Efficient Simulation-Based Minimum Distance Estimation and Indirect Inference

Nickl, Richard and Pötscher, Benedikt M.

Statistical Laboratory, University of Cambridge, Department of Statistics, University of Vienna

March 2009

Efficient Simulation-Based Minimum Distance Estimation and Indirect Inference

Richard Nickl
Statistical Laboratory, University of Cambridge
and
Benedikt M. Pötscher
Department of Statistics, University of Vienna

March 8, 2009

Abstract

Given a random sample from a parametric model, we show how indirect inference estimators based on appropriate nonparametric density estimators (i.e., simulation-based minimum distance estimators) can be constructed that, under mild assumptions, are asymptotically normal with variance-covarince matrix equal to the Cramér-Rao bound.

1 Introduction

Suppose we observe a random sample $X_1, \ldots, X_n$ from a distribution P, and we are in the classical situation where one maintains a *parametric* model $\mathcal{M} = \{P(\theta) : \theta \in \Theta\}$ of probability measures $P(\theta)$, indexed by the set $\Theta \subseteq \mathbb{R}^b$, for statistical inference. Under the assumption of correct specification of the parametric model, i.e., $P = P(\theta_0)$ for a (unique) $\theta_0 \in \Theta$, the maximum likelihood estimator (MLE) is a natural estimator of θ_0 (as well as of $P(\theta_0)$), since it is *asymptotically efficient* under standard regularity conditions.

There are several reasons, however, why maximum likelihood might nevertheless not be the method of choice, and alternatives, that ideally are also *asymptotically efficient,* are of interest.

A first such reason is rather classical (e.g., Huber (1972), Beran (1977), Millar (1981) Donoho and Liu (1988), Lindsay (1994)) and comes from robustness considerations: A good estimator for θ_0 should be robust against misspecifications of $\mathcal{M}$. A lesson from the above-mentioned literature is the following: If one wants an estimator of θ_0 that is robust against perturbations of $P(\theta_0)$ in some metric $\chi(\cdot, \cdot)$, then one should rather use 'minimum distance estimators' of the following form: if $\tilde{P}_n$ is a suitable (typically nonparametric) χ-consistent estimator of P, estimate θ by the minimizer over Θ of

$$Q_n(\theta) := \chi(\tilde{P}_n, P(\theta)). \tag{1}$$

Under several assumptions, Beran (1977) showed the interesting result that, if χ is the Hellinger-distance, and if $\tilde{P}_n$ is some kernel density estimator, such minimum-distance estimators are not only robust, but actually simultaneously *asymptotically efficient*, so that they outperform the MLE in this sense. We will discuss the asymptotic efficiency aspect of his result in more detail below.

A second, more practical reason against the use of the MLE that has arisen in recent applications in econometrics and biostatistics is related to the fact that in these applications analytic expressions for the densities in the parametric model, and hence for the likelihood function, are not available (or intractable for numerical purposes). For example, the data may be modeled by an equation of the form $X_i = g(\varepsilon_i, \theta_0)$, but the implied parametric density may not be analytically tractable, e.g., because g is complicated or ε_i is high-dimensional. The same problem occurs naturally also in estimation of dynamic nonlinear models including stochastic differential equations, we refer to Smith (1993), Gourieroux, Monfort and Renault (1993), Gallant and Tauchen (1996), Gallant and Long (1997) and the monograph Gourieroux and Monfort (1996) for several concrete examples. This problem has led to a growing literature about so-called *indirect inference* methods, where other estimators than the MLE are suggested, often based on simulations, see the just mentioned references and Jiang and Turnbull (2004). From a conceptual point of view, the main idea behind the indirect inference approach can be phrased as follows:

1. For each $\theta \in \Theta$, simulate a sample $X_1(\theta), ..., X_k(\theta)$ of size k from the distribution $P(\theta)$ (which is often possible in the examples alluded to above, e.g., by perusing the equations defining the model).

2. Based on each simulated sample *as well as* on the true data, compute estimators $\tilde{P}_k(\theta)$ and $\tilde{P}_n$ 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. Choose a suitable metric χ on $\mathcal{M}^{aux}$, and estimate θ_0 by minimizing over Θ the objective function
$$\mathcal{Q}_{n,k}(\theta) := \chi(\tilde{P}_n, \tilde{P}_k(\theta)). \tag{2}$$

In most of the indirect inference literature, the auxiliary model $\mathcal{M}^{aux}$ is also finite-dimensional (so that one in fact estimates a finite-dimensional parameter in Step 2 rather than the probability measure directly), and the resulting procedure can be shown to be consistent and asymptotically normal (under standard regularity conditions, see Gourieroux and Monfort (1996)). However, the procedure is asymptotically efficient 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 indirect inference estimators based on suitable nonparametric estimators $\tilde{P}_n$ *and* $\tilde{P}_k(\theta)$ *with common choices for the tuning parameters ('sieve'-dimensions), including rate-optimal choices, are asymptotically efficient in the sense that they are asymptotically normal with asymptotic variance equal to the Cramér-Rao bound.* To the best of our knowledge, no proof of this fact was known before, although there are some related results that need mentioning. We comment on the literature in some detail below, but first wish to discuss the main ideas behind our results. [Robustness issues, misspecification of $\mathcal{M}$, as well as uniformity in the asymptotic normality result are not treated explicitly in this paper; for the latter two issues in a related context see Gach (2009).]

From the discussion so far it transpires that indirect inference estimators from (2) are *minimum distance estimators*, with the important (and nontrivial) modification that $P(\theta)$ in (1) is replaced by an *estimator* based on simulations from $P(\theta)$. It is therefore of interest to first briefly revisit Beran's (1977) asymptotic efficiency result: For simplicity, consider the Fisher-metric $\chi_F(f,g)^2 := \int (f-g)^2 p_0^{-1}$, where p_0 is the density of P, instead of the Hellinger distance. [Note that the Fisher-metric is closely related to the Hellinger distance when f and g are near p_0.] If $\hat{\theta}_n$ is the minimizer of Q_n in (1), then, after a suitable Taylor expansion, asymptotic efficiency of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ essentially reduces to proving two separate results: The first is to prove asymptotic normality for the gradient of (1) at θ_0, namely

$$\sqrt{n} \int s(\theta_0) d(\tilde{P}_n - P(\theta_0)), \tag{3}$$

where the 'influence function' $s(\theta_0)$ equals $\nabla_\theta p(\theta_0) p_0^{-1}$. Note that $s(\theta_0)$ coincides with the *efficient* influence function in this problem, showing that $\chi = \chi_F$ is a natural choice. The second step is to control the remainder term in the Taylor expansion, which essentially requires convergence of $\tilde{P}_n$ to $P = P(\theta_0)$ (in the sense of L^p-convergence of the respective densities for certain values of p). Beran (1977) implicitly proved these two results under relatively restrictive conditions if $\tilde{P}_n$ is a kernel density estimator with certain bandwidths, and if χ is the Hellinger metric. It is typically not sensible (and for the most interesting metrics χ in fact not possible) to take $\tilde{P}_n$ to be the *empirical measure* itself, but rather $\tilde{P}_n$ should be some smoothed version of it. In this case, one cannot directly apply a standard central limit theorem to (3). However, recent results in empirical process theory (Nickl (2007), Giné and Nickl (2008, 2009b)) establish exactly such limit theorems for various density estimators. Furthermore, these limit theorems also hold for density estimators that simultaneously deliver optimal convergence rates in L^p-type loss functions, which is potentially relevant for good control of the remainder term. (We should note that this simultaneous optimality property is related to what Bickel and Ritov (2003) label the 'plug-in property' of the density estimator $\tilde{P}_n$, cf. also Section 3 in Nickl (2007) for more discussion.) Using similar methods we first prove a Beran-type result (Theorem 2), under quite weak (if not sharp) conditions, for the case where $\chi = \chi_F$ (but with the unknown p_0 replaced by an estimator), and where the underlying nonparametric estimator is based on a $\mathcal{L}^2$-projection of the empirical measure onto spaces of piecewise polynomials spanned by dyadic B-splines.

Once asymptotic normality of the minimum distance estimator in (1) is established, the question arises how the simulation step in (2) should be approached. Here two proof strategies arise:

1. The first method is to show that the objective function $\mathcal{Q}_{n,k}$ *with* simulations is stochastically close, uniformly over Θ, to the objective function Q_n where no simulation is performed. If

$$\sup_{\theta \in \Theta} |\mathcal{Q}_{n,k}(\theta) - Q_n(\theta)| \tag{4}$$

has a sufficiently fast rate of convergence to zero (in probability), then it is not difficult to show, using a result from Gach (2009), that the asymptotic distribution of the simulated indirect inference estimator obtained from minimizing (2) is the *same* as the one of the classical minimum distance estimator discussed in the previous paragraph. It turns out that proving that the expression in (4) has a sufficiently fast rate of convergence to zero can be done by deriving sharp bounds for the stochastic processes

$$\left\{ \sqrt{n} \int f d(\tilde{P}_k(\theta) - P(\theta)) \right\}_{\theta \in \Theta, f \in \mathcal{F}},$$

where $\mathcal{F}$ is a relevant class of functions, and again we can apply recent techniques from empirical processes here (cf. Nickl (2007), Giné and Nickl (2008, 2009b) together with moment inequalities in Giné and Koltchinskii (2006)). We prove that if one performs simulations of order $k >> n^2$, then the indirect inference estimators are asymptotically equivalent to the classical minimum distance estimators. A main advantage of this proof strategy is that *no* differentiability properties of the objective function $\mathcal{Q}_{n,k}$ have to be used, and that in turn a large class of simulation mechanisms is admissible. More importantly, this proof strategy allows for the presumably critical condition $\tau > 1/2$ on the underlying density p_0, where τ is the index governing the regularity of p_0.

2. The method of proof described above works if many simulations are performed ($k >> n^2$). However, this condition is not intrinsic to the problem, and the case where the number of simulations k is of a smaller order than n^2 is also of interest. In particular, in the case where $k/n \to \kappa$, $0 < \kappa < \infty$, one has to expect that the asymptotic variance of simulated indirect inference estimators is inflated by the factor $(1+1/\kappa)$. If one is interested in these cases, the (comparably) 'brute force' methods described in the previous paragraph cannot be used. Alternatively, one can try to apply the usual M-estimation asymptotic normality proof to the criterion function $\mathcal{Q}_{n,k}(\theta)$. Among other things this requires differentiation of the simulated estimators $P_k(\theta)$ with respect to θ. Since $P_k(\theta)$ is constructed by applying an *approximate identity* to the empirical measure from the simulated sample, the proofs become more delicate in this case. [Differentiating an approximate identity $h^{-1}K(X(\theta)/h)$ w.r.t. θ introduces a 'penalty' of an additional h^{-1} from the chain rule.] We are able, nevertheless, to establish asymptotic normality of the simulated indirect inference estimator with these simulation sizes as well, under slightly stronger conditions (on the underlying density and the simulation mechanism), and with the expected inflation of variances if $\lim_n k/n < \infty$. Again, the empirical process techniques mentioned in the previous paragraphs, together with some facts from approximation theory, are central to our proofs.

We should comment on some related literature. Related papers are Gallant and Long (1997) and Fermanian and Salanié (2004). The first paper studies the case where $\tilde{P}_n$ is based on nonparametric MLEs 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). Fermanian and Salanié (2004) propose different (but somewhat related) procedures, and establish asymptotic efficiency of their estimators under several high level conditions, which, as they admit themselves, are very stringent. Even in the simplest model they consider, they need to have simulations of order $k \sim n^6$, and the nonparametric estimators considered seem to be only sensible if the true density is very smooth. There are also some other related recent papers on this topic, Altissimo and Mele (2009) and Carrasco, Chernov, Florens, Ghysels (2007), whose proofs, however, are incomplete or incorrect.

The outline of the paper is as follows: After some preliminaries in Section 2, we introduce the model and assumptions, define the auxiliary spline projection estimators as well as the indirect inference estimator in Section 3 and present the main result (Theorem 1) on asymptotic efficiency of the indirect inference estimator. Some basic facts on dyadic splines are summarized in Section 4. Section 5 is devoted to the proof of Theorem 1. Section 6 develops auxiliary convergence rate results for the auxiliary spline projection estimators needed in the proof of Theorem 1. Section 7 establishes a uniform central limit theorem for spline projection estimators that is also essential in the proof of the main result. Three appendices contain further technical results on Besov spaces, projections onto Schoenberg spaces, and moment inequalities for empirical processes.

2 Preliminaries and Notation

We denote the Euclidean norm of a vector $x \in \mathbb{R}^b$ by $\|x\|$ and the associated operator norm of a matrix A by $\|A\|$. With $\mathcal{L}^p := \mathcal{L}^p([0,1], \lambda)$, $1 \leq p < \infty$, we denote the vector space of Borel-measurable p-fold integrable real-valued functions on $[0,1]$, where λ denotes Lebesgue measure on $[0,1]$, the (semi)norm on $\mathcal{L}^p$ being denoted by $\|h\|_p$. Furthermore, $\|h\|_\infty$ stands for the supremum norm (*not* the essential supremum norm) of a real-valued function h defined on $[0,1]$. If H is a vector- or matrix-valued function on $[0,1]$ then $\|H\|_p$ is shorthand for $\|\|H\|\|_p$ and similarly for the supremum norm. By L^∞ we denote the space of all bounded Borel-measurable real-valued functions on $[0,1]$ endowed with the supremum norm. For a (measurable) real-valued function g on $\mathbb{R}$ and $1 \leq p < \infty$ we write $\|g\|_{p,\mathbb{R}}$ to denote its $\mathcal{L}^p$-(semi)norm (w.r.t. Lebesgue measure on $\mathbb{R}$); and we write $\|g\|_{\infty,\mathbb{R}}$ for the supremum norm (not the essential supremum norm). For sequences a_n and b_n of positive real numbers we write $a_n \sim b_n$ to denote the fact that the sequence a_n/b_n is bounded away from zero and infinity.

We next introduce Besov spaces. For a function $g : \mathbb{R} \to \mathbb{R}$ and $z \in \mathbb{R}$, the difference operator Δ_z is defined by $\Delta_z g(\cdot) = g(\cdot + z) - g(\cdot)$ and inductively by $\Delta_z^a g(\cdot) = \Delta_z(\Delta_z^{a-1} g(\cdot))$ for integer $a \geq 2$. For $h : [0,1] \to \mathbb{R}$, we define $\Delta_z^a(h)(x)$ as above if $x, x + az \in [0,1]$, and set $\Delta_z^a(h)(x) = 0$ otherwise. For $0 < s < \infty$ we define function spaces $\mathcal{B}_s$ on $[0,1]$ as follows.

Definition 1 *For $s \in (0,\infty)$, $a \in (s,\infty) \cap \mathbb{N}$, and $h \in \mathcal{L}^2$ define*

$$\|h\|_{s,2} := \|h\|_2 + \sup_{0 \neq |z| < 1} |z|^{-s} \|\Delta_z^a(h)\|_2.$$

Define further

$$\mathcal{B}_s := \mathcal{B}^s_{2\infty} = \{h \in \mathcal{L}^2 : \|h\|_{s,2} < \infty\}.$$

The space $\mathcal{B}_s$ does not depend on a in the sense that different choices of $a > s$ result in equivalent (semi)norms. For definiteness we shall always choose a to be the smallest integer larger than s in the sequel. It is well-known (Proposition 7 in Appendix A) that for $s > 1/2$ every function in $\mathcal{B}_s$ is λ-almost everywhere equal to a (uniquely determined) *continuous* function in $\mathcal{B}_s$. It thus proves useful to define for $s > 1/2$ the Banach-space $(\mathsf{B}_s, \|\cdot\|_{s,2})$ where $\mathsf{B}_s = \mathcal{B}_s \cap \mathsf{C}([0,1])$ and $\mathsf{C}([0,1])$ denotes the set of continuous real-valued functions on $[0,1]$.

A little reflection shows that $\mathcal{B}_s$ is just the usual Besov (or generalized Lipschitz) space $\mathcal{B}^s_{2\infty}$ as, e.g., defined in Chapter 2, Section 10 of DeVore and Lorentz (1993) (with the only difference that there $\mathcal{B}_s$ is viewed as a space of equivalence classes of functions). The space $\mathcal{B}_s$ contains the classical Sobolev space of order s as a subset. Recall that for *integer* s the Sobolev space of order $s > 0$ is given by

$$\mathcal{W}_2^s = \left\{h \in \mathcal{L}^2 : D_w^i h \in \mathcal{L}^2 \text{ for } 0 \leq i \leq s,\ i \text{ integer}\right\},$$

where D_w denotes the weak differential operator. Then for integer $s > 0$

$$\|h\|_{s,2} \leq C(s) \sum_{0 \leq i \leq s} \|D_w^i h\|_2 \tag{5}$$

holds for some universal constant $C(s)$ and all h in the Sobolev space of order s; cf. p.46 and p.52f in DeVore and Lorentz (1993). Some further properties of Besov spaces and their relationship to splines that we shall need in the sequel are summarized in Appendix A.

3 Main Results

Let $X_1, \ldots, X_n$ be independent and identically distributed (i.i.d.) on a compact interval in $\mathbb{R}$ with law P and Lebesgue-density p_0. Without loss of generality we shall take this interval to be $[0,1]$. We assume that a parametric model $\mathcal{P}_\Theta$ is given, i.e., $\mathcal{P}_\Theta = \{p(\theta) : \theta \in \Theta\}$, where the *functions* $p(\theta) : [0,1] \to \mathbb{R}$ are probability densities and the parameter space Θ is a subset of $\mathbb{R}^b$. The probability measure on $[0,1]$ corresponding to $p(\theta)$ will be denoted by $P(\theta)$. We consider here the case where direct likelihood methods for estimation of θ cannot be used for the reasons outlined in the introduction. Suppose, however, that it is feasible to obtain for each $\theta \in \Theta$ simulated data $X_i(\theta)$ via

$$X_i(\theta) = \rho(V_i, \theta), \qquad i = 1, ..., k,$$

that are distributed i.i.d. with density $p(\theta)$ and that are independent of the original sample. [The simulation mechanism may result from an equation for the data as described in Section 1, but may also be obtained in some other way.] More precisely, we assume that the random variables V_i driving the simulation mechanism are i.i.d. with values in some measurable space $(\mathcal{V}, \mathfrak{V})$, the distribution on $\mathcal{V}$ induced by V_i being denoted by μ; furthermore, we assume that for every $\theta \in \Theta$, the $\mathfrak{V}$-measurable function $\rho(\cdot, \theta) : \mathcal{V} \to [0,1]$ is such that the law of $\rho(V_i, \theta)$ has density $p(\theta)$; and that the collection of random variables $\{V_i\}$ is independent of the collection $\{X_i\}$. As the main result depends only on the *distribution* of the random variables X_i and V_i, we can assume without loss of generality that the original data X_i as well as the variables V_i are defined as the respective coordinate projections on the product probability space $([0,1]^\infty \times \mathcal{V}^\infty, \mathfrak{B}_{[0,1]}^\infty \otimes \mathfrak{V}^\infty, P^\infty \otimes \mu^\infty)$; we shall denote by Pr the product probability measure $P^\infty \otimes \mu^\infty$. The basic framework outlined above will be maintained throughout the rest of the paper.

We next construct auxiliary estimators for p_0 from the original data as well as from the simulated data. The estimator of p_0 based on the original data is a spline projection estimator based on B-splines of order $r_* \geq 1$ and is given by

$$p_{n,j,r_*}(y) = \sum_{l=-r_*+1}^{2^j-1} \hat{\gamma}_{lj}^{(r_*)} N_{lj}^{(r_*)}(y)$$

with

$$\hat{\gamma}_{lj}^{(r_*)} = \sum_{m=-r_*+1}^{2^j-1} 2^j g_j^{(r_*)lm} \int_{[0,1]} N_{mj}^{(r_*)}(x) dP_n(x).$$

Here $N_{lj}^{(r_*)}$ denote the B-spline basis functions forming a basis for the Schoenberg space $\mathcal{S}_j(r_*)$ and the coefficients $g_j^{(r_*)lm}$ are the elements of 2^{-j} times the inverse of the Gram matrix of the B-spline basis $N_{lj}^{(r_*)}$; see Section 4 for definitions. Furthermore, $P_n = n^{-1} \sum_{i=1}^n \delta_{X_i}$ denotes the empirical measure of the original data. The positive integer j represents a tuning parameter that governs the dimension of the approximating space ('sieve') spanned by the B-spline basis. Similarly, from each simulated data set $X_i(\theta)$, we construct estimators for $p(\theta)$ based on order-r B-splines via

$$p_{k,J,r}(\theta)(y) = \sum_{l=-r+1}^{2^J-1} \hat{\gamma}_{lJ}^{(r)}(\theta) N_{lJ}^{(r)}(y) \tag{6}$$

with

$$\hat{\gamma}_{lJ}^{(r)}(\theta) = \sum_{m=-r+1}^{2^J-1} 2^J g_J^{(r)lm} \int_{[0,1]} N_{mJ}^{(r)}(x) dP_k(\theta)(x) \tag{7}$$

and $P_k(\theta) = k^{-1} \sum_{i=1}^k \delta_{X_i(\theta)}$. Note that r_* and r need not take the same value, nor need j and J. [For example, $r = 4$ would correspond to using cubic splines for the construction of $p_{k,J,r}(\theta)$, while $r_* = 1$ would correspond to using the Haar basis for the construction of p_{n,j,r_*}.] In the sequel we shall often write $p_{k,J,r}(\theta, y)$ for $p_{k,J,r}(\theta)(y)$ and similarly $p(\theta, x)$ for $p(\theta)(x)$.

