Non-Parametric Maximum Likelihood Density Estimation and Simulation-Based Minimum Distance Estimators

Gach, Florian and Pötscher, Benedikt M.

University of Vienna

16 December 2010

Non-Parametric Maximum Likelihood Density Estimation and Simulation-Based Minimum Distance Estimators*

Florian Gach and Benedikt M. Pötscher
Department of Statistics, University of Vienna

December 16, 2010

Abstract

Indirect inference estimators (i.e., simulation-based minimum distance estimators) in a parametric model that are based on auxiliary non-parametric maximum likelihood density estimators are shown to be asymptotically normal. If the parametric model is correctly specified, it is furthermore shown that the asymptotic variance-covariance matrix equals the Cramér-Rao bound. These results are based on uniform-in-parameters convergence rates and a uniform-in-parameters Donsker-type theorem for non-parametric maximum likelihood density estimators.

1 Introduction

Suppose $X_1, \ldots, X_n$ are independent and identically distributed (i.i.d.) random variables with law $\mathbb{P}$. Furthermore, we are given a *parametric model* $\mathcal{P}_\Theta = \{p_\theta : \theta \in \Theta\}$ of probability density functions p_θ and $\Theta \subseteq \mathbb{R}^m$. Assume for the moment that $\mathcal{P}_\Theta$ is correctly specified and identifiable in the sense that there is a unique $\theta_0 \in \Theta$ such that p_{θ_0} is a density of $\mathbb{P}$. A standard method of estimation of θ is then the maximum likelihood method, which under appropriate regularity conditions is known to lead to *asymptotically efficient* estimators. However, in a number of models, e.g., in econometrics and biostatistics, the maximum likelihood method may not be feasible as no closed form expressions for the densities p_θ, and thus for the likelihood, are available. For example, the data may be modeled by an equation of the form $X_i = g(\varepsilon_i, \theta_0)$ where ε_i are i.i.d. with a known distribution but the implied parametric densities are not analytically tractable because g is complicated or ε_i is high-dimensional. A similar problem naturally also occurs in the estimation of dynamic nonlinear models; see Smith (1993), Gouriéroux, Monfort and Renault (1993), Gallant and Tauchen (1996), Gouriéroux and Monfort (1996), and Gallant and Long (1997) for several concrete examples. This has led to the development of alternative estimation methods like the so-called *indirect inference method*, see the just mentioned references as well as Jiang and Turnbull (2004). Ideally, these estimation methods should also be asymptotically efficient. In our context these methods can be described in a nutshell as follows:

1. Simulate a random sample $X_1(\theta), ..., X_k(\theta)$ of size k from the density p_θ for $\theta \in \Theta$. [This is often possible in the examples alluded to above, e.g., by perusing the equations defining the model. Note that then only the disturbances $\varepsilon_1, \ldots, \varepsilon_k$ have to be simulated once and $X_i(\theta)$ can be computed from $g(\varepsilon_i, \theta)$ for any given θ.]

---

*This paper is based on the doctoral thesis of the first author written under the supervision of the second author.

2. Based on the simulated sample *as well as* on the true data, compute auxiliary estimators $\tilde{p}_k(\theta)$ and $\hat{p}_n$, respectively, in a not necessarily correctly-specified but *numerically tractable auxiliary* model $\mathcal{M}^{aux}$. [For example, by maximum likelihood if $\mathcal{M}^{aux}$ is finite-dimensional.]

3. With a suitable choice of a distance χ then estimate θ_0 by minimizing over Θ the objective function

$$\mathbb{Q}_{n,k}(\theta) := \chi(\hat{p}_n, \tilde{p}_k(\theta)). \tag{1}$$

In most of the indirect inference literature, the auxiliary model $\mathcal{M}^{aux}$ is assumed to be finite-dimensional indexed by a vector $\beta \in B \subseteq \mathbb{R}^l$, say, and one then in fact minimizes a distance between $\hat{\beta}_n$, the maximum likelihood estimator in the auxiliary model computed from the original data, and $\tilde{\beta}_k(\theta)$, the maximum likelihood estimator in the auxiliary model computed from the simulated sample $X_1(\theta), ..., X_k(\theta)$. The resulting indirect inference estimator can be shown to be consistent and asymptotically normal (under standard regularity conditions, see Gouriéroux and Monfort (1996)). However, the indirect inference estimator is asymptotically efficient (in the sense of having the Cramér-Rao bound as its asymptotic variance-covariance matrix) only if $\mathcal{M}^{aux}$ happens to be *correctly specified*. This assumption is certainly restrictive and often unnatural if $\mathcal{M}^{aux}$ is of fixed finite dimension. Therefore Gallant and Long (1997) suggested that choosing $\mathcal{M}^{aux}$ with dimension increasing in sample size should result in estimators that are asymptotically efficient, the idea being that this essentially amounts to choosing an infinite-dimensional auxiliary model $\mathcal{M}^{aux}$, for which the assumption of correct specification is much less restrictive.

In the present paper we show in some generality that the suggestion in Gallant and Long (1997) is indeed correct, namely that the indirect inference estimator for θ is asymptotically normal with the Cramér-Rao bound as its asymptotic variance-covariance matrix if the auxiliary estimators $\tilde{p}_k(\theta)$ and $\hat{p}_n$ in Step 2 are chosen to be non-parametric maximum likelihood (NPML) estimators obtained from optimizing the non-parametric likelihood over suitable bounded subsets of a Sobolev-space and if the size k of the simulated sample is of order larger than n^2. Furthermore, we show that asymptotic normality persist even if the originally given model P_Θ is misspecified.

We now comment on some related literature in the area of indirect inference: Gallant and Long (1997) studies the case where the density estimators are based on non-parametric maximum likelihood estimators over sieves spanned by Hermite-polynomials, but their limiting result is only informative if the sieve dimension stays bounded (so that efficiency of the estimator is only established if the true density is a *finite* linear combination of Hermite-polynomials) bringing one back into the realm of finite-dimensional auxiliary models. Fermanian and Salanié (2004) propose a different procedure and establish asymptotic efficiency of their estimators under several high-level conditions, which, as they admit themselves, are very stringent. For example, even in the simplest model they consider, they need to have simulations of order $k \sim n^6$. Nickl and Pötscher (2010) consider the case where $\tilde{p}_k(\theta)$ and $\hat{p}_n$ are spline projection estimators and establish asymptotic normality and asymptotic efficiency if the parametric model $\mathcal{P}_\Theta$ is correctly specified. There are also some other related recent papers on this topic, Altissimo and Mele (2009) and Carrasco, Chernov, Florens, and Ghysels (2007), whose proofs, however, we were not able to follow.

In the present paper we shall use for χ the Fisher-metric, hence the objective function defining the indirect inference estimator will be given by

$$\mathbb{Q}_{n,k}(\theta) = \int (\hat{p}_n - \tilde{p}_k(\theta))^2 \hat{p}_n^{-1}.$$

It transpires that the indirect inference estimators considered in the present paper can be viewed as *minimum distance estimators* with the important (and nontrivial) modification that p_θ has been replaced by an estimator $\tilde{p}_k(\theta)$ based on the simulated data. In that sense our results can be viewed as an extension of Beran's (1977) asymptotic efficiency result for classical minimum distance estimators to the case of *simulation-based minimum distance estimators*, the simulation step introducing considerable additional complexity into the proofs.

In order to establish the above mentioned results for the indirect inference estimator a careful study of several aspects of the NPML-estimators $\tilde{p}_k(\theta)$ and $\hat{p}_n$ is required. In particular, it turns out to be crucial to establish the weak convergence of the stochastic process

$$(\theta, f) \mapsto \sqrt{k} \int (\tilde{p}_k(\theta) - p_\theta) f \tag{2}$$

to a Gaussian process in $\ell^\infty(\Theta \times \mathcal{F})$ where $\mathcal{F}$ is an appropriate class of functions. This result can be seen to imply a uniform-in-θ version of a Donsker-type result for NPML-estimators obtained recently by Nickl (2007). In the course of establishing this weak convergence result it is also necessary to derive rates of convergence for

$$\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2} \tag{3}$$

where the norm is a suitable Sobolev-norm.

The outline of the paper is as follows: After some preliminaries in Section 2, we introduce the model and assumptions in Section 3. In Section 4.1 we derive existence and uniqueness of the NPML-estimator while rates of convergence as indicated in (3) are given in Section 4.2. Donsker-type theorems like (2) are the subject of Section 4.3. In contrast to Nickl (2007), we avoid an assumption that requires all densities to be bounded away from zero in our results as far as possible. Section 5 introduces simulation-based minimum distance estimators (i.e., indirect inference estimators) based on auxiliary NPML-estimators and establishes asymptotic normality of these estimators even if the originally given parametric model $\mathcal{P}_\Theta$ is misspecified. If $\mathcal{P}_\Theta$ is correctly specified, it is furthermore shown that the estimator is asymptotically efficient in the sense that its asymptotic variance-covariance matrix equals the Cramér-Rao bound. Some proofs and technical results are collected in the appendices.

2 Preliminaries and Notation

For Λ a non-empty set and f a real-valued function on Λ, define $\|f\|_\Lambda = \sup_{x \in \Lambda} |f(x)|$ and let $\ell^\infty(\Lambda)$ denote the Banach space of all bounded real-valued functions on Λ, equipped with the sup-norm $\|\cdot\|_\Lambda$. If $\mathcal{D}$ is a (non-empty) subset of $\ell^\infty(\Lambda)$ we shall write $(\mathcal{D}, \|\cdot\|_\Lambda)$ to denote the metric space $\mathcal{D}$ with the induced metric $\|f - g\|_\Lambda$. For $(\Lambda, \mathcal{A})$ a (non-empty) measurable space, let $\mathcal{L}^0(\Lambda, \mathcal{A})$ denote the vector space of all $\mathcal{A}$-measurable real-valued functions on Λ and define the Banach space $\mathsf{L}^\infty(\Lambda, \mathcal{A}) = \mathcal{L}^0(\Lambda, \mathcal{A}) \cap \ell^\infty(\Lambda)$, again equipped with the sup-norm. For $f \in \mathcal{L}^0(\Lambda, \mathcal{A})$ and μ a non-negative measure on $(\Lambda, \mathcal{A})$, define $\|f\|_{2,\mu} = \left[\int_\Lambda f^2 d\mu \right]^{1/2}$ and set $\mathcal{L}^2(\Lambda, \mathcal{A}, \mu) = \left\{ f \in \mathcal{L}^0(\Lambda, \mathcal{A}) : \|f\|_{2,\mu} < \infty \right\}$. For the measure space $(\Omega, \mathcal{B}(\Omega), \lambda)$, where Ω is a (non-empty) measurable subset of the real line $\mathbb{R}$ with associated Borel σ-field $\mathcal{B}(\Omega)$ and where λ is Lebesgue measure, we shall simplify notation and write $\mathcal{L}^0(\Omega)$, $\mathcal{L}^2(\Omega)$, $\mathsf{L}^\infty(\Omega)$, and $\|\cdot\|_2$ for $\mathcal{L}^0(\Omega, \mathcal{B}(\Omega))$, $\mathcal{L}^2(\Omega, \mathcal{B}(\Omega), \lambda)$, $\mathsf{L}^\infty(\Omega, \mathcal{B}(\Omega))$, and $\|\cdot\|_{2,\lambda}$, respectively. Furthermore, we shall write a.e. instead of λ-a.e. For any (non-empty) metric space (T, d), we denote by $\mathcal{B}(T, d)$, or simply $\mathcal{B}(T)$, its Borel σ-field and by $\mathsf{C}(T, d)$, or simply $\mathsf{C}(T)$, the Banach space of all bounded, d-continuous real-valued functions on T, equipped with the sup-norm.

We shall denote by $\|\cdot\|$ the 2-norm on Euclidean space. For two real-valued functions f and g on $(0, \infty)$, we shall 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 $\varepsilon > 0$. It will also prove useful to define $\log \infty = \infty$ and $\log 0 = -\infty$, thus making the logarithm a continuous function from $[0, \infty]$ to $[-\infty, \infty]$.

Let $(\Lambda_0, \mathcal{A}_0, P_0)$, $(\Lambda_n, \mathcal{A}_n, P_n)$, $n \geq 1$, be probability spaces. Suppose $Y_0 : \Lambda_0 \to T$ is an $\mathcal{A}_0$-$\mathcal{B}(T, d)$-measurable mapping and $Y_n : \Lambda_n \to T$ are (not necessarily measurable) mappings, where (T, d) is a metric space. We say that Y_n converges weakly to Y_0 in (T, d), denoted by $Y_n \rightsquigarrow Y_0$, if the outer integrals $\int_{\Lambda_n}^* g(Y_n) dP_n$ converge to $\int_{\Lambda_0} g(Y_0) dP_0$ for every $g \in \mathsf{C}(T, d)$; furthermore, Y_n is said to converge weakly to a Borel probability measure L on $(T, \mathcal{B}(T, d))$, denoted by $Y_n \rightsquigarrow L$, if $\int_{\Lambda_n}^* g(Y_n) dP_n$ converges to $\int_T g \, dL$ for every $g \in \mathsf{C}(T, d)$. We say that Y_n converges to $\tau \in T$ in outer P_n-probability if $P_n^*(d(Y_n, \tau) > \varepsilon)$ converges to 0 for all $\varepsilon > 0$. If Y_n are real-valued and r_n is 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 $(\Lambda_n, \mathcal{A}_n, P_n)$ are the n-fold products of a single probability space $(\Lambda, \mathcal{A}, P)$, that is, $(\Lambda_n, \mathcal{A}_n, P_n) = (\Lambda^n, \mathcal{A}^n, P^n)$, 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 $Y_n = O_{P^n}^*(r_n)$.

2.1 Hölder and Sobolev Spaces

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

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

Here $f^{(\alpha)}$ denotes the classical derivative of f of order α, and $\lfloor s \rfloor$ denotes the integer part of s. For any non-integer $s > 0$, define the Hölder space $\mathsf{C}^s(\Omega)$ as 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 $f^{(s)}$ is uniformly continuous. Note that $\mathsf{C}^0(\Omega)$ thus is the space of bounded and uniformly continuous functions on Ω.

For Ω and s as above and functions $f, g \in \mathcal{L}^2(\Omega)$, let

$$\langle f | g \rangle_{s,2} = \begin{cases} \sum_{0 \leq \alpha \leq \lfloor s \rfloor} \langle f^{(\alpha)w} | g^{(\alpha)w} \rangle_2 \\ \quad + \int_\Omega \int_\Omega \frac{(f^{(\lfloor s \rfloor)w}(x) - f^{(\lfloor s \rfloor)w}(y))(g^{(\lfloor s \rfloor)w}(x) - g^{(\lfloor s \rfloor)w}(y))}{|x-y|^{1+2(s-\lfloor s \rfloor)}} d\lambda(x) d\lambda(y) \\ \hspace{7cm} \text{if } s \text{ is non-integer,} \\ \\ \sum_{0 \leq \alpha \leq s} \langle f^{(\alpha)w} | g^{(\alpha)w} \rangle_2 \\ \hspace{7cm} \text{otherwise,} \end{cases}$$

and set $\|f\|_{s,2} = \sqrt{\langle f | f \rangle_{s,2}}$. Here, $f^{(\alpha)w}$ denotes the weak derivative of f of order α, and $\langle \cdot | \cdot \rangle_2$ is the usual inner product on $\mathcal{L}^2(\Omega)$. Define $\mathcal{W}_2^s(\Omega)$ as the space of all $f \in \mathcal{L}^2(\Omega)$ 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 $s > 1/2$ and such Ω, 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 $\mathcal{U}_{s,B}$, and its translates $g + \mathcal{U}_{s,B}$ by $\mathcal{U}_{s,B}(g)$. The next proposition collects some properties of Sobolev spaces; see Appendix A for a proof.

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

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

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

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

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

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

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

(c) If $0 \le r < s$, 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 $\mathsf{W}_2^r(\Omega)$.

(d) If $\mathcal{F}$ is a (non-empty) bounded subset of some Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$ such that $\inf_{x\in\Omega, f\in\mathcal{F}} |f(x)| > 0$ holds, then $\{1/f : f \in \mathcal{F}\}$ is also a bounded subset of $\mathsf{W}_2^s(\Omega)$.

2.2 Covering Numbers and Metric Entropy

Let (T, d) be a metric space. Let $0 < \varepsilon < \infty$ and let X be a (non-empty) totally bounded subset of T. Then we denote by $N(\varepsilon, X, T, d)$ the covering number of X, i.e., the minimal number of closed balls in T of radius ε needed to cover X; we define the metric entropy of X as

$$H(\varepsilon, X, T, d) = \log N(\varepsilon, X, T, d).$$

If T is a normed space with norm $\|\cdot\|$, we shall write in abuse of notation $N(\varepsilon, X, T, \|\cdot\|)$ and similarly for the metric entropy.

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

$$[l, u] = \{f \in \mathcal{L}^0(\Lambda, \mathcal{A}) : l(x) \le f(x) \le u(x) \text{ for all } x \in \Lambda\}$$

is called a bracket and $\|u - l\|_{2,\mu}$ its $\mathcal{L}^2(\mu)$-bracketing size. For $0 < \varepsilon < \infty$ and $\mathcal{F}$ a (non-empty) subset of $\mathcal{L}^0(\Lambda, \mathcal{A})$, we define $N_{[\ ]}(\varepsilon, \mathcal{F}, \|\cdot\|_{2,\mu})$ to be the minimal number of brackets of $\mathcal{L}^2(\mu)$-bracketing size less than or equal to 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

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

Furthermore, for $0 < \eta < \infty$ the $\mathcal{L}^2(\mu)$-bracketing metric integral $I_{[\ ]}(\eta, \mathcal{F}, \|\cdot\|_{2,\mu})$ of $\mathcal{F}$ is given by

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

3 The Framework and Assumptions

From now on let Ω be a non-empty bounded, open interval in $\mathbb{R}$. We consider i.i.d. random variables $(X_i)_{i\in\mathbb{N}}$ that take their values in $(\Omega, \mathcal{B}(\Omega))$ and have common law $\mathbb{P}$, with $X_1, \ldots, X_n$ representing the data at sample size n. Furthermore, let Θ be a (non-empty) compact subset of $\mathbb{R}^m$ and let $\mathcal{P}_\Theta = \{p_\theta : \theta \in \Theta\}$ be a parametric family of probability density functions p_θ on Ω. The law $\mathbb{P}$ may or may not correspond to a density in $\mathcal{P}_\Theta$. We assume that there is a way of simulating synthetic data according to the densities in the class $\mathcal{P}_\Theta$ in the following

sense: There is a probability space $(V, \mathcal{V}, \mu)$ and a function $\rho : V \times \Theta \to \Omega$, which is $\mathcal{V}$-$\mathcal{B}(\Omega)$-measurable in its first argument, such that for every $\theta \in \Theta$ the law of $\rho(\cdot, \theta)$ under μ has density p_θ. Consequently, if $(V_i)_{i \in \mathbb{N}}$ is a sequence of i.i.d. random variables with values in $(V, \mathcal{V})$ and law μ, then $X_i(\theta) = \rho(V_i, \theta)$ is an i.i.d. sequence with law having density p_θ, simultaneously so for all $\theta \in \Theta$. We shall also always assume that the process $(V_i)_{i \in \mathbb{N}}$ is independent of $(X_i)_{i \in \mathbb{N}}$. [As indicated in the Introduction, the simulation mechanism ρ may derive form an underlying equation model, but it may also arise in some other way.] In the application to indirect inference in Section 5 we shall estimate θ by matching a non-parametric estimator for (the density of) $\mathbb{P}$ obtained from the data $X_1, \ldots, X_n$ with a non-parametric estimator for p_θ obtained from the synthetic data $X_1(\theta), \ldots, X_k(\theta)$. We stress that construction of the synthetic data requires only one simulation, and not a separate simulation for every θ. For convenience we shall from now on assume that the random variables X_i and V_i are the respective coordinate projections on the measurable space $(\Omega^{\mathbb{N}} \times V^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}} \otimes \mathcal{V}^{\mathbb{N}})$ equipped with the product measure $\mathrm{Pr} := \mathbb{P}^{\mathbb{N}} \otimes \mu^{\mathbb{N}}$. We note, however, that all results of the paper hold also without this assumption; see Remark 19. Furthermore, 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.

The density estimators we shall consider will be NPML-estimators over non-parametric models (called auxiliary models in Section 5) of the form

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

where $t > 1/2$, $0 \leq \zeta < \infty$, and $0 < D < \infty$. Some important properties of $\mathcal{P}(t, \zeta, D)$ that will be used repeatedly are summarized in the subsequent propositions, the proofs of which can be found in Appendix A.

Proposition 2 *Suppose $t > 1/2$, $0 \leq \zeta < \infty$, and $0 < D < \infty$.*

(a) The following statements are equivalent: (i) $\zeta \leq \lambda(\Omega)^{-1} \leq D^2$; (ii) the constant density $\lambda(\Omega)^{-1}$ belongs to $\mathcal{P}(t, \zeta, D)$; (iii) $\mathcal{P}(t, \zeta, D)$ is non-empty.

(b) Suppose $\zeta \leq \lambda(\Omega)^{-1} \leq D^2$. Then the following statements are equivalent: (i) $\zeta = \lambda(\Omega)^{-1}$ or $\lambda(\Omega)^{-1} = D^2$; (ii) the constant density $\lambda(\Omega)^{-1}$ is the only element of $\mathcal{P}(t, \zeta, D)$; (iii) $\mathcal{P}(t, \zeta, D)$ is a singleton.

(c) Suppose $\zeta \leq \lambda(\Omega)^{-1} \leq D^2$. Then $\mathcal{P}(t, \zeta, D)$ 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 $1/2 < s < t$.

In the following let H_t denote the closed affine hyperplane given by $\mathsf{H}_t = \left\{ f \in \mathsf{W}_2^t(\Omega) : \int_\Omega f \, d\lambda = 1 \right\}$ endowed with the relative topology it inherits from $\mathsf{W}_2^t(\Omega)$. Note that $\mathcal{P}(t, \zeta, D) \subseteq \mathsf{H}_t$ holds.

Proposition 3 [1] *Suppose $t > 1/2$ and $0 \leq \zeta \leq \lambda(\Omega)^{-1} \leq D^2 < \infty$.*

(a) An element $p \in \mathcal{P}(t, \zeta, D)$ is an interior point of $\mathcal{P}(t, \zeta, D)$ relative to H_t if and only if (i) $\|p\|_{t,2} < D$ and (ii) $\inf_{x \in \Omega} p(x) > \zeta$ hold.

(b) A (non-empty) subset $\mathcal{P}'$ of $\mathcal{P}(t, \zeta, D)$ is uniformly interior to $\mathcal{P}(t, \zeta, D)$ relative to H_t (meaning that there exists a $\delta > 0$ such that for every $p \in \mathcal{P}'$ the set $\mathcal{U}_{t,\delta}(p) \cap \mathsf{H}_t \subseteq \mathcal{P}(t, \zeta, D)$) if and only if (i) $\sup_{p \in \mathcal{P}'} \|p\|_{t,2} < D$ and (ii) $\inf_{x \in \Omega, p \in \mathcal{P}'} p(x) > \zeta$ hold.

