathbb{Q}_n}{\partial\theta}(\theta^*_{0,n}) &= 2\sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n})^2 v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n})\frac{p_{\theta^*_{0,n}}}{\hat{p}_n p^2_{0,n}}d\lambda \\ &\quad -2\sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n}) v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n})\frac{p_{\theta^*_{0,n}}}{p^2_{0,n}}d\lambda \\ &\quad -2\sqrt{n}\int_\Omega (p_{0,n} - p_{\theta^*_{0,n}}) v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n})\frac{1}{p_{0,n}}d\lambda \\ &= \mathrm{I} + \mathrm{II} + \mathrm{III}. \end{aligned}$$

Note that since $\theta^*_{0,n}$ converges to $\theta^*_{0,\infty}$, there is an index n° such that $\theta^*_{0,n}$ lies in Θ° for every $n \geq n^\circ$.

Convergence of I: By Condition 3, the set

$$\left\{ v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n}) : n \geq n^\circ \right\}$$

is bounded in $\mathsf{W}^s_2(\Omega)$ and thus bounded by

$$C_s \sup\left\{ \left\| v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n}) \right\|_{s,2} : n \geq n^\circ \right\}$$

in the sup-norm. Together with

$$\left\| \frac{p_{\theta^*_{0,n}}}{\hat{p}_n p^2_{0,n}} \right\|_\Omega \leq \zeta^{-3} C_t D$$

we see that Expression I is bounded by

$$2C_s \sup\left\{ \left\| v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n}) \right\|_{s,2} : n \geq n^\circ \right\} \zeta^{-3} C_t D\sqrt{n}\|\hat{p}_n - p_{0,n}\|_2^2$$

and thus converges to 0 in outer $\mathbb{P}(p_{0,n})^n$-probability by applying Theorem 33 with $r = 0$.

Convergence of II: We will use Part 3 of Theorem 34 with $\mathcal{F} = \{f_n : n \geq n^\circ\} \cup \{f_\infty\}$, where

$$f_n := -2v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n})\frac{p_{\theta^*_{0,n}}}{p^2_{0,n}}$$

and

$$f_\infty := -2v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p^2_{0,\infty}}.$$

For this, we first show that f_n belongs to $\mathsf{W}^r_2(\Omega)$, where $r = \min(s,t)$: Since $\mathsf{W}^r_2(\Omega)$ is a multiplication algebra by Part 1 of Proposition 1, it suffices to show that each factor of f_n belongs to $\mathsf{W}^r_2(\Omega)$: $-2v'\frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,n}) \in \mathsf{W}^s_2(\Omega)$ by Condition 3, and $p_{\theta^*_{0,n}} \in \mathsf{W}^t_2(\Omega)$ by Assumption P.1; hence these factors lie in $\mathsf{W}^r_2(\Omega)$ by Part 3 of Proposition 1; $1/p_{0,n}$ lies in $\mathsf{W}^r_2(\Omega)$ by Parts 3 and 4 of Proposition 1. The same arguments also apply to the function f_∞ and show that $f_\infty \in \mathsf{W}^r_2(\Omega)$. Using Assumption P.1 and Condition 3, we thus see that $\mathcal{F}$ is a bounded subset of $\mathsf{W}^r_2(\Omega)$.

To see that f_n converges pointwise to f_∞, use that $\theta^*_{0,n}$ converges to $\theta^*_{0,\infty}$ together with Part 1 of Proposition 5, the continuity of $\frac{\partial p}{\partial \theta}(x,\theta)$ in the second argument, and the hypothesis that $p_{0,n}$ converges pointwise to $p_{0,\infty}$.

Finally, we conclude from Part 3 of Theorem 34 that

$$II = \sqrt{n}\int_\Omega (\hat{p}_n - p_{0,n}) f_n d\lambda \rightsquigarrow N(0, \mathbb{P}(p_{0,\infty})(f - \mathbb{P}(p_{0,\infty})f)^2).$$

Convergence of III: Since $\theta^*_{0,n}$ minimizes $Q_{p_{0,n}}$ and, for $n \geq n^\circ$, lies in the interior of Θ, we have that $\frac{\partial Q_{p_{0,n}}}{\partial \theta}(\theta^*_{0,n}) = 0$ for every $n \geq n^\circ$. Since, up to the factor $\sqrt{n}$, Expression *III* is equal to $v' \frac{\partial Q_{p_{0,n}}}{\partial \theta}(\theta^*_{0,n})$ by Lemma 66, it is 0 for $n \geq n^\circ$.

By the Cramér-Wold device, $\sqrt{n}\frac{\partial \mathbb{Q}_n}{\partial \theta}(\theta^*_{0,\infty})$ asymptotically follows a central normal distribution with variance-covariance matrix given by

$$4\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p^2_{\theta^*_{0,\infty}}}{p^3_{0,\infty}} d\lambda - 4\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}} d\lambda \int_\Omega \frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}} d\lambda. \tag{38}$$

Combine that $-H_n^{-1}$ converges in outer probability to (37) with (38) to obtain

$$\sqrt{n}(\hat{\theta}_n - \theta^*_{0,n}) \rightsquigarrow N(0, \Sigma(\theta^*_{0,\infty})),$$

where

$$\begin{aligned}
&\Sigma(\theta^*_{0,\infty}) \\
&\quad = \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}} d\lambda - \int_\Omega (p_{0,\infty} - p_{\theta^*_{0,\infty}})\frac{\partial^2 p}{\partial \theta \partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}} d\lambda\right]^{-1} \\
&\qquad \times \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p^2_{\theta^*_{0,\infty}}}{p^3_{0,\infty}} d\lambda \right. \\
&\qquad\qquad \left. - \int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}} d\lambda \int_\Omega \frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}} d\lambda\right] \\
&\qquad \times \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}} d\lambda \right. \\
&\qquad\qquad \left. - \int_\Omega (p_{0,\infty} - p_{\theta^*_{0,\infty}})\frac{\partial^2 p}{\partial \theta \partial \theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}} d\lambda\right]^{-1}.
\end{aligned}$$

If there is a parameter $\theta_{0,\infty} \in \Theta$ such that $p_{0,\infty}$ equals $p_{\theta_{0,\infty}}$ a.e.,, then $\theta^*_{0,\infty}$ coincides with $\theta_{0,\infty}$ in view of Condition 1, and $\Sigma(\theta^*_{0,\infty})$ simplifies to

$$\Sigma(\theta_{0,\infty}) = \left[\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_{0,\infty})\frac{\partial p}{\partial \theta'}(\,\cdot\,,\theta_{0,\infty})\frac{1}{p_{0,\infty}} d\lambda\right]^{-1}.$$

[Here, we have used that $\int_\Omega \frac{\partial p}{\partial \theta}(\,\cdot\,,\theta_{0,\infty}) d\lambda = 0$, which follows from Assumption P.5 and the inequality $\int_\Omega |f| d\lambda \leq \lambda(\Omega)^{1/2}\|f\|_2$.] We conclude that in this case

$$\sqrt{n}(\hat{\theta}_n - \theta^*_{0,n}) \rightsquigarrow N(0, I^{-1}(\theta_{0,\infty})) \quad \text{as } n \to \infty.$$

□

Theorem 38. *Let Assumptions P.1, P.5, R.2, and Z be satisfied. Let $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t,\zeta,D)$ and $p_{0,n}$ converge pointwise to $p_{0,\infty}$. For each $n \in \mathbb{N}$, let $\theta^*_{0,n} \in \Theta$ be a minimizer of $Q_{p_{0,n}}$; let $\theta^*_{0,\infty} \in \Theta$ be the unique minimizer of $Q_{p_{0,\infty}}$, and suppose it belongs to the interior Θ° of Θ. Further, let the following conditions be satisfied:*

1. *The density function $p_{0,\infty}$ satisfies the strict inequalities*
$$\inf_{x\in\Omega} p_{0,\infty}(x) > \zeta \quad \text{and} \quad \|p_{0,\infty}\|_{t,2} < D.$$
2. *The $m \times m$ matrix with (i,j)-th entry*
$$\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta_i\partial\theta_j}(\theta^*_{0,\infty})$$
is positive definite.
3. *The set*
$$\left\{\frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta^*_{0,\infty}) : i = 1,\ldots,m\right\} \cup \left\{\frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta^*_{0,n}) : \theta^*_{0,n} \in \Theta^\circ,\, i = 1,\ldots,m\right\}$$
is bounded in some Sobolev space $\mathsf{W}^s_2(\Omega)$ of order $s > 1/2$.

In addition, either (i) *let Assumption P.2 be satisfied, and choose $k(n)$ such that $\liminf_{n\to\infty} k(n)/n^{2+1/t} > 0$; or* (ii) *let Assumption P.3 be satisfied, and choose $k(n)$ such that $\liminf_{n\to\infty} k(n)/n^{2+\alpha}$ for some (arbitrarily small) $\alpha > 0$. Then*
$$\sqrt{n}(\hat\theta_{n,k(n)} - \theta^*_{0,n}) \rightsquigarrow N(0, \Sigma(\theta^*_{0,\infty})) \quad \text{as } n \to \infty,$$
where $\Sigma(\theta^*_{0,\infty})$ is positive definite and is given by
$$\begin{aligned}
&\Sigma(\theta^*_{0,\infty})\\
&= \left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda - \int_\Omega (p_{0,\infty} - p_{\theta^*_{0,\infty}})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda\right]^{-1}\\
&\quad\times\left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p^2_{\theta^*_{0,\infty}}}{p^3_{0,\infty}}d\lambda\right.\\
&\qquad\left. - \int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}}d\lambda \int_\Omega \frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{p_{\theta^*_{0,\infty}}}{p_{0,\infty}}d\lambda\right]\\
&\quad\times\left[\int_\Omega \frac{\partial p}{\partial\theta}(\,\cdot\,,\theta^*_{0,\infty})\frac{\partial p}{\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda\right.\\
&\qquad\left. - \int_\Omega (p_{0,\infty} - p_{\theta^*_{0,\infty}})\frac{\partial^2 p}{\partial\theta\partial\theta'}(\,\cdot\,,\theta^*_{0,\infty})\frac{1}{p_{0,\infty}}d\lambda\right]^{-1}.
\end{aligned}$$

Here, weak convergence refers to the underlying probability spaces $(\Omega^n \times V^{k(n)}, \mathcal{B}(\Omega)^n \otimes \mathcal{V}^{k(n)}, \mathbb{P}(p_{0,n})^n \otimes \mu^{k(n)})$. If there is a parameter $\theta_{0,\infty} \in \Theta$ such that $p_{0,\infty}$ equals $p_{\theta_{0,\infty}}$ a.e., then $\theta^*_{0,\infty} = \theta_{0,\infty}$ and
$$\sqrt{n}(\hat\theta_{n,k(n)} - \theta^*_{0,n}) \rightsquigarrow N(0, I^{-1}(\theta_{0,\infty})) \quad \text{as } n \to \infty,$$
where $I^{-1}(\theta_{0,\infty})$ denotes the inverse of the Fisher information matrix.

Proof. The key step lies in exploiting the closeness of the random objective functions $\mathbb{Q}_{n,k(n)}$ and $\mathbb{Q}_n$ to prove that

$$\sqrt{n}(\hat{\theta}_{n,k(n)} - \hat{\theta}_n) = o^*_{\mathbb{P}(p_{0,n})^n \otimes \mu^{k(n)}}(1) \quad \text{as } n \to \infty.$$

This will be achieved through Lemma 30: By Condition 2, the matrix $\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta\partial\theta'}(\theta^*_{0,\infty})$ is positive definite. Use this together with the continuity of $\frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta\partial\theta'}$ on Θ° (see Lemma 66) and the fact that, for all $i, j = 1, \ldots, m$,

$$\sup_{\theta\in\Theta^\circ} \left| \frac{\partial^2 \mathbb{Q}_n}{\partial\theta_i\partial\theta_j}(\theta) - \frac{\partial^2 Q_{p_{0,\infty}}}{\partial\theta_i\partial\theta_j}(\theta) \right| = o^*_{\mathbb{P}(p_{0,n})^n}(1) \quad \text{as } n \to \infty$$

(see Proposition 67) to conclude that there are events E_n that have inner $\mathbb{P}(p_{0,n})^n$-probability tending to 1 as $n \to \infty$ such that

$$\inf_{\theta\in U} \left[y' \frac{\partial^2 \mathbb{Q}_n}{\partial\theta\partial\theta'}(\theta) y \right] \geq K\|y\|^2 \quad \text{for all } y \in \mathbb{R}^m$$

holds true for some open, convex neighbourhood U of $\theta^*_{0,\infty}$, $U \subseteq \Theta^\circ$, and some constant $K > 0$, where neither U nor K depends on n or the data. By Propositions 35 and 36, $\hat{\theta}_n$ and $\hat{\theta}_{n,k(n)}$ both come to lie in U on events E'_n that have inner $\mathbb{P}(p_{0,n})^n \otimes \mu^{k(n)}$-probability going to 1 as $n \to \infty$. For the rest of the proof we restrict our reasoning to the events $E_n \cap E'_n$, and note that they have inner $\mathbb{P}(p_{0,n})^n \otimes \mu^{k(n)}$-probability tending to 1 as $n \to \infty$. Since, by Part 1 of Proposition 20 and Part 1 of Proposition 25, $\hat{\theta}_n$ and $\hat{\theta}_{n,k(n)}$, respectively, are minimizers of the objective functions $\mathbb{Q}_n$ and $\mathbb{Q}_{n,k(n)}$, we may apply Lemma 30 with $f = \mathbb{Q}_{n,k(n)}|U$, $g = \mathbb{Q}_n|U$, $u = \hat{\theta}_{n,k(n)}$, and $v = \hat{\theta}_n$ to infer the inequality

$$\|\hat{\theta}_{n,k(n)} - \hat{\theta}_n\| \leq 2K^{-1/2}\sqrt{\|\mathbb{Q}_{n,k(n)} - \mathbb{Q}_n\|_U}.$$

It follows from Part 3 of Proposition 49 and the choice of $k(n)$ that

$$\sqrt{n}(\hat{\theta}_{n,k(n)} - \hat{\theta}_n) = o^*_{\mathbb{P}(p_{0,n})^n \otimes \mu^{k(n)}}(1) \quad \text{as } n \to \infty$$

holds under (i) as well as under (ii). The proof now follows from Theorem 31, upon noting that Assumption R.2 implies Assumption P.4 (cf. Part 1 of Proposition 5). □

We are now ready to state the following asymptotic efficiency result for simulation-based II-estimators.

Theorem 39. *Let Assumptions P.3, P.5, R.2, and Z be satisfied. Let $\theta_{0,n} \in \Theta$ and $\theta_{0,n}$ converge to $\theta_{0,\infty}$, where $\theta_{0,\infty}$ is assumed to lie in the interior of Θ. Further, let the following conditions be satisfied:*

1. *The $m \times m$ matrix with (i,j)-th entry*

$$\frac{\partial^2 Q_{p_{\theta_{0,\infty}}}}{\partial\theta_i\partial\theta_j}(\theta_{0,\infty})$$

is positive definite.

2. The set

$$\left\{\frac{\partial p}{\partial \theta_i}(\,\cdot\,,\theta_{0,\infty}) : i=1,\dots,m\right\} \cup \left\{\frac{\partial p}{\partial \theta_i}(\,\cdot\,,\theta_{0,n}) : \theta_{0,n} \in \Theta^\circ,\ i=1,\dots,m\right\}$$

is bounded in some Sobolev space $\mathsf{W}_2^s(\Omega)$ *of order* $s > 1/2$.

If $k(n)$ *is of asymptotic order* $n^{2+\alpha}$ *for some (arbitrarily small)* $\alpha > 0$, *then*

$$\sqrt{n}(\hat\theta_{n,k(n)} - \theta_{0,n}) \rightsquigarrow N(0, I^{-1}(\theta_{0,\infty})) \quad \text{as } n \to \infty,$$

where $I^{-1}(\theta_{0,\infty})$ *denotes the inverse of the Fisher information matrix and weak convergence refers to the underlying probability spaces* $(\Omega^n \times V^{k(n)}, \mathcal{B}(\Omega)^n \otimes \mathcal{V}^{k(n)}, \mathbb{P}(p_{\theta_{0,n}})^n \otimes \mu^{k(n)})$.