The idea behind indirect inference is that, given the parametric model is correctly specified in the sense that $p_0 = p(\theta_0)$ λ-almost everywhere for some $\theta_0 \in \Theta$, the particular value of θ corresponding to the simulation-based estimator $p_{k,J,r}(\theta)$ closest to p_{n,j,r_*} (in an appropriate metric) should provide a reasonable estimator $\hat{\theta}_{n,k}$ of θ_0, since p_{n,j,r_*} will estimate $p_0 = p(\theta_0)$ (λ-a.e.) consistently (under appropriate assumptions and choices of j, J, and k). That is, as explained in Section 1, the estimator $\hat{\theta}_{n,k}$ can be viewed as a simulation-based version of a minimum distance estimator.

To implement this idea we introduce the indirect inference objective function measuring closeness of p_{n,j,r_*} and $p_{k,J,r}(\theta)$

$$\mathcal{Q}_{n,k}(\theta) := \mathcal{Q}_{n,k,j,J,r_*,r}(\theta) = \begin{cases} \int_0^1 (p_{n,j,r_*} - p_{k,J,r}(\theta))^2 p_{n,j,r_*}^{-1} d\lambda & \text{on the event } A_n \\ 0 & \text{otherwise} \end{cases}, \tag{8}$$

where $A_n = \{p_{n,j_n,r_*}(y) > 0 \text{ for every } y \in [0,1]\}$, which is measurable as is easily seen. Note that $\mathcal{Q}_{n,k}(\theta) : [0,1]^\infty \times \mathcal{V}^\infty \to \mathbb{R}$ is $\mathfrak{B}_{[0,1]}^\infty \otimes \mathfrak{V}^\infty$-measurable for every $\theta \in \Theta$ as a consequence of Tonelli's Theorem since p_{n,j,r_*} and $p_{k,J,r}(\theta)$ are both jointly measurable (w.r.t. the combined data and the argument y) and since A_n is measurable. Furthermore, since all functions involved are piecewise polynomials with dyadic breakpoints, the integral featuring in the definition of $\mathcal{Q}_{n,k}(\theta)$ can be computed in a numerically efficient way.

Remark 1 *(i) We have chosen to assign $\mathcal{Q}_{n,k}(\theta)$ the value zero on the complement of A_n for convenience. Since the event A_n will be seen to have probability approaching* 1 *under our assumptions, this particular assignment is irrelevant for asymptotic considerations. However, from a more practical point of view, one might want to use the objective function $\int_{p_{n,j,r_*}>0} (p_{n,j,r_*} - p_{k,J,r}(\theta))^2 p_{n,j,r_*}^{-1} d\lambda$ instead, which clearly coincides with $\mathcal{Q}_{n,k}$ on A_n.*

(ii) In principle, auxiliary estimators other than spline projection estimators could be used in the definition of $\mathcal{Q}_{n,k}(\theta)$. We do not pursue this in this paper but see Gach (2009). We note that standard kernel density estimators are inappropriate here because of boundary effects.

An *indirect inference estimator* $\hat{\theta}_{n,k} := \hat{\theta}_{n,k,j,J,r_*,r}$ is now defined to be any measurable function that satisfies

$$\inf_{\theta \in \Theta} \mathcal{Q}_{n,k}(\theta) = \mathcal{Q}_{n,k}(\hat{\theta}_{n,k}). \tag{9}$$

For the sake of simplicity, we shall use the abbreviation $\mathcal{Q}_{n,k}$ to denote $\mathcal{Q}_{n,k,j,J,r_*,r}$ as well as $\mathcal{Q}_{n,k,j_n,J_k,r_*,r}$, the precise meaning always being clear from the context. [A similar comment applies to $\hat{\theta}_{n,k}$, as well as to Q_n and $\hat{\theta}_n$ defined later in Section 5.2.] That such an estimator exists is shown in the next proposition, the proof of which can be found in Appendix B.

Proposition 1 *Suppose Θ is compact in $\mathbb{R}^b$ and that the simulation mechanism $\rho(v, \cdot)$ is continuous on Θ for every $v \in \mathcal{V}$. Furthermore, assume that $r_* \geq 1$ and $r \geq 2$ hold. Then there exists a $\mathfrak{B}_{[0,1]}^\infty \otimes \mathfrak{V}^\infty$-measurable mapping $\hat{\theta}_{n,k}$ satisfying (9).*

We now introduce the following assumptions on the parametric model that will be used to prove the main result.

Assumption P1: (i) The parameter space Θ is a compact subset of $\mathbb{R}^b$. There exists a $\theta_0 \in \Theta$ such that $p_0 = p(\theta_0)$ λ-almost everywhere. Furthermore, $p(\theta) = p(\theta_0)$ λ-almost everywhere implies $\theta = \theta_0$. The mapping $\theta \mapsto p(\theta, x)$ is continuous on Θ for every $x \in [0,1]$. The density $p(\theta_0)$ is positive on $[0,1]$.

(ii) $\mathcal{P}_\Theta$ is a bounded subset of B_τ for some $\tau > 1/2$.

(iii) θ_0 is an interior point of Θ. There is an open ball $B(\theta_0) \subseteq \Theta$ with center θ_0 such that the map $\theta \mapsto p(\theta, x)$ is twice continuously differentiable on $B(\theta_0)$ for every $x \in [0,1]$. Furthermore,

$$\int_0^1 \sup_{\theta \in B(\theta_0)} \|\nabla_\theta p(\theta, x)\|^2 \, dx < \infty, \quad \int_0^1 \sup_{\theta \in B(\theta_0)} \left\|\nabla_\theta^2 p(\theta, x)\right\| dx < \infty,$$

and $\int_0^1 \nabla_\theta p(\theta_0, x) \nabla_\theta p(\theta_0, x)' p(\theta_0, x)^{-1} dx$ is positive definite. [Here ∇_θ denotes the gradient w.r.t. θ written as a column vector and ∇_θ^2 denotes the matrix of second derivatives.]

(iv) For some $\varsigma > 1/2$

$$\frac{\partial p(\theta_0, \cdot)}{\partial \theta_q} \in \mathsf{B}_\varsigma$$

holds for every $q = 1, ..., b$.

Assumption P1(i) is a standard assumption that implies consistency of the maximum likelihood estimator. In particular, it expresses the fact that the parametric model is correctly specified and that the true parameter value is identifiable. Assumption P1(iii) in conjunction with P1(i) is a typical assumption used to establish asymptotic normality of the maximum likelihood estimator and the information matrix equality. Assumption P1(ii) requires the parametric density functions to behave "regularly" as functions of x (uniformly in θ), the condition being quite weak: Note that if τ is close to 1/2 the density functions are not even required to be differentiable, all that is required is essentially that the functions are "$\mathcal{L}^2$-Hölder continuous" of order τ, uniformly over θ. [Given compactness of Θ, a sufficient condition for Assumption P1(ii) is that $\mathcal{P}_\Theta \subseteq \mathsf{B}_\tau$ for some $\tau > 1/2$ and that the map $\theta \to p(\theta)$ from Θ to B_τ is continuous; in fact, continuity of the map $\theta \to \|p(\theta)\|_{\tau,2}$ already suffices. A simple sufficient condition for this (with $\tau = 1$) is continuity of $\theta \to \|p(\theta)\|_2$ and $\theta \to \|D_w p(\theta)\|_2$ on Θ, cf. (5).] In a similar vein, Assumption P1(iv) imposes an analogous weak regularity condition on the derivative of $p(\theta)$ (w.r.t. θ) at $\theta = \theta_0$.

For parts of the main result we will need to supplement assumption P1 by the following assumption.

Assumption P2: **(i)** The set $\left\{\frac{\partial p(\theta, \cdot)}{\partial \theta_q} : q = 1, \ldots, b, \ \theta \in B(\theta_0)\right\}$ is a relatively compact subset of $\mathcal{L}^2$ where $B(\theta_0)$ is defined in Assumption P1.

(ii) The set $\left\{\frac{\partial^2 p(\theta, \cdot)}{\partial \theta_q \partial \theta_{q'}} : q, q' = 1, \ldots, b, \ \theta \in B(\theta_0)\right\}$ is a bounded subset of $\mathcal{L}^2$, i.e.,

$$\sup_{\theta \in B(\theta_0)} \int_0^1 \left\|\nabla_\theta^2 p(\theta, x)\right\|^2 dx < \infty.$$

These assumptions are not restrictive. For example, Assumption P2(i) is satisfied if the indicated set of functions is a bounded subset of a Besov space $\mathcal{B}_s$ with s only satisfying $s > 0$, which is a very weak condition.

We also need assumptions on the simulation mechanism ρ. The basic assumption will be that the function ρ satisfies a Hölder continuity condition in θ (Assumption R(i)). For some of the results we shall need an additional assumption including twice differentiability in a neighborhood of θ_0 (Assumption R(ii)).

Assumption R: (i) The function ρ is uniformly Hölder in θ, more precisely, for some $0 < L < \infty$ and some $0 < \alpha \leq 1$

$$\sup_{v \in \mathcal{V}} \left| \rho(v, \theta) - \rho(v, \theta') \right| \leq L \left\| \theta - \theta' \right\|^{\alpha}$$

holds for all θ, $\theta' \in \Theta$.

(ii) There is an open ball $B(\theta_0) \subseteq \Theta$ with center θ_0 such that the map $\theta \to \rho(v, \theta)$ is twice continuously differentiable on $B(\theta_0)$ for every $v \in \mathcal{V}$ and

$$\sup_{v \in \mathcal{V}, \theta \in B(\theta_0)} \| \nabla_\theta \rho(v, \theta) \| < \infty, \qquad \sup_{v \in \mathcal{V}, \theta \in B(\theta_0)} \| \nabla_\theta^2 \rho(v, \theta) \| < \infty.$$

Furthermore, for some $0 < L' < \infty$ and some $0 < \beta \leq 1$

$$\sup_{v \in \mathcal{V}} \left\| \nabla_\theta^2 \rho(v, \theta) - \nabla_\theta^2 \rho(v, \theta') \right\| \leq L' \left\| \theta - \theta' \right\|^{\beta}$$

holds for all θ, $\theta' \in B(\theta_0)$.

Assumptions on the parametric model $\mathcal{P}_\Theta$ and assumptions on the simulation mechanism ρ are of course interrelated. For example, one could in principle only impose appropriate assumptions on ρ and then deduce the existence of a $\mathcal{P}_\Theta$ with the required properties from those assumptions; see Gach (2009) for some discussion. However, as this does not seem to lead to a transparent catalogue of assumptions, we have chosen to formulate the assumptions in the form given above.

We now first establish consistency of the indirect inference estimator. The assumptions used for the consistency result in the subsequent proposition are stronger than what is actually needed for such a result, but we do not strive for utmost generality in the consistency result as this is not the main focus of the paper. The proof is given in Section 5.1.

Proposition 2 *Suppose Assumptions P1(i),(ii) and R(i) are satisfied and that $r_* \geq 2$ and $r \geq 2$ hold. If $j_n \to \infty$ as $n \to \infty$ and $J_k \to \infty$ as $k \to \infty$ in such a way that for some $\delta > 1/2$ we have $\sup_{n \geq 1} 2^{j_n(2\delta+1)}/n < \infty$ and $\sup_{k \geq 1} J_k 2^{J_k(2\delta+1)}/k < \infty$, then*

$$\hat{\theta}_{n,k} \to \theta_0 \text{ in } \Pr\text{-probability as } n \wedge k \to \infty.$$

We note that the condition on j_n is, e.g., satisfied if $2^{j_n} \sim n^{\psi}$ with $0 < \psi < 1/2$. A similar comment applies to J_k. In particular, the 'textbook'-choice $\psi = 1/(2\tau + 1)$ with τ from Assumption P1(ii) is covered.

For the main result we need to distinguish several cases characterized by the behavior of the number $k(n) \in \mathbb{N}$ of simulated data as a function of sample size n:

Assumption S1: $\lim_{n \to \infty} k(n)/n^2 = \infty$.
Assumption S2: $\lim_{n \to \infty} k(n)/n = \infty$.
Assumption S3: $\lim_{n \to \infty} k(n)/n = \kappa$ for some $0 < \kappa < \infty$.

The theorem given below is the main result and shows that, under appropriate conditions on the resolution levels j_n and J_k, the indirect inference estimator $\hat{\theta}_{n,k}$ is asymptotically normal and has the same limiting distribution as the maximum likelihood estimator provided the number $k(n)$ of simulated data grows sufficiently fast as a function of sample size n. This is established under the quite weak assumption R(i) if $k(n)$ grows faster than n^2. If $k(n)$ is only required to grow faster than n, the same result is obtained under somewhat stronger assumptions (Assumption

R, $\tau > 3/2$, $r \geq 4$). Under the latter assumptions, the theorem also shows that in case $k(n)$ behaves asymptotically like n, the indirect inference estimator is still asymptotically normal but its asymptotic variance covariance matrix is then inflated by a factor $1 + 1/\kappa$, where $\kappa = \lim_{n\to\infty} k(n)/n$. We also note that the condition $\tau < r_* \wedge r$ in the subsequent theorem is virtually no restriction as discussed in Remark 2 below. The proof of the subsequent theorem is deferred to Section 5.

Theorem 1 *Suppose* $r \geq 2$ *and* $r_* \geq 2$ *hold and Assumption P1 is satisfied for some* $1/2 < \tau < r_* \wedge r$. *Suppose that* $2^{j_n} \sim n^{1/(2\tau+1)}$ *and* $2^{J_{k(n)}} \sim k(n)^{1/(2\tau+1)}$.

a. Suppose one of the following two conditions holds:

1. Assumptions R(i) and S1 hold.

2. Assumptions P2, R, and S2 hold, and that $\tau > 3/2$, $r \geq 4$ *are satisfied.*

Then

$$\sqrt{n}\left(\hat{\theta}_{n,k(n)} - \theta_0\right) \to^d N(0, I(\theta_0))$$

as $n \to \infty$ *where* $I(\theta_0) = \left(\int_0^1 \nabla_\theta p(\theta_0, x) \nabla_\theta p(\theta_0, x)' p(\theta_0, x)^{-1} dx\right)^{-1}$ *is the Cramér-Rao bound.*

b. Suppose Assumptions P2, R, and S3 hold for some $0 < \kappa < \infty$, *and that* $\tau > 3/2$, $r \geq 4$ *are satisfied. Then*

$$\sqrt{n}\left(\hat{\theta}_{n,k(n)} - \theta_0\right) \to^d N\left(0, (1 + 1/\kappa) I(\theta_0)\right)$$

as $n \to \infty$.

We note that the rates of increase for 2^{j_n} and $2^{J_{k(n)}}$ specified in the above theorem are precisely the rate-optimal choices based on mean integrated squared error. As already alluded to prior to the theorem, in Part a of the theorem there is a trade-off between the stringency of assumptions on the model and the simulation mechanism on the one hand and the assumptions on the rate of increase of $k(n)$ (Assumptions S1 versus S2) on the other hand. While the particular form of the trade-off is a consequence of two different methods of proof employed for Part a1 and Part a2 (and thus may in principle be an artefact), it seems plausible that some sort of trade-off is intrinsic to the problem.

Remark 2 *(i) The condition* $\tau < r_* \wedge r$ *in the above theorem is not really a restriction on* $\mathcal{P}_\Theta$ *and can always be achieved in the following sense: If Assumption P1 holds with* $\tau \geq r_* \wedge r$, *it holds with* τ *replaced by any* τ' *satisfying* $1/2 < \tau' < r_* \wedge r$ *as well, since* B_τ *is continuously imbedded in* $\mathsf{B}_{\tau'}$ *for* $\tau' \leq \tau$. *Consequently, the above theorem can be applied with* τ' *replacing* τ *(requiring also* $\tau' > 3/2$ *for Parts a2 and b). [The restriction* $\tau < r_* \wedge r$ *in the theorem simply expresses the fact that the rate of increase of* j_n *and* J_k *is not only governed by the degree of "regularity"* τ *of the densities in* $\mathcal{P}_\Theta$, *but also by the degrees of "regularity" of the splines used to estimate* p_0 *and* $p(\theta)$, *respectively, i.e., by* r_* *and* r.*]*

(ii) The argument underlying (i) also shows that $2^{j_n} \sim n^{1/(2\tau'+1)}$ *and* $2^{J_{k(n)}} \sim k(n)^{1/(2\tau'+1)}$ *are feasible in Theorem 1 as it stands as long as* $1/2 < \tau' \leq \tau$ *(and* $\tau' > 3/2$ *for Parts a2 and b) are satisfied. A careful examination of the proof shows that the range for* 2^{j_n} *and* $2^{J_{k(n)}}$, *under which the conclusion of the theorem holds, is actually somewhat wider. However, we abstain from providing such results as they quickly get unwieldy.*

(iii) If in Part a2 of Theorem 1 the Assumption S2 is strengthened by assuming a particular growth-rate for $k(n)$ *such as, e.g.,* $k(n) = n^\delta$, $1 < \delta \leq 2$, *this can be used to relax the assumption* $\tau > 3/2$. *We refrain from presenting such results.*

(iv) If $k(n)$ *is such that* $0 < \liminf k(n)/n < \infty$, *but* $\limsup k(n)/n = \infty$, *then the distribution* $\sqrt{n}\left(\hat{\theta}_{n,k(n)} - \theta_0\right)$ *does not possess a limit, but 'oscillates' between accumulation points of the form* $N(0, I(\theta_0))$ *and* $N(0, (1 + 1/\kappa) I(\theta_0))$ *where now* $\kappa = \liminf_{n\to\infty} k(n)/n$.

(v) A result similar to Part a1 of Theorem 1 can be proved in case $r^* = 1$. *Since this requires a separate proof, we do not give such a result for the sake of brevity.*

Under Assumption P1 the expression $\Psi(\theta) = \int_0^1 \nabla_\theta p(\theta) \nabla_\theta p(\theta)' p(\theta)^{-1} d\lambda$ depends continuously on θ by dominated convergence. Hence, $\Psi(\bar{\theta})^{-1}$ is a consistent estimator for $I(\theta_0)$ for every consistent estimator $\bar{\theta}$. However, this observation is not very helpful in the context of indirect inference as then expressions for the density $p(\theta)$ are typically not available. An alternative consistent estimator that is feasible to compute is described in the next proposition which is proved in Section 5.5. In the following proposition let $\bar{\theta}_{n,k}$ stand for an arbitrary consistent estimator that depends on the original data and perhaps also on the simulated data. Of course, under the assumptions of Proposition 2 we may take $\bar{\theta}_{n,k} = \hat{\theta}_{n,k}$.

Proposition 3 *Suppose Assumptions P1(i)-(iii), P2(i), and R(ii) hold. Suppose further that* $\bar{\theta}_{n,k} \to \theta_0$ *in probability as* $n \wedge k \to \infty$. *Assume* $r'_* \geq 2$ *and* $r' \geq 3$. *If* $j'_n \to \infty$ *as* $n \to \infty$ *and* $J'_k \to \infty$ *as* $k \to \infty$ *in such a way that for some* $\delta > 1/2$ *we have* $\sup_{n \geq 1} 2^{j'_n(2\delta+1)}/n < \infty$ *and also* $J'_k 2^{3J'_k}/k \to 0$, *then*

$$\left(\int_0^1 \nabla_\theta p_{k,J'_k,r'}(\bar{\theta}_{n,k}) \nabla_\theta p_{k,J'_k,r'}(\bar{\theta}_{n,k})' p^{-1}_{n,j_n,r'_*} d\lambda \right)^{-1}$$

is well-defined on an event that has probability converging to 1, *and is a consistent estimator for* $I(\theta_0)$ *as* $n \wedge k \to \infty$.

Observe that the condition on j'_n is satisfied if $2^{j'_n} \sim n^\psi$ with $0 < \psi < 1/2$; similarly, the condition on J'_k is satisfied if $2^{J'_k} \sim n^\psi$ with $0 < \psi < 1/3$. The reason for allowing r' to differ from r in Theorem 1, is to be able to construct a consistent estimator for $I(\theta_0)$ also in cases where $r = 2$. Allowing J'_k to be different from J_k has the advantage of avoiding a constraint on τ.