(c) Suppose $\zeta < \lambda(\Omega)^{-1} < D^2$ holds. Then the constant density $\lambda(\Omega)^{-1}$ is interior to $\mathcal{P}(t, \zeta, D)$ relative to H_t. Moreover, the interior of $\mathcal{P}(t, \zeta, D)$ relative to H_t is dense in $\mathcal{P}(t, \zeta, D)$ (w.r.t. the $\mathsf{W}_2^t(\Omega)$-topology).

---

[1] An obvious extension of Theorem V.2.1 in Dunford and Schwartz (1966) to affine spaces shows that in our setting the notion of an element being interior relative to $\mathcal{H}$ coincides with the notion of internality of that element (relative to $\mathcal{H}$).

We emphasis that for the rest of the paper t, ζ, and D will be treated as fixed (although at arbitrary values) satisfying the constraints $t > 1/2$ and $0 \leq \zeta < \lambda(\Omega)^{-1} < D^2 < \infty$ (thus excluding only the trivial cases where $\mathcal{P}(t, \zeta, D)$ is empty or the singleton $\{\lambda(\Omega)^{-1}\}$). Many results will hold under the natural condition $\zeta \geq 0$, but for some results we shall have to assume the stronger requirement $\zeta > 0$. In that context we note that if D^2 is sufficiently close to $\lambda(\Omega)^{-1}$, then $\mathcal{P}(t, 0, D)$ coincides with $\mathcal{P}(t, \zeta, D)$ for sufficiently small $\zeta > 0$, cf. Remark 29 in Appendix A.

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 (b) of Proposition 1. We further note the fact that in $\mathcal{P}(t, \zeta, D)$ pointwise convergence is equivalent to convergence in all Sobolev norms of order smaller than t, as well as to convergence in the sup-norm, as shown in the next proposition.

Proposition 4 *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 $0 \leq r < t$.*

Apart from the maintained assumptions laid out at the beginning of this section, we will make frequent use of the assumptions listed below. We start with assumptions on the probability measure $\mathbb{P}$ governing the data.

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

In the following we treat the probability density $p_\blacktriangle$ as a *function* from Ω to $\mathbb{R}$, that is, we let $p_\blacktriangle$ denote a fixed representative of the Radon-Nikodym derivative of $\mathbb{P}$ with respect to λ. Recall also that $\mathbb{P}$ need not correspond to an element of $\mathcal{P}_\Theta$, hence $p_\blacktriangle$ need not be a.e. equal to an element of $\mathcal{P}_\Theta$.

Assumption D.1 *Assumption D holds and the density function $p_\blacktriangle$ belongs to $\mathcal{P}(t, \zeta, D)$.*

Assumption D.2 *Assumption D holds and the density function $p_\blacktriangle$ satisfies the strict inequality*

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

Clearly, if $\zeta > 0$, then Assumption D.1 implies Assumption D.2. In light of Proposition 3, the next assumption just states that $p_\blacktriangle$ is an interior point of $\mathcal{P}(t, \zeta, D)$ relative to H_t.

Assumption D.3 *Assumption D.1 holds and the strict inequalities*

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

are satisfied.

We note here, however, that even under Assumption D.3 the NPML-estimator is *never* an interior point of $\mathcal{P}(t, \zeta, D)$ relative to H_t as shown in Section 4; this leads to a number of complications as discussed prior to Lemma 15 in Section 4.3.

Next are assumptions on the class $\mathcal{P}_\Theta$. We will often write $p(x, \theta)$ for $p_\theta(x)$, and we stress that $p(x, \theta)$ is a *function* from $\Omega \times \Theta$ to $\mathbb{R}$.

Assumption P.1 $\mathcal{P}_\Theta \subseteq \mathcal{P}(t, \zeta, D)$.

Assumption P.2 *The strict inequality*

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

holds true.

Clearly, if $\zeta > 0$ then Assumption P.1 implies Assumption P.2.

Assumption P.3 *Assumption P.1 holds and the strict inequalities*

$$\inf_{\Omega \times \Theta} p(x, \theta) > \zeta \quad and \quad \sup_{\theta \in \Theta} \|p_\theta\|_{t,2} < D$$

are satisfied.

Assumption P.3 states that $\mathcal{P}_\Theta$ is uniformly interior to $\mathcal{P}(t, \zeta, D)$ relative to H_t, cf. Proposition 3. If $\mathcal{P}_\Theta$ happens to be a $\|\cdot\|_{t,2}$-compact subset of $\mathcal{P}(t, \zeta, D)$ (which in light of compactness of Θ is, e.g., the case if the map $\theta \to p_\theta$ is $\|\cdot\|_{t,2}$-continuous), Assumption P.3 is clearly equivalent to $\inf_{x \in \Omega} p(x, \theta) > \zeta$ and $\|p_\theta\|_{t,2} < D$ for every $\theta \in \Theta$ (i.e., equivalent to $\mathcal{P}_\Theta$ belonging to the interior of $\mathcal{P}(t, \zeta, D)$ relative to H_t).

Occasionally we shall also need to refer to the following assumption. However, note that Assumption P.1 together with Assumption R.1 below already imply this assumption, cf. Proposition 30 in Appendix A.

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

Remark 5 If Assumption P.1 is satisfied, then in view of Proposition 4 the following are equivalent: (i) Assumption P.4; (ii) $\theta \mapsto p_\theta$ is continuous as a mapping from Θ into the space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_{s,2})$ for every s satisfying $0 \le s < t$; (iii) $\theta \mapsto p_\theta$ is continuous as a mapping from Θ into the space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$.

Next are assumptions on the simulation mechanism $\rho(v, \theta)$. Apart from the already assumed measurability of $\rho(v, \theta)$ in its first argument, we will need assumptions to control its behaviour in the second argument. We note that Assumption R.2 below is weaker than the corresponding Assumption R.2 in Gach (2010), but we have been able to obtain the same conclusions as in Gach (2010) by refining the proofs.

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, $0 < \gamma \le 1$, and some measurable function $R : V \to (0, \infty)$, the simulation mechanism $\rho : V \times \Theta \to \Omega$ satisfies*

$$|\rho(v, \theta') - \rho(v, \theta)| \le R(v) \|\theta' - \theta\|^\gamma$$

for all $v \in V$ and all $\theta, \theta' \in \Theta$, with the function R satisfying $\int_V R^a d\mu < \infty$ for some $a > 0$.

Assumptions on the class $\mathcal{P}_\Theta$ and on the simulation mechanism $\rho(v, \theta)$ are obviously closely related. In principle, the assumptions on $\mathcal{P}_\Theta$ could be substituted for by assumptions on $\rho(v, \theta)$. [Conversely, the existence of a simulation mechanism having certain required properties can in principle be deduced from suitable assumptions on $\mathcal{P}_\Theta$.] However, the interrelation between assumptions on $\mathcal{P}_\Theta$ and on $\rho(v, \theta)$ is complicated and intricate, and hence we prefer to work with the two sets of assumptions as given above. For some results concerning the relationship between these two sets of assumptions see Proposition 30 in Appendix A.

4 Non-Parametric Maximum Likelihood Estimators

We now introduce NPML-estimators, called auxiliary estimators in Section 5. Define the (non-parametric) log-likelihood function based on the given data $X_1, \ldots, X_n$ as

$$L_n(p) := L_n(p; X_1, \ldots, X_n) = \frac{1}{n} \sum_{i=1}^{n} \log p(X_i)$$

for $p \in \mathcal{P}(t, \zeta, D)$, and based on the simulated data $X_1(\theta) = \rho(V_1, \theta), \ldots, X_k(\theta) = \rho(V_k, \theta)$ as

$$L_k(\theta, p) := L_k(\theta, p; V_1, \ldots, V_k) = \frac{1}{k} \sum_{i=1}^{k} \log p(\rho(V_i, \theta))$$

for $p \in \mathcal{P}(t, \zeta, D)$ and $\theta \in \Theta$. Note that $L_k(\theta, p) = L_k(p; X_1(\theta), \ldots, X_k(\theta)) = k^{-1} \sum_{i=1}^{k} \log p(X_i(\theta))$ holds. In view of our convention for the logarithm, both functions $L_n(f)$ and $L_k(\theta, f)$ are in fact well-defined and take their values in $[-\infty, \infty)$ for any non-negative real-valued function f on Ω.

An NPML-estimator for given $X_1, \ldots, X_n$ is defined as an element $\hat{p}_n(\cdot) := \hat{p}_n(\cdot; X_1, \ldots, X_n)$ of $\mathcal{P}(t, \zeta, D)$ satisfying

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

Similarly, an NPML-estimator for given $X_1(\theta), \ldots, X_k(\theta)$ is an element $\tilde{p}_k(\theta)(\cdot) := \tilde{p}_k(\theta)(\cdot; V_1, \ldots, V_k)$ of $\mathcal{P}(t, \zeta, D)$ satisfying

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

Clearly we have

$$\tilde{p}_k(\theta)(\cdot; V_1, \ldots, V_k) = \hat{p}_k(\cdot; X_1(\theta), \ldots, X_k(\theta)). \tag{4}$$

In this section we investigate existence, uniqueness, consistency, rates of convergence, and uniform central limit theorems for NPML-estimators. The results obtained here go beyond Nickl (2007) in three respects: First, we show not only existence but also *uniqueness* of the NPML-estimators. Second, we allow for non-parametric models $\mathcal{P}(t, \zeta, D)$ where the lower bound for the densities, i.e., ζ, can be equal to 0 and extend the consistency and rate results for the NPML-estimator w.r.t. the Sobolev-norms $\| \cdot \|_{s,2}$ with $s < t$ in Nickl (2007) to this case. We furthermore also establish *inconsistency* of the NPML-estimator in the $\| \cdot \|_{t,2}$-norm. Third, we prove that the consistency and rate results in Nickl (2007) for $\hat{p}_n$ hold for the NPML-estimators $\tilde{p}_k(\theta)$ even *uniformly* over the parameter space Θ (provided that $\zeta > 0$). Finally, we prove a *uniform* Donsker-type theorem which extends Theorem 3 in Nickl (2007) and shows that, for appropriate classes $\mathcal{F}$, the stochastic process $(\theta, f) \mapsto \sqrt{k} \int_{\Omega} (\tilde{p}_k(\theta) - p_\theta) f \, d\lambda$ converges weakly in $\ell^\infty(\Theta \times \mathcal{F})$ to a Gaussian process.

4.1 Existence, Uniqueness, and Consistency of NPML-Estimators

In the following theorem we show that the NPML-estimators defined above exist, are unique, and are measurable (cf. also Lemma 36 in Appendix D).

Theorem 6 *(a) There exists a unique $\hat{p}_n \in \mathcal{P}(t, \zeta, D)$ such that*

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

9

holds. The resulting mapping $\hat{p}_n : \Omega^n \to \mathcal{P}(t, \zeta, D)$ is measurable with respect to the σ-fields $\mathcal{B}(\Omega)^n$ and $\mathcal{B}(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$. Moreover, $\hat{p}_n$ always satisfies $\|\hat{p}_n\|_{t,2} = D$.

(b) 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. The resulting mapping $\tilde{p}_k(\theta) : V^k \to \mathcal{P}(t, \zeta, D)$ is measurable with respect to the σ-fields $\mathcal{V}^k$ and $\mathcal{B}(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$. Moreover, $\tilde{p}_k(\theta)$ always satisfies $\|\tilde{p}_k(\theta)\|_{t,2} = D$. Furthermore, if Assumption R.1 is satisfied, then, for arbitrary fixed values of the underlying simulated variables $V_1, \ldots, V_k$, $\theta \mapsto \tilde{p}_k(\theta)$ is continuous when viewed as a mapping from Θ into the space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$.

Proof. (a) Let $x_1, \ldots, x_n$ be given points in Ω. The existence of a maximizer of $L_n(p) = L_n(p; x_1, \ldots, x_n)$ follows from the fact that L_n is continuous on the compact space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$ by Part (b1) of Proposition 31 in Appendix B with $\mathcal{F} = \mathcal{P}(t, \zeta, D)$ and by Proposition 2. We next establish uniqueness: Denote by S the set of all $p \in \mathcal{P}(t, \zeta, D)$ that maximize L_n, and note that S is non-empty as just shown. Since L_n is a concave function on the convex set $\mathcal{P}(t, \zeta, D)$ with values in $[-\infty, \infty)$, a standard argument shows that S is convex. If S is a subset of the Sobolev *sphere* of radius D we are done, as then S must be a singleton since the Sobolev norm $\|\cdot\|_{t,2}$, being a Hilbert norm, is strictly convex. Suppose now S is not a subset of the Sobolev sphere of radius D and let $p \in S$ with $\|p\|_{t,2} < D$. Then there is some $z \in \Omega$ with $p(z) > \zeta$ since the maintained assumption $\zeta < \lambda^{-1}(\Omega)$ implies that $\zeta \notin \mathcal{P}(t, \zeta, D)$. 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 $q(x_i) > p(x_i)$ whenever $x_i = x_1$ and q coincides with p on the remaining (if any) observations x_j with $x_j \neq x_1$. This will contradict the maximizing property of p (noting that the case $L_n(q) = L_n(p) = -\infty$ is impossible in view of $\lambda(\Omega)^{-1} \in \mathcal{P}(t, \zeta, D)$ and $L_n(p) \geq L_n(\lambda(\Omega)^{-1}) > -\infty$). The existence of such a q can be seen as follows: Choose $\varepsilon > 0$ such that $I := [z - 2\varepsilon, z + 2\varepsilon]$, $\bar{U} := [x_1 - 2\varepsilon, x_1 + 2\varepsilon]$, and $\{x_j : x_j \neq x_1\}$ 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 $U := (x_1 - 2\varepsilon, x_1 + 2\varepsilon)$, there is a compactly supported 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 $[-1, 1]$, integrates to 0, and is contained in $\mathsf{W}_2^t(\Omega)$ since it is C^∞ and has compact support in Ω. Since $\|p\|_{t,2} < D$ and $\inf_{x \in I} p(x) > \zeta$, we can find a scalar $\beta > 0$ such that $\|\beta g\|_{t,2} \leq D - \|p\|_{t,2}$ and $\beta \leq \inf_{x \in I} p(x) - \zeta$. Let $q = p + \beta g$ and observe that $\|q\|_{t,2} \leq \|p\|_{t,2} + \|\beta g\|_{t,2} \leq D$. Further, $q(x) \geq \zeta$ for every $x \in \Omega$, which can be seen as follows: For $x \in \Omega \setminus I$ we have that $g(x) \geq 0$, and hence $q(x) \geq p(x) \geq \zeta$. If $x \in I$, then $q(x) \geq p(x) - \beta \geq p(x) - \inf_{x \in I} p(x) + \zeta \geq \zeta$, where the first inequality holds because $g(x) \geq -1$ for every $x \in \Omega$, the second inequality holds by the choice of β, and the third one does so since $x \in I$ and therefore $p(x) - \inf_{x \in I} p(x) \geq 0$. It follows that $q \in \mathcal{P}(t, \zeta, D)$. Since $\beta > 0$ and $g(x_1) = 1$, $q(x_i) > p(x_i)$ whenever $x_i = x_1$. Furthermore, 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 S is a subset of the Sobolev sphere of radius D. We thus have established uniqueness as well as $\|\hat{p}_n\|_{t,2} = D$.

To see that $\hat{p}_n : \Omega^n \to \mathcal{P}(t, \zeta, D)$ is measurable, we apply Lemma A3 in Pötscher & Prucha (1997), making use of Proposition 31(a),(b1) in Appendix B. [Because L_n potentially can attain

10

the value $-\infty$, we apply this lemma to the *real-valued* function $\arctan(L_n)$ rather than to L_n, where we use the usual convention $\arctan(-\infty) = -\pi/2$.]

(b) The same arguments as above establish existence, uniqueness, and measurability of $\tilde{p}_k(\theta)$, as well as $\|\tilde{p}_k(\theta)\|_{t,2} = D$, for any fixed $\theta \in \Theta$. To see that the mapping $\theta \mapsto \tilde{p}_k(\theta)$ is continuous as claimed, apply Lemma 34 in Appendix B 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 (b2) of Proposition 31 in Appendix B with $\mathcal{F} = \mathcal{P}(t,\zeta,D)$. ∎

Remark 7 (i) The mapping $\hat{p}_n : \Omega \times \Omega^n \to \mathbb{R}$ is continuous in the first argument and $\mathcal{B}(\Omega)^n$-measurable in the second argument. Since Ω is separable, $\hat{p}_n$ is consequently jointly measurable. Similarly, the mappings $\tilde{p}_k(\theta) : \Omega \times V^k \to \mathbb{R}$ are jointly measurable for all $\theta \in \Theta$.

(ii) For any $x_1, \ldots, x_n$ in Ω, we have that $\hat{p}_n(x_i) = \hat{p}_n(x_i; x_1, \ldots, x_n) > 0$ for $i = 1, \ldots, n$. This follows from the observation made in the above proof that $L_n(\hat{p}_n) > -\infty$ must hold. By a similar argument we have that $\tilde{p}_k(\theta)(\rho(v_i,\theta)) = \tilde{p}_k(\theta)(\rho(v_i,\theta); v_1, \ldots, v_k) > 0$ for $i = 1, \ldots, k$ and for every $\theta \in \Theta$.

We next turn to consistency of the NPML-estimators. Theorem 6 already shows that $\hat{p}_n$ cannot be consistent in the $\|\cdot\|_{t,2}$-norm as $\|\hat{p}_n\|_{t,2} = D$ always holds and $\mathcal{P}(t,\zeta,D)$ contains densities with $\|\cdot\|_{t,2}$-norm less than D (under our assumptions on ζ and D). A similar remark applies to $\tilde{p}_k(\theta)$. However, this does not preclude consistency of the NPML-estimators in other norms as we show next. To this end define for any *non-negative* measurable function f on Ω and for any $\theta \in \Theta$

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

and

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

provided the respective integral is defined. If $f \in \mathsf{L}^\infty(\Omega)$, then both functions are well-defined and take their values in $[-\infty, \infty)$. We note that 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 simple facts which are proved in Appendix B.

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

(b) If Assumption D.1 is satisfied, then $p_\blacktriangle$ is the unique maximizer of the function $L(\cdot)$ over $\mathcal{P}(t,\zeta,D)$.

(c) If $p_\theta \in \mathcal{P}(t,\zeta,D)$ for a given $\theta \in \Theta$, then p_θ is the unique maximizer of the function $L(\theta, \cdot)$ over $\mathcal{P}(t,\zeta,D)$.

The consistency result is now given below. Under the additional assumption that ζ is positive, Part (a) of the subsequent theorem already follows from Proposition 6 in Nickl (2007).

Theorem 9 *(a) Let Assumption D.1 be satisfied. Then*

$$\lim_{n \to \infty} \|\hat{p}_n - p_\blacktriangle\|_{s,2} = 0 \quad \mathbb{P}\text{-a.s.}$$

for every s, $0 \le s < t$; in particular, $\lim_{n \to \infty} \|\hat{p}_n - p_\blacktriangle\|_\Omega = 0$ $\mathbb{P}$-a.s.

(b) Let $p_\theta \in \mathcal{P}(t,\zeta,D)$ for a given $\theta \in \Theta$. Then, for the given θ,

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

11

for every s, $0 \le s < t$; in particular, $\lim_{k\to\infty} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega = 0$ μ-a.s.

(c) Let Assumptions P.1, P.2, and R.1 be satisfied. Then

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

for every s, $0 \le s < t$; in particular, $\lim_{k\to\infty} \sup_{\theta\in\Theta} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega = 0$ μ-a.s.

Proof. (a) In view of Part (c) of Proposition 1, we may restrict ourselves to the case $1/2 < s < t$. Note that $|L(p_\blacktriangle)| < \infty$ by Assumption D.1 and Part (a) of Lemma 8; also note that the random variables $\log p_\blacktriangle(X_i)$ are $\mathbb{P}$-a.s. real-valued. By Kolmogorov's strong law of large numbers we then have

$$\lim_{n\to\infty} |L_n(p_\blacktriangle) - L(p_\blacktriangle)| = 0 \quad \mathbb{P}\text{-a.s.} \tag{5}$$

Let ε_l be positive real numbers that converge monotonously to 0 as $l \to \infty$. Apply the uniform law of large numbers in Part (d1) of Proposition 31 in Appendix B with $\mathcal{F} = \{p + \varepsilon_l : p \in \mathcal{P}(t, \zeta, D)\}$ to see that

$$\lim_{n\to\infty} \sup_{p\in\mathcal{P}(t,\zeta,D)} |L_n(p + \varepsilon_l) - L(p + \varepsilon_l)| = 0 \quad \mathbb{P}\text{-a.s.} \tag{6}$$

for every $l \in \mathbb{N}$. In the following arguments we fix an arbitrary element of the probability 1 event where the statements in (5) and (6) hold true. We now prove that $\|\hat{p}_n - p_\blacktriangle\|_{s,2}$ converges to 0 by showing that any subsequence $\hat{p}_{n'}$ of $\hat{p}_n$ has another subsequence converging to $p_\blacktriangle$ in the Sobolev norm $\|\cdot\|_{s,2}$. 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. Now use Assumption D.1, the definition of $\hat{p}_{n''}$ as maximizer, and the monotonicity of the logarithm to obtain

$$\begin{aligned} L_{n''}(p_\blacktriangle) &\le L_{n''}(\hat{p}_{n''}) \le L_{n''}(\hat{p}_{n''} + \varepsilon_l) \\ &\le L(\hat{p}_{n''} + \varepsilon_l) + \sup_{p\in\mathcal{P}(t,\zeta,D)} |L_{n''}(p + \varepsilon_l) - L(p + \varepsilon_l)|. \end{aligned} \tag{7}$$

The first term on the r.h.s. of (7) converges to $L(p^* + \varepsilon_l)$ since $\|\hat{p}_{n''} - p^*\|_{s,2}$, and hence also $\|\hat{p}_{n''} - p^*\|_\Omega$, converges to 0 and since $L(\cdot + \varepsilon_l)$ is sup-norm continuous on $\mathcal{P}(t, \zeta, D)$ by Part (c1) of Proposition 31 in Appendix B. The supremum on the r.h.s. of (7) goes to 0 and $L_{n''}(p_\blacktriangle)$ converges to $L(p_\blacktriangle)$ in view of (5) and (6). It follows that

$$L(p_\blacktriangle) \le L(p^* + \varepsilon_l). \tag{8}$$

The sequence of functions $\log(p^* + \varepsilon_l)$ is monotonously non-increasing in l with pointwise limit $\log p^*$, and is bounded above by the integrable function $\log(p^* + \varepsilon_1)$. Using the theorem of monotone convergence, we conclude from (8) that $L(p_\blacktriangle) \le L(p^*)$. Hence, $p^* = p_\blacktriangle$ by Part (b) of Lemma 8.