Proof. Note that $\theta_{0,n}$ and $\theta_{0,\infty}$, respectively, are the unique minimizers of the functions $Q_{p_{\theta_{0,n}}}$ and $Q_{p_{\theta_{0,\infty}}}$. Since $p_{\theta_{0,n}}$ converges pointwise to $p_{\theta_{0,\infty}}$ in view of Part 1 of Proposition 5, we can apply Theorem 38 with $p_{0,n} = p_{\theta_{0,n}}$ and $p_{0,\infty} = p_{\theta_{0,\infty}}$ to complete the proof. $\square$

8 On the computation of auxiliary maximum likelihood estimators

This section provides a numerically feasible computation of $\hat{p}_n$, at least on events whose probability tends to 1. Recall from Section 5.1 that $\hat{p}_n$ is the unique solution of maximizing

$$L_n(p) = \frac{1}{n}\sum_{i=1}^{n} \log p(X_i)$$

over $\mathsf{W}_2^t(\Omega)$ under the constraints

$$\begin{aligned}\int_\Omega p\,d\lambda &= 1,\\ \|p\|_{t,2} &\leq D,\\ \inf_{x\in\Omega} p(x) &\geq \zeta,\end{aligned}$$

where $0 \leq \zeta \leq \lambda(\Omega)^{-1} \leq D^2$. We know from Remark 7(iii) that $\hat{p}_n(X_i) > 0$ for all $i = 1, \ldots, n$, so that dividing by these terms in the subsequent theorem does not make any difficulties on *any* event. As is discussed below, the following theorem provides the key for computing $\hat{p}_n$.

Theorem 40. *Let Assumptions D, D.1, and the strict inequalities $\zeta < \lambda(\Omega)^{-1} < D^2$ and $\inf_{x\in\Omega} p_0(x) > \zeta$ be satisfied. Then there are events that have probability tending to 1 such that $\hat{p}_n$ satisfies*

$$\frac{1}{n}\sum_{i=1}^{n} \frac{X_i^*}{\hat{p}_n(X_i)} - \frac{\lambda(\Omega) - \frac{1}{n}\sum_{i=1}^{n}\frac{1}{\hat{p}_n(X_i)}}{\lambda(\Omega)D^2 - 1}\left(\hat{p}_n(\,\cdot\,) - D^2\right) - 1 = 0, \tag{39}$$

where X_i^* denotes the Riesz representer of the point evaluation $\delta_{X_i} : \mathsf{W}_2^t(\Omega) \to \mathbb{R}$, given by $\delta_{X_i}(f) = f(X_i)$.

Proof. Throughout the proof derivatives are always Fréchet derivatives taken with respect to the Sobolev norm.

Step 1: By Remark 12(ii), the inequality $\inf_{x\in\Omega} \hat{p}_n(x) > \zeta$ holds on events that have probability tending to 1. Therefore, $\hat{p}_n$ necessarily maximizes L_n over

$$\left\{p \in \mathsf{W}_2^t(\Omega) : \int_\Omega p\,d\lambda = 1,\ \|p\|_{t,2} = D,\ \inf_{x\in\Omega} p(x) > \zeta\right\},$$

upon noting that the constraint $\|p\|_{t,2} \leq D$ is active by Remark 7(i). Put differently, $\hat{p}_n$ maximizes $L_n : \mathcal{U} \to \mathbb{R}$ under the side conditions

$$\begin{aligned}\int_\Omega p\,d\lambda - 1 &= 0,\\ \|p\|_{t,2}^2 - D^2 &= 0,\end{aligned}$$

where $\mathcal{U} := \{p \in \mathsf{W}_2^t(\Omega) : \inf_{x\in\Omega} p(x) > \zeta\}$ is easily seen to be open with respect to the sup-norm topology on $\mathsf{W}_2^t(\Omega)$. It follows from Part 2 of Proposition 1 that $\mathcal{U}$ is open with respect to the topology that is induced by the Sobolev norm.

Step 2: To solve the optimization problem in Step 1 with help of Lagrange multipliers, we first define $G_1(p) = \int_\Omega p\, d\lambda - 1$, $G_2(p) = \|p\|_{t,2}^2 - D^2$, and $G(p) = (G_1(p), G_2(p))'$. We then have that

$$DG_1(p)(f) = \int_\Omega f d\lambda \tag{40}$$

since $\int_\Omega(\,\cdot\,)d\lambda$ is a continuous linear functional on $\mathsf{W}_2^t(\Omega)$ in view of

$$\int_\Omega |f| d\lambda \leq \lambda(\Omega)^{1/2}\|f\|_2 \leq \lambda(\Omega)^{1/2}\|f\|_{t,2}.$$

Since $\langle\,\cdot\,|\,\cdot\,\rangle_{t,2}$ is bilinear and continuous with respect to the Sobolev norm, G_2 is differentiable with respect to the Sobolev norm with derivative given by

$$DG_2(p)(f) = 2\langle p|f\rangle_{t,2}. \tag{41}$$

Step 3: We claim that there is some $h \in \mathsf{W}_2^t(\Omega)$ such that $DG_2(\hat{p}_n)(h) = 0$ and $DG_1(\hat{p}_n)(h) \neq 0$. Suppose, on the contrary, that $DG_2(\hat{p}_n)(h) = 0$ implies $DG_1(\hat{p}_n)(h) = 0$ for all $h \in \mathsf{W}_2^t(\Omega)$, meaning that $\ker DG_2(\hat{p}_n) \subseteq \ker DG_1(\hat{p}_n)$. Then $DG_1(\hat{p}_n)$ factors over $DG_2(\hat{p}_n)$. That is,

$$DG_1(\hat{p}_n) = \gamma DG_2(\hat{p}_n) \tag{42}$$

for some $\gamma \in \mathbb{R}$, and $\gamma \neq 0$ since $DG_1(\hat{p}_n)$ is clearly surjective. In view of (40) and (41), (42) can be rewritten as

$$\langle 1|f\rangle_{t,2} = 2\gamma\langle \hat{p}_n|f\rangle_{t,2} \quad \text{for all } f \in \mathsf{W}_2^t(\Omega).$$

Consequently, $2\gamma\hat{p}_n = 1$, so that $\hat{p}_n$ is constant. But we know from Remark 7(i) that this is impossible.

Step 4: Next, we show that $DG(\hat{p}_n) : \mathsf{W}_2^t(\Omega) \to \mathbb{R}^2$ is surjective. To this end, let $(v, w)' \in \mathbb{R}^2$. Since $\hat{p}_n$ is not constantly equal to 0, $DG_2(\hat{p}_n)$ is surjective, and hence there is some $g \in \mathsf{W}_2^t(\Omega)$ such that $DG_2(\hat{p}_n)(g) = w$. Let $h \in \mathsf{W}_2^t(\Omega)$ such that $DG_2(\hat{p}_n)(h) = 0$ and $DG_1(\hat{p}_n)(h) \neq 0$. We know from Step 2 that such a function exists. Clearly, there is some $c \in \mathbb{R}$ with $DG_1(\hat{p}_n)(ch) = v$. But then $f = g + ch$ satisfies $DG(\hat{p}_n)(f) = (v, w)'$.

Step 5: It follows from Theorem 43.D in Zeidler (1985) that there are Lagrange multipliers $\alpha, \beta \in \mathbb{R}$ such that

$$DL_n(\hat{p}_n)(f) - \alpha DG_1(\hat{p}_n)(f) - \beta DG_2(\hat{p}_n)(f) = 0 \tag{43}$$

holds true for all $f \in \mathsf{W}_2^t(\Omega)$. Computing the derivatives in (43) leads to

$$\frac{1}{n}\sum_{i=1}^n \frac{f(X_i)}{\hat{p}_n(X_i)} - \alpha\int_\Omega f d\lambda - 2\beta\langle \hat{p}_n|f\rangle_{t,2} = 0 \tag{44}$$

for all $f \in \mathsf{W}_2^t(\Omega)$. Since $\mathsf{W}_2^t(\Omega)$ is a reproducing kernel Hilbert space, we may reformulate (44) in terms of Riesz representers and get

$$\frac{1}{n}\sum_{i=1}^n \frac{X_i^*}{\hat{p}_n(X_i)} - \alpha - 2\beta\hat{p}_n = 0, \tag{45}$$

upon noting that the Riesz representer of $DG_1(\hat{p}_n)$ is the constant function 1, whereas that of $DG_2(\hat{p}_n)$ is $2\hat{p}_n$. By taking the Sobolev inner product both with 1 and $\hat{p}_n$, we can solve for α and β:

$$\begin{aligned} 2\beta &= \frac{\lambda(\Omega) - \frac{1}{n}\sum_{i=1}^n \frac{1}{\hat{p}_n(X_i)}}{\lambda(\Omega)D^2 - 1}; \\ \alpha &= 1 - 2\beta D^2. \end{aligned}$$

Plugging this into (45) proves the theorem. □

Remark 41. **(i)** Evaluation of (39) at $X_1, \ldots, X_n$ leads to the following finite-dimensional system of equations in the variables $Z_1 = p(X_1), \ldots, Z_n = p(X_n)$:

$$\frac{1}{n}\sum_{i=1}^n \frac{X_i^*(X_j)}{Z_i} - \frac{\lambda(\Omega) - \frac{1}{n}\sum_{i=1}^n \frac{1}{Z_i}}{\lambda(\Omega)D^2 - 1}(Z_j - D^2) - 1 = 0, \quad j = 1, \ldots, n,$$

where the Z_i are positive. Once $Z_1, \ldots, Z_n$ are known, we may compute $\hat{p}_n$ directly from (39), which leads to the formula

$$\hat{p}_n(x) = \left(\frac{1}{n}\sum_{i=1}^n \frac{X_i^*(x)}{Z_i} - 1\right) \frac{\lambda(\Omega)D^2 - 1}{\lambda(\Omega) - \frac{1}{n}\sum_{i=1}^n \frac{1}{Z_i}} + D^2$$

for all $x \in \Omega$.

(ii) In case $t = 1$ and $t = 2$, explicit formulas for the Riesz representers X_i^* are provided in Thomas-Agnan (1996); see Proposition 2 in combination with Corollary 2 and 3 there.

9 Appendix

Appendix A: Properties of AML-objective functions

Lemma 42.

1. (a) *$x \mapsto \log f(x)$ is $\mathcal{B}(\Omega)$-$\mathcal{B}([-\infty,\infty))$-measurable for any non-negative, $\mathcal{B}(\Omega)$-measurable real-valued function f.*

   (b) *$v \mapsto \log f(\rho(v,\theta))$ is $\mathcal{V}$-$\mathcal{B}([-\infty,\infty))$-measurable for every $\theta \in \Theta$ and every non-negative, $\mathcal{B}(\Omega)$-measurable real-valued function f.*

2. *Let $\mathcal{F}$ be a set of non-negative, bounded real-valued functions on Ω.*

   (a) *Then for every $x \in \Omega$, $f \mapsto \log f(x)$ is a continuous mapping from $(\mathcal{F}, \|\cdot\|_\Omega)$ to $[-\infty,\infty)$.*

   (b) *If the elements $f \in \mathcal{F}$ are also continuous and Assumption R.1 is satisfied, then for every $v \in V$, $(\theta, f) \mapsto \log f(\rho(v,\theta))$ is a continuous mapping from $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$ to $[-\infty,\infty)$.*

Proof. **Part 1:** Part 1(a) is clear as f is $\mathcal{B}(\Omega)$-$\mathcal{B}([0,\infty))$-measurable by hypothesis and the extended logarithm is $\mathcal{B}([0,\infty))$-$\mathcal{B}([-\infty,\infty))$-measurable.

For Part 1(b), additionally use that $\rho : V \times \Theta \to \Omega$ is $\mathcal{V}$-$\mathcal{B}(\Omega)$-measurable in the first argument for every $\theta \in \Theta$.

Part 2: Fix $x \in \Omega$. Let $f_l, f \in \mathcal{F}$ be such that $\|f_l - f\|_\Omega$ converges to 0. Then $f_l(x)$ converges to $f(x)$. Since setting $\log 0 = -\infty$ continuously extends the logarithm to the interval $[0,\infty)$, $\log f_l(x)$ converges to $\log f(x)$. This shows Part 2(a).

To prove Part 2(b), let $\theta_l, \theta \in \Theta$ and $f_l, f \in \mathcal{F}$ be such that $|\theta_l - \theta|$ and $\|f_l - f\|_\Omega$ converge to 0. Use the triangle inequality to obtain

$$\begin{aligned}
&|f_l(\rho(v,\theta_l)) - f(\rho(v,\theta))| \\
&\qquad \le |f_l(\rho(v,\theta_l)) - f(\rho(v,\theta_l))| + |f(\rho(v,\theta_l)) - f(\rho(v,\theta))| \\
&\qquad \le \|f_l - f\|_\Omega + |f(\rho(v,\theta_l)) - f(\rho(v,\theta))|. \qquad (46)
\end{aligned}$$

The first expression on the r.h.s. of (46) converges to 0 by hypothesis. Making use of Assumption R.1 and the continuity of f, the second one converges to 0 as well. Now use the continuity of the extended logarithm on $[0,\infty)$. □

Proposition 43.

1. (a) *$L_n(f)$ is $\mathcal{B}(\Omega)^n$-$\mathcal{B}([-\infty,\infty))$-measurable for any non-negative, $\mathcal{B}(\Omega)$-measurable real-valued function f.*

   (b) *$L_k(\theta, f)$ is $\mathcal{V}^k$-$\mathcal{B}([-\infty,\infty))$-measurable for every $\theta \in \Theta$ and every non-negative, $\mathcal{B}(\Omega)$-measurable real-valued function f.*

2. *Let $\mathcal{F}$ be a set of non-negative, bounded real-valued functions on Ω.*

   (a) *Then, $L_n(f)$ is a continuous mapping from $(\mathcal{F}, \|\cdot\|_\Omega)$ into $[-\infty,\infty)$.*

   (b) *If the elements $f \in \mathcal{F}$ are also continuous and Assumption R.1 is satisfied, then $L_k(\theta, f)$ is a continuous mapping from $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$ into $[-\infty, \infty)$.*

3. *Let $\mathcal{F}$ be a set of non-negative, bounded, $\mathcal{B}(\Omega)$-measurable real-valued functions on Ω.*

   (a) *Then, $L(f)$ is an upper semi-continuous mapping from $(\mathcal{F}, \|\cdot\|_\Omega)$ into $[-\infty, \infty)$.*

   (b) *If the elements $f \in \mathcal{F}$ are also continuous and Assumption R.1 is satisfied, then $L(\theta, f)$ is an upper semi-continuous mapping from $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$ into $[-\infty, \infty)$.*

4. *Let $\mathcal{F}$ be a set of non-negative, bounded, $\mathcal{B}(\Omega)$-measurable real-valued functions on Ω that are uniformly bounded away from 0.*

   (a) *Then, $L(f)$ is a continuous real-valued function on $(\mathcal{F}, \|\cdot\|_\Omega)$.*

   (b) *If the elements $f \in \mathcal{F}$ are also continuous and Assumption R.1 is satisfied, then $L(\theta, f)$ is a continuous real-valued function on $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$.*

5. *Let $\mathcal{F}$ be a sup-norm compact set of non-negative, $\mathcal{B}(\Omega)$-measurable real-valued functions on Ω that are uniformly bounded away from 0. Then we have the following uniform laws of large numbers:*

   (a) $\lim_{n\to\infty} \sup_{f\in\mathcal{F}} |L_n(f) - L(f)| = 0 \quad$ *a.s.*

   (b) *If the elements $f \in \mathcal{F}$ are also continuous and Assumption R.1 is satisfied, then*

$$\lim_{k\to\infty} \sup_{\Theta\times\mathcal{F}} |L_k(\theta, f) - L(\theta, f)| = 0 \quad a.s.$$

Proof. Part 1: Parts 1(a) and 1(b) follow from Parts 1(a) and 1(b) of Lemma 42, respectively.

Part 2: Parts 2(a) and 2(b) follow directly from Parts 2(a) and 2(b) of Lemma 42, respectively.