4 Dyadic Splines

Let $T_j = \{t_l := l2^{-j} : l = 1, \ldots, 2^j - 1\}$ be a dyadic set of knots in $[0, 1]$, where $j \in \mathbb{N}$, the set of nonnegative integers. A function $S : [0, 1] \to \mathbb{R}$ is a (dyadic) spline of order $r \geq 2$ if on each of the intervals $[0, t_1)$, (t_l, t_{l+1}) for $l = 1, \ldots, 2^j - 2$, and $(t_{2^j-1}, 1]$, it is a polynomial of degree not larger than $r - 1$, and on at least one of the intervals it is a polynomial of degree exactly $r - 1$. The Schoenberg spaces $\mathcal{S}_j(r)$ considered here consist of all splines of order less than or equal to r that are $r - 2$ times continuously differentiable on $[0, 1]$ (using one-sided derivatives on the boundary of $[0, 1]$). For $r = 1$ we define the Schoenberg space $\mathcal{S}_j(1)$ to be the space of all functions $S : [0, 1] \to \mathbb{R}$ that are constant on the intervals $[0, t_1)$, $[t_l, t_{l+1})$ for $l = 1, \ldots, 2^j - 2$, and $[t_{2^j-1}, 1]$. The Schoenberg spaces are linear spaces of dimension $2^j + r - 1$. For $r \geq 2$ the B-spline basis for $\mathcal{S}_j(r)$ is given by $\{N^{(r)}_{lj} : l = -r + 1, \ldots, 0, 1, \ldots, 2^j - 1\}$ with

$$N^{(r)}_{lj}(x) = N^{(r)}(2^j x - l) \qquad \text{for } x \in [0, 1],$$

where $N^{(r)}$ is the B-spline-function (of order r) given by the r-fold convolution

$$N^{(r)}(u) = \mathbf{1}_{[0,1)} * \ldots * \mathbf{1}_{[0,1)}(u) \qquad \text{for } u \in \mathbb{R};$$

cf., e.g., Chapter 5 in DeVore and Lorentz (1993). In case $r = 1$ we set

$$N^{(1)}_{lj}(x) = N^{(1)}(2^j x - l) \qquad \text{for } x \in [0, 1],$$

for $l = 0, 1, \ldots, 2^j - 2$, where $N^{(1)}(u) = \mathbf{1}_{[0,1)}(u)$, but we set

$$N_{lj}^{(1)}(x) = \mathbf{1}_{[0,1]}(2^j x - l) \qquad \text{for } x \in [0,1]$$

if $l = 2^j - 1$. The B-spline basis functions $N_{lj}^{(r)}$ are nonnegative, bounded by 1 in absolute value, and form a partition of unity, i.e.,

$$\sum_{l=-r+1}^{2^j-1} N_{lj}^{(r)}(x) = 1 \qquad \text{for } x \in [0,1], \tag{10}$$

for every $j, r \in \mathbb{N}$.

The Schoenberg space $\mathcal{S}_j(r)$ is a finite-dimensional linear subspace of $\mathcal{L}^2$. The ortho-projection $\pi_j^{(r)}$ from $\mathcal{L}^2$ onto $\mathcal{S}_j(r)$ is given by

$$\pi_j^{(r)}(f) = \sum_{l=-r+1}^{2^j-1} \gamma_{lj}^{(r)}(f) N_{lj}^{(r)}$$

where

$$\gamma_{lj}^{(r)}(f) = \sum_{m=-r+1}^{2^j-1} 2^j g_j^{(r)lm} \int_0^1 N_{mj}^{(r)}(x) f(x) dx$$

and $g_j^{(r)lm}$ is the (l,m)-element of the inverse of the $(2^j + r - 1) \times (2^j + r - 1)$ matrix

$$G_j^{(r)} = \left(\int_0^{2^j} N^{(r)}(u-l) N^{(r)}(u-m) du \right)_{l,m}.$$

Note that $G_j^{(r)}$ is a symmetric bandmatrix with bandwidth r. The projection can now also be written as

$$\pi_j^{(r)}(f)(y) = \int_0^1 K_j^{(r)}(x,y) f(x) dx \tag{11}$$

with the kernel given by

$$K_j^{(r)}(x,y) = 2^j \sum_{l=-r+1}^{2^j-1} \sum_{m=-r+1}^{2^j-1} g_j^{(r)lm} N^{(r)}(2^j x - m) N^{(r)}(2^j y - l).$$

We shall frequently need to bound the maximal row-sum of the absolute values of the elements of the inverse of $G_j^{(r)}$, i.e., the ℓ^∞-operator norm of the inverse of $G_j^{(r)}$. For this we use the following special case of a result in Shadrin (2001, Theorem I and Section 4.2).

Proposition 4 *For every $r \in \mathbb{N}$ there exist constants $0 < d_r < \infty$ (independent of j) such that for every $j \in \mathbb{N}$*

$$\left\| \left(G_j^{(r)} \right)^{-1} \right\|_{\infty \to \infty} \leq d_r$$

where $\|\cdot\|_{\infty \to \infty}$ denotes the ℓ^∞-operator norm on $\mathbb{R}^{2^j+r-1}$.

We furthermore note that for $r \geq 2$ the Schoenberg space $\mathcal{S}_j(r)$ is contained in the Sobolev space of order $r-1$, and thus is also contained in B_{r-1}. In fact, for every $r \geq 1$ we have that $\mathcal{S}_j(r)$ is contained in B_s for $s \leq r - 1/2$ (DeVore and Lorentz (1993), Chap. 12, Lemma 3.1). Some approximation properties of splines that we shall use in the sequel are summarized in Appendix A.

For the spline projection estimators defined in Section 3 we make the useful observation that for every $J \geq 1$ and $r \geq 1$

$$\|p_{k,J,r}(\theta)\|_\infty \leq 2^J d_r (2^J + r - 1) \tag{12}$$

holds uniformly in $\theta \in \Theta$, $k \geq 1$, and $v_1, \ldots, v_k \in \mathcal{V}$. [To see this note that the B-spline basis functions are uniformly bounded by 1 and that the coefficients satisfy $\left|\hat{\gamma}_{lJ}^{(r)}(\theta)\right| \leq 2^J d_r$ uniformly in $\theta \in \Theta$, $k \geq 1$, $-r+1 \leq l \leq 2^J - 1$, and $v_1, \ldots, v_k \in \mathcal{V}$ by Proposition 4.] The analogous relation is true for $\|p_{n,j,r_*}\|_\infty$, as well as for $\|Ep_{k,J,r}(\theta)\|_\infty$ and $\|Ep_{n,j,r_*}\|_\infty$.

5 Proofs

We shall use repeatedly in this section the fact that $\xi_0 := \inf_{x\in[0,1]} p(\theta_0, x) > 0$ under Assumptions P1(i),(ii) (as $p(\theta_0)$ is continuous and positive on $[0,1]$ under these assumptions).

5.1 Proof of Proposition 2

Define the function

$$Q(\theta) = \int_0^1 (p(\theta_0) - p(\theta))^2 p^{-1}(\theta_0) d\lambda, \tag{13}$$

which is real-valued and is continuous in θ by dominated convergence, observing that $\xi_0 > 0$ and that Assumption P1(ii) implies sup-norm boundedness of $\mathcal{P}_\Theta$ in view of the discussion following Proposition 7 in Appendix A. The unique minimizer of $Q(\theta)$ over Θ is θ_0 in view of the identifiability assumption made in Assumption P1(i). To establish consistency, it is hence sufficient to prove

$$\sup_{\theta\in\Theta} |\mathcal{Q}_{n,k}(\theta) - Q(\theta)| \to 0$$

in probability as $n \wedge k \to \infty$. Note that this supremum is measurable as $\mathcal{Q}_{n,k}(\theta)$ and $Q(\theta)$ are continuous and Θ is separable. [For continuity of $\mathcal{Q}_{n,k}$ see the proof of Proposition 1 in Appendix B.] Consider the set $A_n^* = \left\{\inf_{y\in[0,1]} p_{n,j_n,r_*}(y) \geq \xi_0/2\right\}$, which is clearly measurable. Since $\xi_0 > 0$ as noted above, Corollary 2 (applied with $t = \delta \wedge \tau \wedge 1$ and noting that $p(\theta_0)$ is a continuous version of p_0 in view of Assumption P1(i)) implies that $\Pr(A_n^*) \to 1$ as $n \to \infty$. A simple calculation now shows that on the event A_n^* (since $A_n^* \subseteq A_n$)

$$\begin{aligned}
\mathcal{Q}_{n,k}(\theta) - Q(\theta) \quad &= \quad \int_0^1 (p_{n,j_n,r_*} - p(\theta_0)) \left[1 - \frac{p(\theta)^2}{p_{n,j_n,r_*} p(\theta_0)}\right] d\lambda + \int_0^1 (p_{k,J_k,r}(\theta) - p(\theta))^2 p_{n,j_n,r_*}^{-1} \\
&\quad +2 \int_0^1 (p_{k,J_k,r}(\theta) - p(\theta)) \left[\frac{p(\theta)}{p_{n,j_n,r_*}} - 1\right] d\lambda
\end{aligned}$$

holds. On A_n^* we can then obtain the bound

$$\begin{aligned}\sup_{\theta\in\Theta}|\mathcal{Q}_{n,k}(\theta)-Q(\theta)| \quad\leq\quad & \|p_{n,j_n,r_*}-p(\theta_0)\|_\infty\left(1+2\xi_0^{-2}\sup_{\theta\in\Theta}\|p(\theta)\|_\infty^2\right)\\ &+2\xi_0^{-1}\sup_{\theta\in\Theta}\|p_{k,J_k,r}(\theta)-p(\theta)\|_\infty^2\\ &+\sup_{\theta\in\Theta}\|p_{k,J_k,r}(\theta)-p(\theta)\|_\infty\left(2+4\xi_0^{-1}\sup_{\theta\in\Theta}\|p(\theta)\|_\infty\right).\end{aligned}$$

The sup-norm boundedness of $\mathcal{P}_\Theta$ together with Corollaries 1 and 2 (applied with $t=\delta\wedge\tau\wedge 1$) then complete the proof.

5.2 An Intermediate Result

Consider the objective function

$$Q_n(\theta):=Q_{n,j,r_*}(\theta)=\begin{cases}\int_0^1\left(p_{n,j,r_*}-p(\theta)\right)^2 p_{n,j,r_*}^{-1}\,d\lambda & \text{on the event } A_n\\ 0 & \text{otherwise}\end{cases}, \tag{14}$$

corresponding to the 'ideal' case $k=\infty$. Let $\hat{\theta}_n:=\hat{\theta}_{n,j,r_*}$ denote an arbitrary measurable minimizer of (14) over Θ. [The existence of such an estimator is established in Proposition 10 in Appendix B.]

Theorem 2 *Suppose $r_*\geq 2$ holds and Assumption P1 is satisfied with $1/2<\tau<r_*$. If $2^{j_n}\sim n^{1/(2\tau+1)}$, then, as $n\to\infty$,*

$$\sqrt{n}\left(\hat{\theta}_n-\theta_0\right)\to^d N\left(0,I(\theta_0)\right).$$

Proof. Consistency of $\hat{\theta}_n$ follows from Proposition 11 in Appendix B by choosing δ in that proposition sufficiently close to 1/2. It follows that $\hat{\theta}_n\in B(\theta_0)$ with probability tending to 1, and hence $\hat{\theta}_n$ belongs to the interior of Θ with probability tending to 1. In the following we work only on the intersection of the event $\left\{\hat{\theta}_n\in B(\theta_0)\right\}$ with $A_n^*=\left\{\inf_{y\in[0,1]}p_{n,j_n,r_*}(y)\geq\xi_0/2\right\}$ which also has probability converging to 1 as a consequence of Corollary 2 (applied with some t satisfying $1/2<t\leq\tau\wedge 1$). Note that $\|p_{n,j_n,r_*}\|_\infty<\infty$ holds, and that $\left\|p_{n,j_n,r_*}^{-1}\right\|_\infty\leq 2/\xi_0$ on the event A_n^*. Furthermore, by Assumption P1(ii) the function $p(\theta)$ is bounded, uniformly in θ, cf. Proposition 7 and the attending discussion in Appendix A. Assumption P1(iii) and dominated convergence then show that $Q_n(\theta)$ is twice continuously differentiable on the open ball $B(\theta_0)$ with derivatives given by

$$\nabla_\theta Q_n(\theta)=-2\int_0^1\left(p_{n,j_n,r_*}-p(\theta)\right)p_{n,j_n,r_*}^{-1}\nabla_\theta p(\theta)d\lambda,$$

$$\nabla_\theta^2 Q_n(\theta)=2\int_0^1 p_{n,j_n,r_*}^{-1}\nabla_\theta p(\theta)\nabla_\theta p(\theta)'d\lambda-2\int_0^1\left(p_{n,j_n,r_*}-p(\theta)\right)p_{n,j_n,r_*}^{-1}\nabla_\theta^2 p(\theta)d\lambda, \tag{15}$$

and these derivatives are measurable functions for every $\theta\in B(\theta_0)$. Since $\hat{\theta}_n$ is an interior maximizer of Q_n (on the event considered), we have that $\nabla_\theta Q_n(\hat{\theta}_n)=0$. Consequently, a standard Taylor expansions gives

$$0=\nabla_\theta Q_n(\hat{\theta}_n)=\nabla_\theta Q_n(\theta_0)+\nabla_\theta^2 Q_n^*(\hat{\theta}_n-\theta_0), \tag{16}$$

where the i-th row of $\nabla^2_\theta Q_n^*$ equals the corresponding row of $\nabla^2_\theta Q_n$ evaluated at a mean-value $\tilde{\theta}_n^{(i)}$ which may depend on the row-index (measurability of $\tilde{\theta}_n^{(i)}$ being no concern here). We now first establish that $n^{1/2}\nabla_\theta Q_n(\theta_0)$ is asymptotically normal with mean zero and variance-covariance matrix $4\int_0^1 \nabla_\theta p(\theta_0)\nabla_\theta p(\theta_0)' p^{-1}(\theta_0) d\lambda$. To this end write $(-1/2)n^{1/2}\nabla_\theta Q_n(\theta_0)$ as

$$\sqrt{n}\int_0^1 (p_{n,j_n,r_*} - p(\theta_0))\, p(\theta_0)^{-1}\nabla_\theta p(\theta_0) d\lambda +$$
$$\sqrt{n}\int_0^1 (p_{n,j_n,r_*} - p(\theta_0))\, (p^{-1}_{n,j_n,r_*} - p(\theta_0)^{-1})\nabla_\theta p(\theta_0) d\lambda,$$

both terms being measurable. The first term in the above display now converges to the required limit by Theorem 4 (applied with $t = \tau$, and some s satisfying $1/2 < s < 1$, $s \leq \varsigma \wedge \tau$) and the Cramér-Wold device: To see this, observe that $p_0 \in \mathcal{B}_t$ by Assumption P1(i),(ii) (since $p_0 = p(\theta_0)$ λ-a.e.). Furthermore, for every $\alpha \in \mathbb{R}^b$, $\alpha \neq 0$, the function $f = p(\theta_0)^{-1}\alpha'\nabla_\theta p(\theta_0)$ belongs to $\mathsf{B}_{\varsigma\wedge\tau}$ as a consequence of Assumption P1(ii),(iv) and Proposition 7 in Appendix A. Hence $\mathcal{F} = \{f\} \subseteq \mathsf{B}_s$. The conditions on j_n in Theorem 4 follow from the assumption on j_n in the current theorem. Finally note that $P(f) = 0$ under Assumption P1. The second term in the above display is bounded in norm (on the event A_n^*) by

$$\begin{aligned} & n^{1/2}\int_0^1 (p_{n,j_n,r_*} - p(\theta_0))^2\, p(\theta_0)^{-1} p^{-1}_{n,j_n,r_*} \left\|\nabla_\theta p(\theta_0)\right\| d\lambda \\ \leq\; & (2/\xi_0^2) \sup_{x\in[0,1]} \left\|\nabla_\theta p(\theta_0, x)\right\| n^{1/2} \left\|p_{n,j_n,r_*} - p(\theta_0)\right\|_2^2, \end{aligned}$$

noting that $\left\|p(\theta_0)^{-1}\right\|_\infty \leq \xi_0^{-1}$, and that $\frac{\partial}{\partial\theta_q}p(\theta_0)$ is bounded on $[0,1]$ for every q since it belongs to B_ς with $\varsigma > 1/2$ by Assumption P1(iv). By Lemma 3 the r.h.s in the above display is $O_p(n^{-1/2}2^{j_n} + n^{1/2}2^{-2j_n\tau})$ which is $o_p(1)$ because of $\tau > 1/2$.

Next we show that $\nabla^2_\theta Q_n^*$ converges to the positive definite matrix $\nabla^2_\theta Q(\theta_0)$ in (outer) probability. To this end we first show that $\nabla^2_\theta Q_n(\theta)$ converges to $\nabla^2_\theta Q(\theta)$ uniformly over $B(\theta_0)$ in probability where $Q(\theta)$ has been defined in (13). By Assumption P1 and dominated convergence we have that $Q(\theta)$ is twice continuously differentiable on $B(\theta_0)$ with

$$\nabla^2_\theta Q(\theta) = 2\int_0^1 p(\theta_0)^{-1}\nabla_\theta p(\theta)\nabla_\theta p(\theta)' d\lambda - 2\int_0^1 (p(\theta_0) - p(\theta))\, p(\theta_0)^{-1}\nabla^2_\theta p(\theta) d\lambda.$$

We now see that

$$\begin{aligned} & \nabla^2_\theta Q_n(\theta) - \nabla^2_\theta Q(\theta) \\ =\; & 2\int_0^1 (p^{-1}_{n,j_n,r_*} - p(\theta_0)^{-1})\nabla_\theta p(\theta)\nabla_\theta p(\theta)' - 2\int_0^1 (p_{n,j_n,r_*} - p(\theta))\, (p^{-1}_{n,j_n,r_*} - p(\theta_0)^{-1})\nabla^2_\theta p(\theta) \\ & +2\int_0^1 (p(\theta_0) - p_{n,j_n,r_*})\, p(\theta_0)^{-1}\nabla^2_\theta p(\theta) \end{aligned}$$

and we obtain (the supremum being measurable because of continuity of $\nabla^2_\theta Q_n$ and $\nabla^2_\theta Q$ on $B(\theta_0)$)

$$
\begin{aligned}
& \sup_{\theta\in B(\theta_0)} \left\|\nabla^2_\theta Q_n(\theta) - \nabla^2_\theta Q(\theta)\right\| && (17) \\
\leq \;& 2\left\|p_{n,j_n,r_*} - p(\theta_0)\right\|_\infty \sup_{\theta\in B(\theta_0)} \left[\int_0^1 p^{-1}_{n,j_n,r_*} p(\theta_0)^{-1} \left\|\nabla_\theta p(\theta)\right\|^2 d\lambda \right.\\
& \left. + \int_0^1 \left|p_{n,j_n,r_*} - p(\theta)\right| p^{-1}_{n,j_n,r_*} p(\theta_0)^{-1} \left\|\nabla^2_\theta p(\theta)\right\| d\lambda + \int_0^1 p(\theta_0)^{-1}\left\|\nabla^2_\theta p(\theta)\right\| d\lambda\right] \\
\leq \;& \left\|p_{n,j_n,r_*} - p(\theta_0)\right\|_\infty \left[4\xi_0^{-2}\int_0^1 \sup_{\theta\in B(\theta_0)} \left\|\nabla_\theta p(\theta)\right\|^2 d\lambda \right.\\
& \left. + \left(4\xi_0^{-2}\left(\left\|p_{n,j_n,r_*}\right\|_\infty + \sup_{\theta\in B(\theta_0)}\left\|p(\theta)\right\|_\infty\right) + 2\xi_0^{-1}\right)\int_0^1 \sup_{\theta\in B(\theta_0)}\left\|\nabla^2_\theta p(\theta)\right\| d\lambda\right] = o_p(1),
\end{aligned}
$$

by Assumption P1 and Corollary 2 (applied with a t satisfying $1/2 < t \leq \tau \wedge 1$). Since $\nabla^2_\theta Q(\theta)$ is continuous at θ_0 as shown above and since $\hat{\theta}_n$ is consistent, convergence of $\nabla^2_\theta Q^*_n$ to $\nabla^2_\theta Q(\theta_0)$ in (outer) probability follows.

The central limit theorem for the score together with the convergence result for $\nabla^2_\theta Q^*_n$ just established delivers now the desired result: rewrite (16) as

$$0 = n^{1/2}\nabla_\theta Q_n(\theta_0) + \nabla^2_\theta Q(\theta_0) n^{1/2}(\hat{\theta}_n - \theta_0) + \left(\nabla^2_\theta Q^*_n - \nabla^2_\theta Q(\theta_0)\right) n^{1/2}(\hat{\theta}_n - \theta_0),$$

observe that $\nabla^2_\theta Q(\theta_0)$ is positive definite by Assumption P1(iii), and that the third term on the r.h.s. is of lower order than the second one. This implies that $n^{1/2}(\hat{\theta}_n - \theta_0)$ is stochastically bounded, and the desired result then easily follows. ■

For the same reasons as given in Remark 2, the condition $\tau < r_*$ in the above theorem is not really a restriction. Furthermore, examining the proof shows that the conclusions of the theorem also hold for other choices of 2^{j_n}: e.g., the theorem (without the condition $\tau < r_*$) holds for $2^{j_n} \sim n^\nu$ with ν satisfying $1/\left(2\left((\tau\wedge r_*) + (\varsigma\wedge\tau\wedge 1)\right)\right) < \nu < 1/2$.