(b) Follows analogously as Part (a) with $p_\blacktriangle$ replaced by p_θ.

(c) As in the proof of Part (a), we may restrict ourselves to the case $1/2 < s < t$. Define $\zeta^\# = \inf_{\Omega\times\Theta} p(x, \theta)$. By hypothesis, $\zeta^\# > 0$, and $\mathcal{P}(t, \zeta^\#, D)$ is non-empty as it contains $\mathcal{P}_\Theta$. We may now apply Part (d2) of Proposition 31 in Appendix B with $\mathcal{F} = \mathcal{P}(t, \zeta^\#, D)$ to get

$$\lim_{k\to\infty} \sup_{\Theta\times\mathcal{P}(t,\zeta^\#,D)} |L_k(\theta, p) - L(\theta, p)| = 0 \quad \mu\text{-a.s.} \tag{9}$$

Let ε_l be as in the proof of Part (a). For each $l \in \mathbb{N}$, Part (d2) of Proposition 31 in Appendix B with $\mathcal{F} = \{p + \varepsilon_l : p \in \mathcal{P}(t, \zeta, D)\}$ implies that

$$\lim_{k\to\infty} \sup_{\theta\in\Theta} \sup_{p\in\mathcal{P}(t,\zeta,D)} |L_k(\theta, p + \varepsilon_l) - L(\theta, p + \varepsilon_l)| = 0 \quad \mu\text{-a.s.} \tag{10}$$

12

In the following arguments we fix an arbitrary element of the probability 1 event where (9) and (10) hold. Assume that $\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ does not converge to 0. 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

$$\|\tilde{p}_{k'}(\theta_{k'}) - p_{\theta_{k'}}\|_{s,2} > \eta. \tag{11}$$

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 Proposition 30 in Appendix A and Remark 5. This is in contradiction to (11) and therefore in contradiction to the assumption that $\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ 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

$$
\begin{aligned}
L_{k''}(\theta_{k''}, p_{\theta_{k''}}) &\leq L_{k''}(\theta_{k''}, \tilde{p}_{k''}(\theta_{k''})) \leq L_{k''}(\theta_{k''}, \tilde{p}_{k''}(\theta_{k''})) + \varepsilon_l) \\
&\leq L(\theta_{k''}, \tilde{p}_{k''}(\theta_{k''}) + \varepsilon_l) \\
&\quad + \sup_{\theta \in \Theta} \sup_{p \in \mathcal{P}(t, \zeta, D)} |L_{k''}(\theta, p + \varepsilon_l) - L(\theta, p + \varepsilon_l)|.
\end{aligned}
\tag{12}
$$

The first term on the r.h.s. of (12) converges to $L(\theta^*, p^* + \varepsilon_l)$ since $\theta_{k''}$ converges to θ^*, $\|\tilde{p}_{k''}(\theta_{k''}) - p^*\|_{s,2}$, and 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 (c2) of Proposition 31 in Appendix B. Recall that the supremum on the r.h.s. of (12) goes to 0 in view of (10). Further, the supremum on the r.h.s. of the inequality

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

converges to 0 by (9). The second term on the r.h.s. goes to 0 as $\theta_{k''}$ converges to θ^*, $\|p_{\theta_{k''}} - p_{\theta^*}\|_{s,2}$, and hence also $\|p_{\theta_{k''}} - p_{\theta^*}\|_\Omega$, converges to 0, and $L(\theta, p)$ is a continuous function on $\Theta \times (\mathcal{P}(t, \zeta^\#, D), \|\cdot\|_\Omega)$ by Part (c2) of Proposition 31 in Appendix B. Hence, the l.h.s. of (12) goes to $L(\theta^*, p_{\theta^*})$. It follows that

$$L(\theta^*, p_{\theta^*}) \leq L(\theta^*, p^* + \varepsilon_l). \tag{13}$$

The sequence of functions $\log(p^* + \varepsilon_l)(\rho(\cdot, \theta^*))$ is monotonously non-increasing in l 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 (13), we conclude that $L(\theta^*, p_{\theta^*}) \leq L(\theta^*, p^*)$. Hence, $p^* = p_{\theta^*}$ by Part (c) of Lemma 8. ∎

Remark 10 For later use we note the following: (i) Let Assumption D.1 be satisfied, and suppose $\chi \geq 0$ satisfies $\inf_{x \in \Omega} p_\blacktriangle(x) > \chi$. It follows from Part (a) of Theorem 9 that there are events $A_n \in \mathcal{B}(\Omega)^n$ that have $\mathbb{P}^n$-probability tending to 1 as $n \to \infty$ on which $\inf_{x \in \Omega} \hat{p}_n(x) > \chi$ holds.

(ii) Let $p_\theta \in \mathcal{P}(t, \zeta, D)$ for a given $\theta \in \Theta$ be satisfied, and suppose $\chi(\theta) \geq 0$ satisfies $\inf_{x \in \Omega} p(x, \theta) > \chi(\theta)$ for the given θ. It follows from Part (b) of Theorem 9 that for the given θ there are events $B_k(\theta) \in \mathcal{V}^k$ that have μ^k-probability tending to 1 as $k \to \infty$ on which $\inf_{x \in \Omega} \tilde{p}_k(\theta)(x) > \chi(\theta)$ holds.

(iii) Let Assumptions P.1 and R.1 be satisfied, and suppose $\chi \geq 0$ satisfies $\inf_{\Omega \times \Theta} p(x, \theta) > \chi$. It follows from Part (c) of Theorem 9 that there are events $B_k \in \mathcal{V}^k$ that have μ^k-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.

4.2 Rates of Convergence for NPML-Estimators

Following ideas of van de Geer (1993), Nickl (2007, Proposition 6) obtained convergence rates for the NPML-estimator $\hat{p}_n$ in various Sobolev-norms as

$$\|\hat{p}_n - p_\blacktriangle\|_{s,2} = O^*_{\mathbb{P}}(n^{-(t-s)/(2t+1)}) \tag{14}$$

for every $0 \leq s \leq t$, provided Assumption D.1 and $\zeta > 0$ hold. Modulo measure-theoretic nuisances, this immediately gives an analogous result for $\|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ for each $\theta \in \Theta$. [The complication here is that the result in Nickl (2007) is proved for data generating processes defined as coordinate projections on a product space, which is not the case for $X_i(\theta)$; cf. the proof of Part (b) of the subsequent proposition.] In Section 4.3 below, however, we shall need convergence rates for $\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$, i.e., convergence rates that hold uniformly w.r.t. $\theta \in \Theta$. Before we turn to these uniform results, we provide an extension of Nickl's (2007) rate result in that we avoid the restriction $\zeta > 0$. Note that Assumption D.2 already follows from Assumption D.1 in case $\zeta > 0$.

Proposition 11 *(a) Under Assumptions D.1 and D.2 we have $\|\hat{p}_n - p_\blacktriangle\|_{s,2} = O_{\mathbb{P}}(n^{-(t-s)/(2t+1)})$ for every $0 \leq s \leq t$. (b) If $p_\theta \in \mathcal{P}(t, \zeta, D)$ and $\inf_{x \in \Omega} p(x, \theta) > 0$ hold for a given $\theta \in \Theta$, then $\|\tilde{p}_k(\theta) - p_\theta\|_{s,2} = O_\mu(k^{-(t-s)/(2t+1)})$ for every $0 \leq s \leq t$ and the given θ.*

Proof. (a) Measurability of $\|\hat{p}_n - p_\blacktriangle\|_{s,2}$ is established in Proposition 37 in Appendix D. The result is trivial in case $s = t$ since $\mathcal{P}(t, \zeta, D)$ is a bounded subset of $\mathsf{W}_2^t(\Omega)$. Hence assume $s < t$. If $\zeta > 0$, the result follows from Proposition 6 in Nickl (2007). Now suppose $\zeta = 0$. By Assumption D.2 we can then choose $\chi > 0 = \zeta$ such that $\inf_{x \in \Omega} p_\blacktriangle(x) > \chi$ holds. By Remark 10(i) we have that $\hat{p}_n \in \mathcal{P}(t, \chi, D)$ on events $A_n \in \mathcal{B}(\Omega)^n$ that have probability tending to 1 as $n \to \infty$. Since $\mathcal{P}(t, \chi, D) \subseteq \mathcal{P}(t, \zeta, D)$, the NPML-estimator $\hat{p}_n$ over $\mathcal{P}(t, \zeta, D)$ coincides with the NPML-estimator over the smaller set $\mathcal{P}(t, \chi, D)$ on these events, and the latter estimator satisfies (14) by Proposition 6 in Nickl (2007).

(b) In view of (4) and since $(x_1, \ldots, x_k) \mapsto \hat{p}_k(\cdot; x_1, \ldots, x_k)$ is a measurable mapping from Ω^k into $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$, cf. Theorem 6, $\tilde{p}_k(\theta)$ has the same law as $\hat{p}_k(\cdot; Z_1, \ldots, Z_k)$, where $(Z_1, \ldots, Z_k)$ has the same distribution as $(X_1(\theta), \ldots, X_k(\theta))$ but the Z_i are given by the coordinate projections on $(\Omega^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}})$. Since $\|\cdot\|_\Omega$ and $\|\cdot\|_{s,2}$ for $s \leq t$ generate the same Borel σ-field on $\mathcal{P}(t, \zeta, D)$ (cf. Lemma 36 in Appendix D), $\|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ is measurable and has the same distribution as $\|\hat{p}_k(\cdot; Z_1, \ldots, Z_k) - p_\theta\|_{s,2}$. Now apply the already established Part (a) to $\hat{p}_k(\cdot; Z_1, \ldots, Z_k)$. ∎

In case $s = t$, in fact $\|\hat{p}_n - p_\blacktriangle\|_{s,2} \leq 2D$ and $\|\tilde{p}_k(\theta) - p_\theta\|_{s,2} \leq 2D$ hold under the assumptions of the above proposition. The next proposition is instrumental in proving the uniform-in-θ convergence rate result.

Proposition 12 *Let $\mathcal{F}$ be a (non-empty) bounded subset of $\mathsf{W}_2^s(\Omega)$ with $s > 1/2$. Suppose Assumption R.2 holds.*
(a) Then the $\mathcal{L}^2(\mu)$-bracketing metric entropy of

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

satisfies

$$H_{[\,]}(\varepsilon, \mathcal{F}^*, \|\cdot\|_{2,\mu}) \lesssim \varepsilon^{-1/s}. \tag{15}$$

In particular, $\mathcal{F}^*$ is μ-Donsker.

(b) Suppose the elements of $\mathcal{F}$ are bounded below by some $\chi > 0$. Then the $\mathcal{L}^2(\mu)$-bracketing metric entropy of

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

satisfies

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

We note that in the subsequent uniform-in-θ convergence rate result Assumption P.2 already follows from Assumption P.1 in case $\zeta > 0$.

Theorem 13 *Let Assumptions P.1, P.2, and R.2 be satisfied. Then*

$$\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2} = O_\mu(k^{-(t-s)/(2t+1)}) \quad \text{as } k \to \infty \tag{16}$$

for every $0 \le s < t$. [In case $s = t$, the above supremum is bounded by $2D$.]

Proof. Measurability of $\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ for $0 \le s < t$ is established in Proposition 37 in Appendix D. The claim in parentheses follows since $\tilde{p}_k(\theta) \in \mathcal{P}(t, \zeta, D)$ by construction and $p_\theta \in \mathcal{P}(t, \zeta, D)$ by Assumption P.1. We now distinguish two cases:

Case 1: Assume first that $\zeta > 0$ and $s = 0$. We then verify the conditions of Theorem 40 in Appendix E with $(\Lambda, \mathcal{A}, P) = (V^{\mathbb{N}}, \mathcal{V}^{\mathbb{N}}, \mu^{\mathbb{N}})$, $S = \Theta$, $T = \mathcal{P}(t, \zeta, D)$, $d(p, q) = \|p - q\|_2$, $H_k(\sigma, \tau) = L_k(\theta, p)$, $H(\sigma, \tau) = L(\theta, p)$, $\hat{\tau}_k(\sigma) = \tilde{p}_k(\theta)$, and $\tau(\sigma) = p_\theta$. Condition (43) is satisfied by definition of the NPML-estimators $\tilde{p}_k(\theta)$. Condition (41) follows from the second-order Taylor expansion of $L(\theta, \cdot)$ around the density p_θ: using Proposition 32 in Appendix B we obtain

$$
\begin{aligned}
L(\theta, p) - L(\theta, p_\theta) &= \mathbf{D}L(\theta, p_\theta)(p - p_\theta) + \frac{1}{2}\mathbf{D}^2 L(\theta, \bar{p})(p - p_\theta, p - p_\theta) \\
&= -\frac{1}{2}\int_\Omega \frac{(p - p_\theta)^2}{\bar{p}^2} p_\theta d\lambda \le -\frac{1}{2}\zeta (C_t D)^{-2} \|p - p_\theta\|_2^2,
\end{aligned}
$$

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 $\|\bar{p}\|_\Omega \le C_t D$. This proves condition (41) in Theorem 40 with $C = 2^{-1}\zeta (C_t D)^{-2}$ and $\alpha = 2$, both constants being independent of θ and p.

Next we verify condition (42): set

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

for $\delta > 0$, which is clearly non-empty. Then clearly

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

where E^* denotes the outer expectation. Since we have temporarily assumed $\zeta > 0$, the logarithm is Lipschitz on $[\zeta, \infty)$ with Lipschitz constant ζ^{-1}. This implies 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. Consequently,

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

by Theorem 41 in Appendix E. Since

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

15

we have that

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

Applying Proposition 12(b) with $s = t$ and $\mathcal{F} = \mathcal{P}(t, \zeta, D)$ we get from this inequality

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

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

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

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 (42) in Theorem 40.

Condition (44) in that theorem is satisfied for $\alpha = 2$ and $r_k = k^{t/(2t+1)}$. This gives the desired rate and completes the proof in case $\zeta > 0$ and $s = 0$. Now suppose $\zeta > 0$ but $0 < s < t$. Recall that $\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{t,2} \leq 2D$. The result then follows from the interpolation inequality

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

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).

Case 2: Suppose now $\zeta = 0$ and $0 \leq s < t$. In view of Assumption P.2 we may choose $\chi > 0$ such that $\inf_{\Omega \times \Theta} p(x, \theta) > \chi$. Then, by Remark 10(iii), 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 $\mathcal{P}(t, \chi, D) \subseteq \mathcal{P}(t, \zeta, D)$, we have that on these events $\tilde{p}_k(\theta)$ coincides with the NPML-estimators over the smaller set $\mathcal{P}(t, \chi, D)$. The result now follows from what has already been established in Case 1 since Assumption P.1 (and P.2) is also satisfied with respect to $\mathcal{P}(t, \chi, D)$. ∎

4.3 Donsker-type Theorems for NPML-Estimators

Nickl (2007) established Part (a) of the following Donsker-type result under the additional assumption that $\zeta > 0$ holds. Part (b) is (modulo measure-theoretic nuisances) a simple consequence of Part (a).

Theorem 14 *Let $\mathcal{F}$ be a non-empty bounded subset of $\mathsf{W}_2^s(\Omega)$ for some $s > 1/2$.*
(a) Suppose Assumption D.3 is satisfied. Then, for all real $j > 1/2$,

$$\sup_{f \in \mathcal{F}} \left| \sqrt{n} \int_\Omega (\hat{p}_n - p_\blacktriangle) f d\lambda - \sqrt{n}(\mathbb{P}_n - \mathbb{P}) f \right| = o_{\mathbb{P}}(n^{-(\min(s,t)-j)/(2t+1)}) \tag{17}$$

as $n \to \infty$; in particular, the l.h.s. of the above display is $o_{\mathbb{P}}(1)$ as $n \to \infty$. Consequently, the stochastic process $f \mapsto \sqrt{n} \int_\Omega (\hat{p}_n - p_\blacktriangle) f d\lambda$ converges weakly to a $\mathbb{P}$-Brownian bridge in $\ell^\infty(\mathcal{F})$.
(b) Suppose $p_\theta \in \mathcal{P}(t, \zeta, D)$, $\inf_{x \in \Omega} p(x, \theta) > \zeta$, and $\|p_\theta\|_{t,2} < D$ hold for a given $\theta \in \Theta$. Then, for the given θ, a result analogous to Part (a) holds for the process $f \mapsto \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda$ with $\mathbb{P}_k$ and $\mathbb{P}$, respectively, replaced by $\mathbb{P}_{\theta,k}$ and $\mathbb{P}_\theta$, where $\mathbb{P}_{\theta,k}$ is the empirical measure of $X_1(\theta), \ldots, X_k(\theta)$ and $\mathbb{P}_\theta$ is the probability measure corresponding to p_θ.

16

Proof. (a) Measurability of the l.h.s. of (17) follows from Proposition 38 in Appendix D. For $\zeta > 0$ the result follows immediately from Theorem 3 in Nickl (2007). Now suppose $\zeta = 0$. In view of Assumption D.3 we may choose $\chi > 0$ such that $\inf_{x \in \Omega} p_{\blacktriangle}(x) > \chi$. Then, by Remark 10(i), there are events that have probability tending to 1 on which $\inf_{x \in \Omega} \hat{p}_n(x) > \chi$ holds true. Since $\mathcal{P}(t, \chi, D) \subseteq \mathcal{P}(t, \zeta, D) = \mathcal{P}(t, 0, D)$, we have that on these events $\hat{p}_n$ coincides with the NPML-estimators over the smaller set $\mathcal{P}(t, \chi, D)$. Since $\chi > 0$ and since Assumption D.3 is also satisfied relative to $\mathcal{P}(t, \chi, D)$, the result now follows from what has already been established.

(b) Note that $\mathfrak{X}_k(\check{x}, f)$ and $\sup_{f \in \mathcal{F}} |\mathfrak{X}_k(\check{x}, f) - \mathfrak{Y}_k(\check{x}, f)|$ defined in Proposition 38(a) in Appendix D are Borel measurable on Ω^k. Consequently,

$$\sup_{f \in \mathcal{F}} |\mathfrak{X}_k(X_1(\theta), \ldots, X_k(\theta), f) - \mathfrak{Y}_k(X_1(\theta), \ldots, X_k(\theta), f)|$$

and

$$\sup_{f \in \mathcal{F}} |\mathfrak{X}_k(Z_1, \ldots, Z_k, f) - \mathfrak{Y}_k(Z_1, \ldots, Z_k, f)|$$

have the same distribution, where the Z_i are as in the proof of Proposition 11. Furthermore, it follows that the finite-dimensional distributions of the processes $f \mapsto \mathfrak{X}_k(X_1(\theta), \ldots, X_k(\theta), f)$ and $f \mapsto \mathfrak{X}_k(Z_1, \ldots, Z_k, f)$ coincide. It is easy to see that the maps $f \to \mathfrak{X}_k(\check{x}, f)$ belong to $\mathsf{C}^0(\mathcal{F}, \|\cdot\|_\Omega)$, the space of bounded uniformly continuous functions on $(\mathcal{F}, \|\cdot\|_\Omega)$. Consequently, $\mathfrak{X}_k(\check{x}, \cdot)$ is Borel measurable as a random element in $\mathsf{C}^0(\mathcal{F}, \|\cdot\|_\Omega)$, since the Borel σ-field on this space is generated by the point-evaluations (observe that $(\mathcal{F}, \|\cdot\|_\Omega)$ is totally bounded in view of Lemma 35 in Appendix C). Since $\mathsf{C}^0(\mathcal{F}, \|\cdot\|_\Omega)$ is Polish by total boundedness of $(\mathcal{F}, \|\cdot\|_\Omega)$, the entire laws of the processes $f \mapsto \mathfrak{X}_k(X_1(\theta), \ldots, X_k(\theta), f)$ and $f \mapsto \mathfrak{X}_k(Z_1, \ldots, Z_k, f)$ on $\mathsf{C}^0(\mathcal{F}, \|\cdot\|_\Omega)$, and hence on $\ell^\infty(\mathcal{F})$, coincide. In view of (4), Part (b) now follows from applying the already established Part (a) to $\hat{p}_k(\cdot; Z_1, \ldots, Z_k)$. ∎

The next theorem shows that a weak limit theorem for the stochastic process $(\theta, f) \mapsto \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda$ can be obtained even in the space $\ell^\infty(\Theta \times \mathcal{F})$. A corollary of this is then a uniform-in-θ version of Part (b) of the above theorem. The proof of this theorem largely follows the ideas in Nickl (2007): Loosely speaking, a mean-value expansion of $\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(\cdot)$, analogous to the one in the classical parametric case, shows that this can be represented as the sum of the score evaluated at the true density p_θ, i.e., $\mathbf{D}L_k(\theta, p_\theta)(\cdot)$, plus a second derivative term applied to the estimation error $(\tilde{p}_k(\theta) - p_\theta, \cdot)$. [For given $\theta \in \Theta$, the Fréchet-derivative of L_k with respect to the second argument is here denoted by $\mathbf{D}L_k(\theta, \cdot)$.] The score, evaluated at the true density p_θ and properly scaled, turns out to be an empirical process having a Gaussian limit. The second derivative term turns out to coincide with $-\int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda$ up to negligible terms. [An important ingredient for establishing negligibility are the uniform-in-θ convergence rates for $\tilde{p}_k(\theta)$ in different Sobolev norms that have been established in the previous section.] Apart from a series of technical difficulties not present in the classical parametric case, the major difficulty is then the following: in the classical parametric case the usual assumption that the true parameter belongs to the interior of the parameter space together with consistency implies that the estimator is eventually an interior point, implying that the score evaluated at the maximizer is zero. In the present case, while p_θ is an interior point of $\mathcal{P}(t, \zeta, D)$ relative to H_t as a consequence of the assumptions underlying Theorem 16, the estimator $\tilde{p}_k(\theta)$ is, however, *not* an interior point of the domain $\mathcal{P}(t, \zeta, D)$ (relative to H_t) over which optimization is performed, as shown in Theorem 6; in particular, $\tilde{p}_k(\theta)$ is *not* consistent w.r.t. the $\|\cdot\|_{t,2}$-norm. As a consequence, one can not conclude that the score evaluated at the maximizer is zero. [Trying to save this argument directly by using an $\|\cdot\|_{s,2}$-norm with $s < t$ does not work either: while $\tilde{p}_k(\theta)$ is consistent in the $\|\cdot\|_{s,2}$-norm, p_θ is then *not* an interior point of $\mathcal{P}(t, \zeta, D)$ relative to H_s.] Hence,

17

a different reasoning is needed to show that $\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(\cdot)$, although not necessarily zero, is of sufficiently small order. This is provided in the subsequent lemma, which is essentially a uniform version of Lemma 4 in Nickl (2007). The proof as given below makes use of Proposition 3 which allows us to simplify the arguments given in Nickl (2007). In the following lemma let H_t^0 denote the linear subspace of $\mathsf{W}_2^t(\Omega)$ that is parallel to H_t.