Part 3: Since the elements $f \in \mathcal{F}$ are bounded,

$$\int_\Omega \max(\log f, 0) d\mathbb{P} < \infty$$

and

$$\int_V \max(\log f(\rho(\,\cdot\,, \theta)), 0) d\mu < \infty,$$

and therefore both $L(f)$ and $L(\theta, f)$ are well-defined mappings with values in $[-\infty, \infty)$. To see that $L(f)$ is upper semi-continuous, let $f_l, f \in \mathcal{F}$ be such that $\|f_l - f\|_\Omega$ converges to 0. By Part 2(a) of Lemma 42 and continuity of taking the maximum,

$$|\max(\log f_l(x), 0) - \max(\log f(x), 0)|$$

and

$$|\max(-\log f_l(x), 0) - \max(-\log f(x), 0)|$$

converge to 0 for every $x \in \Omega$. This has two consequences: First,

$$\int_\Omega \liminf_{l\to\infty} \max(-\log f_l, 0)d\mathbb{P} \le \liminf_{l\to\infty} \int_\Omega \max(-\log f_l, 0)d\mathbb{P}$$

by Fatou's lemma, and thus

$$\int_\Omega \max(-\log f, 0)d\mathbb{P} \le \liminf_{l\to\infty} \int_\Omega \max(-\log f_l, 0)d\mathbb{P}. \tag{47}$$

Second,

$$\int_\Omega \max(\log f, 0)d\mathbb{P} = \lim_{l\to\infty} \int_\Omega \max(\log f_l, 0)d\mathbb{P} \tag{48}$$

by the theorem of dominated convergence. Recall that $\int_\Omega \max(\log f, 0)d\mathbb{P} < \infty$ and that $\{\max(\log f_l, 0) : l \in \mathbb{N}\}$ is bounded in $\mathsf{L}^\infty(\Omega)$ since f is bounded, $\|f_l - f\|_\Omega$ converges to 0, and thus $\{f_l : l \in \mathbb{N}\}$ is bounded in $\mathsf{L}^\infty(\Omega)$. From (47) and (48) we conclude that

$$\begin{aligned} L(f) &= \int_\Omega \max(\log f, 0)d\mathbb{P} - \int_\Omega \max(-\log f, 0)d\mathbb{P} \\ &\ge \lim_{l\to\infty} \int_\Omega \max(\log f_l, 0)d\mathbb{P} - \liminf_{l\to\infty} \int_\Omega \max(-\log f_l, 0)d\mathbb{P} \\ &= \limsup_{l\to\infty} L(f_l). \end{aligned}$$

To see that $L(\theta, f)$ is upper semi-continuous, let $\theta_l, \theta \in \Theta$ and $f_l, f \in \mathcal{F}$ be such that $\|\theta_l - \theta\|$ and $\|f_l - f\|_\Omega$ converge to 0. By Part 2(b) of Lemma 42 and continuity of taking the maximum,

$$|\max(\log f_l(\rho(v, \theta_l)), 0) - \max(\log f(\rho(v, \theta)), 0)|$$

and

$$|\max(-\log f_l(\rho(v, \theta_l)), 0) - \max(-\log f(\rho(v, \theta)), 0)|$$

converge to 0 for every $v \in V$. This has two consequences: First,

$$\int_V \liminf_{l\to\infty} \max(-\log f_l(\rho(\,\cdot\,, \theta_l)), 0)d\mu \le \liminf_{l\to\infty} \int_V \max(-\log f_l(\rho(\,\cdot\,, \theta_l)), 0)d\mu$$

by Fatou's lemma, and thus

$$\int_V \max(-\log f(\rho(\,\cdot\,, \theta)), 0)d\mu \le \liminf_{l\to\infty} \int_V \max(-\log f_l(\rho(\,\cdot\,, \theta_l)), 0)d\mu. \tag{49}$$

Second,

$$\int_V \max(\log f(\rho(\,\cdot\,, \theta)), 0)d\mu = \lim_{l\to\infty} \int_V \max(\log f_l(\rho(\,\cdot\,, \theta_l)), 0)d\mu \tag{50}$$

by the theorem of dominated convergence. Recall that $\int_V \max(\log f(\rho(\,\cdot\,, \theta)), 0)d\mu < \infty$ and that $\{\max(\log f_l(\rho(\,\cdot\,, \theta_l)), 0) : l \in \mathbb{N}\}$ is a bounded subset of $\mathsf{L}^\infty(V, \mathcal{V})$ since f is bounded, $\|f_l - f\|_\Omega$ converges to 0, and thus $\{f_l : l \in \mathbb{N}\}$ is bounded in $\mathsf{L}^\infty(\Omega)$. From (49) and (50) we conclude that

$$\begin{aligned} L(\theta, f) &= \int_V \max(\log f(\rho(\,\cdot\,, \theta)), 0)d\mu - \int_V \max(-\log f(\rho(\,\cdot\,, \theta)), 0)d\mu \\ &\ge \lim_{l\to\infty} \int_V \max(\log f_l(\rho(\,\cdot\,, \theta_l)), 0)d\mu - \liminf_{l\to\infty} \int_V \max(-\log f_l(\rho(\,\cdot\,, \theta_l)), 0)d\mu \\ &= \limsup_{l\to\infty} L(\theta_l, f_l). \end{aligned}$$

Part 4: Denote by ξ the lower uniform bound of all elements in $\mathcal{F}$. By hypothesis, $\xi > 0$. To prove Part 4(a), let $f_l, f \in \mathcal{F}$ be such that $\|f_l - f\|_\Omega$ converges to 0. Then $\{f_l : l \in \mathbb{N}\}$ is bounded by some B, $0 < B < \infty$. Since the logarithm is bounded on $[\xi, B]$, the domination condition

$$\int_\Omega \sup_{l \in \mathbb{N}} |\log f_l(x)| \, d\mathbb{P}(x) \quad < \quad \infty$$

is satisfied. By Part 2(a) of Lemma 42, $\log f_l(x)$ converges to $\log f(x)$ for every $x \in \Omega$. Part 4(a) then follows from the theorem of dominated convergence.

To prove Part 4(b), let $\theta_l, \theta \in \Theta$ and $f_l, f \in \mathcal{F}$ be such that $\|\theta_l - \theta\|$ and $\|f_l - f\|_\Omega$ converge to 0. Then $\{f_l : l \in \mathbb{N}\}$ is bounded by some B, $0 < B < \infty$. Since the logarithm is bounded on $[\xi, B]$, the domination condition

$$\int_V \sup_{\theta \in \Theta} \sup_{l \in \mathbb{N}} |\log f_l(\rho(v, \theta))| \, d\mu(v) < \infty$$

is satisfied. By Part 2(b) of Lemma 42, $\log f_l(\rho(v, \theta_l))$ converges to $\log f(\rho(v, \theta))$ for every $v \in V$. Part 4(b) now follows from the theorem of dominated convergence.

Part 5: The uniform law of large numbers stated in Part 5(a) follows from Mourier's strong law of large numbers: Apply Corollary 7.10 in Ledoux & Talagrand (1991) with $B = \mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$, $X(f) = \log f(X_1) - \int_\Omega f d\mathbb{P}$, and $\|\cdot\| = \|\cdot\|_\mathcal{F}$. Note that X has values in $\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$ as can be seen by using Part 2(a) of Lemma 42, Part 4(a), and the sup-norm compactness of $\mathcal{F}$. Since X_1 is a random variable with values in $(\Omega, \mathcal{B}(\Omega))$ and $x \mapsto \log f(x)$ is $\mathcal{B}(\Omega)$-$\mathcal{B}(\mathbb{R})$-measurable by Part 1(a) of Lemma 42, we see that X is measurable with respect to the cylindrical σ-algebra on $\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$. It follows from the equivalence of the cylindrical with the Borel σ-algebra on $\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$ (see, e.g., Section 2.1 in Ledoux & Talagrand (1991), and observe that $\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$ is a separable Banach space) that X is a random variable with respect to the latter one. The integrability condition $\mathrm{E}\,\|X\| < \infty$ translates into the domination condition

$$\int_\Omega \sup_{f \in \mathcal{F}} |\log f(x)| d\mathbb{P}(x) < \infty$$

and holds true since the elements of $\mathcal{F}$ are uniformly bounded and uniformly bounded away from 0 by hypothesis.

The uniform law of large numbers stated in Part 5(b) also follows from Mourier's strong law of large numbers: Apply Corollary 7.10 in Ledoux & Talagrand (1991) with B the separable Banach space of all bounded, continuous functions on $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$, $X(\theta, f) = \log f(\rho(V_1, \theta)) - \int_V \log f(\rho(\,\cdot\,, \theta)) d\mu$, and $\|\cdot\| = \|\cdot\|_{\Theta \times \mathcal{F}}$. Note that by Part 2(b), Part 4(b), and compactness of $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$, X takes its values in the space of (bounded) continuous functions on $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$. Since V_1 is a random variable with values in $(V, \mathcal{V})$ and $\log f(\rho(\,\cdot\,, \theta))$ is $\mathcal{V}$-$\mathcal{B}(\mathbb{R})$-measurable by Part 1(b) of Lemma 42, X is Borel measurable by the equivalence of the the cylindrical with the Borel σ-algebra on the space of bounded, continuous functions on $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$ (see, e.g., Section 2.1 in Ledoux & Talagrand, 1991). The integrability condition $\mathrm{E}\,\|X\| < \infty$ reads as

$$\int_V \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} |\log f(\rho(v, \theta))| d\mu(v) < \infty$$

and holds true since the elements of $\mathcal{F}$ are uniformly bounded and uniformly bounded away from 0 by hypothesis. $\square$

The following proposition provides explicit formulas for the derivatives of $L_n(f)$ and $L(f)$ with respect to f. It can be found as Proposition 3 in Nickl (2007) and is stated here (without proof) for easy reference.

Proposition 44. *Let $\xi > 0$ and define $\mathcal{U} = \{f \in \mathsf{L}^\infty(\Omega) : \inf_{x\in\Omega} f(x) > \xi\}$. Let α be a non-negative integer, $f \in \mathcal{U}$, and $f_1, \ldots, f_\alpha \in \mathsf{L}^\infty(\Omega)$. The α-linear functionals representing the α-th Fréchet derivative of $L_n : \mathcal{U} \to \mathbb{R}$ and $L : \mathcal{U} \to \mathbb{R}$ are given by*

$$\begin{aligned} D^\alpha L_n(f)(f_1, \ldots, f_\alpha) &= (-1)^{\alpha-1}(\alpha-1)!\, \mathbb{P}_n(f^{-\alpha} f_1 \cdots f_\alpha) \\ &= (-1)^{\alpha-1}(\alpha-1)!\, \frac{1}{n}\sum_{i=1}^{n} f^{-\alpha}(X_i) f_1(X_i) \cdots f_\alpha(X_i), \end{aligned}$$

$$\begin{aligned} D^\alpha L(f)(f_1, \ldots, f_\alpha) &= (-1)^{\alpha-1}(\alpha-1)!\, \mathbb{P}(f^{-\alpha} f_1 \cdots f_\alpha) \\ &= (-1)^{\alpha-1}(\alpha-1)! \int_\Omega f^{-\alpha} f_1 \cdots f_\alpha d\mathbb{P}. \end{aligned}$$

The following proposition gives explicit formulas for the derivatives of $L_k(\theta, f)$ and $L(\theta, f)$ with respect to f.

Proposition 45. *Let $\xi > 0$ and $\mathcal{U}$ as in Proposition 44. For $\theta \in \Theta$, $f \in \mathcal{U}$, α some non-negative integer, and $f_1, \ldots, f_\alpha \in \mathsf{L}^\infty(\Omega)$, the α-linear functionals representing the α-th partial Fréchet derivative of $L_k : \Theta \times \mathcal{U} \to \mathbb{R}$ and $L : \Theta \times \mathcal{U} \to \mathbb{R}$, with respect to the second variable, are given by*

$$\begin{aligned} D_2^\alpha L_k(\theta, f)(f_1, \ldots, f_\alpha) &= (-1)^{\alpha-1}(\alpha-1)! \\ &\quad \times \mu_k(f^{-\alpha}(\rho(\,\cdot\,,\theta)) f_1(\rho(\,\cdot\,,\theta)) \cdots f_\alpha(\rho(\,\cdot\,,\theta))) \\ &= (-1)^{\alpha-1}(\alpha-1)! \\ &\quad \times \frac{1}{k}\sum_{i=1}^{k} f^{-\alpha}(X_i(\theta)) f_1(X_i(\theta)) \cdots f_\alpha(X_i(\theta)), \end{aligned}$$

$$\begin{aligned} D_2^\alpha L(\theta, f)(f_1, \ldots, f_\alpha) &= (-1)^{\alpha-1}(\alpha-1)! \\ &\quad \times \mu(f^{-\alpha}(\rho(\,\cdot\,,\theta)) f_1(\rho(\,\cdot\,,\theta)) \cdots f_\alpha(\rho(\,\cdot\,,\theta))) \\ &= (-1)^{\alpha-1}(\alpha-1)! \int_\Omega f^{-\alpha} f_1 \cdots f_\alpha\, p_\theta d\lambda. \end{aligned}$$

Proof. Adapt the proof of Proposition 3 in Nickl (2007) by replacing n by k, X_i by $\rho(U_i, \theta)$, L_n by $L_k(\theta, \cdot)$, $L_{(i)}$ by $L(\theta, \cdot)$, $\mathbb{P}_k$ by μ_k, and $\mathbb{P}$ by μ. The last equality in the display above follows from the change of variable theorem. $\square$

The next result is a uniform version of Lemma 2 in Nickl (2007). It provides rates of convergence for all derivatives of the auxiliary log-likelihood function, which hold uniformly over the parameter space.

Proposition 46. *Let α be a non-negative integer, and let $\mathcal{H}_1, \ldots, \mathcal{H}_\alpha$ be bounded subsets of some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$. If Assumptions R.2 and Z are satisfied, then*

$$\sup_{\Theta\times\mathcal{P}(t,\zeta,D)} \|D_2^\alpha L_k(\theta,p) - D_2^\alpha L(\theta,p)\|_{\mathcal{H}_1\times\cdots\times\mathcal{H}_\alpha} = O_\mu^*(k^{-1/2}) \quad \text{as } k \to \infty.$$

Proof. Note that

$$\begin{aligned}
&\sup_{\Theta\times\mathcal{P}(t,\zeta,D)} \|D_2^\alpha L_k(\theta,p) - D_2^\alpha L(\theta,p)\|_{\mathcal{H}_1\times\cdots\times\mathcal{H}_\alpha} \\
&\quad = (\alpha-1)! \sup_{\theta\in\Theta} \sup_{p\in\mathcal{P}(t,\zeta,D)} \sup_{h_1\in\mathcal{H}_1} \\
&\qquad \cdots \sup_{h_\alpha\in\mathcal{H}_\alpha} (\mu_k-\mu)(p^{-\alpha}(\rho(\,\cdot\,,\theta))h_1(\rho(\,\cdot\,,\theta))\cdots h_\alpha(\rho(\,\cdot\,,\theta))) \\
&\qquad \text{(by Proposition 45)} \\
&\quad = (\alpha-1)!\,\|\mu_k-\mu\|_{\mathcal{G}},
\end{aligned}$$

where

$$\mathcal{G} = \Big\{ p^{-\alpha}(\rho(\,\cdot\,,\theta))h_1(\rho(\,\cdot\,,\theta))\cdots h_\alpha(\rho(\,\cdot\,,\theta)) : \\ \theta\in\Theta,\, p\in\mathcal{P}(t,\zeta,D),\, h_1\in\mathcal{H}_1,\ldots,h_\alpha\in\mathcal{H}_\alpha \Big\}.$$

The classes of functions $\mathcal{P}^*(t,\zeta,D) = \{p(\rho(\,\cdot\,,\theta)) : \theta\in\Theta,\, p\in\mathcal{P}(t,\zeta,D)\}$ and $\mathcal{H}_i^* = \{h_i(\rho(\,\cdot\,,\theta)) : \theta\in\Theta,\, h\in\mathcal{H}_i\}$, $i=1,\ldots,\alpha$, are μ-Donsker by Part 1 of Proposition 14. Since $\zeta > 0$ by Assumption Z, it follows from Example 2.10.9 in van der Vaart and Wellner (1996) that $\{1/p(\rho(\,\cdot\,,\theta)) : p\in\mathcal{P}(t,\zeta,D)\}$ is μ-Donsker. Now, since $\mathcal{G}$ is the product of μ-Donsker classes, it is μ-Donsker by Example 2.10.8 in van der Vaart and Wellner (1996), and hence $\|\mu_k-\mu\|_{\mathcal{G}}$ is bounded in outer probability at rate $k^{-1/2}$ by Prohorov's theorem. □

Appendix B: Properties of II-objective functions

Lemma 47. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous mapping from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let $f:\Omega\to\mathbb{R}$ be an integrable function satisfying $\inf_{x\in\Omega} f(x) > 0$. Then*

$$H(\theta) := \int_\Omega (f-p_\theta)^2 \frac{1}{f}\,d\lambda$$

is a continuous real-valued function on Θ.

Proof. Rewrite the integrand as $f - 2p_\theta + p_\theta^2/f$, and note that each term is integrable by the hypotheses. Hence, H is real-valued. For continuity, let $\theta_l, \theta \in \Theta$ be such that $\|\theta_l - \theta\|$ converges to 0. Letting $c = \inf_{x\in\Omega} f(x)$,

$$\begin{aligned}
|H(\theta_l) - H(\theta)| &= \left| \int_\Omega \frac{p_{\theta_l}^2}{f}\,d\lambda - \int_\Omega \frac{p_\theta^2}{f}\,d\lambda \right| \\
&\le c^{-1} \int_\Omega |p_{\theta_l}^2 - p_\theta^2|\,d\lambda \\
&\le c^{-1} \|p_{\theta_l} - p_\theta\|_2 (\|p_{\theta_l} - p_\theta\|_2 + 2\|p_\theta\|_2).
\end{aligned}$$

Note that the r.h.s., and therefore also the l.h.s., of the previous display goes to 0 as $l \to \infty$. □

Proposition 48.

1. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$.*

   (a) *On the event where $\inf_{x\in\Omega} \hat{p}_n(x) > 0$,*
   $$\mathbb{Q}_n(\theta) = \int_\Omega (\hat{p}_n - p_\theta)^2 \frac{1}{\hat{p}_n} d\lambda$$
   holds and $\mathbb{Q}_n$ is a continuous real-valued function on Θ.

   (b) *If Assumption Z is satisfied, then, for each $(x_1, \dots, x_n) \in \Omega^n$,*
   $$\mathbb{Q}_n(\theta) = \int_\Omega (\hat{p}_n - p_\theta)^2 \frac{1}{\hat{p}_n} d\lambda$$
   holds and $\mathbb{Q}_n$ is a continuous real-valued function on Θ.

2. *Let Assumption R.1 be satisfied.*

   (a) *On the event where $\inf_{x\in\Omega} \hat{p}_n(x) > 0$,*
   $$\mathbb{Q}_{n,k}(\theta) = \int_\Omega (\hat{p}_n - \tilde{p}_k(\theta))^2 \frac{1}{\hat{p}_n} d\lambda$$
   holds and $\mathbb{Q}_{n,k}$ is a continuous real-valued function on Θ.

   (b) *If Assumption Z is satisfied, then, for each $(x_1, \dots, x_n, v_1, \dots, v_k) \in \Omega^n \times V^k$,*
   $$\mathbb{Q}_{n,k}(\theta) = \int_\Omega (\hat{p}_n - \tilde{p}_k(\theta))^2 \frac{1}{\hat{p}_n} d\lambda$$
   holds and $\mathbb{Q}_{n,k}$ is a continuous real-valued function on Θ.

3. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. If Assumptions D and D.2 hold, then Q is a continuous real-valued function on Θ.*

Proof. Parts 1 and 3 are immediate consequences of Lemma 47.

We next prove Part 2(a). Since $\hat{p}_n$ and $\tilde{p}_k(\theta)$ belong to $\mathcal{P}(t, \zeta, D)$ by construction, these densities are sup-norm bounded by $C_t D$. Hence, $\mathbb{Q}_{n,k}$ is real-valued whenever $\inf_{x\in\Omega} \hat{p}_n(x) > 0$. Since the map $\theta \mapsto \tilde{p}_k(\theta)$ is continuous by Part 2 of Theorem 6, continuity of $\mathbb{Q}_{n,k}$ then follows from the theorem of dominated convergence.

Part 2(b) immediately follows from Part 2(a). □

Proposition 49. *We have the following uniform convergence properties of II-objective functions:*

1. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let further Assumptions D, D.1, and D.2 be satisfied. Then*
   $$\sup_{\theta\in\Theta} |\mathbb{Q}_n(\theta) - Q(\theta)| = o^*_{\mathbb{P}}(1) \quad \text{as } n \to \infty.$$

2. *Let Assumptions P.1, P.2, R.1, D, D.1, and D.2 be satisfied. Then*

$$\sup_{\theta\in\Theta} |\mathbb{Q}_{n,k}(\theta) - Q(\theta)| = o^*_{\Pr}(1) \quad \textit{as } \min(n,k) \to \infty.$$

3. *Let Assumptions P.1, R.2, and Z be satisfied. Then*

$$\sup_{n\in\mathbb{N}} \sup_{\Omega^n} \sup_{\theta\in\Theta} |\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_n(\theta)| = O^*_{\mu}(k^{-t/(2t+1)}) \quad \textit{as } k \to \infty.$$

If Assumption P.1 is strengthened to P.3, then

$$\sup_{n\in\mathbb{N}} \sup_{\Omega^n} \sup_{\theta\in\Theta} |\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_n(\theta)| = O^*_{\Pr}(k^{-1/2}) \quad \textit{as } k \to \infty.$$

Proof. Part 1: Set $\chi = 2^{-1}\inf_{x\in\Omega} p_0(x)$ and observe that $\chi > 0$ by Assumption D.2. In view of Remark 12(ii) there is a sequence of events A_n that have probability converging to 1 as $n \to \infty$ such that $\inf_{x\in\Omega} \hat{p}_n(x) > \chi$. On these events we then have

$$\begin{aligned}
\sup_{\theta\in\Theta} |\mathbb{Q}_n(\theta) - Q(\theta)| &= \sup_{\theta\in\Theta} \left| \int_\Omega \frac{p_\theta^2}{\hat{p}_n} d\lambda - \int_\Omega \frac{p_\theta^2}{p_0} d\lambda \right| \\
&\leq \chi^{-2} \sup_{\theta\in\Theta} \|p_\theta\|_2^2 \, \|\hat{p}_n - p_0\|_\Omega.
\end{aligned}$$

Since Θ is compact, the assumptions on $\mathcal{P}_\Theta$ imply that $\sup_{\theta\in\Theta} \|p_\theta\|_2 < \infty$. Part 1 of Theorem 11 now completes the proof.

Part 2: Let $\chi = 2^{-1}\inf_{x\in\Omega} p_0(x)$, and observe that $\chi > 0$ by Assumption D.2. By Remark 12(ii) there is a sequence of events A_n that have probability tending to 1 as $n \to \infty$ on which $\inf_{x\in\Omega} \hat{p}_n(x) > \chi$. On A_n we have, substracting and adding $\int_\Omega \frac{p_\theta^2}{\hat{p}_n} d\lambda$,

$$\begin{aligned}
\sup_{\theta\in\Theta} |\mathbb{Q}_{n,k}(\theta) - Q(\theta)| &\leq \sup_{\theta\in\Theta} \left| \int_\Omega \frac{\tilde{p}_k(\theta)^2}{\hat{p}_n} d\lambda - \int_\Omega \frac{p_\theta^2}{p_0} d\lambda \right| \\
&\leq \sup_{\theta\in\Theta} \left| \int_\Omega (\tilde{p}_k(\theta) - p_\theta) \frac{\tilde{p}_k(\theta) + p_\theta}{\hat{p}_n} d\lambda + \int_\Omega p_\theta^2 \left(\frac{1}{\hat{p}_n} - \frac{1}{p_0} \right) d\lambda \right| \\
&\leq 2\chi^{-1} C_t D \lambda(\Omega) \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega \\
&\qquad + \chi^{-2} D^2 \sup_{\theta\in\Theta} \|p_\theta\|_2^2 \, \|\hat{p}_n - p_0\|_\Omega.
\end{aligned}$$

The result then follows from Parts 1 and 3 of Theorem 11.

Part 3: Note that $\tilde{p}_k(\theta) \in \mathcal{P}(t, \zeta, D)$ by construction and $p_\theta \in \mathcal{P}(t, \zeta, D)$ by Assumption P.1. Hence, these densities are sup-norm bounded uniformly in θ (and $v_1, \ldots, v_k \in V$ in case of $\tilde{p}_k(\theta)$). Observe now that

$$\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_n(\theta) = \int_\Omega (\tilde{p}_k(\theta) - p_\theta) \frac{\tilde{p}_k(\theta) + p_\theta}{\hat{p}_n} d\lambda.$$

Using Assumption Z, Part 4 of Proposition 1 applied to $\{\hat{p}_n : x_1, \ldots, x_n \in \Omega, \, n \in \mathbb{N}\}$ shows that $\{1/\hat{p}_n : x_1, \ldots, x_n \in \Omega, \, n \in \mathbb{N}\}$ is bounded in $\mathsf{W}_2^t(\Omega)$. By Assumption P.1 and the construction of $\tilde{p}_k(\theta)$, it follows from Part 1 of Proposition 1 that

$$\left\{ \frac{\tilde{p}_k(\theta) + p_\theta}{\hat{p}_n} : \theta \in \Theta, \, x_1, \ldots, x_n \in \Omega, \, v_1, \ldots, v_k \in V, \, n, k \in \mathbb{N} \right\} \tag{51}$$

is contained in a Sobolev ball $\mathcal{U}_{t,B}$ for some B satisfying $0 < B < \infty$. The first claim then follows from Part 1 of Theorem 16 with $s = 0$, where we have made use of the inequality $\int_\Omega |f| d\lambda \leq \lambda(\Omega)^{1/2} \|f\|_2$ and the fact that the set in (51) is bounded in the sup-norm. If Assumption P.1 is strengthened to P.3, we may apply Part 3 of Theorem 17 with $\mathcal{F}$ equal to the set given in (51) to obtain the second claim. □

Remark 50. If Assumption Z holds, then the events A_n in Parts 1 and 2 of the above proof are the entire sample space and $\mathbb{Q}_n - Q$, respectively $\mathbb{Q}_{n,k} - Q$, is continuous on Θ. By separability of Θ, the measurability of the respective suprema then follows.

Lemma 51.

1. *Let Assumptions P.1 and P.5 be satisfied. Then, on the event* $\inf_{x\in\Omega} \hat{p}_n(x) > 0$, *the following is true: For every* θ *in the interior* Θ° *of* Θ, *the objective function* $\mathbb{Q}_n$ *is twice continuously partially differentiable on* Θ° *and*

$$\begin{aligned}\frac{\partial \mathbb{Q}_n}{\partial\theta_i}(\theta) &= -2\int_\Omega (\hat{p}_n - p_\theta)\frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta)\frac{1}{\hat{p}_n}d\lambda\\ \frac{\partial^2 \mathbb{Q}_n}{\partial\theta_i\partial\theta_j}(\theta) &= 2\int_\Omega \frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta)\frac{\partial p}{\partial\theta_j}(\,\cdot\,,\theta)\frac{1}{\hat{p}_n}d\lambda\\ &\qquad -2\int_\Omega (\hat{p}_n - p_\theta)\frac{\partial^2 p}{\partial\theta_i\partial\theta_j}(\,\cdot\,,\theta)\frac{1}{\hat{p}_n}d\lambda,\end{aligned}$$

where $i = 1,\ldots,m$ *and* $j = 1,\ldots,m$.

2. *Let Assumptions P.1, P.5, D, and D.2 be satisfied. Then, for every* θ *in the interior* Θ° *of* Θ, Q *is twice continuously partially differentiable on* Θ° *and*

$$\begin{aligned}\frac{\partial Q}{\partial\theta_i}(\theta) &= -2\int_\Omega (p_0 - p_\theta)\frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta)\frac{1}{p_0}d\lambda\\ \frac{\partial^2 Q}{\partial\theta_i\partial\theta_j}(\theta) &= 2\int_\Omega \frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta)\frac{\partial p}{\partial\theta_j}(\,\cdot\,,\theta)\frac{1}{p_0}d\lambda\\ &\qquad -2\int_\Omega (p_0 - p_\theta)\frac{\partial^2 p}{\partial\theta_i\partial\theta_j}(\,\cdot\,,\theta)\frac{1}{p_0}d\lambda,\end{aligned}$$

where $i = 1,\ldots,m$ *and* $j = 1,\ldots,m$.

Proof. Part 1: Use the domination conditions listed in Assumption P.5 in combination with Assumption P.1 and the hypothesis that $\inf_{x\in\Omega} \hat{p}_n(x) > 0$ to interchange differentiation and integration. For continuity of the partial derivatives, use the theorem of dominated convergence.

Part 2: Use the domination conditions listed in Assumption P.5 in combination with Assumptions P.1 and D.2 to interchange differentiation and integration. For continuity of the partial derivatives, use the theorem of dominated convergence. □

Proposition 52. *Let Assumptions P.1, P.5, Z, D, and D.1 be satisfied. Then, for all* $i, j = 1,\ldots,m$,

$$\sup_{\theta\in\Theta^\circ}\left|\frac{\partial^2 \mathbb{Q}_n}{\partial\theta_i\partial\theta_j}(\theta) - \frac{\partial^2 Q}{\partial\theta_i\partial\theta_j}(\theta)\right| = o_{\mathbb{P}}(1) \quad \textit{as } n\to\infty,$$

where Θ° *is the interior of* Θ.

Proof. Make use of Lemma 51 to see that

$$
\begin{aligned}
&\sup_{\theta\in\Theta^\circ}\left|\frac{\partial^2\mathbb{Q}_n}{\partial\theta_i\partial\theta_j}(\theta)-\frac{\partial^2 Q}{\partial\theta_i\partial\theta_j}(\theta)\right|\\
&\qquad= 2\sup_{\theta\in\Theta^\circ}\left|\int_\Omega(\hat p_n-p_0)\frac{\partial p}{\partial\theta_i}(\,\cdot\,,\theta)\frac{\partial p}{\partial\theta_j}(\,\cdot\,,\theta)\frac{1}{\hat p_n p_0}d\lambda\right.\\
&\qquad\qquad\left.+\int_\Omega(\hat p_n-p_0)\frac{\partial^2 p}{\partial\theta_i\partial\theta_j}(\,\cdot\,,\theta)\frac{p_\theta}{\hat p_n p_0}d\lambda\right|\\
&\qquad\le 2\zeta^{-2}\left[\int_\Omega\sup_{\theta\in\Theta^\circ}\left|\frac{\partial p}{\partial\theta_i}(x,\theta)\frac{\partial p}{\partial\theta_j}(x,\theta)\right|d\lambda(x)\right.\\
&\qquad\qquad\left.+C_tD\int_\Omega\sup_{\theta\in\Theta^\circ}\left|\frac{\partial^2 p}{\partial\theta_i\partial\theta_j}(x,\theta)\right|d\lambda(x)\right]\|\hat p_n-p_0\|_\Omega.
\end{aligned}
$$

By Assumption P.5 and the fact that $\|\hat p_n-p_0\|_\Omega$ converges to 0 in probability by Part 1 of Theorem 11, the l.h.s. in the last display does so too. □

Remark 53. Consider the auxiliary model $\mathcal{P}(t,\zeta,D)$ with $\zeta=0$. The assertion of the preceding proposition still holds true in outer probability when Assumptions P.1, P.5, D, D.1, and D.2 are satisfied and $\frac{\partial^2\mathbb{Q}_n}{\partial\theta\partial\theta'}(\theta)$ is replaced by the matrix

$$
\mathbb{H}_n(\theta)\quad=\quad\begin{cases}\frac{\partial^2\mathbb{Q}_n}{\partial\theta\partial\theta'}(\theta) & \text{on the event } \inf_{x\in\Omega}\hat p_n(x)>0,\\ 0_{m\times m} & \text{otherwise.}\end{cases}
$$

Appendix C: High-level theorems on uniform rates of convergence

The next two theorems give conditions under which random maximizers are uniformly r_k-consistent for the respective limiting maximizers when k tends to infinity. In Theorem 54 the uniformity is with respect to some parameter set, whereas in Theorem 55 the uniformity is with respect to some set of probability measures. Both theorems are modifications of Theorem 3.2.5 in van der Vaart & Wellner (1996).