5.3 Proof of Part a1 of Theorem 1

We first provide an auxiliary result that relates the objective function $\mathcal{Q}_{n,k}(\theta)$ to the somewhat simpler objective function $Q_n(\theta)$ studied in the preceding section. Note that k is not linked to n in the subsequent proposition.

Proposition 5 *Suppose $r \geq 2$ and $r_* \geq 2$ hold and Assumptions P1(i),(ii) are satisfied for some $1/2 < \tau < r_* \wedge r$. Suppose further that Assumption R(i) is satisfied and that $2^{j_n} \sim n^{1/(2\tau+1)}$ and $2^{J_k} \sim k^{1/(2\tau+1)}$. Then for every $\varepsilon > 0$ there exists a positive real number $M(\varepsilon)$ and a natural number $N(\varepsilon)$ such that*

$$\Pr\left(k^{1/2}\sup_{\theta\in\Theta}\left|\mathcal{Q}_{n,k}(\theta) - Q_n(\theta)\right| > M(\varepsilon)\right) < \varepsilon \qquad (18)$$

holds for all $n \geq N(\varepsilon)$ and all $k \geq 1$.

Proof. First note that the supremum in (18) is measurable since $\mathcal{Q}_{n,k}(\theta)$ and $Q_n(\theta)$ are continuous in θ as noted before, cf. Section 5.1. For given $\varepsilon > 0$ choose $N(\varepsilon)$ large enough such that for $n \geq N(\varepsilon)$ we have $\Pr(A_n^*) > 1 - \varepsilon$ where $A_n^* = \left\{\inf_{y\in[0,1]} p_{n,j_n,r_*}(y) \geq \xi_0/2\right\}$. This is possible by Corollary 2. A simple calculation shows that on the event A_n^*

$$\mathcal{Q}_{n,k}(\theta) - Q_n(\theta) = \int_0^1 (p_{k,J_k,r}(\theta) - p(\theta)) \left[\frac{p_{k,J_k,r}(\theta) + p(\theta)}{p_{n,j_n,r_*}} - 2\right]$$

holds. Choose s to satisfy $1/2 < s < \tau \wedge 1$. Applying Corollaries 1 and 2 (with $t = s$) shows that for the given $\varepsilon > 0$ there exists a positive finite D such that the events

$$A_{n,k}^{**} = \left\{\sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta)\|_{s,2} \leq D, \|p_{n,j_n,r_*}\|_{s,2} \leq D\right\}$$

have probability not less than $1 - \varepsilon$ for every $k \geq 1$ and $n \geq 1$. Applying Proposition 7 in Appendix A, we conclude that there exists a finite positive D', depending only on D, ξ_0, and $\sup_{\theta\in\Theta} \|p(\theta)\|_{s,2}$ (which is finite by Assumption P1(ii) and continuous embedding of B_τ in B_s), such that on $A_n^* \cap A_{n,k}^{**}$

$$\sup_{\theta\in\Theta} \left\|(p_{k,J_k,r}(\theta) + p(\theta))p_{n,j_n,r_*}^{-1} - 2\right\|_{s,2} \leq D'$$

holds. Thus for every $M > 0$, all $k \geq 1$, and all $n \geq N(\varepsilon)$

$$\begin{aligned}
&\Pr\left(\sqrt{k}\sup_{\theta\in\Theta} |\mathcal{Q}_{n,k}(\theta) - Q_n(\theta)| > M\right) \\
\leq\ &\Pr\left(\left\{\sqrt{k}\sup_{\theta\in\Theta}\sup_{\|f\|_{s,2}\leq D'} \left|\int_0^1 (p_{k,J_k,r}(\theta) - p(\theta)) f\, d\lambda\right| > M\right\} \cap A_n^* \cap A_{n,k}^{**}\right) + 2\varepsilon \\
\leq\ &\Pr\left(\left\{\sqrt{k}\sup_{\theta\in\Theta} \|P_{k,J_k,r}(\theta) - P(\theta)\|_{\mathcal{F}} > M\right\}\right) + 2\varepsilon
\end{aligned}$$

where $\mathcal{F}$ denotes $\{f \in \mathsf{B}_s : \|f\|_{s,2} \leq D'\}$ and $\|\cdot\|_{\mathcal{F}}$ is defined before Theorem 3. Choose an s' satisfying $1/2 < s' < s$. Then Theorem 3 (applied with $t = \tau$) implies for every $k \geq 1$

$$\begin{aligned}
\sqrt{k}\sup_{\theta\in\Theta} \|P_{k,J_k,r}(\theta) - P(\theta)\|_{\mathcal{F}} &\leq \sqrt{k}\sup_{\theta\in\Theta} \|P_{k,J_k,r}(\theta) - P_k(\theta)\|_{\mathcal{F}} + \sqrt{k}\sup_{\theta\in\Theta} \|P_k(\theta) - P(\theta)\|_{\mathcal{F}} \\
&= O_p\left(\sqrt{k}2^{-J_k(\tau+s)} + 2^{-J_k(s-s')} + 1\right) = O_p(1).
\end{aligned}$$

[Measurability of the suprema on the r.h.s. in the first line of the above display is established in the proof of Theorem 3. The argument given there also establishes measurability of the supremum on the l.h.s.] This completes the proof (noting that the l.h.s. in the above display is certainly a *real-valued* random variable for every k). ■

The closeness of $\mathcal{Q}_{n,k}$ and Q_n expressed in the previous result translates into closeness of the minimizers of these functions with the help of the following simple but useful lemma which is taken from Gach (2009). Note that M_2 below is smooth but M_1 need not be so. This is relevant as $\mathcal{Q}_{n,k}$ is not guaranteed to be smooth under the assumptions of Part a1 of Theorem 1, whereas Q_n is in view of Assumption P1.

Lemma 1 *Let U be a nonempty convex open subset of $\mathbb{R}^b$. Suppose we are given functions $M_1 : U \to \mathbb{R}$ and $M_2 : U \to \mathbb{R}$, such that M_2 is twice partially differentiable on U with Hessian satisfying*

$$\inf_{x \in U} y' \nabla_x^2 M_2(x) y \geq c \|y\|^2 \tag{19}$$

for every $y \in \mathbb{R}^b$ and some $0 < c < \infty$. If $m_1 \in U$ and $m_2 \in U$ minimize M_1 and M_2 over U, respectively, we have

$$\|m_1 - m_2\| \leq 2c^{-1/2} \sqrt{\sup_{u \in U} |M_1(u) - M_2(u)|}$$

where $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^b$.

Proof. Assume that minimizers m_1 and m_2 exist, since otherwise there is nothing to prove. [By convexity of U and the assumption on the Hessian the minimizer m_2 is unique.] Since m_2 is a minimizer of the twice partially differentiable function M_2 on the convex open set U, we have

$$M_2(m_1) = M_2(m_2) + 2^{-1}(m_1 - m_2)' \nabla_x^2 M_2(\tilde{m})(m_1 - m_2)$$

(using a pathwise Taylor series expansion) where $\tilde{m}$ lies in the convex hull of $\{m_1, m_2\}$. We conclude from the assumption on the Hessian that

$$\|m_1 - m_2\| \leq (2c^{-1})^{1/2} \sqrt{|M_2(m_1) - M_2(m_2)|}. \tag{20}$$

Observe next that

$$M_1(m_1) - M_2(m_2) \leq M_1(m_2) - M_2(m_2) \leq \sup_{u \in U} |M_1(u) - M_2(u)|$$

and

$$M_1(m_1) - M_2(m_2) \geq M_1(m_1) - M_2(m_1) \geq - \sup_{u \in U} |M_1(u) - M_2(u)|$$

so that

$$|M_1(m_1) - M_2(m_2)| \leq \sup_{u \in U} |M_1(u) - M_2(u)|.$$

Consequently,

$$|M_2(m_1) - M_2(m_2)| \leq |M_2(m_1) - M_1(m_1)| + |M_1(m_1) - M_2(m_2)| \leq 2 \sup_{u \in U} |M_1(u) - M_2(u)|,$$

which, when plugged into (20), proves the lemma. ■

The proof of Part a1 of Theorem 1 is now as follows: Let $U \subseteq B(\theta_0)$ be a sufficiently small open ball around θ_0 such that the smallest eigenvalues of $\nabla_\theta^2 Q(\theta)$ are bounded away from zero by a positive constant, η say, uniformly in $\theta \in U$. Such an U exists, since $\nabla_\theta^2 Q(\theta)$ is continuous on $B(\theta_0)$, as shown in Section 5.2, and since $\nabla_\theta^2 Q(\theta_0)$ is positive definite by Assumption P1. Now apply Lemma 1 with $M_1 = \mathcal{Q}_{n,k(n)}$, $M_2 = Q_n$, and the set U just mentioned. Note that condition (19) is then satisfied for $M_2 = Q_n$ and $c = \eta/2$ on an event E_n that has probability converging to 1 in view of the choice of U and since it was shown in the proof of Theorem 2 that $\nabla_\theta^2 Q_n(\theta)$ converges to $\nabla_\theta^2 Q(\theta)$ uniformly on $B(\theta_0)$ in probability. Observe also that Proposition 5 implies

$$\sup_{\theta \in \Theta} |\mathcal{Q}_{n,k(n)}(\theta) - Q_n(\theta)| = O_p(k(n)^{-1/2}).$$

Taken together, this implies

$$\left\| \hat{\theta}_{n,k(n)} - \hat{\theta}_n \right\| = O_p(k(n)^{-1/4}), \tag{21}$$

which is $o_p(n^{-1/2})$ in view of Assumption S1. Part a1 of Theorem 1 now follows from asymptotic normality of $\sqrt{n}\left(\hat{\theta}_n - \theta_0\right)$ which has already been established in Theorem 2.

5.4 Proof of the Remaining Parts of Theorem 1

Observe first that it suffices to show that every subsequence n_i of n contains a further subsequence $n_{i(l)}$ along which the claimed asymptotic normality result holds. Given n_i, we may choose the subsequence $n_{i(l)}$ in such a way that $\lim_{l\to\infty} k(n_{i(l)})/n_{i(l)}^2$ exists (possibly being ∞) since the extended real line is compact. But the sequence $k(n_{i(l)})$ can be viewed as the subsequence $\bar{k}(n_{i(l)})$ of a sequence $\bar{k}(n)$ for which $\lim_{n\to\infty} \bar{k}(n)/n^2$ exists (and necessarily equals $\lim_{l\to\infty} k(n_{i(l)})/n_{i(l)}^2$). This shows that for the proof we may assume without loss of generality that $\lim_{n\to\infty} k(n)/n^2$ exists (possibly being ∞). In the case where this limit is infinite, the results then follow from Part a1 which has already been proved in Section 5.3. Thus we may assume without loss of generality not only that the limit of $k(n)/n^2$ exists, but also that

$$\lim_{n\to\infty} k(n)/n^2 < \infty. \tag{22}$$

We shall make this assumption for the remainder of this section.

Under Assumption R and if $r \geq 4$ the mapping

$$\theta \mapsto p_{k,J,r}(\theta, y) = \sum_{l=-r+1}^{2^J-1} \sum_{m=-r+1}^{2^J-1} 2^J g_J^{(r)lm} \left(k^{-1} \sum_{i=1}^{k} N_{mJ}^{(r)}(\rho(V_i, \theta)) \right) N_{lJ}^{(r)}(y)$$

is twice continuously differentiable on $B(\theta_0)$ for every y and every realization of $V_1, \ldots, V_k$ by the chain rule. Similarly as in the proof of Theorem 2, it suffices to work only on the event $A_n^* \cap \left\{ \hat{\theta}_{n,k(n)} \in B(\theta_0) \right\}$ which has probability converging to 1 in view of Proposition 2 (applied with $\delta > 1/2$ sufficiently close to $1/2$) and Corollary 2 (applied with some t satisfying $1/2 < t \leq \tau \wedge 1$). Note that $\left\| p(\theta_0)^{-1} \right\|_\infty \leq \xi_0$, and that $\left\| p_{n,j_n,r_*}^{-1} \right\|_\infty \leq 2/\xi_0$ holds on the before mentioned event; we shall use these facts repeatedly in the sequel. Using this, (12), boundedness of $N_{mJ}^{(r)}$ and of its first two derivatives as well as Assumption R, one concludes from the dominated convergence theorem that also the objective function $\mathcal{Q}_{n,k}$ defined in (8) is twice continuously differentiable on the neighborhood $B(\theta_0)$ with derivatives (measurable for every $\theta \in B(\theta_0)$)

$$\nabla_\theta \mathcal{Q}_{n,k}(\theta) = -2 \int_0^1 (p_{n,j_n,r_*} - p_{k,J_k,r}(\theta)) p_{n,j_n,r_*}^{-1} \nabla_\theta p_{k,J_k,r}(\theta) d\lambda,$$

$$\begin{aligned} \nabla_\theta^2 \mathcal{Q}_{n,k}(\theta) &= 2 \int_0^1 p_{n,j_n,r_*}^{-1} \nabla_\theta p_{k,J_k,r}(\theta) \nabla_\theta p_{k,J_k,r}(\theta) d\lambda \\ &\quad -2 \int_0^1 (p_{n,j_n,r_*} - p_{k,J_k,r}(\theta)) p_{n,j_n,r_*}^{-1} \nabla_\theta^2 p_{k,J_k,r}(\theta) d\lambda. \end{aligned} \tag{23}$$

Since $\hat{\theta}_{n,k(n)}$ is an interior maximizer of $\mathcal{Q}_{n,k(n)}$ (on the event considered), we clearly have that $\nabla_\theta \mathcal{Q}_{n,k(n)}(\hat{\theta}_{n,k(n)}) = 0$. Consequently, a standard Taylor expansions gives

$$0 = \nabla_\theta \mathcal{Q}_{n,k(n)}(\hat{\theta}_{n,k(n)}) = \nabla_\theta \mathcal{Q}_{n,k(n)}(\theta_0) + \nabla_\theta^2 \mathcal{Q}_{n,k(n)}^*(\hat{\theta}_{n,k(n)} - \theta_0), \tag{24}$$

where the i-th row of $\nabla_\theta^2 \mathcal{Q}_{n,k(n)}^*$ equals the corresponding row of $\nabla_\theta^2 \mathcal{Q}_{n,k(n)}$ evaluated at a mean-value $\tilde{\theta}_{n,k(n)}^{(i)}$ which may depend on the row-index (measurability of the mean-value being of no concern). We next show that $\sqrt{n}\nabla_\theta \mathcal{Q}_{n,k(n)}(\theta_0)$ is asymptotically normal and that $\nabla_\theta^2 \mathcal{Q}_{n,k(n)}^*$ converges in (outer) probability to the positive definite matrix $\nabla_\theta^2 Q(\theta_0)$. The asymptotic normality of $\sqrt{n}\left(\hat{\theta}_{n,k(n)} - \theta_0\right)$ then follows along the same lines as in the last paragraph of the proof of Theorem 2.

Step 1: CLT for the score $\sqrt{n}\nabla_\theta \mathcal{Q}_{n,k(n)}(\theta_0)$.

We decompose the score as follows:

$$\begin{aligned}
& \nabla_\theta \mathcal{Q}_{n,k(n)}(\theta_0) \\
= \; & -2\int_0^1 (p_{n,j_n,r_*} - p(\theta_0))p(\theta_0)^{-1}\nabla_\theta p(\theta_0) d\lambda \\
& +2\int_0^1 (p_{k(n),J_{k(n)},r}(\theta_0) - p(\theta_0))p(\theta_0)^{-1}\nabla_\theta p(\theta_0) d\lambda \\
& +2\int_0^1 (p_{n,j_n,r_*} - p_{k(n),J_{k(n)},r}(\theta_0)) \left(p(\theta_0)^{-1}\nabla_\theta p(\theta_0) - p_{n,j_n,r_*}^{-1}\nabla_\theta p_{k(n),J_{k(n)},r}(\theta_0)\right) d\lambda \\
= \; & I + II + III,
\end{aligned}$$

with each of the terms being measurable. We further observe that the terms I and II are independent by construction of the simulation mechanism.

About Term I: As shown in the proof of Theorem 2

$$\sqrt{n} I \to^d N(0,\Sigma)$$

where

$$\Sigma = 4\int_0^1 \nabla_\theta p(\theta_0)\nabla_\theta p(\theta_0)' p(\theta_0)^{-1} d\lambda.$$

About Term II: Exactly the same argument as given in the proof of Theorem 2 for term I, except for using Theorem 3 instead of Theorem 4, establishes that

$$\sqrt{k(n)} II \to^d N(0,\Sigma).$$

But then

$$\sqrt{n} II = \sqrt{n/k(n)}\sqrt{k(n)} II \to^d N\left(0, \frac{1}{\kappa}\Sigma\right)$$

under Assumption S3, and $\sqrt{n} II$ converges to zero in probability under Assumption S2.

About Term III: By Cauchy-Schwarz and the triangle inequality we have the bound

$$\begin{aligned}
\|III\| \leq \; & 2\left\|p_{n,j_n,r_*} - p_{k(n),J_{k(n)},r}(\theta_0)\right\|_2 \left[\left\|\left(p(\theta_0)^{-1} - p_{n,j_n,r_*}^{-1}\right)\nabla_\theta p(\theta_0)\right\|_2 \right. \\
& \left. + \left\|p_{n,j_n,r_*}^{-1}\left(\nabla_\theta p_{k(n),J_{k(n)},r}(\theta_0) - \nabla_\theta p(\theta_0)\right)\right\|_2\right] \\
\leq \; & 2\left\|p_{n,j_n,r_*} - p_{k(n),J_{k(n)},r}(\theta_0)\right\|_2 \left[(2/\xi_0^2)\left\|(p_{n,j_n,r_*} - p(\theta_0))\nabla_\theta p(\theta_0)\right\|_2 \right. \\
& \left. + (2/\xi_0)\left\|\nabla_\theta p_{k(n),J_{k(n)},r}(\theta_0) - \nabla_\theta p(\theta_0)\right\|_2\right] \\
\leq \; & (4/\xi_0)\left[\|p_{n,j_n,r_*} - p(\theta_0)\|_2 + \left\|p(\theta_0) - p_{k(n),J_{k(n)},r}(\theta_0)\right\|_2\right] \times \\
& \left[(1/\xi_0)\|p_{n,j_n,r_*} - p(\theta_0)\|_2 \|\nabla_\theta p(\theta_0)\|_\infty + \left\|\nabla_\theta p_{k(n),J_{k(n)},r}(\theta_0) - \nabla_\theta p(\theta_0)\right\|_2\right]
\end{aligned}$$

with $\|\nabla_\theta p(\theta_0)\|_\infty$ being finite in view of Assumption P1(iv) and Proposition 7 in Appendix A. The r.h.s. of the above display is now

$$O_p\left(\left(\sqrt{\frac{2^{j_n}}{n}}+2^{-j_n\tau}+\sqrt{\frac{2^{J_{k(n)}}}{k(n)}}+2^{-J_{k(n)}\tau}\right)\left(\sqrt{\frac{2^{j_n}}{n}}+2^{-j_n\tau}+\sqrt{\frac{2^{3J_{k(n)}}}{k(n)}}+2^{-J_{k(n)}s}\right)\right)$$

for every $0 < s < r$, $s \leq \varsigma$ in view of Assumptions P1 and R as well as Lemmata 3 and 4. Fixing such an $s > 1/2$, the expression in the above display is seen to be $o_p(n^{-1/2})$ under the assumptions of Part a2 or Part b (in particular, $\tau > 3/2$), showing that $\sqrt{n}III$ is asymptotically negligible.