Lemma 15 *Suppose Assumptions P.3 and R.2 are satisfied and $\zeta > 0$ holds. Let $\mathcal{G}$ be a non-empty bounded subset of $\mathsf{H}_t^0 \subseteq \mathsf{W}_2^t(\Omega)$. Then*

$$\sup_{\theta \in \Theta} \sup_{g \in \mathcal{G}} |\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(g)| = o_\mu(k^{-(t-j)/(2t+1)-1/2}) \tag{18}$$

for every real $j > 1/2$.

Proof. Measurability of the l.h.s. of (18) follows from Proposition 38(c) in Appendix D. W.l.o.g. we may assume $1/2 < j < t$. By Assumption P.3 and Proposition 3(b) we can find $\delta > 0$ small enough such that

$$p_\theta + w \in \mathcal{P}(t, \zeta, D)$$

holds for every $\theta \in \Theta$ and every $w \in \mathcal{U}_{t,\delta} \cap \mathsf{H}_t^0$. Note that δ does not depend on θ. Since $\tilde{p}_k(\theta)$ maximizes $L_k(\theta, \cdot)$ (which is differentiable in view of Proposition 32 as $\zeta > 0$ is assumed) over $\mathcal{P}(t, \zeta, D)$ we conclude that

$$\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(p_\theta + w - \tilde{p}_k(\theta)) \leq 0$$

holds for all $\theta \in \Theta$ and all $w \in \mathcal{U}_{t,\delta} \cap \mathsf{H}_t^0$. This implies

$$\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(w) \leq \mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(\tilde{p}_k(\theta) - p_\theta)$$

for all $\theta \in \Theta$ and $w \in \mathcal{U}_{t,\delta} \cap \mathsf{H}_t^0$. Since $\mathcal{U}_{t,\delta} \cap \mathsf{H}_t^0$ is invariant under multiplication by -1, we obtain

$$\begin{aligned}
\sup_{\theta \in \Theta} \sup_{w \in \mathcal{U}_{t,\delta} \cap \mathsf{H}_t^0} |\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(w)| \quad &\leq \quad \sup_{\theta \in \Theta} |\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(\tilde{p}_k(\theta) - p_\theta)| \\
&\leq \quad \sup_{\theta \in \Theta} |(\mathbf{D}L_k(\theta, \tilde{p}_k(\theta)) - \mathbf{D}L(\theta, \tilde{p}_k(\theta)))(\tilde{p}_k(\theta) - p_\theta)| \\
&\quad + \sup_{\theta \in \Theta} |(\mathbf{D}L(\theta, \tilde{p}_k(\theta)) - \mathbf{D}L(\theta, p_\theta))(\tilde{p}_k(\theta) - p_\theta)| \\
&\leq \quad \sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{j,2} \sup_{\Theta \times \mathcal{P}(t,\zeta,D)} \|\mathbf{D}L_k(\theta, p) - \mathbf{D}L(\theta, p)\|_{\mathcal{U}_{j,1}} \\
&\quad + \zeta^{-1} \sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_2^2,
\end{aligned}$$

where we have repeatedly used Proposition 32, in particular to establish that $\mathbf{D}L(\theta, p_\theta))(\tilde{p}_k(\theta) - p_\theta) = 0$. Now use Theorem 13 and Proposition 33 with $\alpha = 1$ and $\mathcal{H}_1 = \mathcal{U}_{j,1}$ to conclude that the r.h.s. of the last display is

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

since $j > 1/2$. A fortiori this holds for all $j > 1/2$ and thus proves the result for the case where $\mathcal{G}$ is contained in $\mathcal{U}_{t,\delta} \cap \mathsf{H}_t^0$. Since (18) is homogenous w.r.t. scaling of $\mathcal{G}$ and δ does not depend on $\mathcal{G}$, this inclusion can, however, always be achieved by rescaling. ∎

We note that the lemma can easily be extended to the case $\zeta = 0$ by making use of Remark 10(iii). The main result is now the following.

Theorem 16 *Suppose Assumptions P.3 and R.2 are satisfied. Let $\mathcal{F}$ be a non-empty bounded subset of $\mathsf{W}_2^s(\Omega)$ for some $s > 1/2$. Then:*

(a) For all real $j > 1/2$,

$$\sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda - \sqrt{k}(\mu_k - \mu) f(\rho(\cdot, \theta)) \right| = o_\mu(k^{-(\min(s,t)-j)/(2t+1)}) \qquad (19)$$

as $k \to \infty$; in particular, the l.h.s. of the above display is $o_\mu(1)$ as $k \to \infty$.

(b) There exists a zero-mean Gaussian process $\mathbb{G}$ indexed by $\Theta \times \mathcal{F}$ with bounded sample paths such that the stochastic process $(\theta, f) \mapsto \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda$ converges weakly to $\mathbb{G}(\theta, f)$ in $\ell^\infty(\Theta \times \mathcal{F})$. The process $\mathbb{G}$ is measurable as a mapping with values in $\ell^\infty(\Theta \times \mathcal{F})$, has separable range, and has sample paths that are uniformly continuous with respect to the pseudo-metric $d((\theta, f), (\theta', g)) = \left(\mathrm{Var}[\mathbb{G}(\theta, f) - \mathbb{G}(\theta', g)] \right)^{1/2}$. Its covariance function is given by

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

(c)

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

Proof. Part (a): Measurability of the l.h.s. of (19) follows from Proposition 38(b) in Appendix D.

Step 1: We first consider the case $\zeta > 0$. Let $\mathcal{G}$ be a non-empty bounded subset of H_t^0. Applying the pathwise mean-value theorem to the function $\mathbf{D}L_k(\theta, \cdot)(g)$, adding and subtracting a term, and using Proposition 32 leads to

$$\begin{aligned}
\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(g) &= \mathbf{D}L_k(\theta, p_\theta)(g) + \mathbf{D}^2 L_k(\theta, \bar{p}_k(\theta))(\tilde{p}_k(\theta) - p_\theta, g) \\
&= (\mu_k - \mu)(p_\theta^{-1} g)(\rho(\cdot, \theta)) + \mathbf{D}^2 L(\theta, p_\theta)(\tilde{p}_k(\theta) - p_\theta, g) \\
&\quad + \left[\mathbf{D}^2 L_k(\theta, \bar{p}_k(\theta)) - \mathbf{D}^2 L(\theta, p_\theta) \right] (\tilde{p}_k(\theta) - p_\theta, g),
\end{aligned}$$

where $\bar{p}_k(\theta) = \xi \tilde{p}_k(\theta) + (1 - \xi) p_\theta$ for some $\xi \in (0, 1)$; note that $\bar{p}_k(\theta) \in \mathcal{P}(t, \zeta, D)$ by convexity. In the above display we have also made use of the fact that $\mu(p_\theta^{-1} g)(\rho(\cdot, \theta)) = 0$ since $g \in \mathsf{H}_t^0$. Again adding and subtracting a term and using Proposition 32 this leads to

$$\begin{aligned}
\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(g) &= (\mu_k - \mu)(p_\theta^{-1} g)(\rho(\cdot, \theta)) - \int_\Omega p_\theta^{-1}(\tilde{p}_k(\theta) - p_\theta) g d\lambda \\
&\quad + \left[\mathbf{D}^2 L_k(\theta, \bar{p}_k(\theta)) - \mathbf{D}^2 L(\theta, \bar{p}_k(\theta)) \right] (\tilde{p}_k(\theta) - p_\theta, g) \\
&\quad + \int_\Omega \bar{p}_k^{-2}(\theta) p_\theta^{-1}(\bar{p}_k^2(\theta) - p_\theta^2)(\tilde{p}_k(\theta) - p_\theta) g d\lambda.
\end{aligned}$$

Consequently, for every real j with $1/2 < j < t$ we obtain

$$\begin{aligned}
&\sup_{\theta \in \Theta} \sup_{g \in \mathcal{G}} \left| \int_\Omega p_\theta^{-1}(\tilde{p}_k(\theta) - p_\theta) g d\lambda - (\mu_k - \mu)(p_\theta^{-1} g)(\rho(\cdot, \theta)) \right| \\
\leq\ & \sup_{\theta \in \Theta} \sup_{g \in \mathcal{G}} |\mathbf{D}L_k(\theta, \tilde{p}_k(\theta))(g)| + \\
& \sup_{\theta \in \Theta} \sup_{g \in \mathcal{G}} \left| \left[\mathbf{D}^2 L_k(\theta, \bar{p}_k(\theta)) - \mathbf{D}^2 L(\theta, \bar{p}_k(\theta)) \right] (\tilde{p}_k(\theta) - p_\theta, g) \right| \\
& + \sup_{\theta \in \Theta} \sup_{g \in \mathcal{G}} \left| \int_\Omega \bar{p}_k^{-2}(\theta) p_\theta^{-1}(\bar{p}_k^2(\theta) - p_\theta^2)(\tilde{p}_k(\theta) - p_\theta) g d\lambda \right| \\
=\ & I + II + III, \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad (20)
\end{aligned}$$

19

where $I = o_\mu(k^{-(t-j)/(2t+1)-1/2})$ by Lemma 15. We next bound expressions II and III:
Clearly,

$$II \leq \sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{j,2} \sup_{\Theta \times \mathcal{P}(t,\zeta,D)} \left\|\mathbf{D}^2 L(\theta,p) - \mathbf{D}^2 L_k(\theta,p)\right\|_{\mathcal{U}_{j,1} \times \mathcal{G}}$$

The first supremum in the above display is $O_\mu(k^{-(t-j)/(2t+1)})$ by Theorem 13. Since $\mathcal{G}$ is bounded in $\mathsf{W}_2^t(\Omega)$ and hence also in $\mathsf{W}_2^j(\Omega)$ as $j < t$ (cf. Proposition 1), and since $\mathcal{U}_{j,1}$ is clearly bounded in $\mathsf{W}_2^j(\Omega)$, the second supremum in the above display is $O_\mu(k^{-1/2})$ by Proposition 33, when applied with $\alpha = 2$, $\mathcal{H}_1 = \mathcal{U}_{j,1}$, and $\mathcal{H}_2 = \mathcal{G}$. This shows that the expression II is $O_\mu(k^{-(t-j)/(2t+1)-1/2})$ for every real j with $1/2 < j < t$.

Next, observe that $|\bar{p}_k(\theta) - p_\theta| = \xi |\tilde{p}_k(\theta) - p_\theta| \leq |\tilde{p}_k(\theta) - p_\theta|$ and that $\bar{p}_k(\theta) \geq \zeta$, $p_\theta \geq \zeta$ as these functions belong to $\mathcal{P}(t,\zeta,D)$. Hence

$$III \leq 2\zeta^{-3} C_t^2 DG \sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_2^2,$$

where $G < \infty$ is a $\|\cdot\|_{t,2}$-norm bound for $\mathcal{G}$. (Here we have repeatedly used Proposition 1(b)). Theorem 13 then shows that expression III is $O_\mu(k^{-2t/(2t+1)})$. Putting things together we obtain that the l.h.s. of (20) is $O_\mu^*(k^{-(t-j)/(2t+1)-1/2})$ for every real j with $1/2 < j < t$, and hence a fortiori for every real $j > 1/2$. Consequently,

$$\sup_{\theta \in \Theta} \sup_{g \in \mathcal{G}} \sqrt{k} \left| \int_\Omega p_\theta^{-1}(\tilde{p}_k(\theta) - p_\theta) g d\lambda - (\mu_k - \mu)\left(p_\theta^{-1} g\right)(\rho(\cdot,\theta)) \right| = o_\mu^*(k^{-(t-j)/(2t+1)}) \qquad (21)$$

for every real $j > 1/2$.

Let now $\mathcal{F}$ be a nonempty bounded subset of $\mathsf{W}_2^t(\Omega)$ and let $B < \infty$ denote a $\|\cdot\|_{t,2}$-norm bound for $\mathcal{F}$. Define $\pi_{\theta'}(f) = (f - \int_\Omega f p_{\theta'} d\lambda) p_{\theta'}$ for any $f \in \mathsf{W}_2^t(\Omega)$ and $\theta' \in \Theta$. Then, using Proposition 1(a) and the fact that $p_{\theta'} \in \mathcal{P}(t,\zeta,D)$ by Assumption P.3, gives

$$
\begin{aligned}
\sup_{\theta' \in \Theta} \sup_{f \in \mathcal{F}} \|\pi_{\theta'}(f)\|_{t,2} &\leq M_t \sup_{\theta' \in \Theta} \sup_{f \in \mathcal{F}} \left[\left\| f - \int_\Omega f p_{\theta'} \, d\lambda \right\|_{t,2} \|p_{\theta'}\|_{t,2}\right] \\
&\leq M_t D \left[B + \sup_{f \in \mathcal{F}} \|f\|_\Omega \|1\|_{t,2}\right] \\
&\leq M_t DB (1 + C_t \lambda(\Omega)^{1/2}) < \infty.
\end{aligned}
\qquad (22)
$$

This shows that the set
$$\mathcal{G}(\Theta, \mathcal{F}) = \left\{\pi_{\theta'}(f) : f \in \mathcal{F}, \theta' \in \Theta\right\}$$

is a nonempty bounded subset of $\mathsf{W}_2^t(\Omega)$. In fact, it is a subset of H_t^0 by definition of $\pi_{\theta'}$. It is now easy to see that applying (21) to $\mathcal{G}(\Theta, \mathcal{F})$ implies (19) in the case $s = t$. The case $s > t$ immediately follows, since every nonempty bounded subset of $\mathsf{W}_2^s(\Omega)$ with $s > t$ can also be viewed as a nonempty bounded subset of $\mathsf{W}_2^t(\Omega)$ by Proposition 1(c). This proves Part (a) in case $\zeta > 0$ and $s \geq t$.

Step 2: We now consider the case where $\zeta > 0$ and $1/2 < s < t$. For every $f \in \mathcal{F}$ let $u_k(f) \in \mathsf{W}_2^t(\Omega)$ be the approximators defined in the proof of Proposition 1 in Nickl (2007). They have the following properties:

$$\sup_{f \in \mathcal{F}} \|u_k(f)\|_{t,2} = O(k^{(t-s)/(2t+1)}) \quad \text{as } k \to \infty, \qquad (23)$$

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

$$\sup_{f \in \mathcal{F}} \|f - u_k(f)\|_{r,2} = O(k^{-(s-r)/(2t+1)}) \quad \text{as } k \to \infty. \tag{24}$$

We have that

$$\sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) f d\lambda - \sqrt{k}(\mu_k - \mu) f(\rho(\cdot, \theta)) \right|$$

$$\le \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta)(f - u_k(f)) d\lambda \right|$$

$$+ \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| \sqrt{k}(\mu_k - \mu)(f(\rho(\cdot, \theta)) - u_k(f)(\rho(\cdot, \theta))) \right|$$

$$+ \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) u_k(f) d\lambda - \sqrt{k}(\mu_k - \mu) u_k(f)(\rho(\cdot, \theta)) \right|$$

$$= IV + V + VI. \tag{25}$$

We now derive bounds for each of the above expressions:

Using (24) with $r = 0$, the Cauchy-Schwarz inequality, and Theorem 13 we obtain

$$IV \le \sqrt{k} \sup_{f \in \mathcal{F}} \|f - u_k(f)\|_2 \sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_2 = O_\mu(k^{-(s-1/2)/(2t+1)}).$$

Next, choose an arbitrary real j such that $1/2 < j < s$ and observe that

$$V = \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| \sqrt{k}(\mu_k - \mu)(f - u_k(f))(\rho(\cdot, \theta)) \right|$$

$$\le \left(\sup_{\theta \in \Theta} \sup_{h \in \mathcal{U}_{j,1}} \left| \sqrt{k}(\mu_k - \mu) h(\rho(\cdot, \theta)) \right| \right) \sup_{f \in \mathcal{F}} \|f - u_k(f)\|_{j,2}$$

$$= \|\sqrt{k}(\mu_k - \mu)\|_{\mathcal{U}_{j,1}^*} \sup_{f \in \mathcal{F}} \|f - u_k(f)\|_{j,2}, \tag{26}$$

where

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

Since $j > 1/2$, the class of functions $\mathcal{U}_{j,1}^*$ is μ-Donsker by Proposition 12(a), hence

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

in view of Prohorov's theorem, measurability following from Proposition 38. Making use of (24), it follows that the r.h.s. of (26), and hence Expression V, is $O_\mu(k^{-(s-j)/(2t+1)})$.

Finally note that Expression VI is bounded by

$$\sup_{\theta \in \Theta} \sup_{h \in \mathcal{U}_{t,1}} \left| \sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta) h d\lambda - \sqrt{k}(\mu_k - \mu) h(\rho(\cdot, \theta)) \right| \sup_{f \in \mathcal{F}} \|u_k(f)\|_{t,2}.$$

Since $\mathcal{U}_{t,1}$ is a nonempty bounded subset of $\mathsf{W}_2^t(\Omega)$ and since Part (a) has already been established in *Step 1* for such sets of functions, the first term on the r.h.s. of the last display is $o_\mu(k^{-(t-j)/(2t+1)})$, and using (23), we conclude that

$$VI = o_\mu(k^{-(s-j)/(2t+1)}).$$

21

The above bounds imply that the l.h.s. of (25) is $O_\mu(k^{-(s-j)/(2t+1)})$ for all $1/2 < j < s$, and hence is $o_\mu(k^{-(s-j)/(2t+1)})$ for all $j > 1/2$. This completes the proof of Part (a) of the theorem in case $\zeta > 0$.

Step 3: We next consider the case $\zeta = 0$. In view of Assumption P.3 we may choose $\chi > 0$ such that $\inf_{\Omega \times \Theta} p(x, \theta) > \chi$. Then, by Remark 10(iii), 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 $\mathcal{P}(t, \chi, D) \subseteq \mathcal{P}(t, 0, D) = \mathcal{P}(t, \zeta, D)$, we have that on these events $\tilde{p}_k(\theta)$ coincides with the NPML-estimators over the smaller set $\mathcal{P}(t, \chi, D)$. Part (a) in case $\zeta = 0$ now follows from what has already been established in the preceding two steps (applied to the NPML-estimator based on $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$ and noting that Assumption P.3 is also satisfied relative to $\mathcal{P}(t, \chi, D)$).

Part (b): In view of Part (a) 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 $f \in \mathcal{F}$, where $\mathcal{F}^* = \{f(\rho(\cdot, \theta)) : \theta \in \Theta, f \in \mathcal{F}\}$. Note that the resulting mapping $H : \ell^\infty(\mathcal{F}^*) \to \ell^\infty(\Theta \times \mathcal{F})$ is continuous since H is linear and

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

for all $\varphi \in \ell^\infty(\mathcal{F}^*)$. In fact, H is an isometry. Since $\mathcal{F}^*$ is μ-Donsker by Proposition 12(a), $\sqrt{k}(\mu_k - \mu)$ converges weakly in $\ell^\infty(\mathcal{F}^*)$ to a μ-Brownian bridge $\mathbb{G}^*$, that is, $\mathbb{G}^*$ is a mean-zero Gaussian process indexed by $\mathcal{F}^*$, which is measurable as a mapping with values in $\ell^\infty(\mathcal{F}^*)$, has covariance function

$$\mathrm{Cov}[\mathbb{G}^*(f(\rho(\,\cdot\,, \theta))), \mathbb{G}^*(g(\rho(\,\cdot\,, \theta')))]$$
$$= \int_V \left(f(\rho(\cdot, \theta)) - \int_V f(\rho(\cdot, \theta)) d\mu \right) \left(g(\rho(\cdot, \theta')) - \int_V g(\rho(\cdot, \theta')) d\mu \right) d\mu,$$

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

$$d^*(f(\rho(\cdot, \theta)), g(\rho(\cdot, \theta'))) = \left(\mathrm{Var}[\mathbb{G}^*(f(\rho(\,\cdot\,, \theta))) - \mathbb{G}^*(g(\rho(\,\cdot\,, \theta')))] \right)^{1/2}.$$

Since the empirical process $\sqrt{k}(\mu_k - \mu)$ indexed by $\mathcal{F}^*$ is mapped into the process $(\theta, f) \mapsto \sqrt{k}(\mu_k - \mu) f(\rho(\cdot, \theta))$ by the map H, the continuous mapping theorem shows that the latter process converges weakly in $\ell^\infty(\Theta \times \mathcal{F})$ to $\mathbb{G} := H(\mathbb{G}^*)$. The properties of $\mathbb{G}$ claimed in the theorem follow easily from the corresponding properties of the μ-Brownian bridge $\mathbb{G}^*$ and the fact that H is an isometry.

Part (c): Follows directly from Part (b) in view of Prohorov's theorem, with measurability again following from Proposition 38(b) in Appendix D. ∎

We next obtain a corollary showing that $\sqrt{k} \int_\Omega (\tilde{p}_k(\theta) - p_\theta)(\cdot) d\lambda$ converges in $\ell^\infty(\mathcal{F})$ to $\mathbb{G}(\theta)$ *uniformly* over Θ, where $\mathbb{G}(\theta)(f) := \mathbb{G}(\theta, f)$ for all $f \in \mathcal{F}$. For this we recall the following definitions: Let (S, d) be a metric space. For probability spaces $(\Lambda_1, \mathcal{A}_1, P_1)$, $(\Lambda_2, \mathcal{A}_2, P_2)$ and mappings $Y_1 : \Lambda_1 \to S$, $Y_2 : \Lambda_2 \to S$ such that Y_2 is $\mathcal{A}_2$-$\mathcal{B}(S, d)$-measurable and has separable range define an analogue of the dual bounded Lipschitz metric by

$$\beta_{(S,d)}(Y_1, Y_2) = \sup \left\{ \left| \int_{\Lambda_1}^* h(Y_1) dP_1 - \int_{\Lambda_2} h(Y_2) dP_2 \right| : \|h\|_{BL(S,d)} \leq 1 \right\},$$

where $\int^*$ denotes the outer integral and $\|\cdot\|_{BL(S,d)}$ denotes the bounded Lipschitz norm; cf. the definition on p. 115 in Dudley (1999). By Theorem 3.6.4 in Dudley (1999), $Y_n \rightsquigarrow Y$ (where Y is measurable and has separable range) if and only if

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

Corollary 17 *Let the hypotheses of Theorem 16 be satisfied. Then, for every $\theta \in \Theta$, $\mathbb{G}(\theta) = \mathbb{G}(\theta, \cdot)$ is a measurable mapping with values in $\ell^\infty(\mathcal{F})$ that has separable range. Furthermore,*

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

[In fact, $\mathbb{G}(\theta)$ is a P_θ-Brownian bridge where P_θ denotes the probability measure corresponding to p_θ.]