Theorem 54. *Let $(\Lambda,\mathcal{A},P)$ be a probability space, S a non-empty set, and let T be a non-empty set together with a non-negative function $d: T\times T\to\mathbb{R}$. We consider a sequence of real-valued stochastic processes $(H_k(s,t): s\in S, t\in T)$ defined on $(\Lambda,\mathcal{A})$ and a function $H: S\times T\to\mathbb{R}$ with the property that for every $s\in S$ there exists a $t(s)\in T$ such that, for all $t\in T$,*

$$H(s,t)-H(s,t(s))\le -Cd^\alpha(t,t(s)), \tag{52}$$

where $C,\alpha>0$ are constants neither depending on s nor t. Suppose, for all $\delta>0$,

$$\mathrm{E}_P^*\sup_{s\in S}\sup_{\substack{t\in T,\\ d(t,t(s))\le\delta}}\left|\sqrt{k}(H_k-H)(s,t)-\sqrt{k}(H_k-H)(s,t(s))\right|\le\varphi_k(\delta) \tag{53}$$

is satisfied for real-valued functions φ_k such that $\delta\mapsto\delta^{-\beta}\varphi_k(\delta)$ is non-increasing in δ for some $\beta<\alpha$. Assume further that, for every $s\in S$, $\hat t_k(s):\Lambda\to T$ satisfies

$$H_k(s,\hat t_k(s))\ge H_k(s,t)\quad \textit{for all } t\in T, \tag{54}$$

and let r_k be a sequence of positive reals such that

$$\sup_{k\in\mathbb{N}} \frac{r_k^\alpha \varphi_k(r_k^{-1})}{\sqrt{k}} < \infty. \tag{55}$$

Then, for every $s \in S$, $t(s)$ is a maximizer of $H(s,\cdot)$, and we have that

$$\sup_{s\in S} d(\hat{t}_k(s), t(s)) = O_P^*(r_k^{-1}) \quad \text{as } k \to \infty.$$

Proof. We have to prove that

$$\lim_{N\to\infty} \limsup_{k\to\infty} P^*\left(r_k \sup_{s\in S} d(\hat{t}_k(s), t(s)) > 2^N\right) = 0.$$

For $k, j \in \mathbb{N}$, set $V_{k,j} = \{(s,t) : 2^{j-1} < r_k d(t, t(s)) \leq 2^j\}$. Then

$$r_k \sup_{s\in S} d(\hat{t}_k(s), t(s)) > 2^N$$

implies that there is some $s_0 \in S$ such that $r_k d(\hat{t}_k(s_0), t(s_0)) > 2^N$, which in turn gives $(s_0, \hat{t}_k(s_0)) \in V_{k,j_0}$ for some $j_0 > N$. Combine this with (52) and (54) to get

$$\begin{aligned}
&(H_k - H)(s_0, \hat{t}_k(s_0)) - (H_k - H)(s_0, t(s_0)) \\
&\quad = H_k(s_0, \hat{t}_k(s_0)) - H_k(s_0, t(s_0)) + H(s_0, t(s_0)) - H(s_0, \hat{t}_k(s_0)) \\
&\quad \geq C d^\alpha(\hat{t}_k(s_0), t(s_0)) \\
&\quad > C r_k^{-\alpha} 2^{\alpha j_0 - \alpha}
\end{aligned}$$

This implies

$$\begin{aligned}
&P^*\left(r_k \sup_{s\in S} d(\hat{t}_k(s), t(s)) > 2^N\right) \\
&\quad \leq \sum_{j>N} P^*\left(\sup_{(s,t)\in V_{k,j}} \left|\sqrt{k}(H_k - H)(s,t) - \sqrt{k}(H_k - H)(s, t(s))\right| \geq C\sqrt{k} r_k^{-\alpha} 2^{\alpha j - \alpha}\right).
\end{aligned}$$

Via Markov's inequality (for outer probability) and (53), the r.h.s. in the previous display can be bounded by

$$\sum_{j>N} \frac{\varphi_k(2^j r_k^{-1}) r_k^\alpha}{C\sqrt{k} 2^{\alpha j - \alpha}} \leq \sum_{j>N} \frac{2^{\beta j} \varphi_k(r_k^{-1}) r_k^\alpha}{C\sqrt{k} 2^{\alpha j - \alpha}} \leq \frac{2^\alpha}{C} \sup_{k\in\mathbb{N}} \frac{r_k^\alpha \varphi_k(r_k^{-1})}{\sqrt{k}} \sum_{j>N} 2^{(\beta-\alpha)j},$$

where the last expression does not depend on k. Here, the lower inequality follows from $\varphi_k(c\delta) \leq c^\beta \varphi_k(\delta)$ for $c \geq 1$. Since $\beta < \alpha$, $\sum_{j>N} 2^{(\beta-\alpha)j}$ converges to 0 as $N \to \infty$, and, by (55), this completes the proof. $\square$

Theorem 55. *For $(\Lambda, \mathcal{A})$ some measurable space and T some non-empty set, let $(H_n(t) : t \in T)$ be a sequence of stochastic processes defined on $(\Lambda, \mathcal{A})$. Let $\mathfrak{P}$ be a set of probability measures on $(\Lambda, \mathcal{A})$ and $d : T \times T \to \mathbb{R}$ a non-negative function. Suppose that for every $P \in \mathfrak{P}$ there exists a function $H_P : T \to \mathbb{R}$ and $t_P \in T$ such that*

$$H_P(t) - H_P(t_P) \leq -C d^\alpha(t, t_P) \tag{56}$$

holds true for all $t \in T$ with constants $C, \alpha > 0$ that neither depend on t nor P. For all $\delta > 0$,

$$\sup_{P\in\mathfrak{P}} \mathrm{E}_P^* \sup_{\substack{t\in T,\\ d(t,t_P)\leq\delta}} \left|\sqrt{n}(H_n - H_P)(t) - \sqrt{n}(H_n - H_P)(t_P)\right| \leq \varphi_n(\delta) \tag{57}$$

shall be satisfied for real-valued functions φ_n such that $\delta \mapsto \delta^{-\beta}\varphi_n(\delta)$ is non-increasing in δ for some $\beta < \alpha$. Assume further that there are $\hat{t}_n : \Lambda \to T$ satisfying

$$H_n(\hat{t}_n) \geq H_n(t) \quad \textit{for all } t \in T, \tag{58}$$

and let r_n be a sequence of positive reals such that

$$\sup_{n\in\mathbb{N}} \frac{r_n^\alpha \varphi_n(r_n^{-1})}{\sqrt{n}} < \infty. \tag{59}$$

Then, for every $P \in \mathfrak{P}$, t_P is a maximizer of H_P, and we have that

$$\lim_{M\to\infty} \limsup_{n\to\infty} \sup_{P\in\mathfrak{P}} P^*(r_n d(\hat{t}_n, t_P) > M) = 0.$$

Proof. That t_P is a maximizer of H_P follows immediately from (56), so it remains to prove that

$$\lim_{N\to\infty} \limsup_{n\to\infty} \sup_{P\in\mathfrak{P}} P^*\left(r_n d(\hat{t}_n, t_P) > 2^N\right) = 0.$$

For $n, j \in \mathbb{N}$ and $P \in \mathfrak{P}$, set $V_{n,j,P} = \{t \in T : 2^{j-1} < r_n d(t, t_P) \leq 2^j\}$. Then $r_n d(\hat{t}_n, t_P) > 2^N$ implies that $\hat{t}_n$ belongs to $V_{n,j_0,P}$ for some $j_0 > N$. Hence

$$\begin{aligned}
&(H_n - H_P)(\hat{t}_n) - (H_n - H_P)(t_P)\\
&\quad = H_n(\hat{t}_n) - H_n(t_P) + H_P(t_P) - H_P(\hat{t}_n)\\
&\quad \geq C d^\alpha(\hat{t}_n, t_P)\\
&\quad > C r_n^{-\alpha} 2^{\alpha j_0 - \alpha}
\end{aligned}$$

by (56), (58), and the definition of $V_{n,j_0,P}$. This implies

$$\begin{aligned}
&P^*\left(r_n d(\hat{t}_n, t_P) > 2^N\right)\\
&\quad \leq \sum_{j>N} P^*\left(\sup_{t\in V_{n,j,P}} \left|\sqrt{n}(H_n - H_P)(t) - \sqrt{n}(H_n - H_P)(t_P)\right| \geq C\sqrt{n} r_n^{-\alpha} 2^{\alpha j - \alpha}\right).
\end{aligned}$$

By Markov's inequality (for outer probability) and (57), the r.h.s. of this inequality can be bounded by

$$\sum_{j>N} \frac{\varphi_n(2^j r_n^{-1}) r_n^\alpha}{C\sqrt{n} 2^{\alpha j - \alpha}} \leq \sum_{j>N} \frac{2^{\beta j} \varphi_n(r_n^{-1}) r_n^\alpha}{C\sqrt{n} 2^{\alpha j - \alpha}} \leq \frac{2^\alpha}{C} \sup_{n\in\mathbb{N}} \frac{r_n^\alpha \varphi_n(r_n^{-1})}{\sqrt{n}} \sum_{j>N} 2^{(\beta-\alpha)j}$$

where the last expression does neither depend on $n \in \mathbb{N}$ nor $P \in \mathfrak{P}$. Note that the lower inequality follows from $\varphi_n(c\delta) \leq c^\beta \varphi_n(\delta)$ for $c \geq 1$. Since $\beta < \alpha$, the sum $\sum_{j>N} 2^{(\beta-\alpha)j}$ converges to 0 as $N \to \infty$. Using (59) completes the proof. $\square$

Appendix D: An entropy bound for empirical processes

Let $(S, \mathcal{A}, P)$ be a probability space, $Y_1, \ldots, Y_n$ be the coordinate projections of the product space $(S^n, \mathcal{A}^n, P^n)$, and denote by P_n the empirical measure associated with $Y_1, \ldots, Y_n$. In the theory of empirical processes there exist well-known upper bounds for expressions of the form $\mathrm{E}^* \|\sqrt{n}(P_n - P)\|_{\mathcal{F}}$ over classes of functions $\mathcal{F}$ that are bounded in the sup-norm. We will now derive such bounds separately for finite and infinite $\mathcal{F}$. Furthermore, we will provide *explicit* universal constants in each case.

The following lemma originated from a convexity argument of Pisier (1983). In the form presented it is a refinement of (2.5.5) in van der Vaart & Wellner (1996).

Lemma 56. *Let $f_1, \ldots, f_N$ be bounded, $\mathcal{A}$-measurable real-valued functions on S. Then the maximal inequality*

$$\begin{aligned} &\mathrm{E} \max_{1 \le j \le N} \left|\sqrt{n}(P_n - P) f_j\right| \\ &\quad \le 8 \left[\max_{1 \le j \le N} \sqrt{P|f_j - Pf_j|^2} \sqrt{1 + \log N} + \frac{\max_{1 \le j \le N} \|f_j\|_S}{\sqrt{n}} (1 + \log N) \right] \end{aligned}$$

holds true.

Proof. Let $\eta = \max_{1 \le j \le N} \sqrt{P|f_j - Pf_j|^2}$ and $B = \max_{1 \le j \le N} \|f_j\|_S$. Since the assertion of the lemma is satisfied when $\eta = 0$, as is easily seen, assume that $\eta > 0$. We first derive a tail bound for $\sqrt{n}(P_n - P) f_j$, which will be independent of j. To this end, let $\xi_{j,i} = f_j(Y_i) - Pf_j$ for any fixed j, $1 \le j \le N$, and note that

$$\sqrt{n}(P_n - P) f_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{j,i}.$$

The random variables $\xi_{j,1}, \ldots, \xi_{j,n}$ are i.i.d. with expectation 0 and variance $P|f_j - Pf_j|^2 \le \eta^2$; and they are bounded in the sup-norm by $2\|f_j\|_S \le 2B$. Consequently, for every $y > 0$,

$$\Pr\left(\left| \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{j,i} \right| \ge y \right) \le 2 \exp\left[-\frac{1}{2} \frac{y^2}{\eta^2 + \frac{2B}{3\sqrt{n}} y} \right]$$

by Bernstein's inequality; see, e.g., 1.3.2 in Dudley (1999). Note that the above inequality trivially holds for $y = 0$. Letting $\xi_j = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \xi_{j,i}$, it follows that, for every $y \ge 0$,

$$\Pr(|\xi_j| \ge y) \le \begin{cases} 2 \exp\left[-y^2/4\eta^2\right] & \text{if } 0 \le y \le \frac{\eta^2}{2B/3\sqrt{n}}, \\ 2 \exp\left[-\frac{y}{8B/3\sqrt{n}}\right] & \text{if } y \ge \frac{\eta^2}{2B/3\sqrt{n}}. \end{cases} \tag{60}$$

Note that the r.h.s. in the previous display is well-defined and does not depend on j.

Next, we split $|\xi_j|$ into

$$\begin{aligned} |\xi_j| &= |\xi_j| 1_{\left\{|\xi_j| \le \frac{\eta^2}{2B/3\sqrt{n}}\right\}} + |\xi_j| 1_{\left\{|\xi_j| > \frac{\eta^2}{2B/3\sqrt{n}}\right\}} \\ &=: \xi_{\mathrm{I},j} + \xi_{\mathrm{II},j}, \end{aligned}$$

and prove a moment bound for each of the two summands. First, for every $\lambda_1 > 2$, we have

$$\begin{aligned}
\mathrm{E}\exp\left[\frac{\xi_{\mathrm{I},j}^2}{\eta^2\lambda_1^2}\right] &= 1+\mathrm{E}\int_{[0,\xi_{\mathrm{I},j}^2]}\frac{1}{\eta^2\lambda_1^2}\exp\left[\frac{y}{\eta^2\lambda_1^2}\right]dy\\
&= 1+\int_{[0,\infty)}\frac{1}{\eta^2\lambda_1^2}\exp\left[\frac{y}{\eta^2\lambda_1^2}\right]\Pr(\xi_{\mathrm{I},j}\geq\sqrt{y})\,dy\\
&\qquad\text{(by Tonelli's theorem)}\\
&\leq 1+\int_{[0,\infty)}\frac{2}{\eta^2\lambda_1^2}\exp\left[-\frac{y}{\eta^2}\left(\frac{1}{4}-\frac{1}{\lambda_1^2}\right)\right]dy\\
&\qquad\text{(by (60) and the definition of } \xi_{\mathrm{I},j}\text{)}\\
&= 1+\frac{8}{\lambda_1^2-4} \qquad (61)\\
&\qquad\text{(because } \lambda_1>2\text{)}.
\end{aligned}$$

Second, for every $\lambda_2 > 4$, we have that

$$\begin{aligned}
&\mathrm{E}\exp\left[\frac{\xi_{\mathrm{II},j}}{2B/3\sqrt{n}\times\lambda_2}\right]\\
&=1+\mathrm{E}\int_{[0,\xi_{\mathrm{II},j}]}\frac{1}{2B/3\sqrt{n}\times\lambda_2}\exp\left[\frac{y}{2B/3\sqrt{n}\times\lambda_2}\right]dy\\
&=1+\int_{[0,\infty)}\frac{1}{2B/3\sqrt{n}\times\lambda_2}\exp\left[\frac{y}{2B/3\sqrt{n}\times\lambda_2}\right]\Pr(\xi_{\mathrm{II},j}\geq y)\,dy\\
&\quad\text{(by Tonelli's theorem)}\\
&\leq 1+\int_{[0,\infty)}\frac{1}{B/3\sqrt{n}\times\lambda_2}\exp\left[-\frac{y}{2B/3\sqrt{n}}\left(\frac{1}{4}-\frac{1}{\lambda_2}\right)\right]dy\\
&\quad\text{(by (63) below)}\\
&=1+\frac{8}{\lambda_2-4} \qquad (62)\\
&\quad\text{(because } \lambda_2>4\text{)}.
\end{aligned}$$

The inequality

$$\Pr(\xi_{\mathrm{II},j}\geq y)\leq 2\exp\left[-\frac{y}{8B/3\sqrt{n}}\right] \qquad (63)$$

is satisfied for every $y \geq 0$: It is trivial for $y = 0$. If $0 < y \leq \frac{\eta^2}{2B/3\sqrt{n}}$, then the event $\{\xi_{\mathrm{II},j} \geq y\}$ is contained in the event $\left\{|\xi_j| > \frac{\eta^2}{2B/3\sqrt{n}}\right\}$ by definition of $\xi_{\mathrm{II},j}$, and therefore

$$\begin{aligned}
\Pr(\xi_{\mathrm{II},j}\geq y) &\leq \Pr\left(|\xi_j|>\frac{\eta^2}{2B/3\sqrt{n}}\right)\\
&\leq 2\exp\left[-\frac{\frac{\eta^2}{2B/3\sqrt{n}}}{8B/3\sqrt{n}}\right]\\
&\qquad\text{(in view of (60))}\\
&\leq 2\exp\left[-\frac{y}{8B/3\sqrt{n}}\right].
\end{aligned}$$

Finally, if $y > \frac{\eta^2}{2B/3\sqrt{n}}$, then (63) holds true in view of (60) and the inequality $\xi_{\mathrm{II},j} \le |\xi_j|$.