This completes Step 1 and shows that

$$\sqrt{n}\nabla_\theta \mathcal{Q}_{n,k(n)}(\theta_0) \to^d N\left(0,(1+\kappa^{-1})\Sigma\right)$$

under the assumptions of Part b, whereas under the assumptions of Part a2

$$\sqrt{n}\nabla_\theta \mathcal{Q}_{n,k(n)}(\theta_0) \to^d N\left(0,\Sigma\right).$$

Step 2: Convergence of second order derivatives.
We have

$$\left\|\nabla^2_\theta \mathcal{Q}^*_{n,k(n)} - \nabla^2_\theta Q(\theta_0)\right\| \leq \left\|\nabla^2_\theta \mathcal{Q}^*_{n,k(n)} - \nabla^2_\theta Q^\dagger_n\right\| + \left\|\nabla^2_\theta Q^\dagger_n - \nabla^2_\theta Q(\theta_0)\right\|$$

where $\nabla^2_\theta Q^\dagger_n$ is the matrix $\nabla^2_\theta Q_n$ row-wise evaluated at the mean-values $\tilde{\theta}^{(i)}_{n,k(n)}$. In view of (17), consistency of $\hat{\theta}_{n,k(n)}$, and continuity of $\nabla^2_\theta Q$ at θ_0, the second term on the r.h.s. above converges to zero in (outer) probability. We now show the same for the first term on the r.h.s. in the above display: Note that the argument leading to (21) is also valid under the current assumptions, and therefore we can conclude from (21), (22), and Theorem 2 that $\left\|\hat{\theta}_{n,k(n)} - \theta_0\right\| = O_p(k(n)^{-1/4})$. Consequently, it suffices to show that

$$\sup_{\theta\in B(\theta_0),\|\theta-\theta_0\|\leq Mk(n)^{-1/4}} \left\|\nabla^2_\theta \mathcal{Q}_{n,k(n)}(\theta) - \nabla^2_\theta Q_n(\theta)\right\| \to 0$$

in probability for every $0 < M < \infty$, the above supremum being measurable (as the functions involved are continuous). Now, by (23) and (15)

$$\begin{aligned}
&\frac{1}{2}\left(\nabla^2_\theta \mathcal{Q}_{n,k(n)}(\theta) - \nabla^2_\theta Q_n(\theta)\right) = \int_0^1 (p_{n,j_n,r_*} - p(\theta))p^{-1}_{n,j_n,r_*}\left(\nabla^2_\theta p(\theta) - \nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)\right)d\lambda \\
&- \int_0^1 (p(\theta) - p_{k(n),J_{k(n)},r}(\theta))p^{-1}_{n,j_n,r_*}\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)d\lambda \\
&+ \int_0^1 p^{-1}_{n,j_n,r_*}\left(\nabla_\theta p_{k(n),J_{k(n)},r}(\theta)\nabla_\theta p_{k(n),J_{k(n)},r}(\theta)' - \nabla_\theta p(\theta)\nabla_\theta p(\theta)'\right)d\lambda = I - II + III.
\end{aligned}$$

About Term I: By the Cauchy-Schwarz and the triangle inequalities

$$\begin{aligned}
\|I\| \quad \leq \quad & 2\xi_0^{-1}\left[\|p_{n,j_n,r_*} - p(\theta_0)\|_2 + \|p(\theta_0) - p(\theta)\|_2\right] \times \\
& \left[\left\|\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta) - E\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)\right\|_2 + \left\|\nabla^2_\theta p(\theta) - E\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)\right\|_2\right].
\end{aligned}$$

The first term on the r.h.s. of the above display is $O_p(n^{-\tau/(2\tau+1)})$ in view of Lemma 3 and the choice of j_n. For the second term, observe that in view of Assumption P1(iii) we have $p(\theta,x)-p(\theta_0,x)=\nabla_\theta p(\breve{\theta}(x),x)'(\theta-\theta_0)$ by the pathwise mean value theorem, and hence

$$\|p(\theta_0)-p(\theta)\|_2 \leq \left(\int_0^1 \sup_{\theta\in B(\theta_0)} \|\nabla_\theta p(\theta,x)\|^2\, dx\right)^{1/2} \|\theta-\theta_0\| = O(\|\theta-\theta_0\|)$$

holds for all $\theta\in B(\theta_0)$. In view of Lemma 5 and the choice of $J_{k(n)}$, the supremum over $B(\theta_0)$ of the third term is $O_p(k(n)^{(2-\tau)/(2\tau+1)}\sqrt{\log k(n)})$. Furthermore, note that

$$E\frac{\partial^2 p_{k(n),J_{k(n)},r}(\theta)}{\partial\theta_i\partial\theta_{i'}} = \pi^{(r)}_{J_{k(n)}}\left(\frac{\partial^2 p(\theta)}{\partial\theta_i\partial\theta_{i'}}\right) \tag{25}$$

holds for $\theta\in B(\theta_0)$. [This is proved analogously as (38) in Section 6, making use of the dominance assumptions on $\nabla^2_\theta p$ in Assumption P1, the uniform boundedness assumption on the derivatives of ρ in assumption R(ii), the boundedness of the B-spline basis functions and their first two derivatives (as $r\geq 4$ holds), as well as using that $\frac{\partial^2 p(\theta)}{\partial\theta_i\partial\theta_{i'}}\in\mathcal{L}^2$ in view of Assumption P2(ii).] The above established relation, together with the fact that the spectral matrix norm is bounded by the Frobenius norm, implies that the supremum over $B(\theta_0)$ of the fourth term is bounded by

$$\sup_{\theta\in B(\theta_0)}\sum_{i,i'=1}^{b}\left\|\frac{\partial^2 p(\theta)}{\partial\theta_i\partial\theta_{i'}} - \pi^{(r)}_{J_{k(n)}}\left(\frac{\partial^2 p(\theta)}{\partial\theta_i\partial\theta_{i'}}\right)\right\|_2 \leq \sup_{\theta\in B(\theta_0)}\sum_{i,i'=1}^{b}\left\|\frac{\partial^2 p(\theta)}{\partial\theta_i\partial\theta_{i'}}\right\|_2 < \infty$$

the last inequality following from Assumption P2(ii). Consequently, in view of (22),

$$\begin{aligned} &\sup_{\theta\in B(\theta_0),\|\theta-\theta_0\|\leq Mk(n)^{-1/4}} \|I\| \\ \leq\ & \left[O_p(n^{-\tau/(2\tau+1)})+O(k(n)^{-1/4})\right]\left[O_p(k(n)^{(2-\tau)/(2\tau+1)}\sqrt{\log k(n)})+const\right] = o_p(1)\end{aligned}$$

under either the assumptions of Part a2 or Part b (since $\tau > 3/2 > 4/3$).

About Term II: By the Cauchy-Schwarz and the triangle inequalities

$$\begin{aligned} &\sup_{\theta\in B(\theta_0)}\|II\| \leq 2\xi_0^{-1}\sup_{\theta\in B(\theta_0)}\left\|p(\theta)-p_{k(n),J_{k(n)},r}(\theta)\right\|_2 \times \\ &\sup_{\theta\in B(\theta_0)}\left[\left\|\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)-E\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)\right\|_2+\left\|E\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)\right\|_2\right] \\ =\ & O_p(k(n)^{-\tau/(2\tau+1)}\sqrt{\log k(n)})\left[O_p(k(n)^{(2-\tau)/(2\tau+1)}\sqrt{\log k(n)})+const\right] \end{aligned}\tag{26}$$

where we have made use of Lemmata 3 and 5; and we have used the bound

$$\sup_{\theta\in B(\theta_0)}\left\|E\nabla^2_\theta p_{k(n),J_{k(n)},r}(\theta)\right\|_2 \leq \sup_{\theta\in B(\theta_0)}\sum_{i,i'=1}^{b}\left\|\frac{\partial^2 p(\theta)}{\partial\theta_i\partial\theta_{i'}}\right\|_2 < \infty$$

which follows from (25) and Assumption P2(ii). The r.h.s. of (26) is now $o_p(1)$ since $\tau > 3/2 > 1$.

About Term III: By the Cauchy-Schwarz and the triangle inequalities

$$\|III\| \leq 2\xi_0^{-1}\left\|\nabla_\theta p_{k(n),J_{k(n)},r}(\theta)-\nabla_\theta p(\theta)\right\|_2\left[\left\|\nabla_\theta p_{k(n),J_{k(n)},r}(\theta)-\nabla_\theta p(\theta)\right\|_2+2\left\|\nabla_\theta p(\theta)\right\|_2\right].$$

Now

$$\sup_{\theta\in B(\theta_0)} \left\| \nabla_\theta p_{k(n),J_{k(n)},r}(\theta) - E\nabla_\theta p_{k(n),J_{k(n)},r}(\theta) \right\|_2 = O_p(k(n)^{(1-\tau)/(2\tau+1)}\sqrt{\log k(n)}) = o_p(1)$$

by Lemma 5 and since $\tau > 3/2 > 1$. Furthermore,

$$\begin{aligned} \sup_{\theta\in B(\theta_0)} \left\| E\nabla_\theta p_{k(n),J_{k(n)},r}(\theta) - \nabla_\theta p(\theta) \right\|_2 &\leq \sum_{i=1}^{b} \sup_{\theta\in B(\theta_0)} \left\| E\frac{\partial p_{k(n),J_{k(n)},r}(\theta)}{\partial\theta_i} - \frac{\partial p(\theta)}{\partial\theta_i} \right\|_2 \\ &= \sum_{i=1}^{b} \sup_{\theta\in B(\theta_0)} \left\| \pi^{(r)}_{J_{k(n)}} \left(\frac{\partial p(\theta)}{\partial\theta_i} \right) - \frac{\partial p(\theta)}{\partial\theta_i} \right\|_2, \end{aligned}$$

the last equality holding as shown in (38) in Section 6. By Proposition 8 in Appendix A and Assumption P2(i) the r.h.s. in the above display is now $o(1)$. Taken together, this provides a bound for $\sup_{\theta\in B(\theta_0),\|\theta-\theta_0\|\leq Mk(n)^{-1/4}} \|III\|$ which converges to zero in probability. This completes the proof of Step 2.

5.5 Proof of Proposition 3

Since $\bar\theta_{n,k} \to \theta_0$ by assumption, since $\Phi(\theta) := \int_0^1 \nabla_\theta p(\theta)\nabla_\theta p(\theta)' p(\theta_0)^{-1} d\lambda$ is continuous on the neighborhood $B(\theta_0)$ of θ_0 by dominated convergence and Assumption P1(iii), and since $\Phi(\theta_0)$ is positive definite by the same assumption, it suffices to show that, uniformly over $B(\theta_0)$, the expression $\hat\Phi(\theta) = \int_0^1 \nabla_\theta p_{k,J'_k,r'}(\theta)\nabla_\theta p_{k,J'_k,r'}(\theta)' p^{-1}_{n,j'_n,r'_*} d\lambda$ converges to $\Phi(\theta)$ in probability as $n\wedge k\to\infty$. Note that $\hat\Phi(\theta)$ is well-defined on the event A_n^* which has probability converging to 1 in view of Corollary 2. In the sequel we only work on that event. Now

$$\begin{aligned} \left|\hat\Phi(\theta) - \Phi(\theta)\right| &\leq \left| \int_0^1 \nabla_\theta p(\theta)\nabla_\theta p(\theta)' \left(p^{-1}_{n,j'_n,r'_*} - p(\theta_0)^{-1} \right) d\lambda \right| \\ &\quad + \left| \int_0^1 \left(\nabla_\theta p_{k,J'_k,r'}(\theta)\nabla_\theta p_{k,J'_k,r'}(\theta)' - \nabla_\theta p(\theta)\nabla_\theta p(\theta)' \right) p^{-1}_{n,j'_n,r'_*} d\lambda \right| \\ &\leq 2\xi_0^{-2} \left\| p_{n,j_n,r_*} - p(\theta_0) \right\|_\infty \int_0^1 \sup_{\theta\in B(\theta_0)} \left\| \nabla_\theta p(\theta) \right\|^2 d\lambda \\ &\quad + 2\xi_0^{-1} \left\| \nabla_\theta p_{k,J_k,r'}(\theta) - \nabla_\theta p(\theta) \right\|_2 \left[\left\| \nabla_\theta p_{k,J_k,r'}(\theta) - \nabla_\theta p(\theta) \right\|_2 + 2\left\| \nabla_\theta p(\theta) \right\|_2 \right]. \end{aligned}$$

The first term on the r.h.s. is independent of θ and converges to zero in probability by Corollary 2. The supremum over $B(\theta_0)$ of the second term converges to zero by essentially repeating the argument that has been used in the very last step of the proof of Theorem 1.

6 Rates of Convergence for Spline Projection Estimators

This section contains the main stochastic bounds used to control remainder terms in the proofs in Section 5. We first collect some simple facts about the B-splines $N^{(r)}$ that will repeatedly be used in this section:

$$\left\| N^{(r)} \right\|_{\infty,\mathbb{R}} \leq 1, \quad \left\| N^{(r)} \right\|_{1,\mathbb{R}} = 1, \quad \left\| N^{(r)} \right\|_{2,\mathbb{R}} \leq 1 \quad \text{for } r\geq 1. \tag{27}$$

The first relation is a direct consequence of the definition of $N^{(r)}$, the second one follows since $N(r)$ is – as a convolution of probability densities – a probability density again, and the third relation is a consequence of Young's inequality. Furthermore, it is easy to see that $N^{(r)}$ is continuously differentiable for $r \geq 3$ with derivative $N^{(r)\prime}$ given by

$$N^{(r)\prime} = N^{(r-1)} - N^{(r-1)}(\cdot - 1). \tag{28}$$

For $r = 2$, the B-spline $N^{(2)}$ is Lipschitz and only has a weak derivative $N^{(2)\prime}$ which, in order to have it defined everywhere, will always be taken as $N^{(1)} - N^{(1)}(\cdot - 1)$. The bounds

$$\left\| N^{(r)\prime} \right\|_{\infty,\mathbb{R}} \leq 1, \quad \left\| N^{(r)\prime} \right\|_{1,\mathbb{R}} \leq 2, \quad \left\| N^{(r)\prime} \right\|_{2,\mathbb{R}} \leq 2 \quad \text{for } r \geq 2 \tag{29}$$

are then an immediate consequence of (27), (28), and the fact that $N^{(r-1)}$ is nonnegative. By repeated application of (28) we can obtain bounds for higher-order derivatives, for example, we shall need

$$\left\| N^{(r)\prime\prime} \right\|_{\infty,\mathbb{R}} \leq 2, \quad \left\| N^{(r)\prime\prime} \right\|_{2,\mathbb{R}} \leq 4 \quad \text{for } r \geq 3, \quad \text{and} \quad \left\| N^{(r)\prime\prime\prime} \right\|_{\infty,\mathbb{R}} \leq 4 \quad \text{for } r \geq 4. \tag{30}$$

The above discussion also implies that $N^{(r)}$ for $r \geq 2$, $N^{(r)\prime}$ for $r \geq 3$, and $N^{(r)\prime\prime}$ for $r \geq 4$ are globally Lipschitz on $\mathbb{R}$ with Lipschitz constants bounded by 1, 2, and 4, respectively.

For $f \in \mathcal{S}_j(r)$, $r \geq 3$, we denote in the following by f' its derivative (using one-sided derivatives on the boundary of $[0,1]$); for $r = 2$ we use f' to denote the weak derivative.

Lemma 2 *Let $f = \sum_{l=-r+1}^{2^j-1} \alpha_l N_{lj}^{(r)}$ where α_l are real numbers and $r \geq 1$, i.e., $f \in \mathcal{S}_j(r)$. Then*

$$\|f\|_2 \leq 2^{-j/2} \left(\sum_{l=-r+1}^{2^j-1} \alpha_l^2 \right)^{1/2}, \tag{31}$$

$$\|f'\|_2 \leq 2^{1+j/2} \left(\sum_{l=-r+1}^{2^j-1} \alpha_l^2 \right)^{1/2} \qquad \textit{for } r \geq 2, \tag{32}$$

and

$$\|f''\|_2 \leq 2^{2+3j/2} \left(\sum_{l=-r+1}^{2^j-1} \alpha_l^2 \right)^{1/2} \qquad \textit{for } r \geq 3. \tag{33}$$

Furthermore, for every $0 < s' \leq 1$ there exists a finite constant $C_0(s')$ such that for every $r \geq 2$ and f as above

$$\|f\|_{s',2} \leq C_0(s') 2^{j(s'-1/2)} \left(\sum_{l=-r+1}^{2^j-1} \alpha_l^2 \right)^{1/2}. \tag{34}$$

Proof. The first claim is well-known, see, e.g., DeVore and Lorentz (1993), Theorem 5.4.2. To prove (32), use (28) and the fact that $N^{(r-1)}$ vanishes outside of $(0, r-1)$ for $r \geq 3$ and outside of $[0,1)$ for $r = 2$, to obtain (interpreting the equality modulo λ-nullsets in case $r = 2$)

$$\begin{aligned} f'(x) &= 2^j \sum_{l=-r+1}^{2^j-1} \alpha_l N^{(r)\prime}(2^j x - l) = 2^j \sum_{l=-r+1}^{2^j-1} \alpha_l \left[N^{(r-1)}(2^j x - l) - N^{(r-1)}(2^j x - l - 1) \right] \\ &= 2^j \sum_{l=-(r-1)+1}^{2^j-1} \alpha_l N^{(r-1)}(2^j x - l) - 2^j \sum_{l=-(r-1)+1}^{2^j-1} \alpha_{l-1} N^{(r-1)}(2^j x - l) =: f_1 + f_2. \end{aligned}$$

Using (31) for f_1 and f_2, we obtain

$$\|f'\|_2 \leq \|f_1\|_2 + \|f_2\|_2 \leq 2^{1+j/2} \left(\sum_{l=-r+1}^{2^j-1} \alpha_l^2 \right)^{1/2}.$$

The third claim is proved similarly. To prove the final claim, we use the following interpolation inequality: for every $0 < s' \leq 1$ there exists a finite constant $C^*(s')$ such that for every $h \in \mathcal{W}_2^1$

$$\|h\|_{s',2} \leq C^*(s')(\|h\|_2 + \|D_w h\|_2)^{s'} \|h\|_2^{1-s'} \tag{35}$$

holds. [This follows from (5) if $s' = 1$; if $s' < 1$ it follows from Theorem 6.7.1 in DeVore and Lorentz (1993) applied to the intermediate spaces $(\mathbb{R}, \mathbb{R})_{s',\infty}$, $(\mathcal{L}^2, \mathcal{W}_2^1)_{s',\infty}$, and to the operator that maps any real number a into ah, observing that $(\mathcal{L}^2, \mathcal{W}_2^1)_{s',\infty}$ is equal to $\mathcal{B}_{s'}$ up to a equivalence of norms, cf. p.196 in DeVore and Lorentz (1993).] Observe that $f \in \mathcal{W}_2^1$ if $r \geq 2$. Now, using (35) with $h = f$, (31), and (32) completes the proof upon setting $C_0(s') = (2.5)^{s'} C^*(s')$. ■

Lemma 3 *Assume $r \geq 1$ and let $\theta \in \Theta$.*

a. Suppose the density $p(\theta)$ is bounded. Then for all $k \geq 1$ and $J \geq 1$

$$E \|p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)\|_2^2 \leq C_1(\theta, r) \frac{2^J}{k},$$

where $C_1(\theta, r) = (\frac{r+1}{2}) d_r^2 \|p(\theta)\|_\infty$ with d_r defined in Proposition 4. Furthermore, for $r \geq 2$ and $0 < s' \leq 1$

$$E \|p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)\|_{s',2}^2 \leq C_0(s')^2 C_1(\theta, r) \frac{2^{J(2s'+1)}}{k}$$

holds for all $k \geq 1$ and $J \geq 1$, where $C_0(s')$ is given in Lemma 2.

b. If $p(\theta) \in \mathcal{L}^2$, then for every k

$$\lim_{J \to \infty} \|Ep_{k,J,r}(\theta) - p(\theta)\|_2 = 0.$$

If $p(\theta) \in \mathcal{B}_t$ for some $0 < t < r$ then for all $k \geq 1$ and $J \geq 1$

$$\|Ep_{k,J,r}(\theta) - p(\theta)\|_2 \leq 2^{-Jt} c_t' \|p(\theta)\|_{t,2},$$

where c_t' is the constant given in Proposition 8 in Appendix A.

c. If the assumptions of Part a (Part b) hold for (a version of) p_0 and r_* in place of $p(\theta)$ and r, respectively, then the results in Part a (Part b) also apply mutatis mutandis to p_{n,j,r_*}.