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 $f \in \mathcal{F}$. 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. Clearly, $\mathbb{G}(\theta) = H_\theta(\mathbb{G})$ holds. Since $\mathbb{G}$ is a measurable mapping with separable range in $\ell^\infty(\Theta \times \mathcal{F})$ by Part (b) of Theorem 16, this shows that, for every $\theta \in \Theta$, $\mathbb{G}(\theta)$ is measurable with separable range 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

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

The r.h.s., and therefore the l.h.s., of the previous display converges to 0 by Part (b) of Theorem 16. That $\mathbb{G}(\theta)$ is in fact a P_θ-Brownian bridge indexed by $\mathcal{F}$ easily follows from Part (b) of Theorem 16 and the transformation theorem. $\blacksquare$

Remark 18 The statement in Corollary 17 is in fact independent of any distance describing the concept of weak convergence in $\ell^\infty(\mathcal{F})$. More precisely, under the assumptions of Corollary 17, 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,

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

To see this we apply Lemma 39 in Appendix D with $T = \Theta$, $S = \ell^\infty(\mathcal{F})$ equipped with the sup-norm, $Y_k(\theta) = \sqrt{k}\int_\Omega (\tilde{p}_k(\theta) - p_\theta)(\cdot)d\lambda$, $Y(\theta) = \mathbb{G}(\theta)$. Note that $\mathbb{G}(\theta)$ is measurable with separable range in $\ell^\infty(\mathcal{F})$ as shown in Corollary 17. It remains to show that $\mathbb{G}(\theta_k) \rightsquigarrow \mathbb{G}(\theta)$ for any $\theta_k, \theta \in \Theta$ such that θ_k converges to θ. Since $\mathbb{G}(\theta, \cdot)$ is measurable in $\ell^\infty(\mathcal{F})$, it is sufficient to show that $\mathbb{G}(\theta_k)$ converges a.s. to $\mathbb{G}(\theta)$ in $\ell^\infty(\mathcal{F})$, that is

$$\lim_{k\to\infty} \sup_{f\in\mathcal{F}} |\mathbb{G}(\theta_k, f) - \mathbb{G}(\theta, f)| = 0. \tag{27}$$

23

Fix an arbitrary $\varepsilon > 0$. Because the sample paths of $\mathbb{G}$ are uniformly continuous with respect to the pseudo-metric d given in Theorem 16(b), there is a $\delta(\varepsilon) > 0$ such that $|\mathbb{G}(\theta, f) - \mathbb{G}(\theta', g)| < \varepsilon$ whenever $d((\theta, f), (\theta', g)) < \delta(\varepsilon)$. Now, observe that

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

for some finite constant $L_s > 0$ not depending on $f \in \mathcal{F}$. Here, we have used that $\mathcal{F}$ is a bounded subset of $\mathsf{C}^{\min(s-1/2, 1)}(\Omega)$ by Proposition 1(b). The r.h.s. of the last display converges to 0 as $k \to \infty$ in view of the theorem of dominated convergence (use that Assumption R.2 implies Assumption R.1 together with the fact that ρ takes its values in the bounded set Ω). Consequently, $d((\theta_k, f)), (\theta, f)) < \delta(\varepsilon)$ for $k \geq k(\delta(\varepsilon))$, which implies (27).

Remark 19 We have assumed that the processes (X_i) and (V_i) are canonically defined, i.e., are given by the respective coordinate projections of the measurable space $(\Omega^{\mathbb{N}} \times V^{\mathbb{N}}, \mathcal{B}(\Omega)^{\mathbb{N}} \otimes \mathcal{V}^{\mathbb{N}})$. We have made this assumption to be able to freely use results from empirical process theory as well as from Nickl (2007) which typically are formulated in this canonical setting. However, the measurability results in Appendix D show that all results of the paper continue to hold if (X_i) and (V_i) are defined on an arbitrary probability space.

5 Simulation-Based Minimum Distance Estimators

We next study simulation-based minimum distance (indirect inference) estimators when the auxiliary density estimators are the NPML-estimators $\hat{p}_n$ and $\tilde{p}_k(\theta)$ based on the given auxiliary model $\mathcal{P}(t, \zeta, D)$. To this end we define for every $\theta \in \Theta$

$$
\mathbb{Q}_{n,k}(\theta) = \begin{cases} \int_\Omega (\hat{p}_n - \tilde{p}_k(\theta))^2 \hat{p}_n^{-1} d\lambda & \text{if } \hat{p}_n(x) > 0 \text{ for all } x \in \Omega, \\ 0 & \text{otherwise,} \end{cases} \tag{28}
$$

and

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

Note that $\mathbb{Q}_{n,k}$ as well as $\mathbb{Q}_n$ take their values in $[0, \infty]$. By separability of Ω and continuity of $\hat{p}_n$, the set $\{\hat{p}_n(x) > 0 \text{ for all } x \in \Omega\}$ belongs to the σ-field $\mathcal{B}(\Omega)^n$. Since $\hat{p}_n$ and $\tilde{p}_k(\theta)$, respectively, are jointly measurable by Remark 7(i), it follows from Tonelli's theorem that $\mathbb{Q}_{n,k}(\theta)$ is $\mathcal{B}(\Omega)^n \otimes \mathcal{V}^k$-measurable and that $\mathbb{Q}_n(\theta)$ is $\mathcal{B}(\Omega)^n$-measurable for every $\theta \in \Theta$. [Assigning the value 0 on the complement of $\{\hat{p}_n(x) > 0 \text{ for all } x \in \Omega\}$ to both objective functions is arbitrary and irrelevant for the asymptotic considerations to follow.]

A simulation-based minimum distance (SMD) estimator is now a mapping $\hat{\theta}_{n,k} : \Omega^n \times V^k \to \Theta$ that minimizes $\mathbb{Q}_{n,k}$ over Θ whenever the minimum exists (and is defined arbitrarily otherwise). Similarly, a minimum distance (MD) estimator is a mapping $\hat{\theta}_n : \Omega^n \to \Theta$ that minimizes $\mathbb{Q}_n$ over Θ whenever the minimum exists (and is defined arbitrarily otherwise). The MD-estimator is of course only feasible if a closed form expression for p_θ can be found; here it serves as an auxiliary device for proving asymptotic results for the SMD-estimator.

Furthermore, whenever Assumption D.2 is satisfied, we define

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

which takes its values in $[0, \infty]$. In view of convergence of $\hat{p}_n$ to $p_\blacktriangle$ and of $\tilde{p}_k(\theta)$ to p_θ (under the assumptions of Theorem 9), Q can be viewed as the limiting counterpart of both $\mathbb{Q}_{n,k}$ as well as $\mathbb{Q}_n$.

5.1 Consistency of SMD-Estimators

Before turning to consistency, we show that MD- and SMD-estimators in fact minimize their corresponding objective function at least on events that have probability tending to 1. Note that in the following proposition the statement of Part (c) is stronger than the one of Part (b), but also requires additional assumptions.

Proposition 20 *Let Assumption R.1 be satisfied.*

(a) Suppose $\zeta > 0$ holds. Then any SMD-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 an SMD-estimator that is $\mathcal{B}(\Omega)^n \otimes \mathcal{V}^k$-$\mathcal{B}(\Theta)$-measurable.

(b) Suppose $\zeta = 0$ and Assumptions D.1 and D.2 hold. Then there are events $A_n \in \mathcal{B}(\Omega)^n$ having probability converging to 1 as $n \to \infty$ such that, on the events $A_n \times V^k$ and for every $k \in \mathbb{N}$, any SMD-estimator $\hat{\theta}_{n,k}$ minimizes $\mathbb{Q}_{n,k}$.

(c) Suppose $\zeta = 0$ and Assumptions D.1, D.2, P.1, and P.2 hold. Then, for every constant $\chi > 0$ satisfying $\inf_{x \in \Omega} p_\blacktriangle(x) > \chi$ and $\inf_{\Omega \times \Theta} p(x, \theta) > \chi$, there are events $C_{n,k} \in \mathcal{B}(\Omega)^n \otimes \mathcal{V}^k$ that have probability tending to 1 as $\min(n, k) \to \infty$ such that on $C_{n,k}$ any SMD-estimator $\hat{\theta}_{n,k}$ coincides with an SMD-estimator that is obtained from using $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$ as the underlying auxiliary model.

Proof. (a) By Proposition 43(b) in Appendix F, $\mathbb{Q}_{n,k}$ is continuous and real-valued on the compact set Θ for each $(x_1, \ldots, x_n, v_1, \ldots, v_k) \in \Omega^n \times V^k$ implying that any $\hat{\theta}_{n,k}$ is a minimizer for each $(x_1, \ldots, x_n, v_1, \ldots, v_k)$. 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).

(b) By Remark 10(i) there are events $A_n \in \mathcal{B}(\Omega)^n$ that have probability tending to 1 as $n \to \infty$ on which $\inf_{x \in \Omega} \hat{p}_n(x) > 2^{-1} \inf_{x \in \Omega} p_\blacktriangle(x) > 0$. From Proposition 43(b) it follows that $\mathbb{Q}_{n,k}$ is continuous and real-valued on Θ for each $(x_1, \ldots, x_n, v_1, \ldots, v_k) \in A_n \times V^k$. Compactness of Θ completes the proof.

(c) Let χ be as in the proposition. Set $C_{n,k} = A_n \times B_k$, where A_n and B_k are as in Remarks 10(i) and (iii), and observe that $C_{n,k}$ has probability tending to 1 as $\min(n, k) \to \infty$. By Remark 10, 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 NPML-estimators $\hat{p}_n$ and $\tilde{p}_k(\theta)$, respectively, coincide with the corresponding NPML-estimators based on the auxiliary model $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$. Therefore, on $C_{n,k}$, the objective function $\mathbb{Q}_{n,k}$ coincides with the corresponding objective function based on the auxiliary model $\mathcal{P}(t, \chi, D)$, and thus $\hat{\theta}_{n,k}$ coincides with the corresponding SMD-estimator based on the auxiliary model $\mathcal{P}(t, \chi, D)$. ■

The proofs of Parts (a) and (b) of the subsequent proposition are analogous to the proofs of Proposition 20 above. Part (c) follows immediately from compactness of Θ and Lemma 42 in Appendix F.

Proposition 21 *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) Suppose $\zeta > 0$ holds. Then any MD-estimator $\hat{\theta}_n$ minimizes $\mathbb{Q}_n$ for every $(x_1, \ldots, x_n) \in \Omega^n$. Furthermore, there exists an MD-estimator $\hat{\theta}_n$ that is $\mathcal{B}(\Omega)^n$-$\mathcal{B}(\Theta)$-measurable.

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

(c) Suppose Assumption D.2 is 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 that $\theta \mapsto p_\theta$ 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 *(a) Let Assumptions D.1, D.2, P.1, P.2, and R.1 be satisfied. If Q has a unique minimizer θ_0^* over Θ, then any SMD-estimator $\hat{\theta}_{n,k}$ converges to θ_0^* in outer probability as $\min(n, k) \to \infty$.*

(b) 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.1 and D.2 be satisfied. If Q has a unique minimizer θ_0^* over Θ, then any MD-estimator $\hat{\theta}_n$ converges to θ_0^* in outer probability as $n \to \infty$.

Proof. (a) For $\varepsilon > 0$, let
$$c(\varepsilon) = \inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [Q(\theta) - Q(\theta_0^*)].$$

Since Q is continuous by Remark 22, Proposition 30 in Appendix A, and Proposition 43(c) in Appendix F, 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 Proposition 44(b) in Appendix F 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 $\delta < c(\varepsilon)/2$. Since
$$\inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_{n,k}(\theta_0^*)] \geq \inf_{\|\theta - \theta_0^*\| \geq \varepsilon} [Q(\theta) - Q(\theta_0^*)] - 2 \sup_{\theta \in \Theta} |\mathbb{Q}_{n,k}(\theta) - Q(\theta)|$$

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 20 this implies that $\|\hat{\theta}_{n,k} - \theta_0^*\| < \varepsilon$, at least on events having inner probability tending to 1 as $\min(n, k) \to \infty$.

(b) Analogous. ∎

Remark 24 (i) It follows from Proposition 30 in Appendix A together with Remark 22 that the assumptions of Proposition 21(c) are satisfied under the assumptions of Part (a) of the above proposition (and they are trivially satisfied under the assumptions of Part (b)). Consequently, under the assumptions of the above proposition, Q always has a minimizer over Θ. Hence, the assumption in the above proposition that Q has a unique minimizer is in fact only a uniqueness assumption.

(ii) We do not strive for utmost generality in the consistency result for MD-estimators; possible relaxations lie in weakening the assumptions that $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and that θ_0^* is unique.

5.2 Asymptotic Normality of SMD-Estimators

We next show that SMD- and MD-estimators are asymptotically normally distributed, with their asymptotic variance-covariance matrix reaching the Cramér-Rao bound in case the parametric model $\mathcal{P}_\Theta$ is correctly specified. We first prove the result for MD-estimators and then show how this can be carried over to SMD-estimators. To this end we introduce a further assumption which is standard in maximum likelihood theory.

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

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

We note that under the assumptions of the subsequent theorem, as well as under the assumptions of Theorem 27, the function Q always possesses a minimizer (cf. Proposition 21(c) and Remark 22, as well as Proposition 30 in Appendix A in case of Theorem 27); furthermore, the Hessian matrix of $Q(\theta)$ exists for every $\theta \in \Theta^\circ$, cf. Lemma 46 in Appendix F which provides an explicit formula. We shall write $J(\theta)$ for $1/2$ times the Hessian matrix of $Q(\theta)$.

Theorem 25 *Let Assumptions D.3, P.1, P.2, P.4, P.5 be satisfied. Suppose that the minimizer θ_0^* of Q over Θ is unique and belongs to Θ°, and suppose that the matrix $J(\theta_0^*)$ is positive definite. Furthermore, assume that the first-order partial derivatives $\frac{\partial p}{\partial \theta_i}(\cdot, \theta_0^*)$ belong to $\mathsf{W}_2^s(\Omega)$ for some $s > 1/2$ and for all $i = 1, \ldots, m$. Then*

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

where $I(\theta_0^*)$ is given by

$$\int_\Omega \frac{\partial p}{\partial \theta}(\cdot, \theta_0^*) \frac{\partial p}{\partial \theta'}(\cdot, \theta_0^*) p_{\theta_0^*}^2 p_\blacktriangle^{-3} d\lambda - \int_\Omega \frac{\partial p}{\partial \theta}(\cdot, \theta_0^*) p_{\theta_0^*} p_\blacktriangle^{-1} d\lambda \int_\Omega \frac{\partial p}{\partial \theta'}(\cdot, \theta_0^*) p_{\theta_0^*} p_\blacktriangle^{-1} d\lambda,$$

which is well-defined and nonnegative definite. If, additionally, $\mathcal{P}_\Theta$ is correctly specified in the sense that $p_\blacktriangle = p_{\theta_0}$ a.e. for some $\theta_0 \in \Theta$, then $\theta_0^* = \theta_0$ and $I(\theta_0) = J(\theta_0)$ hold, and $I(\theta_0)$ coincides with the Fisher-information matrix.

Proof. Step 1: Assume first that $\zeta > 0$. By Proposition 23(b), $\hat{\theta}_n$ belongs to a sufficiently small open ball, centered at θ_0^* and contained in Θ°, on subsets E_n of the sample space that have inner probability tending to 1 as $n \to \infty$. Consequently,

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

holds on E_n. Applying the mean-value theorem to each component of $\partial \mathbb{Q}_n / \partial \theta$ then yields on E_n

$$\sqrt{n} \frac{\partial \mathbb{Q}_n}{\partial \theta}(\theta_0^*) + J(\theta_0^*) \sqrt{n}(\hat{\theta}_n - \theta_0^*) + (H_n - J(\theta_0^*)) \sqrt{n}(\hat{\theta}_n - \theta_0^*) = 0, \tag{29}$$

where H_n is the Hessian matrix of $\mathbb{Q}_n$ with i-th row evaluated at some mean value $\bar{\theta}_{n,i}$ on the line segment that joins θ_0^* and $\hat{\theta}_n$. Observe that H_n converges to the invertible matrix $J(\theta_0^*)$ in outer probability by Proposition 23, Proposition 47 in Appendix F, and continuity of $J(\theta)$ on Θ°

(cf. Lemma 46 in Appendix F). We next show that the score evaluated at θ_0^* satisfies a central limit theorem. To this end let $v \in \mathbb{R}^m$ be arbitrary, and use Lemma 46(a) to obtain

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

Observe that Expression *III* equals $\sqrt{n}v'(\partial Q/\partial\theta)(\theta_0^*)$ by Lemma 46(b) in Appendix F. Since θ_0^* is an interior minimizer of Q by assumption, Expression *III* is 0.

Convergence of I: By assumption $v'\frac{\partial p}{\partial\theta}(\cdot,\theta_0^*)$ belongs to $\mathsf{W}_2^s(\Omega)$ with $s > 1/2$ and is thus sup-norm bounded by $C_s\|v'\frac{\partial p}{\partial\theta}(\cdot,\theta_0^*)\|_{s,2} < \infty$. Clearly, $\left\|p_{\theta_0^*}\hat{p}_n^{-1}p_{\blacktriangle}^{-2}\right\|_\Omega \le \zeta^{-3}C_t D$ holds in view of Assumption P.1. Hence,

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

Consequently, Expression *I* converges to 0 in outer probability by Proposition 11(a) applied with $s = 0$.

Convergence of II: Set $r = \min(s,t) > 1/2$. Observe that $-2v'(\partial p/\partial\theta)(\cdot,\theta_0^*) \in \mathsf{W}_2^r(\Omega)$ by assumption, that $p_{\theta_0^*} \in \mathsf{W}_2^r(\Omega)$ by Assumption P.1, and that $p_{\blacktriangle} \in \mathsf{W}_2^r(\Omega)$ by Assumption D.1. Since $\zeta > 0$ has been assumed, it follows that

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

belongs to $\mathsf{W}_2^r(\Omega)$ in view of Proposition 1(a),(d). Applying Theorem 14(a) with $\mathcal{F} = \{f\}$ we obtain that II converges in distribution to a centered normal distribution with variance $4v'I(\theta_0^*)v$. By the Cramér-Wold device, $\sqrt{n}(\partial\mathbb{Q}_n/\partial\theta)(\theta_0^*)$ asymptotically follows a centered normal distribution with variance-covariance matrix $4I(\theta_0^*)$. Nonnegative definiteness of $I(\theta_0^*)$ is now an immediate consequence and the asymptotic distribution of $\sqrt{n}(\hat{\theta}_n - \theta_0^*)$ follows easily from (29). The claims under correct specification of the model $\mathcal{P}_\Theta$ follow easily from Lemma 46(b) in Appendix F.

Step 2: Now assume that $\zeta = 0$. Note that $\inf_{x\in\Omega}p_{\blacktriangle}(x) > 0$ and $\inf_{\Omega\times\Theta}p(x,\theta) > 0$ because of Assumptions D.3 and P.2. Let $\chi > 0$ be such that $\inf_{x\in\Omega}p_{\blacktriangle}(x) > \chi$ and $\inf_{\Omega\times\Theta}p(x,\theta) > \chi$. Then it follows from Proposition 21(b) that there are events that have probability tending to 1 such that on these events $\hat{\theta}_n$ coincides with an MD-estimator $\check{\theta}_n$ that is based on $\mathcal{P}(t,\chi,D)$ instead of $\mathcal{P}(t,\zeta,D)$. Since the assumptions of the theorem are also satisfied with $\mathcal{P}(t,\chi,D)$ instead of $\mathcal{P}(t,\zeta,D)$, applying to $\check{\theta}_n$ what has already been established in Step 1 completes the proof. ∎

The following lemma will be instrumental in proving the asymptotic normality result for SMD-estimators.

Lemma 26 *Let $U \subseteq \mathbb{R}^m$ be a (non-empty) open, convex set. Let $f : U \to \mathbb{R}$ and $g : U \to \mathbb{R}$ be functions such that g is twice partially differentiable on U with Hessian satisfying*

$$
\inf_{x\in U} y'\frac{\partial^2 g}{\partial x\partial x'}(x)y \ge K\|y\|^2 \tag{30}
$$

for all $y \in \mathbb{R}^m$ and some $0 < K < \infty$. If u is a minimizer of f over U and v is a minimizer of g over U, then

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

Proof. Suppose that minimizers u and v exist, since otherwise there is nothing to prove. As v is a minimizer of the twice partially differentiable function g on the convex open set U, we have (by a pathwise Taylor series expansion) that

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

where $\bar{v}$ lies in the convex hull of $\{u, v\} \subseteq U$. By (30) we obtain

$$\|u - v\| \leq \sqrt{2}K^{-1/2}\sqrt{|g(u) - g(v)|}. \tag{31}$$

Next, note the inequality

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

which implies

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

which in turn yields

$$|g(u) - g(v)| \leq |g(u) - f(u)| + |f(u) - g(v)| \leq 2\|f - g\|_U.$$

Plugged into (31) this proves the result. ■

The asymptotic normality result for SMD-estimators is now as follows.

Theorem 27 *Let Assumptions D.3, P.1, P.5, R.2 be satisfied. Suppose that the minimizer θ_0^* of Q over Θ is unique and belongs to Θ°, suppose that the matrix $J(\theta_0^*)$ is positive definite, and assume that the first-order partial derivatives $\frac{\partial p}{\partial \theta_i}(\cdot, \theta_0^*)$ belong to $\mathsf{W}_2^s(\Omega)$ for some $s > 1/2$ and for all $i = 1, \ldots, m$. Suppose further that either (i) Assumption P.2 is satisfied and $k(n)$ satisfies $k(n)/n^{2+1/t} \to \infty$ as $n \to \infty$; or (ii) Assumption P.3 is satisfied and $k(n)$ satisfies $k(n)/n^2 \to \infty$ as $n \to \infty$. Then*

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

where $I(\theta_0^*)$ is given as in Theorem 25, is well-defined, and is nonnegative definite. If, additionally, $\mathcal{P}_\Theta$ is correctly specified in the sense that $p_\blacktriangle = p_{\theta_0}$ a.e. for some $\theta_0 \in \Theta$, then $\theta_0^* = \theta_0$ and $I(\theta_0) = J(\theta_0)$ hold, and $I(\theta_0)$ coincides with the Fisher-information matrix.

Proof. Step 1: Assume that $\zeta > 0$. Observe first that the assumptions of the current theorem imply the assumptions of Theorem 25, noting that Assumption P.4 follows from Assumptions P.1 and R.2 in view of Proposition 30 in Appendix A. It hence suffices to prove that

$$\sqrt{n}(\hat{\theta}_{n,k(n)} - \hat{\theta}_n) = o_{\mathrm{Pr}}^*(1) \quad \text{as } n \to \infty. \tag{32}$$

We achieve this by applying Lemma 26 to the objective functions $\mathbb{Q}_{n,k}$ and $\mathbb{Q}_n$: Let U be a sufficiently small open, convex neighbourhood of θ_0^* that is contained in Θ° such that the smallest eigenvalue of $J(\theta)$ is bounded from below by a positive constant for all $\theta \in U$, the

constant not depending on θ. Such a set U exists, since $J(\theta_0^*)$ is positive definite by assumption and $J(\theta)$ is continuous on Θ° by Lemma 46 in Appendix F. Since for all $i, j = 1, \ldots, m$

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

by Proposition 47 in Appendix F, it follows that there are events E_n having probability tending to 1 as $n \to \infty$ such that on E_n

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

holds for some constant $K > 0$ which does not depend on n or the data. By Propositions 23, $\hat{\theta}_n$ and $\hat{\theta}_{n,k(n)}$ belong to U on subsets E_n' of the sample space 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 $n \to \infty$. By Proposition 21(a) and Proposition 20(a) the estimators $\hat{\theta}_n$ and $\hat{\theta}_{n,k(n)}$, respectively, minimize the objective functions $\mathbb{Q}_n$ and $\mathbb{Q}_{n,k(n)}$. Hence, we may apply Lemma 26 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 obtain

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

It follows from Proposition 44(c) in Appendix F and the choice of $k(n)$ that (32) holds under (i) as well as under (ii).