Now,

$$\begin{aligned}
\exp\left[\left(\mathrm{E}\,\frac{\max_{1\le j\le N}\xi_{\mathrm{I},j}}{\eta\lambda_1}\right)^2\right] &\le \mathrm{E}\exp\left[\left(\max_{1\le j\le N}\frac{\xi_{\mathrm{I},j}}{\eta\lambda_1}\right)^2\right] \\
&\qquad \text{(by Jensen's inequality)} \\
&\le \mathrm{E}\max_{1\le j\le N}\exp\left[\frac{\xi_{\mathrm{I},j}^2}{\eta^2\lambda_1^2}\right] \\
&\le \sum_{j=1}^N \mathrm{E}\exp\left[\frac{\xi_{\mathrm{I},j}^2}{\eta^2\lambda_1^2}\right] \\
&\le N\left(1+\frac{8}{\lambda_1^2-4}\right) \\
&\qquad \text{(by (61));}
\end{aligned}$$

hence,

$$\mathrm{E}\max_{1\le j\le N}\xi_{\mathrm{I},j} \le \eta\lambda_1\sqrt{\log N\left(1+\frac{8}{\lambda_1^2-4}\right)}. \tag{64}$$

Further,

$$\begin{aligned}
\exp\left[\mathrm{E}\,\frac{\max_{1\le j\le N}\xi_{\mathrm{II},j}}{2B/3\sqrt{n}\times\lambda_2}\right] &\le \mathrm{E}\exp\left[\frac{\max_{1\le j\le N}\xi_{\mathrm{II},j}}{2B/3\sqrt{n}\times\lambda_2}\right] \\
&\qquad \text{(by Jensen's inequality)} \\
&\le \mathrm{E}\max_{1\le j\le N}\exp\left[\frac{\xi_{\mathrm{II},j}}{2B/3\sqrt{n}\times\lambda_2}\right] \\
&\le \sum_{j=1}^N \mathrm{E}\exp\left[\frac{\xi_{\mathrm{II},j}}{2B/3\sqrt{n}\times\lambda_2}\right] \\
&\le N\left(1+\frac{8}{\lambda_2-4}\right) \\
&\qquad \text{(by (62));}
\end{aligned}$$

hence,

$$\mathrm{E}\max_{1\le j\le N}\xi_{\mathrm{II},j} \le \frac{2B}{3\sqrt{n}}\lambda_2\log N\left(1+\frac{8}{\lambda_2-4}\right). \tag{65}$$

Combine (64) with (65) to get

$$\mathrm{E}\max_{1\le j\le N}|\xi_j| \le \eta\lambda_1\sqrt{\log N\left(1+\frac{8}{\lambda_1^2-4}\right)} + \frac{2B}{3\sqrt{n}}\lambda_2\log N\left(1+\frac{8}{\lambda_2-4}\right).$$

Choosing $\lambda_1 = 2\sqrt{3}$ and $\lambda_2 = 12$ then leads to the inequality

$$\mathrm{E}\max_{1\le j\le N}|\xi_j| \le 8\left[\eta\sqrt{1+\log N} + \frac{B}{\sqrt{n}}(1+\log N)\right],$$

upon noting that $\log 2N \le 1+\log N$. $\square$

Remark 57. The proof of Lemma 56 actually shows that

$$\begin{aligned} &\mathrm{E} \max_{1\le j\le N} |\sqrt{n}(P_n - P)f_j| \\ &\quad \le \max_{1\le j\le N} \sqrt{P|f_j - Pf_j|^2} \inf_{\lambda_1>2} \left[\lambda_1 \sqrt{\log N \left(1 + \frac{8}{\lambda_1^2 - 4}\right)}\right] \\ &\qquad + \frac{2\max_{1\le j\le N} \|f_j\|_S}{3\sqrt{n}} \inf_{\lambda_2>4} \left[\lambda_2 \log N \left(1 + \frac{8}{\lambda_2 - 4}\right)\right]. \end{aligned}$$

The following theorem can be found as Lemma 3.4.2 in van der Vaart & Wellner (1996). We give a complete proof and thereby ensure that the constant in the inequality appearing there does not depend on the underlying probability measure.

Theorem 58. *Let $\mathcal{F}$ be a non-empty class of $\mathcal{A}$-measurable functions on S, which are bounded by B, $0 < B < \infty$, in the sup-norm and by η, $0 < \eta < \infty$, with respect to $\|\cdot\|_{2,P}$. Then*

$$\mathrm{E}^* \|\sqrt{n}(P_n - P)\|_{\mathcal{F}} \le (1696 + 64\sqrt{2})\, I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) \left[1 + \frac{B}{\eta^2\sqrt{n}} I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P})\right].$$

Proof. The notation and way of reasoning draw on the proof of Theorem 6.7 in Giné (2007).

If $N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,P}) = \infty$ for some $\varepsilon > 0$, then it is also infinite for all smaller values of ε. Hence, $I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) = \infty$, and the inequality in the theorem is trivially satisfied. So assume that $N_{[\,]}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,P}) < \infty$ for all $\varepsilon > 0$. Denote by j the unique integer satisfying $2^{-(j+1)} < \eta \le 2^{-j}$. We can find a sequence of partitions $\{A_{k,i} : i = 1, \ldots, N_k\}$ of $\mathcal{F}$ with the following properties:

1. $\sum_{k\ge j+1} 2^{-k}\sqrt{1 + \log N_k} \le 4\, I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P})$
2. $\left\|\sup^*_{g,h\in A_{k,i}} |g(\,\cdot\,) - h(\,\cdot\,)|\right\|_{2,P} \le 2^{-k}$ for all $i = 1, \ldots, N_k$ and $k \ge j+1$. [Here, $\sup^*_{g,h\in A_{k,i}} |g(\,\cdot\,) - h(\,\cdot\,)|$ denotes any measurable envelope function of $\sup_{g,h\in A_{k,i}} |g(\,\cdot\,) - h(\,\cdot\,)|$.]
3. The $\{A_{k,i}\}$ are nested, meaning that $\{A_{t,i}\}$ is a refinement of $\{A_{s,i}\}$ whenever $t \ge s$.

This is achieved as follows: For each integer $k \ge j+1$, choose $N_{[\,]}(2^{-k}, \mathcal{F}, \|\cdot\|_{2,P})$-many brackets that cover $\mathcal{F}$ and disjointify them. This gives a sequence $\{T_{k,i} : i = 1, \ldots, N_{[\,]}(2^{-k}, \mathcal{F}, \|\cdot\|_{2,P})\}$ of partitions of $\mathcal{F}$ satisfying

$$\sum_{k\ge j+1} 2^{-k}\sqrt{1 + H_{[\,]}(2^{-k}, \mathcal{F}, \|\cdot\|_{2,P})} \quad\le\quad 2I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) \tag{66}$$

(by the integral test)

and Property 2 above. For $k \ge j+1$ and

$$s_{j+1} \in \left\{1, \ldots, N_{[\,]}(2^{-(j+1)}, \mathcal{F}, \|\cdot\|_{2,P})\right\}, \ldots, s_k \in \left\{1, \ldots, N_{[\,]}(2^{-k}, \mathcal{F}, \|\cdot\|_{2,P})\right\},$$

set $A_{k,s_{j+1},\ldots,s_k} = \bigcap_{l=j+1}^{k} T_{l,s_l}$. Enumerate $(s_{j+1},\ldots,s_k)$ by natural numbers and call N_k the cardinality of $\{A_{k,i}\}$. Since

$$N_k \le N_{[\,]}(2^{-(j+1)},\mathcal{F},\|\cdot\|_{2,P})\cdots N_{[\,]}(2^{-k},\mathcal{F},\|\cdot\|_{2,P}),$$

Property 1 follows from

$$\begin{aligned}
\sum_{k\ge j+1} 2^{-k}\sqrt{1+\log N_k} &\le \sum_{k\ge j+1} 2^{-k}\sum_{m=j+1}^{k}\sqrt{1+H_{[\,]}(2^{-m},\mathcal{F},\|\cdot\|_{2,P})}\\
&= \sum_{m\ge j+1}\left[\sqrt{1+H_{[\,]}(2^{-m},\mathcal{F},\|\cdot\|_{2,P})}\sum_{k\ge m}2^{-k}\right]\\
&= 2\sum_{m\ge j+1}2^{-m}\sqrt{1+H_{[\,]}(2^{-m},\mathcal{F},\|\cdot\|_{2,P})}\\
&\le 4\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\\
&\qquad \text{(by (66))};
\end{aligned}$$

Properties 2 and 3 are clear by construction.

For $k \ge j+1$ and $i = 1,\ldots,N_k$, choose $f_{k,i} \in A_{k,i}$ and define $\pi_k f = f_{k,i}$ as well as $\Delta_k f(x) = \sup^*_{g,h\in A_{k,i}} |g(x)-h(x)|$ for $f \in A_{k,i}$ and $x \in S$. Note that as f varies in $\mathcal{F}$, $\Delta_k f$ varies in a set of cardinality at most N_k. By Property 3, all $f \in A_{k,i}$ have the same $\pi_{j+1}f,\ldots,\pi_k f,\Delta_{j+1}f,\ldots,\Delta_k f$, and $\Delta_k f \le \ldots \le \Delta_{j+1} f$ holds true. For $f \in \mathcal{F}$ and $x \in S$, define

$$\begin{aligned}
(\tau f)(x) &= \min\{k \ge j+1 : \Delta_k f(x) > \sqrt{n}a_k\}\\
&\qquad \text{(with the convention that } \min\emptyset := \infty),
\end{aligned}$$

where $a_k := 1/(2^{k+1}\sqrt{1+\log N_{k+1}})$. By definition we have the relations

$$\{\tau f = j+1\} = \{\Delta_{j+1} f > \sqrt{n}a_{j+1}\}$$

and, for $k > j+1$,

$$\begin{aligned}
\{\tau f = k\} &= \{\Delta_{j+1} f \le \sqrt{n}a_{j+1},\ldots,\Delta_{k-1} f \le \sqrt{n}a_{k-1},\Delta_k f > \sqrt{n}a_k\},\\
\{\tau f \ge k\} &= \{\Delta_{j+1} f \le \sqrt{n}a_{j+1},\ldots,\Delta_{k-1} f \le \sqrt{n}a_{k-1}\}.
\end{aligned}$$

For every $r \ge j+2$, we have that

$$\begin{aligned}
f &= \pi_{j+1}f + f - \pi_{j+1}f\\
&= \pi_{j+1}f + f - \pi_r f + \sum_{k=j+2}^{r}(\pi_k f - \pi_{k-1}f)\\
&= \pi_{j+1}f + f - \pi_r f + \sum_{k=j+2}^{r}(\pi_k f - \pi_{k-1}f)1_{\{\tau f<k\}} + \sum_{k=j+2}^{r}(\pi_k f - \pi_{k-1}f)1_{\{\tau f\ge k\}}.
\end{aligned}$$

Combine this with

$$\begin{aligned}
&\sum_{k=j+2}^{r}(\pi_k f-\pi_{k-1}f)1_{\{\tau f<k\}}\\
&\quad=\sum_{k=j+2}^{r}\pi_k f 1_{\{\tau f<k\}}-\sum_{k=j+3}^{r}\pi_{k-1}f1_{\{\tau f<k-1\}}-\sum_{k=j+2}^{r}\pi_{k-1}f1_{\{\tau f=k-1\}}\\
&\quad=\pi_r f1_{\{\tau f<r\}}-\sum_{k=j+2}^{r}\pi_{k-1}f1_{\{\tau f=k-1\}}
\end{aligned}$$

and $\pi_r f=\pi_r f1_{\{\tau f<r\}}+\pi_r f1_{\{\tau f\geq r\}}$ to arrive at the following decomposition of f:

$$\begin{aligned}
f &= \pi_{j+1}f+f-\pi_r f1_{\{\tau f\geq r\}}-\sum_{k=j+2}^{r}\pi_{k-1}f1_{\{\tau f=k-1\}}+\sum_{k=j+2}^{r}(\pi_k f-\pi_{k-1}f)1_{\{\tau f\geq k\}}\\
&= \pi_{j+1}f+(f-\pi_{j+1}f)1_{\{\tau f=j+1\}}+(f-\pi_r f)1_{\{\tau f\geq r\}}+\sum_{k=j+2}^{r-1}(f-\pi_k f)1_{\{\tau f=k\}}\\
&\quad+\sum_{k=j+2}^{r}(\pi_k f-\pi_{k-1}f)1_{\{\tau f\geq k\}}.
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathrm{E}^*\left\|\sqrt{n}(P_n-P)\right\|_{\mathcal{F}}\\
&\quad\leq \mathrm{E}\left\|\sqrt{n}(P_n-P)\pi_{j+1}f\right\|_{\mathcal{F}}\\
&\qquad+\mathrm{E}^*\left\|\sqrt{n}(P_n-P)(f-\pi_{j+1}f)1_{\{\tau f=j+1\}}\right\|_{\mathcal{F}}\\
&\qquad+\mathrm{E}^*\left\|\sqrt{n}(P_n-P)(f-\pi_r f)1_{\{\tau f\geq r\}}\right\|_{\mathcal{F}}\\
&\qquad+\mathrm{E}^*\left\|\sqrt{n}(P_n-P)\sum_{k=j+2}^{r-1}(f-\pi_k f)1_{\{\tau f=k\}}\right\|_{\mathcal{F}}\\
&\qquad+\mathrm{E}\left\|\sqrt{n}(P_n-P)\sum_{k=j+2}^{r}(\pi_k f-\pi_{k-1}f)1_{\{\tau f\geq k\}}\right\|_{\mathcal{F}}\\
&\quad=\mathrm{I}+\mathrm{II}+\mathrm{III}+\mathrm{IV}+\mathrm{V}.
\end{aligned}$$

In order to prove the theorem, we now derive bounds for each of these expressions. In doing so, we will make use of the observation that $|f|\leq g$ implies that

$$|(P_n-P)f|\leq P_n g+Pg=(P_n-P)g+2Pg. \tag{67}$$

Bound for I: By Lemma 56 and the assumptions about $\mathcal{F}$,

$$\begin{aligned}
\mathrm{I} &\leq 8\left[\eta\sqrt{1+\log N_{j+1}}+\frac{B}{\sqrt{n}}(1+\log N_{j+1})\right]\\
&\leq 8\left[2\times 2^{-(j+1)}\sqrt{1+\log N_{j+1}}+\frac{2^{2(j+1)}B}{\sqrt{n}}2^{-2(j+1)}(1+\log N_{j+1})\right]\\
&\leq 8\left[8\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})+\frac{64B}{\eta^2\sqrt{n}}\,I^2_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\right]\\
&\qquad \text{(by Property 1)}\\
&\leq 512\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\left[1+\frac{B}{\eta^2\sqrt{n}}\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\right].
\end{aligned}$$

Bound for II: By (67),

$$\mathrm{II}\leq \mathrm{E}\,\|\sqrt{n}(P_n-P)\Delta_{j+1}f\|_{\mathcal{F}}+2\sqrt{n}\|P\Delta_{j+1}f1_{\{\tau f=j+1\}}\|_{\mathcal{F}}.$$

Using Properties 1–2 and Lemma 56, the first term can be bounded by

$$\begin{aligned}
&8\left[2^{-(j+1)}\sqrt{1+\log N_{j+1}}+\frac{2B}{\sqrt{n}}(1+\log N_{j+1})\right]\\
&\quad\leq 8\left[2^{-(j+1)}\sqrt{1+\log N_{j+1}}+\frac{2^{2j+3}B}{\sqrt{n}}2^{-2(j+1)}(1+\log N_{j+1})\right]\\
&\quad\leq 8\left[4\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})+\frac{128B}{\eta^2\sqrt{n}}\,I^2_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\right].
\end{aligned}$$

For the second term observe that

$$\begin{aligned}
P\Delta_{j+1}f1_{\{\tau f=j+1\}} &= P\Delta_{j+1}f1_{\{\Delta_{j+1}f>\sqrt{n}a_{j+1}\}}\\
&\leq \frac{P(\Delta_{j+1}f)^2}{\sqrt{n}a_{j+1}}\\
&\leq \frac{4}{\sqrt{n}}2^{-(j+2)}\sqrt{1+\log N_{j+2}}\\
&\qquad \text{(by Property 2 and the definition of } a_{j+1})\\
&\leq \frac{16}{\sqrt{n}}I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\\
&\qquad \text{(by Property 1)}
\end{aligned}$$

Hence

$$\mathrm{II}\leq 1024\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\left[1+\frac{B}{\eta^2\sqrt{n}}\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\right].$$

Bound for III: It follows from Property 3 that

$$\{\tau f\geq r\}\subseteq\{\Delta_{r-1}f\leq\sqrt{n}a_{r-1}\}\subseteq\{\Delta_r f\leq\sqrt{n}a_{r-1}\}.$$

This together with (67) yields

$$\mathrm{III}\leq \mathrm{E}\,\|\sqrt{n}(P_n-P)\Delta_r f1_{\{\Delta_r f\leq\sqrt{n}a_{r-1}\}}\|_{\mathcal{F}}+2\sqrt{n}\|P\Delta_r f1_{\{\Delta_r f\leq\sqrt{n}a_{r-1}\}}\|_{\mathcal{F}}.$$

By Lemma 56 and Property 2, the first term can be bounded by

$$8\left[2^{-r}\sqrt{1+\log N_r}+\frac{\sqrt{n}}{2^r\sqrt{1+\log N_r}}\frac{1}{\sqrt{n}}(1+\log N_r)\right]=16\times 2^{-r}\sqrt{1+\log N_r},$$

whereas the second term is less or equal than $2\sqrt{n}\sup_{f\in\mathcal{F}}\|\Delta_r f\|_{2,P}\leq 2\sqrt{n}2^{-r}$. Consequently, Expression *III* is bounded by

$$C(r):=16\times 2^{-r}\sqrt{1+\log N_r}+2^{-r+1}\sqrt{n},$$

which converges to 0 as $r\to\infty$.