Proof. In view of Lemma 2, the definition of $p_{k,J,r}(\theta)$, (31) and (34), it suffices to bound

$E\left(\hat{\gamma}_{lJ}^{(r)}(\theta) - E\hat{\gamma}_{lJ}^{(r)}(\theta)\right)^2$ in order to prove Part a. We obtain

$$\begin{aligned}
& E\left(\hat{\gamma}_{lJ}^{(r)}(\theta) - E\hat{\gamma}_{lJ}^{(r)}(\theta)\right)^2 \\
\leq\ & \frac{2^{2J}}{k} E\left(\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} N_{mJ}^{(r)}(\rho(V_i,\theta))\right)^2 = \frac{2^{2J}}{k}\int_0^1 \left(\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} N_{mJ}^{(r)}(x)\right)^2 p(\theta,x)dx \\
\leq\ & \frac{2^{2J}}{k}\|p(\theta)\|_\infty \left\|\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} N_{mJ}^{(r)}\right\|_2^2 \leq \frac{2^J}{k}\|p(\theta)\|_\infty \sum_{m=-r+1}^{2^J-1}\left(g_J^{(r)lm}\right)^2 \\
\leq\ & \frac{2^J}{k}\|p(\theta)\|_\infty \left(\sum_{m=-r+1}^{2^J-1}\left|g_J^{(r)lm}\right|\right)^2 \leq \frac{2^J}{k} d_r^2 \|p(\theta)\|_\infty, \qquad (36)
\end{aligned}$$

where we have used independence, (31), and Proposition 4. This establishes Part a. [Measurability of the $\mathcal{L}^2$-norm is obvious, and measurability of the Besov-norm follows from Appendix B.] Since $Ep_{k,J,r}(\theta) = \pi_J^{(r)}(p(\theta))$, Part b follows from Proposition 8 in Appendix A. Part c is proved completely analogously. ■

Lemma 4 *Assume $r \geq 3$ and let θ be an interior point of Θ such that the partial derivative $\frac{\partial\rho(v,\theta)}{\partial\theta_q}$ at θ exists for every $v \in \mathcal{V}$.*

a. Suppose the density $p(\theta)$ is bounded and $\sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\right| < \infty$. Then for all $k \geq 1$ and $J \geq 1$

$$E\left\|\frac{\partial p_{k,J,r}(\theta)}{\partial\theta_q} - E\frac{\partial p_{k,J,r}(\theta)}{\partial\theta_q}\right\|_2^2 \leq C_2(\theta,r)\frac{2^{3J}}{k},$$

where $C_2(\theta,r) = 2(r+1)d_r^2\|p(\theta)\|_\infty \sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\right|^2$.

b. Suppose there exists an open ball $B(\theta) \subseteq \Theta$ with center θ such that $\frac{\partial p(\cdot,x)}{\partial\theta_q}$ and $\frac{\partial\rho(v,\cdot)}{\partial\theta_q}$ exist on $B(\theta)$ for every $x \in [0,1]$ and $v \in \mathcal{V}$, suppose $\frac{\partial p(\theta,\cdot)}{\partial\theta_q}$ belongs to $\mathcal{B}_s$ for some $0 < s < r$, and that

$$\int_0^1 \sup_{\theta'\in B(\theta)}\left|\frac{\partial p(\theta',x)}{\partial\theta_q}\right| dx < \infty, \qquad \int_{\mathcal{V}} \sup_{\theta'\in B(\theta)}\left|\frac{\partial\rho(v,\theta')}{\partial\theta_q}\right| d\mu(v) < \infty.$$

Then for all $k \geq 1$ and $J \geq 1$

$$\left\|E\frac{\partial p_{k,J,r}(\theta)}{\partial\theta_q} - \frac{\partial p(\theta)}{\partial\theta_q}\right\|_2 \leq 2^{-Js}c_s'\left\|\frac{\partial p(\theta)}{\partial\theta_q}\right\|_{s,2},$$

where the constant c_s' is defined in Proposition 8 in Appendix A. [If $\frac{\partial p(\theta,\cdot)}{\partial\theta_q} \in \mathcal{B}_s$ is weakened to $\frac{\partial p(\theta,\cdot)}{\partial\theta_q} \in \mathcal{L}^2$, then $\lim_{J\to\infty}\left\|E\frac{\partial p_{k,J,r}(\theta)}{\partial\theta_q} - \frac{\partial p(\theta)}{\partial\theta_q}\right\| = 0$ holds.]

Proof. Observe that $p_{k,J,r}$ is differentiable at θ because $r \geq 3$ is assumed. To prove Part a note that

$$\frac{\partial p_{k,J,r}(\theta)}{\partial\theta_q} - E\frac{\partial p_{k,J,r}(\theta)}{\partial\theta_q} = \sum_{l=-r+1}^{2^J-1}\left(\frac{\partial\hat{\gamma}_{lJ}^{(r)}(\theta)}{\partial\theta_q} - E\frac{\partial\hat{\gamma}_{lJ}^{(r)}(\theta)}{\partial\theta_q}\right)N_{lJ}^{(r)},$$

and that the $\mathcal{L}^2$-norm of this expression is measurable by Fubini's Theorem; also note that the expectations in the above display exist since the B-spline basis functions are bounded and since $\sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\right| < \infty$ has been assumed. Now, using the chain rule and (32), we obtain

$$
\begin{aligned}
E\left(\frac{\partial\hat{\gamma}_{lJ}^{(r)}(\theta)}{\partial\theta_q} - E\frac{\partial\hat{\gamma}_{lJ}^{(r)}(\theta)}{\partial\theta_q}\right)^2 &\le \frac{2^{2J}}{k}E\left(\frac{\partial\rho(V_i,\theta)}{\partial\theta_q}\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} N_{mJ}^{(r)\prime}(x)_{|x=\rho(V_i,\theta)}\right)^2 \\
\le \frac{2^{2J}}{k}\sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\right|^2 \int_0^1\left(\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} N_{mJ}^{(r)\prime}(x)\right)^2 p(\theta,x)dx \qquad (37)\\
\le \frac{2^{2J}}{k}\sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\right|^2 \|p(\theta)\|_\infty\left\|\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} N_{mJ}^{(r)\prime}\right\|_2^2 &\le \frac{2^{3J+2}}{k}d_r^2\sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\right|^2\|p(\theta)\|_\infty .
\end{aligned}
$$

An application of Lemma 2 then completes the proof of Part a.

To prove Part b, note that

$$
\begin{aligned}
\int_{\mathcal{V}}\frac{\partial}{\partial\theta_q}N_{mJ}^{(r)}(\rho(v,\theta))d\mu(v) &= \frac{\partial}{\partial\theta_q}\int_{\mathcal{V}}N_{mJ}^{(r)}(\rho(v,\theta))d\mu(v) \\
&= \frac{\partial}{\partial\theta_q}\int_0^1 N_{mJ}^{(r)}(x)p(\theta,x)dx = \int_0^1 N_{mJ}^{(r)}(x)\frac{\partial}{\partial\theta_q}p(\theta,x)dx,
\end{aligned}
$$

where the two-fold interchange of integration and differentiation is permitted by dominated convergence in view of the maintained dominance assumptions on the derivatives of ρ and p as well as the boundedness of the B-spline basis functions and their first derivative. Consequently,

$$
\begin{aligned}
E\frac{\partial p_{k,J,r}(\theta,y)}{\partial\theta_q} &= 2^J\sum_{l=-r+1}^{2^J-1}\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm}\int_{\mathcal{V}}\frac{\partial}{\partial\theta_q}N_{mJ}^{(r)}(\rho(v,\theta))d\mu(v)N_{lJ}^{(r)}(y) \qquad (38)\\
&= 2^J\sum_{l=-r+1}^{2^J-1}\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm}\int_0^1 N_{mJ}^{(r)}(x)\frac{\partial}{\partial\theta_q}p(\theta,x)dxN_{lJ}^{(r)}(y) = \pi_J^{(r)}\left(\frac{\partial}{\partial\theta_q}p(\theta)\right),
\end{aligned}
$$

and Part b now follows immediately from Proposition 8 in Appendix A. ■

Lemma 5 *a. Suppose Assumption R(i) is satisfied, $r \ge 2$, Θ is a bounded subset of $\mathbb{R}^b$, and $\sup_{\theta\in\Theta}\|p(\theta)\|_\infty < \infty$. Then there exist finite positive constants C_3 and C_4, depending only on Θ, b, ρ, r, and $\sup_{\theta\in\Theta}\|p(\theta)\|_\infty$ but not on k and J, such that*

$$E\sup_{\theta\in\Theta}\|p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)\|_2^2 \le C_3\frac{2^J J}{k},$$

holds for all $k \ge 1$ and $J \ge 1$ satisfying $2^J J \le C_4 k$. Furthermore, for $0 < s' \le 1$

$$E\sup_{\theta\in\Theta}\|p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)\|_{s',2}^2 \le C_0(s')^2 C_3\frac{2^{J(2s'+1)}J}{k} \qquad (39)$$

holds for all $k \ge 1$ and $J \ge 1$ satisfying $2^J J \le C_4 k$ where $C_0(s')$ is given in Lemma 2.

b. Suppose Assumption R(ii) is satisfied for some interior point θ_0 of Θ, $\sup_{\theta\in B(\theta_0)}\|p(\theta)\|_\infty < \infty$ and $r \geq 3$ hold. Then there exist finite positive constants C_5 and C_6, depending only on $B(\theta_0)$, b, ρ, r and $\sup_{\theta\in B(\theta_0)}\|p(\theta)\|_\infty$ but not on k and J, such that for every $q = 1, \dots, b$

$$E \sup_{\theta\in B(\theta_0)} \left\| \frac{\partial}{\partial\theta_q} p_{k,J,r}(\theta) - E\frac{\partial}{\partial\theta_q} p_{k,J,r}(\theta) \right\|_2^2 \leq C_5 \frac{2^{3J}J}{k}$$

holds for all $k \geq 1$ and $J \geq 1$ satisfying $2^J J \leq C_6 k$.

c. Suppose the assumptions of Part b are satisfied except that now $r \geq 4$. Then there exist finite positive constants C_7 and C_8, depending only on $B(\theta_0)$, b, ρ, r and $\sup_{\theta\in B(\theta_0)}\|p(\theta)\|_\infty$ but not on k and J, such that for every $q, q' = 1, \dots, b$

$$E \sup_{\theta\in B(\theta_0)} \left\| \frac{\partial^2}{\partial\theta_q\partial\theta_{q'}} p_{k,J,r}(\theta) - E\frac{\partial^2}{\partial\theta_q\partial\theta_{q'}} p_{k,J,r}(\theta) \right\|_2^2 \leq C_7 \frac{2^{5J}J}{k}$$

holds for all $k \geq 1$ and $J \geq 1$ satisfying $2^J J \leq C_8 k$.

Proof. a. By Lemma 2 we have

$$E\sup_{\theta\in\Theta}\|p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)\|_2^2 \leq 2^{-J}\sum_{l=-r+1}^{2^J-1} E\sup_{\theta\in\Theta}\left(\hat\gamma_{lJ}^{(r)}(\theta) - E\hat\gamma_{lJ}^{(r)}(\theta)\right)^2.$$

Note that the suprema in the above display are measurable as the functions over which the suprema are taken depend continuously on θ in view of assumption R(i) and $r \geq 2$. We bound the r.h.s. in the above display by applying the moment inequality given in Proposition 12 in Appendix C: fix an arbitrary l and express the corresponding summand in the above display as

$$E\sup_{\theta\in\Theta}\left(\hat\gamma_{lJ}^{(r)}(\theta) - E\hat\gamma_{lJ}^{(r)}(\theta)\right)^2 = \frac{2^{2J}}{k^2} E\sup_{\theta\in\Theta}\left|\sum_{i=1}^k h_{\theta,l}(V_i)\right|^2 \tag{40}$$

where

$$h_{\theta,l}(v) = \sum_{m=-r+1}^{2^J-1} g_J^{(r)lm}\left[N_{mJ}^{(r)}(\rho(v,\theta)) - EN_{mJ}^{(r)}(\rho(V_i,\theta))\right]$$

and set $\mathcal{H}_{l,J,r} = \{h_{\theta,l} : \theta\in\Theta\}$. Furthermore, set $U = d_r \max\left(2, \sup_{\theta\in\Theta}\|p(\theta)\|_\infty^{1/2}\right)$ and $\sigma^2 = 2^{-J}U^2$. Then $0 < \sigma \leq U$ holds, and using the calculations that have led to (36) we obtain for every $\theta \in \Theta$

$$Eh_{\theta,l}^2(V_i) \leq E\left(\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} N_{mJ}^{(r)}(\rho(v,\theta))\right)^2 \leq 2^{-J}d_r^2\|p(\theta)\|_\infty \leq 2^{-J}d_r^2\sup_{\theta\in\Theta}\|p(\theta)\|_\infty \leq \sigma^2.$$

Furthermore, using (27), we obtain for every $\theta \in \Theta$

$$\sup_{v\in\mathcal{V}}|h_{\theta,l}| \leq 2d_r\left\|N^{(r)}\right\|_{\infty,\mathbb{R}} \leq 2d_r \leq U.$$

We next bound the uniform L^∞-covering numbers of $\mathcal{H}_{l,J,r}$: observe that the elements of $\mathcal{H}_{l,J,r}$ satisfy for θ, $\theta' \in \Theta$

$$\sup_{v\in\mathcal{V}}\left|h_{\theta,l}(v) - h_{\theta',l}(v)\right| \leq 2^{J+1}d_r L\left\|\theta - \theta'\right\|^\alpha, \tag{41}$$

where L, α are the Hölder constants from Assumption R(i) and where we have made use of the fact that $N^{(r)}$ has Lipschitz constant bounded by 1 for $r \geq 2$; cf. the discussion at the beginning of this section. Since Θ is assumed to be bounded in $\mathbb{R}^b$, it can be covered by fewer than M/δ^b open balls with centers $\theta_i \in \Theta$ and radius δ, for $0 < \delta \leq 1$ where M depends only on Θ. By (41), the functions $h_{\theta_i,l}$ in $\mathcal{H}_{l,J,r}$ corresponding to the θ_i's give rise to a covering of $\mathcal{H}_{l,J,r}$ by sup-norm balls of radius $2^{J+1}d_r L\delta^\alpha$. Consequently, the L^∞-covering numbers satisfy

$$N(\mathcal{H}_{l,J,r}, \mathsf{L}^\infty(\mathcal{V}), \varepsilon) \leq M\left(\frac{2^{J+1}d_r L}{\varepsilon}\right)^{b/\alpha} \quad \text{for } 0 < \varepsilon \leq 2^{J+1}d_r L. \tag{42}$$

Replacing M by $M_* = M \max\left(1, (U/(2d_r L))^{b/\alpha}\right)$ in (42), guarantees that (42) then holds for $0 < \varepsilon \leq 2U$, which leads to

$$N(\mathcal{H}_{l,J,r}, \mathsf{L}^\infty(\mathcal{V}), \varepsilon) \leq (AU/\varepsilon)^v \quad \text{for } 0 < \varepsilon \leq 2U, \tag{43}$$

for $v = \max(b/\alpha, 2)$ and $A = \max\left(2^{J+1}M_*^{\alpha/b}d_r LU^{-1}, 2e\right)$, where we have also enforced $v \geq 2$ and $A > e$. Note that, apart from the factor 2^J, A depends only on Θ, b, ρ (via α and L), r (via d_r), and $\sup_{\theta\in\Theta} \|p(\theta)\|_\infty$. Observe that $\mathcal{H}_{l,J,r}$ contains a countable sup–norm dense subset in view of (41) and separability of Θ. Hence the expectation bound in Part a of Proposition 12 in Appendix C applied to this subset and with $b_0 = v^{-1}$ now yields the existence of positive finite constants C_3' and C_4' both depending only on Θ, b, ρ, r, and $\sup_{\theta\in\Theta} \|p(\theta)\|_\infty$, such that for all $J \in \mathbb{N}$ and all $k \geq C_4' 2^J J$

$$E \sup_{\theta\in\Theta} \left|\sum_{i=1}^k h_{\theta,l}(V_i)\right|^2 \leq C_3' k 2^{-J} J. \tag{44}$$

Since this bound does not depend on the summation index l, the proof of the first claim is complete upon setting $C_3 = (r+1)C_3'/2$ and $C_4 = 1/C_4'$. The second claim follows immediately from applying (34) in Lemma 2 to the l.h.s. of (39) and using (40) and (44), the measurability of the supremum in (39) following from Appendix B.

b. Observe that $p_{k,J,r}$ is continuously differentiable on $B(\theta_0)$ because of $r \geq 3$ and Assumption R(ii). Similarly as in Part a we have measurability of the suprema and obtain from Lemma 2

$$\begin{aligned}
& E \sup_{\theta\in B(\theta_0)} \left\|\frac{\partial}{\partial\theta_q} p_{k,J,r}(\theta) - E\frac{\partial}{\partial\theta_q} p_{k,J,r}(\theta)\right\|_2^2 \\
\leq\ & 2^{-J} \sum_{l=-r+1}^{2^J-1} E \sup_{\theta\in B(\theta_0)} \left(\frac{\partial}{\partial\theta_q}\hat{\gamma}_{lJ}^{(r)}(\theta) - E\frac{\partial}{\partial\theta_q}\hat{\gamma}_{lJ}^{(r)}(\theta)\right)^2 \\
=\ & 2^{-J} \sum_{l=-r+1}^{2^J-1} \frac{2^{4J}}{k^2} E \sup_{\theta\in B(\theta_0)} \left|\sum_{i=1}^k h_{\theta,l}^{(1)}(V_i)\right|^2
\end{aligned}$$

where

$$h_{\theta,l}^{(1)}(v) = \frac{\partial\rho(v,\theta)}{\partial\theta_q} \sum_{m=-r+1}^{2^J-1} g_J^{(r)lm} \left[N^{(r)\prime}(2^J\rho(v,\theta) - m) - EN^{(r)\prime}(2^J\rho(V_i,\theta) - m)\right].$$

Set $\mathcal{H}_{l,J,r}^{(1)} = \left\{h_{\theta,l}^{(1)} : \theta \in B(\theta_0)\right\}$ and define

$$U = 2d_r \sup_{\theta\in B(\theta_0)} \sup_{v\in\mathcal{V}} \left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\right| \max\left(1, \sup_{\theta\in B(\theta_0)} \|p(\theta)\|_\infty^{1/2}\right)$$

and $\sigma^2 = 2^{-J}U^2$. Then $0 < \sigma \leq U$ holds (where we exclude the trivial case $U = 0$). Observing that $N_{mJ}^{(r)\prime}(x) = 2^J N^{(r)\prime}(2^J x - m)$ by the chain rule, we obtain, using the same calculations that have led to (37), for $\theta \in B(\theta_0)$

$$Eh_{\theta,l}^{(1)2}(V_i) \leq 2^{-J+2} d_r^2 \sup_{\theta \in B(\theta_0)} \sup_{v \in \mathcal{V}} \left| \frac{\partial \rho(v,\theta)}{\partial \theta_q} \right|^2 \sup_{\theta \in B(\theta_0)} \|p(\theta)\|_\infty \leq \sigma^2.$$

Furthermore, for every $\theta \in B(\theta_0)$

$$\sup_{v \in \mathcal{V}} \left| h_{\theta,l}^{(1)} \right| \leq 2 \sup_{v \in \mathcal{V}} \left| \frac{\partial \rho(v,\theta)}{\partial \theta_q} \right| d_r \left\| N^{(r)\prime} \right\|_{\infty,\mathbb{R}} \leq 2 d_r \sup_{\theta \in B(\theta_0)} \sup_{v \in \mathcal{V}} \left| \frac{\partial \rho(v,\theta)}{\partial \theta_q} \right| \leq U,$$

where we have made use of (29). To bound the uniform L^∞-covering numbers of $\mathcal{H}_{l,J,r}^{(1)}$, observe that the elements of $\mathcal{H}_{l,J,r}^{(1)}$ satisfy for θ, $\theta' \in B(\theta_0)$

$$\begin{aligned}
& \sup_{v \in \mathcal{V}} \left| h_{\theta,l}^{(1)}(v) - h_{\theta',l}^{(1)}(v) \right| \leq \\
& 2 d_r \left\| N^{(r)\prime} \right\|_{\infty,\mathbb{R}} \sup_{\theta \in B(\theta_0)} \sup_{v \in \mathcal{V}} \left\| \nabla_\theta^2 \rho(v,\theta) \right\| \left\| \theta - \theta' \right\| + 2^{J+1} d_r \sup_{\theta \in B(\theta_0)} \sup_{v \in \mathcal{V}} \left\| \nabla_\theta \rho(v,\theta) \right\|^2 \left\| \theta - \theta' \right\| \\
\leq \quad & 2^{J+1} d_r \left\{ \sup_{\theta \in B(\theta_0)} \sup_{v \in \mathcal{V}} \left\| \nabla_\theta^2 \rho(v,\theta) \right\| + \sup_{\theta \in B(\theta_0)} \sup_{v \in \mathcal{V}} \left\| \nabla_\theta \rho(v,\theta) \right\|^2 \right\} \left\| \theta - \theta' \right\| \leq 2^J c_* \left\| \theta - \theta' \right\|,
\end{aligned}$$

where we have made use of (29), of the bound on the Lipschitz constant of $N^{(r)\prime}$ given at the beginning of this section, and of the boundedness of $B(\theta_0)$; the constant c_* is finite and depends only on ρ, r, and $B(\theta_0)$. Proceeding as in the proof of Part a we obtain

$$N(\mathcal{H}_{l,J,r}^{(1)}, \mathsf{L}^\infty(\mathcal{V}), \varepsilon) \leq (AU/\varepsilon)^v \quad \text{for } 0 < \varepsilon \leq 2U,$$

for $v = \max(b, 2)$ and $A = \max\left(2^J M^{1/b} \max(c_* U^{-1}, 1), 2e\right)$ with M only depending on $B(\theta_0)$. Note that, apart from the factor 2^J, A depends only on $B(\theta_0)$, b, ρ, r and $\sup_{\theta \in B(\theta_0)} \|p(\theta)\|_\infty$. Part a of Proposition 12 in Appendix C applied to a countable sup-norm dense subset of $\mathcal{H}_{l,J,r}^{(1)}$ and with $b_0 = v^{-1}$ now yields the existence of positive finite constants C_5' and C_6' depending only on $B(\theta_0)$, b, ρ, r and $\sup_{\theta \in B(\theta_0)} \|p(\theta)\|_\infty$, such that for all $J \in \mathbb{N}$ and all $k \geq C_6' 2^J J$

$$E \sup_{\theta \in \Theta} \left| \sum_{i=1}^k h_{\theta,l}^{(1)}(V_i) \right|^2 \leq C_5' k 2^{-J} J$$

holds. Since this bound does not depend on l, the proof is complete upon setting $C_5 = (r+1)C_5'/2$ and $C_6 = 1/C_6'$.