Step 2: Now assume that $\zeta = 0$. Note that $\inf_{x \in \Omega} p_{\blacktriangle}(x) > 0$ and $\inf_{\Omega \times \Theta} p(x, \theta) > 0$ because of Assumptions D.3 and P.2 (P.3, respectively). Let $\chi > 0$ be such that $\inf_{x \in \Omega} p_{\blacktriangle}(x) > \chi$ and $\inf_{\Omega \times \Theta} p(x, \theta) > \chi$. Then it follows from Proposition 21(b) and Proposition 20(c) 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 SMD-estimator $\check{\theta}_{n,k(n)}$ that is based on $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$. Since the assumptions of the theorem are also satisfied with $\mathcal{P}(t, \chi, D)$ instead of $\mathcal{P}(t, \zeta, D)$, applying to $\check{\theta}_{n,k(n)}$ what has already been established in Step 1 completes the proof. ∎

Remark 28 The preceding theorem was proved by showing that $\hat{\theta}_{n,k(n)}$ and $\hat{\theta}_n$ are sufficiently close and by applying Theorem 25. The reason for going this route instead of directly applying a mean-value expansion to the score $\partial \mathbb{Q}_{n,k(n)}/\partial \theta$ is that this would require knowledge about differentiability properties of the mapping $\theta \mapsto \tilde{p}_{k(n)}(\theta)$, which we were unable to obtain. [The usual approach to establish such differentiability properties via the implicit function theorem is not feasible here since $\tilde{p}_{k(n)}(\theta)$ falls on the boundary of $\mathcal{P}(t, \zeta, D)$ as shown in Proposition 6.] A consequence of the method of proof chosen is that we have to assume at least $k(n)/n^2 \to \infty$. It is likely, that if the more direct method of proof via expansion of the score $\partial \mathbb{Q}_{n,k(n)}/\partial \theta$ can be made to work, this would deliver asymptotic normality under weaker conditions on $k(n)$; see Nickl & Pötscher (2009) where density estimators other than NPML-estimators are used.

A   Appendix: Proofs for Sections 2 and 3

Proof of Proposition 1: For a proof of Part (a) see, e.g., Part 6 of Proposition 2 in Nickl (2007). Part (b) follows from Section 2.7.1, Formula (12), in Triebel (1983) additionally noting 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. For Part (c) see Theorem 1.16.1 in Lions & Magenes (1972). We finally prove Part (d): By the already established Part (b), $\mathcal{F}$ is a bounded subset of $\mathsf{C}^{s-1/2}(\Omega)$.

Hence, every $f \in \mathcal{F}$ has classical derivatives $f^{(\alpha)}$ at least for $0 \leq \alpha \leq \lfloor s \rfloor - 1$ (since $\lfloor s \rfloor - 1 \leq \lfloor s - 1/2 \rfloor$), and each of these derivatives is sup-norm bounded uniformly by the finite constant $c = \sup_{f \in \mathcal{F}} \|f\|_{s-1/2, \Omega}$. Furthermore, since $f^{(\lfloor s \rfloor)_w} \in \mathcal{L}^2(\Omega)$ by the assumptions, it follows from Theorem 2.1.4 in Ziemer (1989) that $f^{(\lfloor s \rfloor - 1)}$ is absolutely continuous, and hence $f^{(\lfloor s \rfloor)}$ exists a.e. and coincides with the weak derivative on a set of full Lebesgue measure. The derivatives of $1/f$ can now be computed for $0 \leq \alpha \leq \lfloor s \rfloor$ by the classical rules for differentiation, leading to

$$(1/f)^{(\alpha)} = -f^{(\alpha)}/f^2 + P_\alpha \left(1/f, f^{(1)}, \ldots, f^{(\alpha-1)} \right), \tag{33}$$

where P_α is a polynomial in α variables which is completely determined solely by α. Note that the equality in (33) holds everywhere for $\alpha < \lfloor s \rfloor$, but may only hold a.e. for $\alpha = \lfloor s \rfloor$. Let $\xi = \inf_{x \in \Omega, f \in \mathcal{F}} f(x)$, which is positive by the assumptions. It follows from the remarks at the beginning of this proof that $\left\| P_\alpha \left(1/f, f^{(1)}, \ldots, f^{(\alpha-1)} \right) \right\|_\Omega$ is bounded by a finite constant uniformly over $f \in \mathcal{F}$ for $0 \leq \alpha \leq \lfloor s \rfloor$. Consequently, the same is true for the $\mathcal{L}^2(\Omega)$-norm of $P_\alpha \left(1/f, f^{(1)}, \ldots, f^{(\alpha-1)} \right)$. Furthermore, $\left\| f^{(\alpha)}/f^2 \right\|_2 \leq \xi^{-2} \left\| f^{(\alpha)} \right\|_2$ which is bounded by a finite constant uniformly in $f \in \mathcal{F}$ for every $0 \leq \alpha \leq \lfloor s \rfloor$, since $\mathcal{F}$ is a bounded subset of $\mathsf{W}_2^s(\Omega)$. This proves the claim in case s is integer. If s is not an integer, we additionally have to establish a bound, uniformly in $f \in \mathcal{F}$, for

$$\int_\Omega \int_\Omega \left(\left(f^{(\lfloor s \rfloor)}/f^2 \right)(x) - \left(f^{(\lfloor s \rfloor)}/f^2 \right)(y) \right)^2 |x-y|^{-(1+2(s-\lfloor s \rfloor))} \, d\lambda(x) d\lambda(y) \tag{34}$$

and

$$\int_\Omega \int_\Omega \frac{\left(P_{\lfloor s \rfloor} \left(1/f, f^{(1)}, \ldots, f^{(\lfloor s \rfloor - 1)} \right)(x) - P_{\lfloor s \rfloor} \left(1/f, f^{(1)}, \ldots, f^{(\lfloor s \rfloor - 1)} \right)(y) \right)^2}{|x-y|^{(1+2(s-\lfloor s \rfloor))}} \, d\lambda(x) d\lambda(y). \tag{35}$$

Since $\mathcal{F}$ is a bounded subset of $\mathsf{C}^{s-1/2}(\Omega)$, it follows that $f^{(\alpha)}$, for $0 \leq \alpha \leq \lfloor s \rfloor - 1$, is Hölder of order at least β with Hölder constant not exceeding c where $\beta = s - 1/2 - \lfloor s - 1/2 \rfloor$ in case $\lfloor s - 1/2 \rfloor = \lfloor s \rfloor - 1$, and $\beta = 1$ in case $\lfloor s - 1/2 \rfloor = \lfloor s \rfloor$ (using the mean value theorem). Since the derivatives involved are also sup-norm bounded uniformly in $f \in \mathcal{F}$ and since $\xi > 0$ holds, it follows that $P_{\lfloor s \rfloor} \left(1/f, f^{(1)}, \ldots, f^{(\lfloor s \rfloor - 1)} \right)$ is Hölder of order β in both cases, with a Hölder constant that can be chosen independently of $f \in \mathcal{F}$. Hence, the integral (35) is bounded by a finite constant, that is independent of $f \in \mathcal{F}$, times $\iint |x-y|^\delta \, d\lambda(x) d\lambda(y)$ which is finite since $\delta = 2\beta - (1 + 2(s - \lfloor s \rfloor))$ satisfies $-1 < \delta \leq 0$ in both cases. The integral in (34) is bounded by

$$2\xi^{-4} \int_\Omega \int_\Omega \left(f^{(\lfloor s \rfloor)}(x) - f^{(\lfloor s \rfloor)}(y) \right)^2 |x-y|^{-(1+2(s-\lfloor s \rfloor))} \, d\lambda(x) d\lambda(y)$$

$$+ 8c^4 \xi^{-8} \int_\Omega \int_\Omega \left(f^{(\lfloor s \rfloor)}(y) \right)^2 |x-y|^\delta \, d\lambda(x) d\lambda(y),$$

using the fact that f is sup-norm bounded and Hölder as mentioned earlier. The first one of these two integrals now is clearly bounded by $2\xi^{-4} \sup_{f \in \mathcal{F}} \|f\|_{s,2}^2$. The second one is bounded by

$$8c^4 \xi^{-8} \sup_{f \in \mathcal{F}} \left\| f^{(\lfloor s \rfloor)} \right\|_2^2 \sup_{y \in \Omega} \int_\Omega |x-y|^\delta \, d\lambda(x) \leq 16c^4 \xi^{-8} \sup_{f \in \mathcal{F}} \|f\|_{s,2}^2 \left[(b-a)/2 \right]^{\delta+1} / (\delta+1) < \infty,$$

where a and b denote the endpoints of Ω. ∎

Proof of Proposition 2: (a) The implications (i) in (ii) and (ii) in (iii) are obvious. If p is an element of $\mathcal{P}(t, \zeta, D)$, we have $1 = \int_\Omega p \, d\lambda \geq \int_\Omega \zeta d\lambda = \zeta \lambda(\Omega)$ showing that $\zeta \leq \lambda(\Omega)^{-1}$.

31

Furthermore, the Cauchy-Schwarz inequality implies $1 = \|p\|_1 \leq \|p\|_2 \|1\|_2 \leq \|p\|_{t,2} \|1\|_2 \leq D\lambda(\Omega)^{1/2}$, which implies $\lambda(\Omega)^{-1} \leq D^2$. Thus (iii) implies (i).

(b) Suppose (i) holds. Then $\lambda(\Omega)^{-1} \in \mathcal{P}(t, \zeta, D)$ by Part (a). Suppose $p \in \mathcal{P}(t, \zeta, D)$. If now $\zeta = \lambda(\Omega)^{-1}$, then $p - \lambda(\Omega)^{-1} \geq 0$. But clearly $\int_\Omega \left(p - \lambda(\Omega)^{-1}\right) d\lambda = 0$, implying that $p = \lambda(\Omega)^{-1}$ λ-a.e., and hence everywhere by continuity of p. If $\lambda(\Omega)^{-1} = D^2$, then $\|p\|_1 = \|p\|_2 \|1\|_2$ follows from the calculations in the proof of Part (a). But this shows that p is λ-a.e., and hence everywhere by continuity of p, proportional to the constant function 1, the proportionality factor necessarily being $\lambda(\Omega)^{-1}$. This proves that (i) implies (ii). That (ii) implies (iii) is trivial. Since the constant density $\lambda(\Omega)^{-1}$ belongs to $\mathcal{P}(t, \zeta, D)$ by Part (a), (iii) is equivalent to (ii). To show that (ii) implies (i), assume that $\zeta < \lambda(\Omega)^{-1} < D^2$. Choose $\varepsilon > 0$ small enough such that $\zeta < \lambda(\Omega)^{-1} - \varepsilon$ holds. Then define f to be the restriction to Ω of the affine function that has the value $\lambda(\Omega)^{-1} - \varepsilon$ at the left endpoint of Ω and $\lambda(\Omega)^{-1} + \varepsilon$ at the right endpoint. By construction $f \in \mathsf{W}_2^t(\Omega)$, integrates to 1, satisfies $\inf_\Omega f \geq \zeta$, and $\|f\|_{t,2} \leq D$ provided ε is small enough. That is, f is a further element of $\mathcal{P}(t, \zeta, D)$, contradicting (ii).

(c) Note that $\mathcal{P}(t, \zeta, D)$ is non-empty by Part (a). 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 $\zeta = 0$.] $\blacksquare$

Proof of Proposition 3: Since (a) is a special case of (b) it suffices to prove the latter: Suppose $\mathcal{P}'$ satisfies (i) and (ii), and choose $\delta > 0$ small enough such that $\delta < D - \sup_{p \in \mathcal{P}'} \|p\|_{t,2}$ and $C_t \delta < \inf_{x \in \Omega, p \in \mathcal{P}'} p(x) - \zeta$ hold, where C_t is the constant appearing in Proposition 1. For every $p \in \mathcal{P}'$ and $f \in \mathsf{W}_2^t(\Omega)$ with $\|f\|_{t,2} \leq \delta$ we then have $\|p + f\|_{t,2} \leq \|p\|_{t,2} + \|f\|_{t,2} \leq \sup_{p \in \mathcal{P}'} \|p\|_{t,2} + \delta < D$ and $\inf_\Omega(p + f) \geq \inf_\Omega p - \sup_\Omega f \geq \inf_{x \in \Omega, p \in \mathcal{P}'} p(x) - C_t \delta > \zeta$ (for the latter using Proposition 1). This shows that $\mathcal{U}_{t,\delta}(p) \cap \mathsf{H}_t$ is a subset of $\mathcal{P}(t, \zeta, D)$ for every $p \in \mathcal{P}'$. Conversely, suppose $\mathcal{P}'$ is uniformly interior to $\mathcal{P}(t, \zeta, D)$ relative to H_t. We first establish (i): Let $\delta > 0$ be the radius figuring in the definition of being uniformly interior and let $p \in \mathcal{P}'$ be arbitrary. Choose a $q \in \mathsf{H}_t$ different from p and define $f = \delta(q - p)/(2\|q - p\|_{t,2})$. [Note that q and hence f may depend on p.] Then $f \neq 0$, $\|f\|_{t,2} = \delta/2 < \delta$, and $\int_\Omega f d\lambda = 0$ hold. Observe that $p + f$ and $p - f$ then both belong to $\mathcal{U}_{t,\delta}(p) \cap \mathsf{H}_t$ and hence to $\mathcal{P}(t, \zeta, D)$, since $\mathcal{U}_{t,\delta}(p) \cap \mathsf{H}_t \subseteq \mathcal{P}(t, \zeta, D)$ by assumption; in particular $\|p + f\|_{t,2} \leq D$ and $\|p - f\|_{t,2} \leq D$ is satisfied. Since the Sobolev-norm originates from an inner product, we have $\|p + f\|_{t,2}^2 + \|p - f\|_{t,2}^2 = 2\left[\|p\|_{t,2}^2 + \|f\|_{t,2}^2\right]$ and thus $\|p\|_{t,2}^2 \leq D^2 - \delta^2/4$. Since this is true for every $p \in \mathcal{P}'$ we obtain (i). We finally prove (ii): Let $x_n \in \Omega$ and $p_n \in \mathcal{P}'$ satisfy $p_n(x_n) \to \inf_{x \in \Omega, p \in \mathcal{P}'} p(x)$. The sequence x_n has a cluster point x_0 in the closure $\bar{\Omega}$ of the interval Ω. There exists a sufficiently small neighborhood A of x_0 in $\bar{\Omega}$ and a C^∞ function h satisfying $h(x) = -1$ for all $x \in A \cap \Omega$ (which is non-empty) as well as $\int_\Omega h d\lambda = 0$. Furthermore, h can be chosen to be bounded with all its derivatives having compact support contained in Ω; consequently, $h \in \mathsf{W}_2^t(\Omega)$. Since $\mathcal{P}'$ is uniformly interior to $\mathcal{P}(t, \zeta, D)$ relative to H_t by assumption, it follows that $p_n + \alpha h \in \mathcal{P}(t, \zeta, D)$ for sufficiently small $\alpha > 0$, where α can be chosen independently of n. Consequently, $\inf_\Omega (p_n + \alpha h) \geq \zeta$ must hold. But this implies $p_n(x_n) \geq \inf_\Omega p_n = \inf_{A \cap \Omega} p_n = \inf_{A \cap \Omega} (p_n - \alpha) + \alpha = \inf_{A \cap \Omega} (p_n + \alpha h) + \alpha \geq \inf_\Omega (p_n + \alpha h) + \alpha \geq \zeta + \alpha$, which in turn implies $\inf_{x \in \Omega, p \in \mathcal{P}'} p(x) \geq \zeta + \alpha > \zeta$. Finally, we prove Part (c): Note that $\lambda(\Omega)^{-1} \in \mathcal{P}(t, \zeta, D)$ by Proposition 2. It is interior to $\mathcal{P}(t, \zeta, D)$ relative to H_t by Part (a) of the current proposition and the assumption $\zeta < \lambda(\Omega)^{-1} < D^2$. The second claim then follows from Theorem V.2.1. in Dunford and Schwartz (1966). $\blacksquare$

Proof of Proposition 4: To show that (v) implies (vi), it suffices, in light of Part (c) 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 (c) 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) and keeping in mind that p and p^* are continuous, 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 (b) 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 (b) 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. ∎

Remark 29 We note that $\mathcal{P}(t, \zeta, D)$ can equivalently be written as

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

because $p - \lambda^{-1}(\Omega)$ and 1 are orthogonal in $\mathsf{W}_2^t(\Omega)$. As a consequence, $\mathcal{P}(t, 0, D) = \mathcal{P}(t, \zeta, D)$ at least for all $0 \leq \zeta \leq \lambda^{-1}(\Omega) - C_t \left(D^2 - \lambda^{-1}(\Omega) \right)^{1/2}$, since $p \in \mathcal{P}(t, 0, D)$ implies $\inf_{x \in \Omega} p(x) \geq \zeta$ for such ζ by Proposition 1(b).

Assumptions on the density functions in the class $\mathcal{P}_\Theta$ 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 30 *If Assumption P.1 is satisfied, then Assumption R.1 implies Assumption P.4. However, in general Assumption R.1 does not imply Assumption P.4.*

Proof. The first claim is proved as follows: Let $F(z, \theta) = \int_{\{x \in \Omega: \, x \leq z\}} p_\theta \, d\lambda$ be 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. Now Assumption R.1 implies that $\rho(\cdot, \theta_n)$ converges to $\rho(\cdot, \theta)$ in distribution under μ. Noting that $F(\cdot, \theta)$ and $F(\cdot, \theta_n)$ are the distribution functions of $\rho(\cdot, \theta)$ and $\rho(\cdot, \theta)$, respectively, as well as noting that $F(\cdot, \theta)$ is continuous in its first argument, it follows that $F(z, \theta_n)$ converges to $F(z, \theta)$ for every $z \in \Omega$. By Assumption P.1 and sup-norm compactness of $\mathcal{P}(t, \zeta, D)$ it follows that every subsequence $p_{\theta_{n'}}$ of p_{θ_n} has a further subsequence $p_{\theta_{n''}}$ that converges to an element $p^* \in \mathcal{P}(t, \zeta, D)$ in the sup-norm. But this clearly implies that $F(z, \theta_{n''})$ converges to $\int_{\{x \in \Omega: \, x \leq z\}} p^* \, d\lambda$ for every $z \in \Omega$. It follows that $p^* = p_\theta$ a.e., hence everywhere on Ω by continuity of p_θ and p^*. This proves the first claim. For a proof of the second claim see Proposition 5 in Gach (2010). ∎

B Appendix: Properties of the Non-Parametric Likelihood Function

Proposition 31

(a) For every non-negative $\mathcal{B}(\Omega)$-measurable real-valued function f the map $(x_1, \ldots, x_n) \mapsto L_n(f; x_1, \ldots, x_n)$ is $\mathcal{B}(\Omega)^n$-$\mathcal{B}([-\infty, \infty))$-measurable, and the map $(v_1, \ldots, v_k) \mapsto L_k(\theta, f; v_1, \ldots, v_k)$ is $\mathcal{V}^k$-$\mathcal{B}([-\infty, \infty))$-measurable for every $\theta \in \Theta$.

(b) Let $\mathcal{F}$ be a set of non-negative bounded real-valued functions on Ω.

(b1) Then, for every $(x_1, \ldots, x_n) \in \Omega^n$, $f \mapsto L_n(f; x_1, \ldots, x_n)$ is a continuous map from $(\mathcal{F}, \|\cdot\|_\Omega)$ to $[-\infty, \infty)$. The same is true for the map $f \mapsto L_k(\theta, f; v_1, \ldots, v_k)$ for every $\theta \in \Theta$ and every $(v_1, \ldots, v_k) \in V^k$.

(b2) If the elements $f \in \mathcal{F}$ are additionally also continuous and Assumption R.1 is satisfied, then, for every $(v_1, \ldots, v_k) \in V^k$, $(\theta, f) \mapsto L_k(\theta, f; v_1, \ldots, v_k)$ is a continuous map from $\Theta \times (\mathcal{F}, \| \cdot \|_\Omega)$ to $[-\infty, \infty)$.

(c) 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.

(c1) Then $L(f)$ is a continuous real-valued function on $(\mathcal{F}, \| \cdot \|_\Omega)$. The same is true for $L(\theta, f)$ for every given $\theta \in \Theta$.

(c2) If the elements $f \in \mathcal{F}$ are additionally also continuous and Assumption R.1 is satisfied, then $L(\theta, f)$ is a continuous real-valued function on $\Theta \times (\mathcal{F}, \| \cdot \|_\Omega)$.

(d) Let $\mathcal{F}$ be a sup-norm compact set of non-negative bounded $\mathcal{B}(\Omega)$-measurable real-valued functions on Ω that are uniformly bounded away from 0.

(d1) Then

$$\lim_{n \to \infty} \sup_{f \in \mathcal{F}} |L_n(f) - L(f)| = 0 \quad \mathbb{P}\text{-}a.s.,$$

and, for every $\theta \in \Theta$,

$$\lim_{k \to \infty} \sup_{f \in \mathcal{F}} |L_k(\theta, f) - L(\theta, f)| = 0 \quad \mu\text{-}a.s.$$

(d2) If the elements $f \in \mathcal{F}$ are additionally also continuous and Assumption R.1 is satisfied, then

$$\lim_{k \to \infty} \sup_{\Theta \times \mathcal{F}} |L_k(\theta, f) - L(\theta, f)| = 0 \quad \mu\text{-}a.s.$$

(In Part (d) we use the convention that the supremum is 0 if $\mathcal{F}$ is empty.)

Proof. (a) The first claim 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 the second claim additionally use that $\rho : V \times \Theta \to \Omega$ is $\mathcal{V}$-$\mathcal{B}(\Omega)$-measurable in the first argument for every $\theta \in \Theta$.