Bound for IV: Using (67),

$$\mathrm{IV}\leq\sum_{k=j+2}^{r-1}\left[\mathrm{E}\,\|\sqrt{n}(P_n-P)\Delta_k f 1_{\{\tau f=k\}}\|_{\mathcal{F}}+2\sqrt{n}\|P\Delta_k f 1_{\{\tau f=k\}}\|_{\mathcal{F}}\right].$$

On $\{\tau f=k\}$, $\Delta_k f\leq\Delta_{k-1}f\leq\sqrt{n}a_{k-1}=\sqrt{n}/(2^k\sqrt{1+\log N_k})$. Further, using Property 2,

$$\begin{aligned}\mathrm{Var}_P[\Delta_k f 1_{\{\tau f=k\}}] &\leq P(\Delta_k f)^2 1_{\{\tau f=k\}}\\ &\leq \sqrt{n}a_{k-1}P\Delta_k f 1_{\{\Delta_k f>\sqrt{n}a_k\}}\\ &\leq \frac{a_{k-1}}{a_k}P(\Delta_k f)^2\leq\frac{2\sqrt{1+\log N_{k+1}}}{\sqrt{1+\log N_k}}2^{-2k}\leq\frac{1+\log N_{k+1}}{1+\log N_k}2^{-2k+1}.\end{aligned}$$

So, by Lemma 56,

$$\begin{aligned}&\mathrm{E}\,\|\sqrt{n}(P_n-P)\Delta_k f 1_{\{\tau f=k\}}\|_{\mathcal{F}}\\ &\leq 8\left[\sqrt{\frac{2^{-2k+1}(1+\log N_{k+1})}{1+\log N_k}}\sqrt{1+\log N_k}+\frac{\sqrt{n}}{2^k\sqrt{1+\log N_k}}\frac{1}{\sqrt{n}}(1+\log N_k)\right]\\ &=8\left(\sqrt{8}\times 2^{-(k+1)}\sqrt{1+\log N_{k+1}}+2^{-k}\sqrt{1+\log N_k}\right)\end{aligned}\tag{68}$$

Next,

$$2\sqrt{n}\|P\Delta_k f 1_{\{\tau f=k\}}\|_{\mathcal{F}}\leq 8\times 2^{-(k+1)}\sqrt{1+\log N_{k+1}}\tag{69}$$

since, using Property 2,

$$P\Delta_k f 1_{\{\tau f=k\}}\leq P\Delta_k f 1_{\{\Delta_k f>\sqrt{n}a_k\}}\leq\frac{P(\Delta_k f)^2}{\sqrt{n}a_k}\leq 2^{-k+1}\frac{\sqrt{1+\log N_{k+1}}}{\sqrt{n}}.$$

Combine (68) with (69) to obtain

$$\begin{aligned}\mathrm{IV} &\leq (8\sqrt{8}+8)\sum_{k=j+2}^{r-1}2^{-(k+1)}\sqrt{1+\log N_{k+1}}+8\sum_{k=j+2}^{r-1}2^{-k}\sqrt{1+\log N_k}\\ &\leq 64(1+\sqrt{2})\,I_{[\,]}(\eta,\mathcal{F},\|\cdot\|_{2,P})\\ &\qquad\text{(by Property 1).}\end{aligned}$$

Bound for V: Property 3 implies that as f varies in $\mathcal{F}$, $(\pi_k f-\pi_{k-1}f)1_{\{\tau f\geq k\}}$ varies in a set of cardinality at most N_k. Each of these functions is bounded in absolute value by

$$\Delta_{k-1}f 1_{\{\Delta_{k-1}f\leq\sqrt{n}a_{k-1}\}}\leq\frac{\sqrt{n}}{2^k\sqrt{1+\log N_k}}$$

and satisfies

$$\mathrm{Var}_P[(\pi_k f - \pi_{k-1} f) 1_{\{\tau f \geq k\}}] \leq P(\pi_k f - \pi_{k-1} f)^2 \leq P(\Delta_{k-1} f)^2 \leq 2^{-2(k-1)}$$

by Property 2. Apply Lemma 56 to conclude that

$$\begin{aligned}
\mathrm{V} &\leq 8 \sum_{k=j+2}^{r} \left[2^{-(k-1)} \sqrt{1 + \log N_k} + \frac{\sqrt{n}}{2^k \sqrt{1 + \log N_k}} \frac{1}{\sqrt{n}} (1 + \log N_k) \right] \\
&= 24 \sum_{k=j+2}^{r} 2^{-k} \sqrt{1 + \log N_k} \\
&\leq 96\, I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) \\
&\qquad \text{(by Property 1)}.
\end{aligned}$$

All in all,

$$\begin{aligned}
&\mathrm{E}_P^* \, \|\sqrt{n}(\mathbb{P}_n - P)\|_{\mathcal{F}} \\
&\quad \leq (1696 + 64\sqrt{2})\, I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) \left[1 + \frac{B}{\eta^2 \sqrt{n}} I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) \right] + C(r).
\end{aligned}$$

Letting $r \to \infty$ completes the proof. □

Appendix E: Convexity of level sets

The following lemma states that the maximum set of a concave mapping that is defined on a convex set is convex. The result is standard in optimization theory (see, e.g., p. 263 f. in Rockafellar, 1970).

Lemma 59. *Let E be a real vector space, C a convex subset of E, and $f : C \to [-\infty, \infty)$ a concave mapping. Denote by S the set of all $c \in C$ that satisfy*

$$f(c) = \sup_{x \in C} f(x).$$

Then S is convex.

Proof. If S is empty, it is convex, and we are done. So let S be non-empty. Pick $c, d \in S$ and set $\alpha = \sup_{x \in C} f(x)$. For $\lambda \in (0, 1)$, we conclude from the concavity of f that

$$\alpha \geq f(\lambda c + (1 - \lambda) d) \geq \lambda f(c) + (1 - \lambda) f(d) = \alpha.$$

Hence, $\lambda c + (1 - \lambda) d \in S$. This shows the convexity of S. □

Appendix F: Continuous selections

The following lemma is a special case of Berge's maximum theorem.

Lemma 60. *Let X be a metric and Y a compact metric space. Let $u : X \times Y \to [-\infty, \infty)$ be a continuous function that for every $x \in X$, has a unique maximizer, say $v(x)$, on the fiber $\{(x, y) : y \in Y\}$. Then the mapping $v : X \to Y$ is continuous.*

Proof. Suppose, on the contrary, that v is not continuous at some point $x \in X$. Then there is a sequence x_n converging to x such that $y_n := v(x_n)$ does not converge to $v(x)$. Hence there is an $\varepsilon > 0$ and a subsequence $y_{n'}$ of y_n such that $y_{n'}$ does not belong to the open ball of radius ε around $v(x)$. Since Y is compact, there is a subsequence $y_{n''}$ of $y_{n'}$ converging to some $y \in Y$. Since the $y_{n''}$ are at least ε-far away from $v(x)$, y has to be different from $v(x)$. Making use of the continuity of u, we see that $u(x_{n''}, y_{n''})$ converges to $u(x, y)$ and $u(x_{n''}, v(x))$ converges to $u(x, v(x))$. Combine this with the strict inequality $u(x, y) < u(x, v(x))$ to conclude that $u(x_{n_0''}, y_{n_0''}) < u(x_{n_0''}, v(x))$ for some index n_0''. This is a contradiction to $y_{n_0''}$ being the unique maximizer of u over the fiber $\{(x_{n_0''}, y) : y \in Y\}$. □

Appendix G: Auxiliary results for Section 7

Proposition 61. *Let $\xi > 0$ and define $\mathcal{U} = \{f \in \mathsf{L}^\infty(\Omega) : \inf_{x\in\Omega} f(x) > \xi\}$. Let α be a non-negative integer, $f \in \mathcal{U}$, and $f_1, \ldots, f_\alpha \in \mathsf{L}^\infty(\Omega)$. For any probability measure P on $(\Omega, \mathcal{B}(\Omega))$, the α-linear functionals representing the α-th Fréchet derivative of $L_P : \mathcal{U} \to \mathbb{R}$ are given by*

$$\begin{aligned} D^\alpha L_P(f)(f_1, \ldots, f_\alpha) &= (-1)^{\alpha-1}(\alpha-1)!\, P(f^{-\alpha} f_1 \cdots f_\alpha) \\ &= (-1)^{\alpha-1}(\alpha-1)! \int_\Omega f^{-\alpha} f_1 \cdots f_\alpha dP. \end{aligned}$$

Proof. The proof is the same as that of the corresponding part of Proposition 3 in Nickl (2007). □

Proposition 62. *Let $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t, \zeta, D)$ be such that $p_{0,n}$ converges pointwise to $p_{0,\infty}$. Let α be a non-negative integer and $\mathcal{H}_1, \ldots, \mathcal{H}_\alpha$ be non-empty bounded subsets of some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$. If Assumption Z is satisfied, then the following statements hold true:*

1. *$\sqrt{n}(D^\alpha L_n(p) - D^\alpha L_{\mathbb{P}(p_{0,n})}(p))(h_1, \ldots, h_\alpha)$ converges weakly in $\ell^\infty(\mathcal{F})$ to a $\mathbb{P}(p_{0,\infty})$-Brownian bridge indexed by $\mathcal{F}$, where*

$$\mathcal{F} := \left\{(\alpha-1)!\, p^{-\alpha} h_1 \cdots h_\alpha : p \in \mathcal{P}(t, \zeta, D),\, h_1 \in \mathcal{H}_1, \ldots, h_\alpha \in \mathcal{H}_\alpha\right\};$$

2.

$$\lim_{M\to\infty} \limsup_{n\to\infty} \mathbb{P}(p_{0,n})^{n*}\left(\sqrt{n} \sup_{p\in\mathcal{P}(t,\zeta,D)} \left\|D^\alpha L_n(p) - D^\alpha L_{\mathbb{P}(p_{0,n})}(p)\right\|_{\mathcal{H}_1\times\cdots\times\mathcal{H}_\alpha} > M\right) = 0. \tag{70}$$

Proof. Part 1: For any probability measure P on $(\Omega, \mathcal{B}(\Omega))$, we have

$$\begin{aligned} &\sup_{p\in\mathcal{P}(t,\zeta,D)} \|D^\alpha L_n(p) - D^\alpha L_P(p)\|_{\mathcal{H}_1\times\cdots\times\mathcal{H}_\alpha} \\ &= (\alpha-1)! \sup_{p\in\mathcal{P}(t,\zeta,D)} \sup_{h_1\in\mathcal{H}_1} \cdots \sup_{h_\alpha\in\mathcal{H}_\alpha} (\mathbb{P}_n - P)(p^{-\alpha} h_1 \cdots h_\alpha) \\ &\qquad \text{(by the first part of Proposition 44 and by Proposition 61)} \\ &= \|\mathbb{P}_n - P\|_{\mathcal{F}}. \end{aligned}$$

The class of functions $\mathcal{F}$ is uniformly Donsker by the following arguments: Let $r = \min(s,t)$. Assumption Z ensures that the densities $p \in \mathcal{P}(t,\zeta,D)$ are bounded below by $\zeta > 0$, and so $\{1/p : p \in \mathcal{P}(t,\zeta,D)\}$ is a bounded subset of $\mathsf{W}_2^t(\Omega)$ by Part 4 of Proposition 1. Consequently, $\{1/p : p \in \mathcal{P}(t,\zeta,D)\}$ is bounded in $\mathsf{W}_2^r(\Omega)$ by Part 3 of Proposition 1. Since $\mathcal{H}_1,\ldots,\mathcal{H}_\alpha$ are also bounded in $\mathsf{W}_2^r(\Omega)$ by Part 3 of Proposition 1, we conclude that $\mathcal{F}$ is a bounded subset of $\mathsf{W}_2^r(\Omega)$, and hence is uniformly Donsker by Lemma 13.

Since, in view of Proposition 3,

$$\begin{aligned}\lim_{n\to\infty} \|\mathbb{P}(p_{0,n}) - \mathbb{P}(p_{0,\infty})\|_{\mathcal{G}} &= \lim_{n\to\infty}\sup_{g\in\mathcal{G}}\left|\int_\Omega g(p_{0,n}-p_{0,\infty})d\lambda\right| \\ &\le \lambda(\Omega)\sup_{g\in\mathcal{G}}\|g\|_\Omega \lim_{n\to\infty}\|p_{0,n}-p_{0,\infty}\|_\Omega \\ &= 0\end{aligned}$$

for any class $\mathcal{G}$ of functions that is bounded in the sup-norm, it follows from Corollary 2.7 in Giné & Zinn (1991) that $\sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))$ converges weakly in $\ell^\infty(\mathcal{F})$ to a $\mathbb{P}(p_{0,\infty})$-Brownian bridge indexed by $\mathcal{F}$.

Part 2: In view of Lemma 1.3.8 in van der Vaart & Wellner (1996) and by continuity of the sup-norm $\|\cdot\|_{\mathcal{F}}$ on $\ell^\infty(\mathcal{F})$, $\|\sqrt{n}(\mathbb{P}_n - \mathbb{P}(p_{0,n}))\|_{\mathcal{F}}$ is asymptotically tight in $\mathbb{R}$, which in turn is equivalent to (70). □

Remark 63. Inspection of the proof of Proposition 62 shows that (70) still holds true when $\mathcal{H}_1,\ldots,\mathcal{H}_\alpha$ are bounded subsets of $\mathsf{L}^\infty(\Omega)$ that are uniformly Donsker. The only important property we need is a permanence result: Any finite product of classes of functions that are uniformly Donsker is itself uniformly Donsker.

Proposition 64.

1. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous mapping from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let Assumption Z be satisfied. Then, for every $q \in \mathcal{P}(t,\zeta,D)$, Q_q is a continuous real-valued function on Θ.*

2. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let Assumption Z be satisfied. Then*

$$\lim_{n\to\infty}\sup_{q\in\mathcal{P}(t,\zeta,D)} \Pr(q)^*\left(\sup_{\theta\in\Theta}|\mathbb{Q}_n(\theta) - Q_q(\theta)| > \eta\right) = 0$$

for every $\eta > 0$.

3. *Let Assumptions P.1, P.2, R.1, and Z be satisfied. Then*

$$\lim_{\min(n,k)\to\infty}\sup_{q\in\mathcal{P}(t,\zeta,D)} \Pr(q)^*\left(\sup_{\theta\in\Theta}|\mathbb{Q}_{n,k}(\theta) - Q_q(\theta)| > \eta\right) = 0$$

for every $\eta > 0$.

Proof. Part 1: Part 1 is an immediate consequence of Lemma 47.

Part 2: A straightforward calculation shows that

$$\mathbb{Q}_n(\theta) - Q_q(\theta) = \int_\Omega (\hat{p}_n - q)\frac{-p_\theta^2}{\hat{p}_n q}d\lambda.$$

Consequently,

$$\sup_{\theta\in\Theta} |\mathbb{Q}_n(\theta) - Q_q(\theta)| \leq \zeta^{-2} \sup_{\theta\in\Theta} \|p_\theta\|_2^2 \, \|\hat{p}_n - q\|_\Omega. \tag{71}$$

We have that $\sup_{\theta\in\Theta} \|p_\theta\|_2 < \infty$ by compactness of Θ and the assumptions on $\mathcal{P}_\Theta$. By Theorem 33,

$$\lim_{n\to\infty} \sup_{q\in\mathcal{P}(t,\zeta,D)} \Pr(q)^* \left(\|\hat{p}_n - q\|_\Omega > \eta\right) = 0$$

for every $\eta > 0$, and so Part 2 is true in view of (71).