c. The proof is similar to the proof of Part b: Observe that $p_{k,J,r}$ is twice continuously differentiable on $B(\theta_0)$ because of $r \geq 4$ and Assumption R(ii). By Lemma 2 we have

$$E \sup_{\theta \in B(\theta_0)} \left\| \frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} p_{k,J,r}(\theta) - E \frac{\partial^2}{\partial \theta_q \partial \theta_{q'}} p_{k,J,r}(\theta) \right\|_2^2 \leq \frac{2^{5J}}{k^2} \sum_{l=-r+1}^{2^J-1} E \sup_{\theta \in B(\theta_0)} \left| \sum_{i=1}^k h_{\theta,l}^{(2)}(V_i) \right|^2$$

where

$$h^{(2)}_{\theta,l}(v) = 2^{-J}\frac{\partial^2\rho(v,\theta)}{\partial\theta_q\partial\theta_{q'}}\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm}\left[N^{(r)\prime}(2^J\rho(v,\theta)-m)-EN^{(r)\prime}(2^J\rho(V_i,\theta)-m)\right]+$$
$$\frac{\partial\rho(v,\theta)}{\partial\theta_q}\frac{\partial\rho(v,\theta)}{\partial\theta_{q'}}\sum_{m=-r+1}^{2^J-1} g_J^{(r)lm}\left[N^{(r)\prime\prime}(2^J\rho(v,\theta)-m)-EN^{(r)\prime\prime}(2^J\rho(V_i,\theta)-m)\right].$$

Set $\mathcal{H}^{(2)}_{l,J,r}=\left\{h^{(2)}_{\theta,l}:\theta\in B(\theta_0)\right\}$, set

$$U = d_r\max\Bigg\{\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla^2_\theta\rho(v,\theta)\right\|+4\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla_\theta\rho(v,\theta)\nabla_\theta\rho(v,\theta)'\right\|,$$
$$\sup_{\theta\in B(\theta_0)}\|p(\theta)\|_\infty^{1/2}\left[2\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla^2_\theta\rho(v,\theta)\right\|^2+32\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla_\theta\rho(v,\theta)\nabla_\theta\rho(v,\theta)'\right\|^2\right]^{1/2}\Bigg\}$$

and $\sigma^2=2^{-J}U^2$. Then $0<\sigma\le U$ holds (where we exclude the trivial case $U=0$), and for $\theta\in B(\theta_0)$ we have

$$Eh^{(2)2}_{\theta,l}(V_i) \le 2^{3-3J}d_r^2\sup_{\theta\in B(\theta_0)}\|p(\theta)\|_\infty\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left|\frac{\partial^2\rho(v,\theta)}{\partial\theta_q\partial\theta_{q'}}\right|^2$$
$$+2^{5-J}d_r^2\sup_{\theta\in B(\theta_0)}\|p(\theta)\|_\infty\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\frac{\partial\rho(v,\theta)}{\partial\theta_{q'}}\right|^2\le\sigma^2,$$

using a calculation similar to the one that has led to (37) and making use of Lemma 2. Similarly, for $\theta\in B(\theta_0)$ we obtain

$$\sup_{v\in\mathcal{V}}\left|h^{(2)}_{\theta,l}(v)\right|\le 2d_r\left\{2^{-J}\sup_{v\in\mathcal{V}}\left|\frac{\partial^2\rho(v,\theta)}{\partial\theta_q\partial\theta_{q'}}\right|\left\|N^{(r)\prime}\right\|_{\infty,\mathbb{R}}+\sup_{v\in\mathcal{V}}\left|\frac{\partial\rho(v,\theta)}{\partial\theta_q}\frac{\partial\rho(v,\theta)}{\partial\theta_{q'}}\right|\left\|N^{(r)\prime\prime}\right\|_{\infty,\mathbb{R}}\right\}\le U,$$

using $\left\|N^{(r)\prime}\right\|_{\infty,\mathbb{R}}\le 1$ and $\left\|N^{(r)\prime\prime}\right\|_{\infty,\mathbb{R}}\le 2$, cf. (29), (30). Furthermore, for θ, $\theta'\in B(\theta_0)$ we get again using (29), (30), the bounds for the Lipschitz constants of $N^{(r)\prime}$ and $N^{(r)\prime\prime}$ given at the beginning of this section, and boundedness of $B(\theta_0)$

$$\sup_{v\in\mathcal{V}}\left|h^{(2)}_{\theta,l}(v)-h^{(2)}_{\theta',l}(v)\right| \le 2^{1-J}d_rL'\left\|\theta-\theta'\right\|^\beta$$
$$+12d_r\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla_\theta\rho(v,\theta)\right\|\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla^2_\theta\rho(v,\theta)\right\|\left\|\theta-\theta'\right\|$$
$$+2^{J+3}d_r\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla_\theta\rho(v,\theta)\right\|\sup_{\theta\in B(\theta_0)}\sup_{v\in\mathcal{V}}\left\|\nabla_\theta\rho(v,\theta)\right\|^2\left\|\theta-\theta'\right\|$$
$$\le 2^Jc_{**}\left\|\theta-\theta'\right\|^\beta$$

with the constant c_{**} being finite and depending only on $B(\theta_0)$, r, ρ. Proceeding as in the proof of Part a we obtain

$$N(\mathcal{H}^{(2)}_{l,J,r},\mathsf{L}^\infty(\mathcal{V}),\varepsilon)\le(AU/\varepsilon)^v\quad\text{for }0<\varepsilon\le 2U,$$

where now $v=\max(b/\beta,2)$ and $A=\max\left(2^JM^{\beta/b}\max(c_{**}U^{-1},1),2e\right)$ with M only depending on $B(\theta_0)$. Again, apart from the factor 2^J, A depends only on $B(\theta_0)$, b, ρ, r, and $\sup_{\theta\in B(\theta_0)}\|p(\theta)\|_\infty$.

Part a of Proposition 12 in Appendix C applied to a countable sup-norm dense subset of $\mathcal{H}^{(2)}_{l,J,r}$ and with $b_0 = v^{-1}$ now yields the existence of positive finite constants C'_7 and C'_8 depending only on $B(\theta_0)$, b, ρ, r, and $\sup_{\theta\in B(\theta_0)} \|p(\theta)\|_\infty$, such that for all $J \in \mathbb{N}$ and all $k \geq C'_8 2^J J$

$$E \sup_{\theta\in\Theta} \left| \sum_{i=1}^{k} h^{(2)}_{\theta,l}(V_i) \right|^2 \leq C'_7 k 2^{-J} J$$

holds. Since this bound does not depend on l, the proof is complete upon setting $C_7 = (r+1)C'_7/2$ and $C_8 = 1/C'_8$. ■

Corollary 1 *Suppose Assumption R(i) is satisfied and $r \geq 2$. Suppose further that Θ is a bounded subset of $\mathbb{R}^b$ and that $\{p(\theta) : \theta \in \Theta\}$ is bounded in B_t for some $1/2 < t \leq 1$. If $J_k \in \mathbb{N}$ satisfies*

$$\sup_{k\geq 1} 2^{J_k(2t+1)} J_k/k < \infty, \tag{45}$$

then $\sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta)\|_{t,2}$ is stochastically bounded, i.e.,

$$\lim_{M\to\infty} \sup_{k\geq 1} \Pr\left(\sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta)\|_{t,2} > M \right) = 0.$$

If (45) holds and $J_k \to \infty$ for $k \to \infty$, then, for every $0 < t' < t$, $\sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta) - p(\theta)\|_{t',2}$ as well as $\sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta) - p(\theta)\|_\infty$ converge to zero in (outer) probability as $k \to \infty$.

Proof. Observe that under (45) we have $2^{J_k} J_k \leq C_4 k$ for k large enough, where C_4 is as in Lemma 5, and that $\{p(\theta) : \theta \in \Theta\}$ is sup-norm bounded. Now, using Lemma 5 together with Ljapunov's inequality as well as Proposition 9 in Appendix A, we arrive, for k large enough, at

$$\begin{aligned}
E \sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta)\|_{t,2} &\leq E \sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta) - Ep_{k,J_k,r}(\theta)\|_{t,2} + \sup_{\theta\in\Theta} \|Ep_{k,J_k,r}(\theta)\|_{t,2} \\
&\leq C_0(t)\sqrt{C_3} 2^{J_k t} \sqrt{\frac{2^{J_k} J_k}{k}} + \sup_{\theta\in\Theta} \|\pi^{(r)}_{J_k}(p(\theta))\|_{t,2} \\
&\leq C_0(t)\sqrt{C_3} \sup_{k\geq 1} 2^{J_k t} \sqrt{\frac{2^{J_k} J_k}{k}} + c''_t \sup_{\theta\in\Theta} \|p(\theta)\|_{t,2} < \infty,
\end{aligned}$$

where we have used the already established fact that $Ep_{k,J_k,r}(\theta) = \pi^{(r)}_{J_k}(p(\theta))$. [Measurability of $\sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta)\|_{t,2}$ follows from Appendix B.] Together with the observation that $E \sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta)\|_{t,2} < \infty$ for every $k \geq 1$, this completes the proof of the first claim. Next, Lemma 5 (applied with $s' = t'$) gives for k large enough (E^* denoting outer expectation)

$$\begin{aligned}
E^* \sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta) - p(\theta)\|_{t',2} &\leq E \sup_{\theta\in\Theta} \|p_{k,J_k,r}(\theta) - Ep_{k,J_k,r}(\theta)\|_{t',2} + \sup_{\theta\in\Theta} \|\pi^{(r)}_{J_k}(p(\theta)) - p(\theta)\|_{t',2} \\
&\leq C_0(t')\sqrt{C_3} 2^{J_k t'} \sqrt{\frac{2^{J_k} J_k}{k}} + 2^{-J_k(t-t')} c'''_{t,t'} \sup_{\theta\in\Theta} \|p(\theta)\|_{t,2},
\end{aligned}$$

where we have used Proposition 9 in Appendix A in the final step. The upper bound now converges to zero as $k \to \infty$. The claim regarding the sup-norm now follows from Proposition 7 in Appendix A. ■

The following corollary is proved analogously using Lemma 3 instead of Lemma 5, with measurability of the relevant quantities following from Appendix B.

Corollary 2 *Suppose $r_* \geq 2$ and that $p_0 \in \mathcal{B}_t$ for some $1/2 < t \leq 1$. If $j_n \in \mathbb{N}$ satisfies*

$$\sup_{n\geq 1} 2^{j_n(2t+1)}/n < \infty, \tag{46}$$

then $\|p_{n,j_n,r_*}\|_{t,2}$ is stochastically bounded, i.e.,

$$\lim_{M\to\infty} \sup_{n\geq 1} \Pr\left(\|p_{n,j_n,r_*}\|_{t,2} > M\right) = 0.$$

If (46) holds and $j_n \to \infty$ for $n \to \infty$, then, for every $0 < t' < t$, $\|p_{n,j_n,r_*} - p_0\|_{t'}$ as well as $\|p_{n,j_n,r_*} - \tilde{p}_0\|_\infty$ converge to zero in probability as $n \to \infty$, where $\tilde{p}_0$ is the continuous version of p_0.

7 Uniform Central Limit Theorems for Spline Projection Estimators

We now study the difference between the random measure $P_{k,J,r}(\theta)$ given by

$$dP_{k,J,r}(\theta)(y) = p_{k,J,r}(\theta, y)dy$$

and $P_k(\theta)$, acting on Besov classes by integration. In the following $\|\nu\|_{\mathcal{F}}$ stands for $\sup_{f\in\mathcal{F}} |\nu(f)|$, where ν is a (signed) measure.

Theorem 3 *Suppose Assumption R(i) is satisfied, $r \geq 2$, Θ is a bounded subset of $\mathbb{R}^b$, and $\{p(\theta) : \theta \in \Theta\}$ is a bounded subset of $\mathcal{B}_t$ for some t, $0 < t < r$. Let $\mathcal{F}$ be a (non-empty) bounded subset of B_s for some s, $1/2 < s < 1$. Then for every $1/2 < s' \leq s$ there is a finite positive constant C_9, depending only on s, s', t, $\mathcal{F}$, Θ, b, α, L, and $\{p(\theta) : \theta \in \Theta\}$ but not on J and k, such that for every $J \geq 1$ and $k \geq 1$*

$$E \sup_{\theta\in\Theta} \|P_{k,J,r}(\theta) - P_k(\theta)\|_{\mathcal{F}} \leq C_9(2^{-J(t+s)} + 2^{-J(s-s')}k^{-1/2}). \tag{47}$$

Furthermore,

$$\sup_{\theta\in\Theta} \|P_k(\theta) - P(\theta)\|_{\mathcal{F}} = O_p(k^{-1/2}) \tag{48}$$

holds. Finally, if $J_k \to \infty$ as $k \to \infty$ satisfies $2^{-J_k(t+s)} = o(k^{-1/2})$, then for every $\theta \in \Theta$

$$\sqrt{k}\,(P_{k,J_k,r}(\theta) - P(\theta)) \rightsquigarrow_{\ell^\infty(\mathcal{F})} G_{P(\theta)},$$

where $G_{P(\theta)}$ is a sample-bounded and sample-continuous generalized $P(\theta)$-Brownian bridge indexed by $\mathcal{F}$. Here $\rightsquigarrow_{\ell^\infty(\mathcal{F})}$denotes convergence in law as defined in Chapter 1 of van der Vaart and Wellner (1996).

Proof. We first note that $\sup_{\theta\in\Theta} \|P_{k,J,r}(\theta) - P_k(\theta)\|_{\mathcal{F}}$ and $\sup_{\theta\in\Theta} \|P_k(\theta) - P(\theta)\|_{\mathcal{F}}$ are measurable since they can be represented as suprema over countable dense subsets of Θ and $\mathcal{F}$ in view of Assumption R(i), $r \geq 2$, and separability of $\mathcal{F}$. For $f \in \mathcal{F}$ we can write, using (6), (7), (11) and symmetry of the projection kernel $K_J^{(r)}$,

$$\begin{aligned}
&(P_{k,J,r}(\theta) - P_k(\theta))(f) = \frac{1}{k}\sum_{i=1}^{k}\left(\int_0^1 f(y)K_J^{(r)}(X_i(\theta), y)dy - f(X_i(\theta))\right) \\
= \;& \frac{1}{k}\sum_{i=1}^{k}(\pi_J^{(r)}(f) - f)(X_i(\theta)) = (P_k(\theta) - P(\theta))(\pi_J^{(r)}(f) - f) + \int_0^1 (\pi_J^{(r)}(f) - f)(y)p(\theta)(y)dy \\
= \;& A + B.
\end{aligned}$$

Consider first term B: Using $f \in \mathcal{L}^2$, $p(\theta) \in \mathcal{L}^2$, self-adjointness and idempotency of the projection $Id - \pi_J^{(r)}$ we obtain

$$\begin{aligned}
\left| \int_0^1 \left(\pi_J^{(r)}(f) - f \right)(y) p(\theta)(y) dy \right| &= \left| \int_0^1 \left(\left(Id - \pi_J^{(r)} \right) f \right)(y) \left(\left(Id - \pi_J^{(r)} \right) p(\theta) \right)(y) dy \right| \\
&\leq \left\| f - \pi_J^{(r)}(f) \right\|_2 \left\| p(\theta) - \pi_J^{(r)}(p(\theta)) \right\|_2 \\
&\leq c_s' c_t' \|f\|_{s,2} \|p(\theta)\|_{t,2} 2^{-J(s+t)}, \qquad (49)
\end{aligned}$$

where we have used Proposition 8 for the last inequality. Consider next the term A: Define for $J \geq 1$ the class of functions

$$\begin{aligned}
\mathcal{F}_{J,r,\rho} &= \left\{ \int_0^1 K_J^{(r)}(\rho(\cdot,\theta), y) f(y) dy - f(\rho(\cdot,\theta)) : f \in \mathcal{F},\, \theta \in \Theta \right\} \\
&= \left\{ (\pi_J^{(r)}(f) - f)(\rho(\cdot,\theta)) : f \in \mathcal{F},\, \theta \in \Theta \right\}, \qquad (50)
\end{aligned}$$

which allows us to write

$$E \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| (P_k(\theta) - P(\theta))(\pi_J^{(r)}(f) - f) \right| = \frac{1}{k} E \sup_{h \in \mathcal{F}_{J,r,\rho}} \left| \sum_{i=1}^k (h(V_i) - Eh(V_i)) \right|. \qquad (51)$$

Choose an arbitrary s' satisfying $1/2 < s' \leq s$ and observe that $(\pi_J^{(r)}(f) - f) \in \mathsf{B}_s \subseteq \mathsf{B}_{s'}$ since $\mathcal{F} \subseteq \mathsf{B}_s$ by assumption and that $\mathcal{S}_J(r) \subseteq \mathsf{B}_s \subseteq \mathsf{B}_{s'}$ in view of $s < 1 < r - 1/2$. Propositions 7 and 9 in Appendix A then give

$$\begin{aligned}
\sup_{h \in \mathcal{F}_{J,r,\rho}} \sup_{v \in \mathcal{V}} |h(v) - Eh(V_i)| &\leq 2 \sup_{h \in \mathcal{F}_{J,r,\rho}} \sup_{v \in \mathcal{V}} |h(v)| \leq 2 \sup_{f \in \mathcal{F}} \left\| \pi_J^{(r)}(f) - f \right\|_\infty \\
&\leq 2 c_{s'} \sup_{f \in \mathcal{F}} \left\| \pi_J^{(r)}(f) - f \right\|_{s',2} \leq 2 c_{s'} c_{s,s'}''' \sup_{f \in \mathcal{F}} \|f\|_{s,2} 2^{-J(s-s')} =: U
\end{aligned}$$

where $U < \infty$ since $\mathcal{F}$ is a (non-empty) bounded subset of B_s. We may assume $U > 0$, the case $U = 0$ being trivial. Since $\mathcal{F}_{J,r,\rho}$ contains a countable sup-norm dense subset in view of Proposition 6 below, we may apply the moment inequality from Proposition 12, part b, in Appendix C to (51) (with U as above, $\sigma = U$, $A' = c^{*s'} / \left(2 c_{s'} c_{s,s'}''' \sup_{f \in \mathcal{F}} \|f\|_{s,2} \right)$, and with $w = 1/s'$) and make use of the entropy bound in Proposition 6 below with $\varepsilon^* = 4 c_{s'} c_{s,s'}''' \sup_{f \in \mathcal{F}} \|f\|_{s,2} \geq 2U$. This gives the bound

$$E \sup_{\theta \in \Theta} \sup_{f \in \mathcal{F}} \left| (P_k(\theta) - P(\theta))(\pi_J^{(r)}(f) - f) \right| \leq 2^{-J(s-s')+1} k^{-1/2} c_{s'} c_{s,s'}''' \sup_{f \in \mathcal{F}} \|f\|_{s,2} b_2$$

where the constant b_2 only depends on A' and w. Together with (49), this proves the bound (47). To prove the second claim, define the class

$$\mathcal{F}_\rho = \{ f(\rho(\cdot,\theta)) : f \in \mathcal{F},\, \theta \in \Theta \} \qquad (52)$$

and note that $\mathcal{F}_\rho$ is uniformly bounded since $\mathcal{F}$ is and that

$$\sup_{\theta \in \Theta} \|P_k(\theta) - P(\theta)\|_{\mathcal{F}} = \frac{1}{k} \sup_{h \in \mathcal{F}_\rho} \left| \sum_{i=1}^k (h(V_i) - Eh(V_i)) \right|.$$

Now (48) follows since $\mathcal{F}_\rho$ is a universal Donsker class by Proposition 6 below. The third claim of the theorem follows immediately from (47) with s' chosen to satisfy $s' < s$, from the assumptions on J_k, and from the universal Donsker property of $\{f(\rho(\cdot,\theta)) : f \in \mathcal{F}\}$ for every θ, which it inherits from $\mathcal{F}_\rho$. ■

Proposition 6 *Suppose Assumption R(i) is satisfied, $r \geq 2$, and Θ is a bounded subset of $\mathbb{R}^b$. Let $\mathcal{F}$ be a (non-empty) bounded subset of B_s, $1/2 < s < 1$. Let $\mathcal{F}_{J,r,\rho}$ and $\mathcal{F}_\rho$ be defined as in (50) and (52). Then for every $1/2 < s' \leq s$ and every $\varepsilon^* > 0$ there exists a (positive) finite constant c^*, depending only on s, s', $\mathcal{F}$, Θ, b, α, L, and ε^* but not on J, such that for every $J \geq 1$*

$$\log N(\mathcal{F}_{J,r,\rho}, \mathsf{L}^\infty(\mathcal{V}), \varepsilon) \leq 2^{-J(s-s')/s'} c^* \varepsilon^{-1/s'} \quad \text{for } 0 < \varepsilon \leq \varepsilon^* \tag{53}$$

holds. Furthermore, for every $\varepsilon^* > 0$ there exists a (positive) finite constant c^{**} (depending only on s, $\mathcal{F}$, Θ, b, α, L, and ε^*) such that

$$\log N(\mathcal{F}_\rho, \mathsf{L}^\infty(\mathcal{V}), \varepsilon) \leq c^{**} \varepsilon^{-1/s} \quad \text{for } 0 < \varepsilon \leq \varepsilon^* \tag{54}$$

holds. In particular, $\mathcal{F}_\rho$ and $\mathcal{F}_{J,r,\rho}$ are universal Donsker classes.