(b) To prove the first claim in Part (b1), fix $(x_1, \ldots, x_n) \in \Omega^n$. Let $f_l, f \in \mathcal{F}$ be such that $\|f_l - f\|_\Omega$ converges to 0. Since setting $\log 0 = -\infty$ continuously extends the logarithm to the interval $[0, \infty)$, $\log f_l(x_i)$ then converges to $\log f(x_i)$ for every i, thus establishing the first claim. The second claim in Part (b1) is proved analogously. To prove Part (b2), fix $(v_1, \ldots, v_k) \in V^k$ and 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 for every i

$$
\begin{aligned}
|f_l(\rho(v_i, \theta_l)) - f(\rho(v_i, \theta))| &\leq |f_l(\rho(v_i, \theta_l)) - f(\rho(v_i, \theta_l))| + |f(\rho(v_i, \theta_l)) - f(\rho(v_i, \theta))| \\
&\leq \|f_l - f\|_\Omega + |f(\rho(v_i, \theta_l)) - f(\rho(v_i, \theta))|.
\end{aligned}
\tag{36}
$$

The first expression on the r.h.s. of (36) converges to 0 by hypothesis. Making use of Assumption R.1 and the continuity of f, the second one converges to 0 as well. Continuity of the extended logarithm on $[0, \infty)$ delivers Part (b2).

(c) To prove the first claim in Part (c1), denote by $\xi > 0$ the lower uniform bound of all elements in $\mathcal{F}$. 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

$$\int_\Omega \sup_{l \in \mathbb{N}} |\log f_l(x)| \, d\mathbb{P}(x) < \infty$$

is satisfied. By the already established Part (b1) (with $n = 1$), $\log f_l(x)$ converges to $\log f(x)$ for every $x \in \Omega$. The first claim then follows from the theorem of dominated convergence. The second claim in Part (c1) is proved in exactly the same manner. To prove Part (c2), 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 the same argument as before, 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 the already established Part (b2) (with $k = 1$), $\log f_l(\rho(v, \theta_l))$ converges to $\log f(\rho(v, \theta))$ for every $v \in V$. Part (c2) then follows from the theorem of dominated convergence.

(d) To prove the first claim in Part (d1), we use Mourier's strong law of large numbers as given in Corollary 7.10 of Ledoux & Talagrand (1991) with the separable Banach space $(B, \|\cdot\|)$ given by $(\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega), \|\cdot\|_\mathcal{F})$ and the mapping X given by $X(f) = \log f(X_1) - \int_\Omega \log f \, d\mathbb{P}$ for $f \in \mathcal{F}$. Note that X has values in $\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$ by using the already established Parts (b1) and (c1) in conjunction with the assumed sup-norm compactness of $\mathcal{F}$. Clearly, $X(f)$ is a random variable for every $f \in \mathcal{F}$, and hence X is measurable with respect to the σ-field on $\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$ that is generated by the point-evaluations. Since this σ-field coincides with the Borel σ-field on $\mathsf{C}(\mathcal{F}, \|\cdot\|_\Omega)$ (see, e.g., Problem 1 in Section 1.7 in van der Vaart & Wellner (1996) and observe that $(\mathcal{F}, \|\cdot\|_\Omega)$ is a compact metric space), X is a Borel random mapping. The integrability condition $\mathrm{E}\|X\| < \infty$ follows from

$$\int_\Omega \sup_{f \in \mathcal{F}} |\log f(x)| d\mathbb{P}(x) < \infty,$$

which is true since the elements of $\mathcal{F}$ are uniformly bounded and uniformly bounded away from 0 by hypothesis. The second claim in Part (d1) is proved completely analogously. Part (d2) is proved in a similar manner: Apply Corollary 7.10 in Ledoux & Talagrand (1991) with B the separable Banach space of all bounded, continuous functions on $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$ equipped with the sup-norm $\|\cdot\|_{\Theta \times \mathcal{F}}$ and with X given by $X(\theta, f) = \log f(\rho(V_1, \theta)) - \int_V \log f(\rho(\cdot, \theta)) d\mu$. Note that by the already established Parts (b2) and (c2) in conjunction with compactness of $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$, X takes its values in the space of (bounded) continuous functions on $\Theta \times (\mathcal{F}, \|\cdot\|_\Omega)$. Again X is a Borel random mapping. The integrability condition $\mathrm{E}\|X\| < \infty$ now follows from

$$\int_V \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} |\log f(\rho(v, \theta))| d\mu(v) < \infty,$$

which is true since the elements of $\mathcal{F}$ are uniformly bounded and uniformly bounded away from 0 by hypothesis. $\blacksquare$

Proof of Lemma 8: (a) It is sufficient to show that

$$\int_\Omega (\log(p_\blacktriangle))^- \, d\mathbb{P} = \int_{\{x \in \Omega : \, p_\blacktriangle(x) > 0\}} (\log(p_\blacktriangle))^- \, p_\blacktriangle d\lambda < \infty.$$

By Assumption D this is equivalent to showing that

$$\int_{\{x \in \Omega : \, 0 < p_\blacktriangle(x) \leq 1\}} h(p_\blacktriangle) d\lambda < \infty, \tag{37}$$

where $h(y)$ is defined by $h(y) = -y \log y$ for every $y \in (0, 1]$. Since $h : (0, 1] \to [0, \infty)$ can be continuously extended to $[0, 1]$ by setting $h(0) = 0$, it is bounded on the compact interval $[0, 1]$, and a fortiori on $(0, 1]$. But this establishes (37) since $\lambda(\Omega) < \infty$ and thus completes the proof for L. The proof for $L(\theta, \cdot)$ is analogous upon observing that

$$\int_V (\log p_\theta(\rho(\cdot, \theta)))^- \, d\mu = \int_{\{x \in \Omega : \, p(x, \theta) > 0\}} (\log(p_\theta))^- \, p_\theta d\lambda \tag{38}$$

35

by the change of variable theorem.

(b) For any $p \in \mathcal{P}(t, \zeta, D)$ different from $p_\blacktriangle$, the set $\{x \in \Omega : p(x) \neq p_\blacktriangle(x) > 0\}$ has positive $\mathbb{P}$-probability since p and $p_\blacktriangle$ are continuous functions on Ω. In view of the already established Part (a) the expression $L(p) - L(p_\blacktriangle)$ is well-defined, and the strict Jensen inequality gives

$$L(p) - L(p_\blacktriangle) = \int_{\{x \in \Omega : p_\blacktriangle(x) > 0\}} \log \frac{p}{p_\blacktriangle} d\mathbb{P} < \log \int_{\{x \in \Omega : p_\blacktriangle(x) > 0\}} \frac{p}{p_\blacktriangle} d\mathbb{P} \leq 0.$$

(c) Follows similarly to Part (b) in view of the representation

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

■

Part (a) of the following proposition is essentially given in Proposition 3 in Nickl (2007). [We note that the set $\mathcal{V}$ defined there is not sup-norm open as implicitly claimed, the apparently intended definition in the notation of Nickl (2007) being $\mathcal{V} = \{d \in \mathsf{L}^\infty(\Omega) : \inf_{x \in \Omega} d(x) > \zeta/2\}$. Inspection of the proof shows that this proposition remains correct for $\zeta = 0$.] The proof for Part (b) is completely analogous.

Proposition 32 *Define* $\mathcal{U} = \{f \in \mathsf{L}^\infty(\Omega) : \inf_{x \in \Omega} f(x) > 0\}$. *Let* α *be a positive integer,* $f \in \mathcal{U}$, *and* $f_1, \ldots, f_\alpha \in \mathsf{L}^\infty(\Omega)$.
(a) The α-th Fréchet derivatives of $L_n : \mathcal{U} \to \mathbb{R}$ *and* $L : \mathcal{U} \to \mathbb{R}$ *are given by*

$$\mathbf{D}^\alpha L_n(f)(f_1, \ldots, f_\alpha) = (-1)^{\alpha-1}(\alpha-1)! \mathbb{P}_n(f^{-\alpha} f_1 \cdots f_\alpha),$$

$$\mathbf{D}^\alpha L(f)(f_1, \ldots, f_\alpha) = (-1)^{\alpha-1}(\alpha-1)! \mathbb{P}(f^{-\alpha} f_1 \cdots f_\alpha).$$

(b) The α-th partial Fréchet derivatives 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, for* $\theta \in \Theta$, *given by*

$$\mathbf{D}^\alpha L_k(\theta, f)(f_1, \ldots, f_\alpha) = (-1)^{\alpha-1}(\alpha-1)! \mu_k(f^{-\alpha}(\rho(\cdot, \theta)) f_1(\rho(\cdot, \theta)) \cdots f_\alpha(\rho(\cdot, \theta))),$$

$$\begin{aligned}\mathbf{D}^\alpha L(\theta, f)(f_1, \ldots, f_\alpha) &= (-1)^{\alpha-1}(\alpha-1)! \mu(f^{-\alpha}(\rho(\cdot, \theta)) f_1(\rho(\cdot, \theta)) \cdots f_\alpha(\rho(\cdot, \theta))) \\ &= (-1)^{\alpha-1}(\alpha-1)! \int_\Omega f^{-\alpha} f_1 \cdots f_\alpha \, p_\theta d\lambda.\end{aligned}$$

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 that hold uniformly in θ and p.

Proposition 33 *Let* α *be a positive 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 Assumption R.2 and* $\zeta > 0$ *are satisfied, then*

$$\sup_{\Theta \times \mathcal{P}(t, \zeta, D)} \|\mathbf{D}^\alpha L_k(\theta, p) - \mathbf{D}^\alpha L(\theta, p)\|_{\mathcal{H}_1 \times \cdots \times \mathcal{H}_\alpha} = O_\mu(k^{-1/2}) \quad \text{as } k \to \infty. \tag{39}$$

Proof. Note that

$$\sup_{\Theta \times \mathcal{P}(t, \zeta, D)} \|\mathbf{D}^\alpha L_k(\theta, p) - \mathbf{D}^\alpha L(\theta, p)\|_{\mathcal{H}_1 \times \cdots \times \mathcal{H}_\alpha} = (\alpha-1)! \|\mu_k - \mu\|_{\mathcal{H}^*}$$

36

by Proposition 32, where $\mathcal{H}^* = \{h(\rho(\cdot, \theta)) : h \in \mathcal{H}, \theta \in \Theta\}$ and

$$\mathcal{H} = \left\{ p^{-\alpha} h_1 \cdot \ldots \cdot h_\alpha : p \in \mathcal{P}(t, \zeta, D), h_1 \in \mathcal{H}_1, \ldots, h_\alpha \in \mathcal{H}_\alpha \right\}.$$

Since $\zeta > 0$, the class $\mathcal{H}$ is a bounded subset of the Sobolev-space $\mathsf{W}_2^r(\Omega)$ with $r = \min(t, s) > 1/2$ by Proposition 1. Measurability of the supremum on the l.h.s. of (39) now follows immediately from Proposition 38 in Appendix D. The class $\mathcal{H}^*$ is μ-Donsker by an application of Proposition 12(a), hence $\|\mu_k - \mu\|_{\mathcal{H}^*}$ is bounded in probability at rate $k^{-1/2}$ by Prohorov's theorem. ∎

The following lemma is a special case of Berge's (1963) maximum theorem.

Lemma 34 *Let X be a metrizable space and Y a compact metrizable space. Let $u : X \times Y \to [-\infty, \infty)$ be a continuous function that has a unique maximizer, say $v(x)$, on the fiber $\{(x, y) : y \in Y\}$ for every $x \in X$. Then the mapping $v : X \to Y$ is continuous.*

C   Appendix: Proofs for Section 4.2

The following lemma is a consequence of Birman and Solomyak (1967), cf. Lorentz, v.Golitschek, and Makovoz (1996), p. 506. It can also be obtained from Theorem 1 in Nickl & Pötscher (2007) via a retraction argument; see Gach (2010).

Lemma 35 *Let $\mathcal{F}$ be a bounded subset of the Sobolev space $\mathsf{W}_2^s(\Omega)$ of order $s > 1/2$. Then the sup-norm metric entropy of $\mathcal{F}$ satisfies*

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

Proof of Proposition 12: (a) Choose a real number $r \leq s$ satisfying $1/2 < r < 3/2$ and $2r - 1 \leq a$, where a is as in Assumption R.2. Then $\mathcal{F}$ can also be viewed as a bounded subset of $\mathsf{W}_2^r(\Omega)$, and hence of $\mathsf{C}^{r-1/2}(\Omega)$, in view of Proposition 1(b),(c). We use this to obtain

$$\sup_{f \in \mathcal{F}} |f(\rho(v, \theta')) - f(\rho(v, \theta))| \leq L_r |\rho(v, \theta') - \rho(v, \theta)|^{r-1/2} \leq L_r \left[R(v) \|\theta' - \theta\|^\gamma \right]^{r-1/2}$$

for some finite constant $L_r > 0$ and all $v \in V$, all $\theta, \theta' \in \Theta$, where we have made use of Assumption R.2. A cover of $\mathcal{F}^*$ is obtained from suitable covers of Θ and $\mathcal{F}$ as follows: Fix $\varepsilon > 0$ and set $\delta(\varepsilon) = (\varepsilon/L_r)^{1/\nu}$, where $\nu := \gamma(r - 1/2)$. To cover Θ, note that it is contained in an m-cube of edge length l and thus in the union of at most $\lceil l\sqrt{m}/\delta(\varepsilon) \rceil^m$-many closed Euclidean balls $B(\theta_i, \delta(\varepsilon))$ with centers $\theta_i \in \Theta$ and radius $\delta(\varepsilon)$, where $\lceil x \rceil$ denotes the smallest integer not less than x. 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 brackets

$$\left[f_j(\rho(\cdot, \theta_i)) - R^{\nu/\gamma}(\cdot)\varepsilon - 2\varepsilon, f_j(\rho(\cdot, \theta_i)) + R^{\nu/\gamma}(\cdot)\varepsilon + 2\varepsilon \right] \tag{40}$$

with $i = 1, \ldots, \lceil l\sqrt{m}/\delta(\varepsilon) \rceil^m$ and $j = 1, \ldots, N(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Omega), \|\cdot\|_\Omega)$ provide a cover of $\mathcal{F}^*$. To see this, 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,

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

37

Now,

$$h(v) \leq f_j(\rho(v,\theta)) + 2\varepsilon \quad \leq \quad f_j(\rho(v,\theta_i)) + |f_j(\rho(v,\theta)) - f_j(\rho(v,\theta_i))| + 2\varepsilon$$
$$\leq \quad f_j(\rho(v,\theta_i)) + R^{\nu/\gamma}(v)\varepsilon + 2\varepsilon$$

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

$$f_j(\rho(v,\theta_i)) - R^{\nu/\gamma}(v)\varepsilon - 2\varepsilon \leq h(v).$$

By construction of r, we have that $\int_V \left(R^{\nu/\gamma}\right)^2 d\mu < \infty$, and hence the $\mathcal{L}^2(\mu)$-bracketing size of any of the brackets in (40) can be bounded by ε times a positive constant c that only depends on R, r, and μ. Using the elementary inequality $\lceil x \rceil^m \leq \max(1, (2x)^m)$ this leads to the relationship

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

Apply Lemma 35 to get

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

which proves (15). The claim that $\mathcal{F}^*$ is μ-Donsker now follows from Ossiander's central limit theorem (see Theorem 7.2.1 in Dudley, 1999) since clearly $\mathcal{F}^* \subseteq \mathcal{L}^2(V, \mathcal{V}, \mu)$ holds.

(b) For any fixed $\varepsilon > 0$, we take for $\mathcal{F}^*$ the cover given in (40). Since the elements of $\mathcal{F}$ are bounded below by $\chi > 0$, the sets

$$\left[\log\max(\chi, f_j(\rho(\cdot,\theta_i)) - R^{\nu/\gamma}(\cdot)\varepsilon - 2\varepsilon), \log(f_j(\rho(\cdot,\theta_i)) + R^{\nu/\gamma}(\cdot)\varepsilon + 2\varepsilon)\right],$$

for $i = 1, \ldots, \lceil l\sqrt{m}/\delta(\varepsilon)\rceil^m$, $j = 1, \ldots, N(\varepsilon, \mathcal{F}, \mathsf{W}_2^s(\Omega), \|\cdot\|_\Omega)$ are non-empty brackets and cover $\log\mathcal{F}^*$. Since the logarithm is Lipschitz on $[\chi, \infty)$ with Lipschitz constant χ^{-1}, the $\mathcal{L}^2(\mu)$-bracketing size of these brackets can be bounded by χ^{-1} times the $\mathcal{L}^2(\mu)$-bracketing size of the corresponding brackets given in (40). Arguing now as in the proof of Part (a) completes the proof. ∎

D Appendix: Measurability Issues and An Auxiliary Result

Lemma 36 *Suppose $t > 1/2$. Then the Borel σ-fields $\mathcal{B}(\mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)$, and $\mathcal{B}(\mathsf{W}_2^t(\Omega), \|\cdot\|_{s,2})$ for $0 \leq s \leq t$ all coincide. In particular, the norms $\|\cdot\|_\Omega$ and $\|\cdot\|_{s,2}$ for $0 \leq s \leq t$ are $\mathcal{B}(\mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)$-measurable.*

Proof. Since the $\|\cdot\|_\Omega$-topology on $\mathsf{W}_2^t(\Omega)$ is coarser than the $\|\cdot\|_{s,2}$-topology on $\mathsf{W}_2^t(\Omega)$, which in turn is coarser than the $\|\cdot\|_{t,2}$-topology on $\mathsf{W}_2^t(\Omega)$ (cf. Proposition 1), it suffices to show that $\mathcal{B}(\mathsf{W}_2^t(\Omega), \|\cdot\|_{t,2}) \subseteq \mathcal{B}(\mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)$. The former σ-field is generated by the collection of all closed $\|\cdot\|_{t,2}$-balls since $(\mathsf{W}_2^t(\Omega), \|\cdot\|_{t,2})$ is separable. As shown in the proof of Lemma 3 in Nickl (2007), these balls are $\|\cdot\|_\Omega$-compact and hence belong to $\mathcal{B}(\mathsf{W}_2^t(\Omega), \|\cdot\|_\Omega)$. ∎

Proposition 37 *(a) The quantities $\|\hat{p}_n - p_\blacktriangle\|_\Omega$, $\|\hat{p}_n - p_\blacktriangle\|_{s,2}$ for $0 \leq s \leq t$, $\|\tilde{p}_k(\theta) - p_\theta\|_\Omega$, and $\|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ for $0 \leq s \leq t$ are random variables.*

(b) Suppose Assumptions P.1 and R.1 are satisfied. Then $\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega$ and $\sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_{s,2}$ for $0 \leq s < t$ are random variables.

Proof. (a) Follows immediately from Theorem 6 and Lemma 36. (b) By Assumption R.1 and Proposition 30 in Appendix A the parameterization $\theta \mapsto p_\theta(x)$ is continuous, and hence is continuous in the $\|\cdot\|_\Omega$- and $\|\cdot\|_{s,2}$-norms ($0 \le s < t$) in view of Assumption P.1 and Proposition 4. By Theorem 6(b) and again Proposition 4 $\theta \mapsto \tilde{p}_k(\theta) - p_\theta$ is then continuous in the same norms. Since Θ is separable, (b) follows from Part (a). ∎

Proposition 38 *Suppose $s > 1/2$.*
(a) Then

$$\mathfrak{X}_n(\breve{x}, f) = \sqrt{n}\left(\int_\Omega \hat{p}_n(\cdot; x_1, \ldots, x_n) f(\cdot) d\lambda - \mathbb{P}(f)\right)$$

and

$$\mathfrak{Y}_n(\breve{x}, f) = n^{-1/2} \sum_{i=1}^n (f(x_i) - \mathbb{P}(f))$$

are Borel measurable on Ω^n for every $f \in \mathsf{W}_2^s(\Omega)$, where $\breve{x}$ denotes $(x_1, \ldots, x_n) \in \Omega^n$. Furthermore, if $\mathcal{F}$ is a non-empty bounded subset of $\mathsf{W}_2^s(\Omega)$, then $\sup_{f \in \mathcal{F}} |\mathfrak{Z}_n(\breve{x}, f)|$ is Borel measurable on Ω^n, where $\mathfrak{Z}_n$ stands for any of $\mathfrak{X}_n$, $\mathfrak{Y}_n$, and $\mathfrak{X}_n - \mathfrak{Y}_n$.
(b) Then

$$\mathfrak{U}_k(\breve{v}, \theta, f) = \sqrt{k} \int_\Omega \left(\tilde{p}_k(\theta)(\cdot; v_1, \ldots, v_k) - p_\theta(\cdot)\right) f(\cdot) d\lambda$$

and

$$\mathfrak{V}_k(\breve{v}, \theta, f) = k^{-1/2} \sum_{i=1}^k \left(f(\rho(v_i, \theta)) - \mu(f(\rho(\cdot, \theta)))\right)$$

are Borel measurable on V^k for every $\theta \in \Theta$ and every $f \in \mathsf{W}_2^s(\Omega)$, where $\breve{v}$ denotes $(v_1, \ldots, v_k) \in V^k$. Furthermore, if Assumption R.1 is satisfied and $\mathcal{F}$ is a non-empty bounded subset of $\mathsf{W}_2^s(\Omega)$, then $\sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} |\mathfrak{V}_k(\breve{v}, \theta, f)|$ is Borel measurable on V^k; if, additionally, Assumption P.1 holds, then $\sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} |\mathfrak{W}_k(\breve{v}, \theta, f)|$ is Borel measurable on V^k, where $\mathfrak{W}_k$ stands for any of $\mathfrak{U}_k$ and $\mathfrak{U}_k - \mathfrak{V}_k$.
(c) Then

$$\mathfrak{T}_k(\breve{v}, \theta, f) = k^{-1} \sum_{i=1}^k \tilde{p}_k^{-1}(\theta)(\rho(v_i, \theta); v_1, \ldots, v_k) f(\rho(v_i, \theta))$$

is Borel measurable on V^k for every $\theta \in \Theta$ and every $f \in \mathsf{W}_2^s(\Omega)$. Furthermore, if Assumption R.1 is satisfied, $\mathcal{F}$ is a non-empty bounded subset of $\mathsf{W}_2^s(\Omega)$, and $\zeta > 0$ holds, then $\sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} |\mathfrak{T}_k(\breve{v}, \theta, f)|$ is Borel measurable on V^k.

Proof. (a) Since $(x_1, \ldots, x_n) \mapsto \hat{p}_n(\cdot; x_1, \ldots, x_n)$ is a measurable map from Ω^n into $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$ by Theorem 6, since the map $p \mapsto \sqrt{n}\left(\int pf d\lambda - \mathbb{P}(f)\right)$ is $\|\cdot\|_\Omega$-continuous on $\mathcal{P}(t, \zeta, D)$ for every $f \in \mathsf{W}_2^s(\Omega)$, and since every f is clearly Borel measurable, we see that $\mathfrak{X}_n(\breve{x}, f)$ as well as $\mathfrak{Y}_n(\breve{x}, f)$ are Borel measurable on Ω^n for every $f \in \mathsf{W}_2^s(\Omega)$. Furthermore, it is easy to see that $\mathfrak{X}_n(\breve{x}, f)$ and $\mathfrak{Y}_n(\breve{x}, f)$, and thus also $\mathfrak{X}_n(\breve{x}, f) - \mathfrak{Y}_n(\breve{x}, f)$, are continuous on $(\mathcal{F}, \|\cdot\|_\Omega)$ for given $\breve{x}$. Since $(\mathcal{F}, \|\cdot\|_\Omega)$ is clearly separable, Borel measurability of the suprema in Part (a) follows.