Part 3: A straightforward calculation shows that

$$\mathbb{Q}_{n,k}(\theta) - Q_q(\theta) = \int_\Omega (\hat{p}_n - q)\frac{-p_\theta^2}{\hat{p}_n q} d\lambda + \int_\Omega (\tilde{p}_k(\theta) - p_\theta)\frac{\tilde{p}_k(\theta) + p_\theta}{\hat{p}_n} d\lambda.$$

Hence,

$$\begin{aligned}\sup_{\theta\in\Theta} |\mathbb{Q}_{n,k}(\theta) - Q_q(\theta)| \;\; &\leq \;\; \zeta^{-2} D^2 \, \|\hat{p}_n - q\|_\Omega \\ &\qquad\qquad + 2\zeta^{-1} C_t D \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega \, ,\end{aligned}$$

where we have used Part 2 of Proposition 1. Theorem 33 states that

$$\lim_{n\to\infty} \sup_{q\in\mathcal{P}(t,\zeta,D)} \Pr(q)^* \left(\|\hat{p}_n - q\|_\Omega > \eta\right) = 0$$

for every $\eta > 0$, whereas Part 3 of Theorem 11 implies that $\sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega$ converges to 0 in probability as $k \to \infty$. This proves Part 3. □

Lemma 65. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous mapping from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let Assumption Z be satisfied. For $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t,\zeta,D)$, let $\theta_{0,n}^* \in \Theta$ be a minimizer of $Q_{p_{0,n}}$, and let $\theta_{0,\infty}^* \in \Theta$ be the unique minimizer of $Q_{p_{0,\infty}}$. If $p_{0,n}$ converges pointwise to $p_{0,\infty}$, then $\theta_{0,n}^*$ converges to $\theta_{0,\infty}^*$.*

Proof. By Part 1 of Proposition 64, the functions $Q_{p_{0,n}}$ and $Q_{p_{0,\infty}}$ are continuous on the compact set Θ and consequently attain their respective infima. We claim that the assertion follows from Lemma 3.1 in Pötscher & Prucha (1997), and apply this lemma to $R_n = Q_{p_{0,n}}$, $\bar{R}_n = Q_{p_{0,\infty}}$, $\hat{\beta}_n = \theta_{0,n}^*$, and $\bar{\beta}_n = \theta_{0,\infty}^*$. We have to check its premises: Condition (3.2) in Pötscher & Prucha (1997) reads as

$$\lim_{n\to\infty} \sup_{\theta\in\Theta} |Q_{p_{0,n}}(\theta) - Q_{p_{0,\infty}}(\theta)| = 0. \tag{72}$$

We have that

$$\begin{aligned}\sup_{\theta\in\Theta} |Q_{p_{0,n}}(\theta) - Q_{p_{0,\infty}}(\theta)| \;\; &= \;\; \sup_{\theta\in\Theta} \left| \int_\Omega (p_{0,n} - p_{0,\infty}) \frac{-p_\theta^2}{p_{0,n} p_{0,\infty}} d\lambda \right| \\ &\leq \;\; \zeta^{-2} \sup_{\theta\in\Theta} \|p_\theta\|_2^2 \, \|p_{0,n} - p_{0,\infty}\|_\Omega;\end{aligned}$$

Since Θ is compact, the assumptions on $\mathcal{P}_\Theta$ imply that $\sup_{\theta\in\Theta} \|p_\theta\|_2 < \infty$. Now, (72) holds in view of Proposition 3.

Next, observe that $\theta_{0,\infty}^*$ is identifiably unique in the sense used in Pötscher & Prucha (1997) because it uniquely minimizes $Q_{p_{0,\infty}}$, and $Q_{p_{0,\infty}}$ is a continuous function on the compact set Θ.

Finally, Condition (3.3) in Pötscher & Prucha (1997) holds by choice of $\theta_{0,n}^*$. □

Lemma 66. *Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous mapping from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Let Assumptions P.1, P.5, and Z be satisfied. Let $q \in \mathcal{P}(t,\zeta,D)$. Then Q_q is twice continuously partially differentiable on Θ° and*

$$\begin{aligned}
\frac{\partial Q_q}{\partial \theta_i}(\theta) &= -2\int_\Omega (q-p_\theta)\frac{\partial p}{\partial \theta_i}(\,\cdot\,,\theta)\frac{1}{q}d\lambda \\
\frac{\partial^2 Q_q}{\partial \theta_i \partial \theta_j}(\theta) &= 2\int_\Omega \frac{\partial p}{\partial \theta_i}(\,\cdot\,,\theta)\frac{\partial p}{\partial \theta_j}(\,\cdot\,,\theta)\frac{1}{q}d\lambda \\
&\qquad -2\int_\Omega (q-p_\theta)\frac{\partial^2 p}{\partial \theta_i \partial \theta_j}(\,\cdot\,,\theta)\frac{1}{q}d\lambda,
\end{aligned}$$

where $i = 1,\dots,m$ and $j = 1,\dots,m$.

Proof. Use Assumptions P.1, P.5, and Z to interchange differentiation and integration. For continuity, use the theorem of dominated convergence. □

Proposition 67. *Let Assumptions P.1, P.5, and Z be satisfied. Let $p_{0,n}, p_{0,\infty} \in \mathcal{P}(t,\zeta,D)$ be such that $p_{0,n}$ converges pointwise to $p_{0,\infty}$. Then*

$$\sup_{\theta\in\Theta^\circ}\left|\frac{\partial^2 \mathbb{Q}_n}{\partial \theta_i \partial \theta_j}(\theta) - \frac{\partial^2 Q_{p_{0,\infty}}}{\partial \theta_i \partial \theta_j}(\theta)\right| = o^*_{\mathbb{P}(p_{0,n})^n}(1) \quad \text{as } n \to \infty$$

for all $i,j = 1,\dots,m$, where the underlying probability spaces are $(\Omega^n, \mathcal{B}(\Omega)^n, \mathbb{P}(p_{0,n})^n)$.

Proof. Combine Part 1 of Lemma 51 and Lemma 66 to obtain

$$\begin{aligned}
&\frac{\partial^2 \mathbb{Q}_n}{\partial \theta_i \partial \theta_j}(\theta) - \frac{\partial^2 Q_{p_{0,\infty}}}{\partial \theta_i \partial \theta_j}(\theta) \\
&\quad = -2\int_\Omega (\hat{p}_n - p_{0,\infty})\frac{\partial p}{\partial \theta_i}(\,\cdot\,,\theta)\frac{\partial p}{\partial \theta_j}(\,\cdot\,,\theta)\frac{1}{p_{0,n}p_{0,\infty}}d\lambda \\
&\qquad -2\int_\Omega (\hat{p}_n - p_{0,\infty})\frac{\partial^2 p}{\partial \theta_i \partial \theta_j}(\,\cdot\,,\theta)\frac{p_\theta}{\hat{p}_n p_{0,\infty}}d\lambda
\end{aligned}$$

for every $\theta \in \Theta^\circ$. Hence,

$$\begin{aligned}
&\sup_{\theta\in\Theta^\circ}\left|\frac{\partial^2 \mathbb{Q}_n}{\partial \theta_i \partial \theta_j}(\theta) - \frac{\partial^2 Q_{p_{0,\infty}}}{\partial \theta_i \partial \theta_j}(\theta)\right| \\
&\quad \le 2\zeta^{-2}\left[\int_\Omega \sup_{\theta\in\Theta^\circ}\left|\frac{\partial p}{\partial \theta_i}(x,\theta)\frac{\partial p}{\partial \theta_j}(x,\theta)\right| d\lambda(x)\right. \\
&\qquad \left. + C_t D\int_\Omega \sup_{\theta\in\Theta^\circ}\left|\frac{\partial^2 p}{\partial \theta_i \partial \theta_j}(x,\theta)\right| d\lambda(x)\right]\|\hat{p}_n - p_{0,\infty}\|_\Omega.
\end{aligned}$$

Using Assumptions P.1 and P.5, it remains to show that $\|\hat{p}_n - p_{0,\infty}\|_\Omega$ converges to 0 in probability. Clearly,

$$\|\hat{p}_n - p_{0,\infty}\|_\Omega \le \|\hat{p}_n - p_{0,n}\|_\Omega + \|p_{0,n} - p_{0,\infty}\|_\Omega.$$

Now, it follows from Theorem 33 that $\|\hat{p}_n - p_{0,n}\|_\Omega$ converges to 0 in outer $\mathbb{P}(p_{0,n})^n$-probability. Further, $\|p_{0,n} - p_{0,\infty}\|_\Omega$ converges to 0 by Proposition 3. This completes the proof. □

References

[1] Altissimo, F. and Mele, A. (2009). Simulated non-parametric estimation of dynamic models. *Rev. Econom. Stud.* **76** 413–450.

[2] Beran, R. (1977). Minimum Hellinger distance estimates for parametric models. *Ann. Statist.* **5** 445–463.

[3] Carrasco, M., Chernov, M., Florens, J.-P., and Ghysels, E. (2007). Efficient estimation of general dynamic models with a continuum of moment conditions. *J. Econometrics* **140** 529–573.

[4] Dridi, R., Guay, A., and Renault, E. (2007). Indirect inference and calibration of dynamic stochastic general equilibrium models. *J. Econometrics* **136** 397–430.

[5] Dudley, R. M. (1999). *Uniform central limit theorems.* Cambridge Studies in Advanced Mathematics **63**. Cambridge University Press, Cambridge.

[6] Gallant, R. and Long, J. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* **84** 125–141.

[7] Giné, E. (2007). *Empirical Processes and some of their applications*, Lecture notes.

[8] Giné, E. and Zinn, J. (1991). Gaussian characterization of uniform Donsker classes of functions. *Ann. Probab.* **19** 758–782.

[9] Gouriéroux, C. and Monfort, A. (1996). *Simulation-Based Econometric Methods.* Oxford University Press, New York.

[10] Gouriéroux, C., Monfort, A., and Renault, E. (1993). Indirect inference. *J. Appl. Econom.* **8** 85–118.

[11] Jiang, W. and Turnbull, B. (2004). The indirect method: inference based on intermediate statistics – a synthesis and examples. *Statist. Sci.* **19** 239–263.

[12] Ledoux, M. and Talagrand, M. (1991). *Probability in Banach Spaces.* Results in Mathematics and Related Areas (3) **23**. Springer-Verlag, Berlin.

[13] Lions, J. L. and Magenes, E. (1972). *Non-homogeneous boundary value problems and applications. Vol. I.* Die Grundlehren der mathematischen Wissenschaften **181**. Springer-Verlag, New York-Heidelberg.

[14] Lombardi, M. and Calzolari, G. (2008). Indirect estimation of α-stable distributions and processes. *Econometrics Journal* **11** 193–208.

[15] McFadden, D. (1989). A method of simulated moments for estimation of discrete response models without numerical integration. *Econometrica* **57** 995–1026.

[16] Nickl, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields* **138** 411–449.

[17] Nickl, R. and Pötscher, B. M. (2007). Bracketing metric entropy rates and empirical central limit theorems for function classes of Besov- and Sobolev-type, *J. Theoret. Probab.* **20** 177–199.

[18] Nickl, R. and Pötscher, B. M. (2009). Efficient simulation-based minimum distance estimation and indirect inference, Submitted article.

[19] Pisier, G. (1983). Some applications of the metric entropy condition to harmonic analysis. *Banach spaces, harmonic analysis and probability. Lecture Notes in Mathematics* **995** 123–154. Springer-Verlag, New York.

[20] Pötscher, B. M. and Prucha, I. R. (1997). *Dynamic nonlinear econometric models. Asymptotic theory.* Springer-Verlag, Berlin.

[21] Rockafellar, R. T. (1970). *Convex analysis.* Princeton University Press.

[22] Smith, A. (1990). Three essays on the solution and estimation of dynamic macroeconomic models. Ph.D. dissertation, Duke University.

[23] Smith, A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions, *J. Appl. Econom.* **8** 63–84.

[24] Thomas-Agnan, C. (1996). Computing a family of reproducing kernels for statistical applications. *Numer. Algorithms* **13** 21–32.

[25] Triebel, H. (1983). *Theory of function spaces.* Monographs in Mathematics **78**. Birkhäuser Verlag, Basel.

[26] Triebel, H. (1995). *Interpolation theory, function spaces, differential operators.* Second edition. Johann Ambrosius Barth, Heidelberg.

[27] van de Geer, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44.

[28] van der Vaart, A. W. and Wellner, J. A. (1996). *Weak convergence and empirical processes.* Springer, New York.

Abstract

We extend the method of indirect inference by using *non-parametric* auxiliary models of densities. Suppose we observe i.i.d. random variables $X_1, \dots, X_n$ and are given a (possibly misspecified) parametric model with parameter set $\Theta \subseteq \mathbb{R}^m$. The auxiliary maximum likelihood estimator given $X_1, \dots, X_n$ is defined as the maximizer of the auxiliary likelihood function and is denoted by $\hat{p}_n$. Similarly, for data $X_1(\theta), \dots, X_{k(n)}(\theta)$ that are simulated according to the parametric model, define $\tilde{p}_{k(n)}(\theta)$ as the maximizer of the corresponding auxiliary likelihood function. We show that $\hat{p}_n$ and $\tilde{p}_{k(n)}(\theta)$ are unique, thereby allowing to define an indirect inference estimator $\hat{\theta}_{n,k(n)}$ as minimizer of an appropriately weighted L^2-distance between $\hat{p}_n$ and $\tilde{p}_{k(n)}(\theta)$. We prove that $\hat{\theta}_{n,k(n)}$ is asymptotically normal if $k(n)$ is chosen of order $n^{2+\alpha}$ for some $\alpha > 0$; and that it is asymptotically efficient under correct specification of the parametric model. We also investigate the asymptotic behaviour of $\hat{\theta}_{n,k(n)}$ under convergent sequences of parameters and again obtain asymptotic efficiency. Finally, we show that the auxiliary maximum likelihood estimators are solutions of finite-dimensional systems of equations, thereby suggesting how to compute them.

Zusammenfassung

Es seien $X_1, \ldots, X_n$ unabhängige, identisch verteilte Zufallsvariablen, und es liege ein (möglicherweise misspezifiziertes) parametrisches Modell mit Parameterraum $\Theta \subseteq \mathbb{R}^m$ vor. In dieser Arbeit erweitern wir die 'Indirect Inference'-Methode durch Verwendung nicht-parametrischer Auxiliarmodelle. Dazu definieren wir den sogenannten auxiliaren Maximum-Likelihood-Schätzer $\hat{p}_n$ als Maximierer der Likelihood-Funktion über das gewählte Auxiliarmodell. Für Daten $X_1(\theta), \ldots, X_{k(n)}(\theta)$, die nach dem gegebenen Modell simuliert werden, definieren wir analog $\tilde{p}_{k(n)}(\theta)$ als Maximierer der entsprechenden Likelihood-Funktion. Wir zeigen zunächst, dass $\hat{p}_n$ und $\tilde{p}_{k(n)}(\theta)$ existieren und eindeutig sind, und definieren dann $\hat{\theta}_{n,k(n)}$ als Minimierer einer geeignet gewichteten L^2-Distanz zwischen $\hat{p}_n$ und $\tilde{p}_k(\theta)$. Wir beweisen, dass $\hat{\theta}_{n,k(n)}$ asymptotisch normalverteilt ist, so $k(n)$ von der Ordnung $n^{2+\alpha}$ mit $\alpha > 0$ gewählt wird. Weiters zeigen wir, dass $\hat{\theta}_{n,k(n)}$ asymptotisch effizient ist, wenn das gegebene Modell korrekt spezifiert ist. Darüber hinaus untersuchen wir das asymptotische Verhalten von $\hat{\theta}_{n,k(n)}$ unter beliebigen konvergenten Parameterfolgen und erhalten wiederum asymptotische Effizienz. Schließlich zeigen wir, dass die auxiliaren Maximum-Likelihood-Schätzer Lösungen von endlich-dimensionalen Gleichungssystemen sind, was deren Berechnung ermöglicht.

LEBENSLAUF

PERSÖNLICHE DATEN

Name	**GACH, Florian**
Adresse	Windmühlgasse 20/7/75 A-1060 Wien
Telefon	(43) 699 10 42 62 62
Email	florian.gach@univie.ac.at
Homepage	http://homepage.univie.ac.at/florian.gach
Staatsangehörigkeit	Österreich
Geburtsdatum	16. August 1980

BERUFLICHE TÄTIGKEITEN

Seit Februar 2007	Universitätsassistent in Ausbildung an der Universität Wien
Februar 2006 - Februar 2007	Wissenschaftlicher Mitarbeiter im Projekt *Mathematik und Evolution – Mathematische und statistische Analyse von ökologischer und genetischer Vielfalt*
2005 - 2007	Externer Lehrender an der Universität Wien
2004 - 2005	Tutor am Institut für Mathematik der Universität Wien

AUSBILDUNG

Juni 2008	Förderpreis der Österreichischen Statistischen Gesellschaft
April 2007	Mag. rer. soc. oec.
2006-2007	Magisterstudium Statistik Magisterarbeit: *The coalescent: A stochastic approach to evolution*
März 2006	Bakk. rer. soc. oec.
2004-2006	Bakkalaureatsstudium Statistik
Februar 2004	Mag. rer. nat.
1998-2004	Magisterstudium Mathematik Diplomarbeit: *Topological versus Bornological Concepts in Infinite Dimensions*
Juli-August 1998	Sommersprachschule am *Centre International d'Etudes Françaises*, Universität von Burgund, Frankreich
Juni 1998	Matura Fachbereichsarbeit in Physik: *Auflösungsvermögen optischer Instrumente*
1990-1998	Bundesgymnasium Fichtnergasse Fichtnergasse 15 A-1130 Wien

KENNTNISSE UND FÄHIGKEITEN

Sprachen	Fließende Sprachfertigkeit in Deutsch (Muttersprache), Englisch und Französisch
Programmieren	Objekt-orientierte Programmierung: C++, C#

WISSENSCHAFTLICHE PUBLIKATIONEN

- Futschik, A. and Gach, F. (2008). On the Inadmissibility of Watterson's Estimate. *Theor. Popul. Biol.* **73** 212-221.
- Gach, F. (2006). A note on closed graph theorems. *Acta Math. Univ. Comenianae* **75** 209-218.

WEITERE INTERESSEN

- Schlagzeug spielen
- Fotografieren