(b) The first claim is proved completely analogous, making also use of the fact that ρ is measurable in its first argument. The second claim is also proved analogously by showing that now $\mathfrak{U}_k(\breve{v}, \theta, f)$ and $\mathfrak{V}_k(\breve{v}, \theta, f)$ are continuous on the separable space $(\Theta \times \mathcal{F}, \|\cdot\| + \|\cdot\|_\Omega)$ for given $\breve{v}$: for $\mathfrak{V}_k$ use that $\theta \mapsto \rho(v, \theta)$ is continuous on Θ by Assumption R.1 and that $\mathcal{F}$ is a sup-norm bounded set of continuous functions. For $\mathfrak{U}_k$ use the fact that $\theta \mapsto \tilde{p}_k(\theta)$ as a mapping

from Θ into the space $(\mathcal{P}(t, \zeta, D), \|\cdot\|_\Omega)$ is continuous by Theorem 6, and that the same is true for p_θ in view of Assumption P.1, Proposition 30 in Appendix A, and Remark 5.

(c) Measurability of $\mathfrak{T}_k(\cdot, \theta, f)$ for $\theta \in \Theta$ and $f \in \mathsf{W}_2^s(\Omega)$ follows from measurability of f and $\rho(\cdot, \theta)$ and Remark 7(i). Continuity of $\mathfrak{T}_k(\breve{v}, \cdot, \cdot)$ on the separable space $(\Theta \times \mathcal{F}, \|\cdot\| + \|\cdot\|_\Omega)$ follows from continuity of $\tilde{p}_k(\theta)(\cdot; v_1, \ldots, v_k)$ and $f(\cdot)$, Assumption R.1, and $\zeta > 0$. ∎

Lemma 39 *Let T be a non-empty compact metric space, and (S, d) a metric space. For every $\tau \in T$ and $n \in \mathbb{N}$, let $Y_n(\tau)$ be a (not necessarily measurable) mapping from a probability space $(\Lambda_n, \mathcal{A}_n, P_n)$ into S; furthermore, for every $\tau \in T$, let $Y(\tau)$ be a measurable mapping from a probability space $(\Lambda, \mathcal{A}, P)$ into S that has separable range. If $Y(\tau_n) \rightsquigarrow Y(\tau)$ whenever $\tau_n, \tau \in T$ are such that τ_n converges to τ, then the following statements are equivalent:*

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

(ii) For any $\tau_n, \tau \in T$ such that τ_n converges to τ, $Y_n(\tau_n) \rightsquigarrow Y(\tau)$.

Proof. We first prove that (i) implies (ii). Suppose that $\tau_n \in T$ converges to $\tau \in T$. Observe that

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

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 $Y(\tau_n) \rightsquigarrow Y(\tau)$. Hence, the l.h.s. converges to 0, which just means that $Y_n(\tau_n) \rightsquigarrow Y(\tau)$ by Theorem 3.6.4 in Dudley (1999).

That (ii) implies (i) is seen as follows: Let $\tau_n \in T$ be a sequence such that the distance between $\beta_{(S,d)}(Y_n(\tau_n), Y(\tau_n))$ and $\sup_{\tau \in T} \beta_{(S,d)}(Y_n(\tau), Y(\tau))$ converges to zero. Since T is compact, we can find for every subsequence of τ_n a subsubsequence, $\tau_{n'}$ say, that converges to some $\tau \in T$. It follows that the l.h.s. of the inequality

$$\beta_{(S,d)}(Y_{n'}(\tau_{n'}), Y(\tau_{n'})) \leq \beta_{(S,d)}(Y_{n'}(\tau_{n'}), Y(\tau)) + \beta_{(S,d)}(Y(\tau_{n'}), Y(\tau))$$

converges to 0 because the r.h.s. does so in view of (ii) (applied to the sequence $\bar{\tau}_n$ given by $\tau_{n'}$ when $n = n'$ and by τ otherwise) and since $Y(\tau_{n'}) \rightsquigarrow Y(\tau)$ by hypothesis. ∎

E Appendix: Uniform Rates of Convergence and Entropy Bounds for Empirical Processes

The subsequent theorem is a uniform version of Theorem 3.2.5 in van der Vaart & Wellner (1996).

Theorem 40 *Let $(\Lambda, \mathcal{A}, P)$ be a probability space, S and T non-empty sets, and let d be a non-negative real-valued function on $T \times T$. Consider a sequence of real-valued stochastic processes $(H_k(\sigma, \tau) : \sigma \in S, \tau \in T)$ defined on $(\Lambda, \mathcal{A})$ and a function $H : S \times T \to \mathbb{R}$ with the property that for every $\sigma \in S$ there exists a $\tau(\sigma) \in T$ such that for all $\tau \in T$*

$$H(\sigma, \tau) - H(\sigma, \tau(\sigma)) \leq -Cd^\alpha(\tau, \tau(\sigma)) \tag{41}$$

holds, where $C, \alpha > 0$ are constants neither depending on σ nor τ. Suppose, for all $\delta > 0$,

$$\mathrm{E}^* \sup_{\sigma \in S} \sup_{\tau \in T, d(\tau, \tau(\sigma)) \leq \delta} \sqrt{k} \, |(H_k - H)(\sigma, \tau) - (H_k - H)(\sigma, \tau(\sigma))| \leq \varphi_k(\delta) \tag{42}$$

is satisfied for real-valued functions φ_k such that for some $\beta < \alpha$ the functions $\delta \mapsto \delta^{-\beta}\varphi_k(\delta)$ are all non-increasing in δ. Assume further that, for every $\sigma \in S$, $\hat{\tau}_k(\sigma) : \Lambda \to T$ satisfies

$$H_k(\sigma, \hat{\tau}_k(\sigma)) \geq H_k(\sigma, \tau) \quad \text{for all } \tau \in T, \tag{43}$$

and let r_k be a sequence of positive reals such that

$$\sup_{k \in \mathbb{N}} \frac{r_k^{\alpha}\varphi_k(r_k^{-1})}{\sqrt{k}} < \infty. \tag{44}$$

Then, for every $\sigma \in S$, $\tau(\sigma)$ is a maximizer of $H(\sigma, \cdot)$, and

$$\sup_{\sigma \in S} d(\hat{\tau}_k(\sigma), \tau(\sigma)) = O_P^*(r_k^{-1}) \quad \text{as } k \to \infty.$$

Proof. We have to show that for every $N \in \mathbb{N}$

$$\lim_{N \to \infty} \limsup_{k \to \infty} P^* \left(r_k \sup_{\sigma \in S} d(\hat{\tau}_k(\sigma), \tau(\sigma)) > 2^N \right) = 0.$$

For $k, j \in \mathbb{N}$, set $V_{k,j} = \{(\sigma, \tau) : 2^{j-1} < r_k d(\tau, \tau(\sigma)) \leq 2^j\}$. Then

$$r_k \sup_{\sigma \in S} d(\hat{\tau}_k(\sigma), \tau(\sigma)) > 2^N$$

implies that there is some $\sigma_0 \in S$ such that $r_k d(\hat{\tau}_k(\sigma_0), \tau(\sigma_0)) > 2^N$, which in turn gives $(\sigma_0, \hat{\tau}_k(\sigma_0)) \in V_{k,j_0}$ for some $j_0 > N$. Combine this with (41) and (43) to get

$$(H_k - H)(\sigma_0, \hat{\tau}_k(\sigma_0)) - (H_k - H)(\sigma_0, \tau(\sigma_0)) \geq Cd^{\alpha}(\hat{\tau}_k(\sigma_0), \tau(\sigma_0)) > Cr_k^{-\alpha}2^{\alpha j_0 - \alpha}.$$

This implies

$$P^* \left(r_k \sup_{\sigma \in S} d(\hat{\tau}_k(\sigma), \tau(\sigma)) > 2^N \right)$$

$$\leq \sum_{j > N} P^* \left(\sup_{(\sigma, \tau) \in V_{k,j}} \left| \sqrt{k}(H_k - H)(\sigma, \tau) - \sqrt{k}(H_k - H)(\sigma, \tau(\sigma)) \right| \geq C\sqrt{k}r_k^{-\alpha}2^{\alpha j - \alpha} \right).$$

Via Markov's inequality (for outer probability) and (42), 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 first inequality follows from $\varphi_k(c\delta) \leq c^{\beta}\varphi_k(\delta)$ for $c \geq 1$. Note that the upper bound is finite by (44) and does not depend on k; since $\sum_{j > N} 2^{(\beta - \alpha)j}$ converges to 0 as $N \to \infty$ as $\beta < \alpha$ holds, the proof is complete. ∎

We next present an upper bound for $E^* \|\sqrt{n}(P_n - P)\|_{\mathcal{F}}$ for sup-norm bounded classes of functions $\mathcal{F}$. This result is essentially well-known, see Lemma 3.4.2 in van der Vaart & Wellner (1996), but we provide *explicit* constants. A proof, under the additional assumption that $Y_1, \ldots, Y_n$ are the coordinate projections on a product space, can be found in Gach (2010); inspection of the proof reveals that this assumption is unnecessary.

Theorem 41 *Suppose $(\Lambda, \mathcal{A}, P)$ is a probability space, $Y_1, \ldots, Y_n$ are i.i.d. with law P, and P_n denotes the empirical measure associated with $Y_1, \ldots, Y_n$. Let $\mathcal{F}$ be a non-empty class of $\mathcal{A}$-measurable functions on Λ, 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*

$$\mathrm{E}^* \|\sqrt{n}(P_n - P)\|_{\mathcal{F}} \leq (1696 + 64\sqrt{2})\, I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) \left[1 + \frac{B}{\eta^2 \sqrt{n}}\, I_{[\,]}(\eta, \mathcal{F}, \|\cdot\|_{2,P}) \right].$$

F   Appendix: Auxiliary Results for SMD-Estimation

Lemma 42 *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*

$$H(\theta) := \int_\Omega (f - p_\theta)^2 f^{-1} d\lambda$$

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 $c = \inf_{x \in \Omega} f(x)$,

$$
\begin{aligned}
|H(\theta_l) - H(\theta)| &= \left| \int_\Omega p_{\theta_l}^2 f^{-1} d\lambda - \int_\Omega p_\theta^2 f^{-1} d\lambda \right| \leq c^{-1} \int_\Omega \left| p_{\theta_l}^2 - p_\theta^2 \right| d\lambda \\
&\leq c^{-1} \|p_{\theta_l} - p_\theta\|_2 (\|p_{\theta_l} - p_\theta\|_2 + 2\|p_\theta\|_2) \to 0 \quad \text{for } l \to \infty.
\end{aligned}
$$

∎

Proposition 43 *(a) Suppose $\mathcal{P}_\Theta \subseteq \mathcal{L}^2(\Omega)$ and $\theta \mapsto p_\theta$ is a continuous map from Θ into $(\mathcal{L}^2(\Omega), \|\cdot\|_2)$. Then, on the event where $\inf_{x \in \Omega} \hat{p}_n(x) > 0$,*

$$\mathbb{Q}_n(\theta) = \int_\Omega (\hat{p}_n - p_\theta)^2 \hat{p}_n^{-1} d\lambda$$

holds and $\mathbb{Q}_n$ is a continuous real-valued function on Θ. [In particular, in case $\zeta > 0$ holds, the above event is the entire sample space Ω^n.]

(b) Let Assumption R.1 be satisfied. Then, on the event where $\inf_{x \in \Omega} \hat{p}_n(x) > 0$,

$$\mathbb{Q}_{n,k}(\theta) = \int_\Omega (\hat{p}_n - \tilde{p}_k(\theta))^2 \hat{p}_n^{-1} d\lambda$$

holds and $\mathbb{Q}_{n,k}$ is a continuous real-valued function on Θ. [In particular, in case $\zeta > 0$ holds, the above event is the entire sample space $\Omega^n \times V^k$.]

(c) 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 Assumption D.2 holds, then Q is a continuous real-valued function on Θ.

Proof. Parts (a) and (c) are immediate consequences of Lemma 42. We next prove Part (b): 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 Theorem 6(b), continuity of $\mathbb{Q}_{n,k}$ then follows from the theorem of dominated convergence. ∎

Proposition 44 *(a) 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.1 and D.2 be satisfied. Then*

$$\sup_{\theta \in \Theta} |\mathbb{Q}_n(\theta) - Q(\theta)| = o_\mathbb{P}^*(1) \quad as \ n \to \infty.$$

(b) Let Assumptions D.1, D.2, P.1, P.2, and R.1 be satisfied. Then

$$\sup_{\theta \in \Theta} |\mathbb{Q}_{n,k}(\theta) - Q(\theta)| = o_{\mathrm{Pr}}^*(1) \quad as \ \min(n, k) \to \infty.$$

(c) Suppose $\zeta > 0$ holds and Assumptions P.1 and R.2 are satisfied. Then

$$\sup_{n \in \mathbb{N}} \sup_{\Omega^n} \sup_{\theta \in \Theta} |\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_n(\theta)| = O_\mu^*(k^{-t/(2t+1)}) \quad as \ k \to \infty.$$

If Assumption P.1 is strengthened to P.3, then

$$\sup_{n \in \mathbb{N}} \sup_{\Omega^n} \sup_{\theta \in \Theta} |\mathbb{Q}_{n,k}(\theta) - \mathbb{Q}_n(\theta)| = O_{\mathrm{Pr}}^*(k^{-1/2}) \quad as \ k \to \infty.$$

Proof. (a) Set $\chi = 2^{-1} \inf_{x \in \Omega} p_\blacktriangle(x)$ and observe that $\chi > 0$ by Assumption D.2. In view of Remark 10(i) 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

$$\sup_{\theta \in \Theta} |\mathbb{Q}_n(\theta) - Q(\theta)| = \sup_{\theta \in \Theta} \left| \int_\Omega \frac{p_\theta^2}{\hat{p}_n} d\lambda - \int_\Omega \frac{p_\theta^2}{p_\blacktriangle} d\lambda \right| \leq \chi^{-2} \sup_{\theta \in \Theta} \|p_\theta\|_2^2 \|\hat{p}_n - p_\blacktriangle\|_\Omega.$$

Since Θ is compact, the assumptions on $\mathcal{P}_\Theta$ imply that $\sup_{\theta \in \Theta} \|p_\theta\|_2 < \infty$. Part (a) of Theorem 9 now completes the proof.

(b) Let χ and A_n be as in the proof of Part (a). On A_n we have

$$
\begin{aligned}
\sup_{\theta \in \Theta} |\mathbb{Q}_{n,k}(\theta) - Q(\theta)| &\leq \sup_{\theta \in \Theta} \left| \int_\Omega \frac{\tilde{p}_k(\theta)^2}{\hat{p}_n} d\lambda - \int_\Omega \frac{p_\theta^2}{p_\blacktriangle} d\lambda \right| \\
&\leq \sup_{\theta \in \Theta} \left| \int_\Omega (\tilde{p}_k(\theta) - p_\theta) \frac{\tilde{p}_k(\theta) + p_\theta}{\hat{p}_n} d\lambda + \int_\Omega p_\theta^2 \left(\frac{1}{\hat{p}_n} - \frac{1}{p_\blacktriangle} \right) d\lambda \right| \\
&\leq 2\chi^{-1} \sup_{\theta \in \Theta} \|\tilde{p}_k(\theta) - p_\theta\|_\Omega + \chi^{-2} D^2 \|\hat{p}_n - p_\blacktriangle\|_\Omega.
\end{aligned}
$$

The result then follows from Parts (a) and (c) of Theorem 9.

(c) 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 $\zeta > 0$, Part (d) 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 (a) of Proposition 1 that

$$\left\{ \frac{\tilde{p}_k(\theta) + p_\theta}{\hat{p}_n} : \theta \in \Theta, \ x_1, \ldots, x_n \in \Omega, \ v_1, \ldots, v_k \in V, \ n, k \in \mathbb{N} \right\} \tag{45}$$

is contained in a Sobolev ball $\mathcal{U}_{t,B}$ for some B satisfying $0 < B < \infty$. The first claim then follows from Theorem 13 with $s = 0$ (note that under $\zeta > 0$ Assumption P.1 implies Assumption P.2), 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 (45) is bounded in the sup-norm. If Assumption P.1 is strengthened to P.3, we may apply Part (c) of Theorem 16 with $\mathcal{F}$ equal to the set given in (45) to obtain the second claim. $\blacksquare$

Remark 45 If $\zeta > 0$ holds, then the events A_n in Parts (a) and (b) 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 46 *(a) Let Assumptions P.1 and P.5 be satisfied. Then, on the event $\inf_{x \in \Omega} \hat{p}_n(x) > 0$, the objective function $\mathbb{Q}_n$ is twice continuously partially differentiable on Θ° with*

$$\frac{\partial \mathbb{Q}_n}{\partial \theta_i}(\theta) = -2 \int_\Omega (\hat{p}_n - p_\theta) \frac{\partial p}{\partial \theta_i}(\cdot, \theta) \hat{p}_n^{-1} d\lambda,$$

$$\frac{\partial^2 \mathbb{Q}_n}{\partial \theta_i \partial \theta_j}(\theta) = 2 \int_\Omega \left(\frac{\partial p}{\partial \theta_i}(\cdot, \theta) \frac{\partial p}{\partial \theta_j}(\cdot, \theta) + \frac{\partial^2 p}{\partial \theta_i \partial \theta_j}(\cdot, \theta) p_\theta \right) \hat{p}_n^{-1} d\lambda,$$

for $i, j = 1, \ldots, m$.

(b) Let Assumptions D.2, P.1, and P.5 be satisfied. Then Q is twice continuously partially differentiable on Θ° with

$$\frac{\partial Q}{\partial \theta_i}(\theta) = -2 \int_\Omega (p_\blacktriangle - p_\theta) \frac{\partial p}{\partial \theta_i}(\cdot, \theta) p_\blacktriangle^{-1} d\lambda,$$

$$\frac{\partial^2 Q}{\partial \theta_i \partial \theta_j}(\theta) = 2 \int_\Omega \left(\frac{\partial p}{\partial \theta_i}(\cdot, \theta) \frac{\partial p}{\partial \theta_j}(\cdot, \theta) + \frac{\partial^2 p}{\partial \theta_i \partial \theta_j}(\cdot, \theta) p_\theta \right) p_\blacktriangle^{-1} d\lambda,$$

for $i, j = 1, \ldots, m$.

Proof. Note that the densities involved are all uniformly bounded by Assumption P.1. Under the respective assumptions, differentiation and integration can be interchanged, leading to the above formulae upon noting that the integral of $\partial^2 p/(\partial \theta_i \partial \theta_j)(\cdot, \theta)$ is zero. Continuity of the partial derivatives follows from the theorem of dominated convergence. ∎

Proposition 47 *Let Assumptions D.1, P.1, and P.5 be satisfied and suppose $\zeta > 0$. Then, for all $i, j = 1, \ldots, m$,*

$$\sup_{\theta \in \Theta^\circ} \left| \frac{\partial^2 \mathbb{Q}_n}{\partial \theta_i \partial \theta_j}(\theta) - \frac{\partial^2 Q}{\partial \theta_i \partial \theta_j}(\theta) \right| = o_\mathbb{P}(1) \quad \text{as } n \to \infty. \tag{46}$$

Proof. Let $b < \infty$ be a bound for all the integrals appearing in Assumption P.5. By Lemma 46 the l.h.s. of (46) is not larger than $2\zeta^{-2} b (1 + C_t D) \|\hat{p}_n - p_\blacktriangle\|_\Omega$, which converges to 0 in probability by Theorem 9(a). Measurability of the supremum in (46) follows from continuity of the second derivatives (Lemma 46) and separability of Θ°. ∎

Remark 48 If $\zeta = 0$ the assertion of the preceding proposition still holds true in outer probability under Assumptions D.1, D.2, P.1, and P.5, if $\partial^2 \mathbb{Q}_n(\theta)/\partial\theta\partial\theta'$ is interpreted as the zero matrix on the event where $\inf_{x \in \Omega} \hat{p}_n(x) = 0$.

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] BERGE, C.(1963). *Topological Spaces.* Oliver and Boyd, Edinburgh.

[4] BIRMAN, M. S. and SOLOMYAK, M.Z. (1967). Piecewise-polynomial approximations of functions of the class W_p^α. *Math. USSR Sbornik* **2**, 295-317.

[5] 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.

[6] DUDLEY, R. M. (1999). *Uniform Central Limit Theorems.* Cambridge Studies in Advanced Mathematics **63**. Cambridge University Press, Cambridge.

[7] DUNFORD, N. and SCHWARTZ, J. T. (1966). *Linear Operators. Part I: General Theory.* Wiley, New York.

[8] FERMANIAN, J. D. and SALANIÉ, B. (2004): A nonparametric simulated maximum likelihood estimation method. *Econometric Theory* **20**, 701-734.

[9] GACH, F.(2010). *Efficiency in Indirect Inference.* Ph.D. Thesis, Department of Statistics, University of Vienna.

[10] GALLANT, R. and LONG, J. (1997). Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* **84** 125–141.

[11] GOURIÉROUX, C. and MONFORT, A. (1996). *Simulation-Based Econometric Methods.* Oxford University Press, New York.

[12] GOURIÉROUX, C., MONFORT, A., and RENAULT, E. (1993). Indirect inference. *J. Appl. Econom.* **8** 85–118.

[13] JIANG, W. and TURNBULL, B. (2004). The indirect method: inference based on intermediate statistics – a synthesis and examples. *Statist. Sci.* **19** 239–263.

[14] LEDOUX, M. and TALAGRAND, M. (1991). *Probability in Banach Spaces.* Results in Mathematics and Related Areas (3) **23**. Springer-Verlag, Berlin.

[15] 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.

[16] LORENTZ, G. G., v.GOLITSCHEK, M., and MAKOVOZ, Y. (1996). *Constructive Approximation: Advanced Problems.* Springer-Verlag.

[17] NICKL, R. (2007). Donsker-type theorems for nonparametric maximum likelihood estimators. *Probab. Theory Related Fields* **138** 411–449.

[18] 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.

[19] NICKL, R. and PÖTSCHER, B. M. (2010). Efficient simulation-based minimum distance estimation and indirect inference, *Math. Meth. Statist.,* forthcoming.

[20] PÖTSCHER, B. M. and PRUCHA, I. R. (1997). *Dynamic Nonlinear Econometric Models. Asymptotic Theory.* Springer-Verlag, Berlin.

[21] SMITH, A. (1993). Estimating nonlinear time-series models using simulated vector autoregressions, *J. Appl. Econom.* **8** 63–84.

[22] TRIEBEL, H. (1983). *Theory of Function Spaces.* Monographs in Mathematics **78**. Birkhäuser Verlag, Basel.

[23] VAN DE GEER, S. (1993). Hellinger-consistency of certain nonparametric maximum likelihood estimators. *Ann. Statist.* **21** 14–44.

[24] VAN DER VAART, A. W. and WELLNER, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer, New York.

[25] ZIEMER, W. P. (1989). *Weakly Differentiable Functions.* Springer, New York.