Proof. Let s' be as in the proposition. By Proposition 9

$$\sup_{f\in\mathcal{F}} \left\| \pi_J^{(r)}(f) - f \right\|_{s',2} \leq 2^{-J(s-s')} c'''_{s,s'} \sup_{f\in\mathcal{F}} \|f\|_{s,2} = 2^{-J(s-s')} D < \infty, \tag{55}$$

where the constant D depends only on s, s', and $\mathcal{F}$. As a consequence,

$$\mathcal{G}_J := \left\{ (\pi_J^{(r)}(f) - f) : f \in \mathcal{F} \right\}$$

is contained in a ball $\mathcal{U}_J$ in $\mathsf{B}_{s'}$ of radius $2^{-J(s-s')}D$. Using entropy bounds for balls in Besov spaces (e.g., Theorem 15.6.1 in Lorentz, v.Golitschek, and Makovoz (1996)) we obtain

$$\log N(\mathcal{G}_J, \mathsf{L}^\infty([0,1]), \varepsilon) \leq 2^{-J(s-s')/s'} c(s, s', \mathcal{F}) \varepsilon^{-1/s'} \quad \text{for } 0 < \varepsilon < \infty$$

where the finite and positive constant $c(s, s', \mathcal{F})$ depends only on s, s', and $\mathcal{F}$ (in particular, it is independent of J). [Setting $p = 2$, $q = \infty$ in Lorentz, v.Golitschek, and Makovoz (1996) we actually obtain the above bound only in the ess-sup norm. However, since $\mathcal{G}_J$ consists of *continuous* functions only and since we can always assume that the centers of the covering ess-sup norm balls belong to $\mathcal{G}_J$ (perhaps at the expense of doubling ε), we immediately obtain the same bound for the supremum-norm.]

To prove the entropy bound for $\mathcal{F}_{J,r,\rho} = \{g(\rho(\cdot,\theta)) : g \in \mathcal{G}_J, \theta \in \Theta\}$ we proceed as follows: Note that the elements of $\mathcal{G}_J$ are Hölder continuous of order $s' - 1/2$ with Hölder constants uniformly bounded by $2^{-J(s-s')} c_1(s', D)$, with $0 < c_1(s', D) < \infty$ depending only on s' and D, since $\mathcal{G}_J \subseteq \mathcal{U}_J \subseteq \mathsf{B}_{s'}$ and since for $1/2 < s' < 1$ the space $\mathsf{B}_{s'}$ is continuously embedded into $\mathsf{C}^{s'-1/2}$, cf. Proposition 7 in Appendix A. Define $\eta = (\alpha(s' - 1/2))^{-1}$ with α defined in Assumption R1. For $0 < \varepsilon \leq 1$ set $\delta = \left(2^{J(s-s')}\varepsilon\right)^\eta$ and cover Θ by δ-balls with centers $\theta_1, \ldots, \theta_{N(\delta,\Theta)}$ where $N(\delta, \Theta)$ satisfies $N(\delta, \Theta) \leq \max(1, M(\Theta)/\delta^b)$ for some constant $M(\Theta)$ only depending on Θ. Let $g_1, \ldots, g_{N(\mathcal{G}_J, \mathsf{L}^\infty([0,1]), \varepsilon)}$ be the centers of $\mathsf{L}^\infty([0,1])$-balls of radius ε covering $\mathcal{G}_j$. We then have for $g(\rho(\cdot,\theta)) \in \mathcal{F}_{J,r,\rho}$ using Assumption R1

$$\begin{aligned}
& \sup_{v\in\mathcal{V}} |g(\rho(v,\theta)) - g_i(\rho(v,\theta_l)| \\
\leq\ & \sup_{v\in\mathcal{V}} |g(\rho(v,\theta)) - g(\rho(v,\theta_l))| + \sup_{v\in\mathcal{V}} |g(\rho(v,\theta_l)) - g_i(\rho(v,\theta_l))| \\
\leq\ & 2^{-J(s-s')} c_1(s', D) \left(L\,|\theta - \theta_l|^\alpha\right)^{s'-1/2} + \sup_{x\in[0,1]} |g(x) - g_i(x)| \leq \left(c_1(s', D) L^{1/\eta} + 1\right)\varepsilon
\end{aligned}$$

for suitable choice of i and l. Consequently, we obtain for $0 < \varepsilon \leq 1$

$$\log N(\mathcal{F}_{J,r,\rho}, \mathsf{L}^\infty(\mathcal{V}), \left(c_1(s', D)L^{1/\eta} + 1\right)\varepsilon) \leq \log N(\mathcal{G}_J, \mathsf{L}^\infty([0,1]), \varepsilon) + \log N(\delta, \Theta)$$

$$\leq \quad c(s, s', \mathcal{F}) \left(2^{J(s-s')}\varepsilon\right)^{-1/s'} + \log^+ \left(M(\Theta)/(2^{J(s-s')}\varepsilon)^{b\eta}\right) \leq c_\bullet 2^{-J(s-s')/s'}\varepsilon^{-1/s'},$$

for a suitable finite constant $c_\bullet$ only depending on s, s', $\mathcal{F}$, Θ, b, and α, but not on J. After a simple substitution, this gives (53) for $0 < \varepsilon \leq c_1(s', D)L^{1/\eta} + 1$. Appropriately adjusting the multiplicative constant in this so-obtained bound gives (53) for all $0 < \varepsilon \leq \varepsilon^*$; note that the adjustment of the constant only introduces an additional dependence on ε^* (but no dependence on J). The entropy bound (54) for $\mathcal{F}_\rho$ is proved in a similar (even simpler) way. The Donsker property of $\mathcal{F}_{J,r,\rho}$ and $\mathcal{F}_\rho$ now follows from (53), (54) and Theorem 2.8.4 in van der Vaart and Wellner (1996), noting that $\mathcal{F}_{J,r,\rho}$ and $\mathcal{F}_\rho$ are uniformly bounded in view of Proposition 7 and that the bracketing covering numbers are dominated by the sup-norm covering numbers. ■

An analogous result holds for the random measure P_{n,j,r_*} given by $dP_{n,j,r_*}(y) = p_{n,j,r_*}(y)dy$. The proof of this result is similar to, in fact simpler than, the proof of Theorem 3 and thus is omitted.

Theorem 4 *Suppose $r_* \geq 2$, and $p_0 \in \mathcal{B}_t$ for some t, $0 < t < r_*$. Let $\mathcal{F}$ be a (non-empty) bounded subset of B_s for some s, $1/2 < s < 1$. Then for every $1/2 < s' \leq s$ there is a finite positive constant C_{10} independent of j (only depending on s, s', t, $\mathcal{F}$, and p_0) such that for every $j \geq 1$ and $k \geq 1$*

$$E\left\|P_{n,j,r_*} - P_n\right\|_{\mathcal{F}} \leq C_{10}(2^{-j(t+s)} + 2^{-j(s-s')}n^{-1/2}).$$

Furthermore, $\|P_n - P\|_{\mathcal{F}} = O_p(n^{-1/2})$ holds. Finally, if $j_n \to \infty$ as $n \to \infty$ satisfies $2^{-j_n(t+s)} = o(n^{-1/2})$, then

$$\sqrt{n}\,(P_{n,j_n,r_*} - P) \rightsquigarrow_{\ell^\infty(\mathcal{F})} G_P,$$

where G_P is a sample-bounded and sample-continuous generalized P-Brownian bridge indexed by $\mathcal{F}$.

A Appendix: Some Properties of Besov Spaces and Approximation by Splines

In the following, we summarize some simple properties of the spaces $\mathcal{B}_s$. For $0 < s \leq 1$ and bounded $f : [0,1] \to \mathbb{R}$ denote by

$$\|f\|_{s,\infty} = \|f\|_\infty + \sup_{x,y\in[0,1],x\neq y} \frac{|f(x) - f(y)|}{|x-y|^s}$$

the usual Hölder norm and denote by C^s the set of all functions f with finite $\|f\|_{s,\infty}$. For simplicity we restrict ourselves to the case $s < 1$ in the following proposition.

Proposition 7 *Let $1/2 < s < 1$.*

a. Every $f \in \mathcal{B}_s$ is λ-a.e. equal to a function $\tilde{f} \in \mathsf{C}^{s-1/2}$ and

$$\left\|\tilde{f}\right\|_\infty \leq \left\|\tilde{f}\right\|_{(s-1/2),\infty} \leq c_s \left\|\tilde{f}\right\|_{s,2} = c_s \|f\|_{s,2}$$

holds for some finite (positive) constant c_s that depends only on s.

b. If $f \in \mathcal{B}_s$ and $h \in \mathcal{B}_s$, then $\|fh\|_{s,2} \leq 2c_s \|f\|_{s,2} \|h\|_{s,2}$. If $h \in \mathcal{B}_s$ satisfies $\zeta := \inf_{x\in[0,1]} h(x) > 0$, then $\|1/h\|_{s,2} \leq \zeta^{-1} + \zeta^{-2} \|h\|_{s,2}$.

Proof. a. Observe that $\mathcal{B}_s$ coincides (up to norm equivalence) with the intermediate space $(\mathcal{L}^2, \mathcal{W}_2^1)_{s,\infty}$ (DeVore and Lorentz (1993), p.196) and hence coincides with the Besov space $\mathcal{B}^{s;2,\infty}((0,1))$ defined in Adams and Fournier (2003) (the fact that the latter is defined on the open unit interval being irrelevant). The claim then follows from applying Theorem 7.37 in Adams and Fournier (2003) (with $m = n = 1$, $j = 0$, $p = 2$, $q = \infty$).

b. Since $s < 1$ by assumption, we may set $a = 1$ in the definition of the Besov (semi)norm. Elementary calculations then show that

$$\|fh\|_{s,2} \leq \|f\|_{s,2} \operatorname{esssup}|h| + \|h\|_{s,2} \operatorname{esssup}|f| \leq 2c_s \|f\|_{s,2} \|h\|_{s,2}$$

in view of Part a. The second claim follows since clearly $\|1/h\|_2 \leq \zeta^{-1}$ and since elementary calculations give $\left\|\Delta_z h^{-1}\right\|_2 \leq \zeta^{-2} \|\Delta_z h\|_2$. ■

The above proposition, together with the continuous embedding of $\mathcal{B}_t$ into $\mathcal{B}_s$ for $t \geq s$ (DeVore and Lorentz (1993), p.56), immediately guarantees for *every* $t > 1/2$ the existence of a constant c_t, $0 < c_t < \infty$, such that for every $f \in \mathcal{B}_t$ there exists a (unique) continuous $\tilde{f}$, λ-a.e. equal to f, such that $\|\tilde{f}\|_\infty \leq c_t \|\tilde{f}\|_{t,2} = c_t \|f\|_{t,2}$. In particular, bounded subsets of B_t, $t > 1/2$, are sup-norm bounded.

As is well known, functions in $\mathcal{B}_s$ can be approximated by elements of the Schoenberg spaces $\mathcal{S}_j(r)$, the error decreasing as j increases. We summarize these facts in the following proposition.

Proposition 8 *Suppose $r \in \mathbb{N}$.*

a. If $h \in \mathcal{L}^2$, then the ortho-projection operator $\pi_j^{(r)}$ from $\mathcal{L}^2$ onto the Schoenberg space $\mathcal{S}_j(r)$ satisfies

$$\lim_{j\to\infty} \|\pi_j^{(r)}(h) - h\|_2 = 0.$$

If $\mathcal{H}$ is a relatively compact subset of $\mathcal{L}^2$, then

$$\lim_{j\to\infty} \sup_{h\in\mathcal{H}} \|\pi_j^{(r)}(h) - h\|_2 = 0.$$

b. If $h \in \mathcal{B}_s$ for some $s \in (0, r)$, then

$$\|\pi_j^{(r)}(h) - h\|_2 \leq 2^{-js} c_s' \|h\|_{s,2},$$

for every $j \in \mathbb{N}$, where the (positive) finite constant c_s' depends only on s.

Proof. To prove the first claim in Part a, observe that by Proposition 2.4.1 and (12.3.2) in DeVore and Lorentz (1993)

$$\|\pi_j^{(r)}(h) - h\|_2 \leq 2C^{(r)} \sup_{0<z\leq 2^{-j}} \|\Delta_z^r(h)\|_2$$

for some universal constant $C^{(r)}$. By continuity of translation in $\mathcal{L}^2(\mathbb{R})$ (cf., e.g., Folland (1999), Proposition 8.5) the right-hand side converges to zero as $j \to \infty$ (note that $\|\Delta_z^r(h)\|_2$ is less than or equal to the corresponding expression that is obtained when h is viewed as a function on $\mathbb{R}$ which is zero outside of $[0,1]$). The second claim in Part a follows since for every $\varepsilon > 0$ and ε-net $\{h_l : 1 \leq l \leq N(\varepsilon)\}$ for $\mathcal{H}$ we have that $\|h - h_l\|_2 \leq \varepsilon$ implies $\|\pi_j^{(r)}(h) - \pi_j^{(r)}(h_l)\|_2 \leq \varepsilon$ and thus

$$\sup_{h\in\mathcal{H}} \|\pi_j^{(r)}(h) - h\|_2 \leq \max_{1\leq l\leq N(\varepsilon)} \|\pi_j^{(r)}(h_l) - h_l\|_2 + 2\varepsilon$$

holds. For the proof of Part b use Proposition 2.4.1 and (12.3.2) in DeVore and Lorentz (1993) (where one sets $p = 2$, $n = 2^j$) together with the definition of the Besov-norm. ■

Proposition 9 *Suppose $r \in \mathbb{N}$. Let $h \in \mathcal{B}_s$ for some $s \in (0, r-1/2)$. Then*

$$\|\pi_j^{(r)}(h)\|_{s,2} \leq c_s'' \|h\|_{s,2},$$

for every $j \in \mathbb{N}$, where the (positive) finite constant c_s'' depends only on s. Furthermore, for every $s' \in (0, s]$

$$\|\pi_j^{(r)}(h) - h\|_{s',2} \leq 2^{-j(s-s')} c_{s,s'}''' \|h\|_{s,2}$$

for every $j \in \mathbb{N}$, where the (positive) finite constant $c_{s,s'}'''$ depends only on s and s'.

Proof. By Theorem 12.3.3. in DeVore and Lorentz (1993) (with $p = 2$, $\lambda = r - 1/2$, $q = \infty$, $\alpha = s$, and $d_{n,r}(\cdot)_2$ defined on p.358 of that reference) we have

$$\begin{aligned}
\|\pi_j^{(r)}(h)\|_{s,2} &= \|\pi_j^{(r)}(h)\|_2 + \sup_{0 \neq |z| < 1} |z|^{-s} \|\Delta_z^r(\pi_j^{(r)}(h))\|_2 \\
&\leq \|h\|_2 + e_s \sup_{n \geq 0} 2^{ns} d_{n,r}(\pi_j^{(r)}(h))_2 \\
&\leq \|h\|_2 + e_s \sup_{n \geq 0} 2^{ns} \|\pi_n^{(r)}(\pi_j^{(r)}(h)) - \pi_j^{(r)}(h)\|_2 \\
&\leq \|h\|_2 + e_s \sup_{0 \leq n < j} 2^{ns} \|\pi_n^{(r)}(h) - \pi_j^{(r)}(h)\|_2 \leq \|h\|_{s,2} + 2e_s c_s' \|h\|_{s,2}
\end{aligned}$$

for some universal constant e_s, where we have used Proposition 8 in the last step. To prove the second claim we argue as before and then use Proposition 8 to obtain

$$\begin{aligned}
\|\pi_j^{(r)}(h) - h\|_{s',2} &\leq \|\pi_j^{(r)}(h) - h\|_2 + e_{s'} \sup_{n \geq 0} 2^{ns'} \|\pi_n^{(r)}(\pi_j^{(r)}(h) - h) - (\pi_j^{(r)}(h) - h)\|_2 \\
&\leq \|\pi_j^{(r)}(h) - h\|_2 + e_{s'} \left[2^{js'} \|\pi_j^{(r)}(h) - h\|_2 + \sup_{n > j} 2^{ns'} \|\pi_n^{(r)}(h) - h\|_2 \right] \\
&\leq 2^{-js} c_s' \|h\|_{s,2} + e_{s'} \left[2^{j(s'-s)} c_s' \|h\|_{s,2} + \sup_{n > j} 2^{n(s'-s)} c_s' \|h\|_{s,2} \right] \\
&\leq 2^{-j(s-s')} (1 + 2e_{s'}) c_s' \|h\|_{s,2}.
\end{aligned}$$

■

B Appendix: Consistency of the Indirect Inference Estimator and Measurability Issues

Proof of Proposition 1. Because of continuity of the B-spline basis functions for $r \geq 2$ and continuity of $\theta \to \rho(v, \theta)$ for every $v \in \mathcal{V}$, the map $\theta \to p_{k,J,r}(\theta)(y)$ is continuous for every $y \in [0,1]$. Furthermore, p_{n,j,r_*} and $p_{k,J,r}(\theta)$ are bounded on $[0,1]$, the latter one uniformly in θ, in view of the discussion surrounding (12). Next note that the set A_n appearing in the definition of $\mathcal{Q}_{n,k}$ coincides with the event $\left\{\inf_{y \in [0,1]} p_{n,j,r_*}(y) > 0\right\}$, since p_{n,j,r_*} is continuous on $[0,1]$ in case $r_* > 1$, and is piecewise constant in case $r_* = 1$. Hence, by the dominated convergence theorem, $\mathcal{Q}_{n,k}$ is continuous (and real-valued) on Θ if $p_{n,j,r_*}(y) > 0$ for every $y \in [0,1]$; and the same conclusion trivially holds in the other case. As mentioned before, $\mathcal{Q}_{n,k}(\theta) : [0,1]^\infty \times \mathcal{V}^\infty \to \mathbb{R}$ is $\mathfrak{B}_{[0,1]}^\infty \otimes \mathfrak{V}^\infty$-measurable for every $\theta \in \Theta$. Since Θ is compact, existence of a measurable minimizer then follows, e.g., from Lemma A3 in Pötscher and Prucha (1997). ■

Proposition 10 *Suppose Θ is compact in $\mathbb{R}^b$, that the map $\theta \to p(\theta, x)$ is continuous on Θ for every $x \in [0,1]$ and that $\sup_{\theta\in\Theta} \|p(\theta)\|_\infty < \infty$. Furthermore, assume that $r_* \geq 1$ holds. Then there exists a $\mathfrak{B}^\infty_{[0,1]} \otimes \mathfrak{V}^\infty$-measurable $\hat{\theta}_n$ that minimizes $Q_n(\theta)$ over Θ. (In fact, $\hat{\theta}_n$ is $\mathfrak{B}^\infty_{[0,1]}$-measurable as it does not depend on the simulations.)*

Proof. Since $\|p_{n,j,r_*}\|_\infty < \infty$ and since on the event A_n also $\inf_{y\in[0,1]} p_{n,j,r_*} > 0$ holds, the assumptions on $p(\theta)$ and the dominated convergence theorem imply that Q_n is real-valued and continuous in θ on the event A_n; and the same conclusion trivially holds on the complement of A_n. Furthermore, $\mathfrak{B}^\infty_{[0,1]} \otimes \mathfrak{V}^\infty$-measurability of $Q_n(\theta) : [0,1]^\infty \times \mathcal{V}^\infty \to \mathbb{R}$ for every $\theta \in \Theta$ follows from Tonelli's Theorem since p_{n,j,r_*} is jointly measurable (and A_n is measurable). Since Θ is compact, existence of a measurable minimizer then follows, e.g., from Lemma A3 in Pötscher and Prucha (1997). ■

Proposition 11 *Suppose Assumptions P1(i),(ii) are satisfied and $r_* \geq 2$ holds. If $j_n \to \infty$ as $n \to \infty$ in such a way that for some $\delta > 1/2$ we have $\sup_{n\geq 1} 2^{j_n(2\delta+1)}/n < \infty$ then*

$$\hat{\theta}_n \to \theta_0 \text{ in } \Pr\text{-probability as } n \to \infty,$$

where $\hat{\theta}_n$ has been defined in Section 5.2.

The proof of this result is completely analogous to the proof of Proposition 2 and is thus omitted.

Remark 3 *(Measurability issues) (i) For every $J \geq 1$, $r \geq 1$, and $\theta \in \Theta$, the expressions $\|p_{k,J,r}(\theta)\|_2$, $\|p_{k,J,r}(\theta)\|_\infty$, and $\|p_{k,J,r}(\theta)\|_{s,2}$ (for $s \leq r - 1/2$) are measurable functions of $v_1, \ldots, v_k$, since the coefficients $\hat{\gamma}^{(r)}_{lJ}(\theta)$ are measurable. This is obvious for the $\mathcal{L}^2$-norm, but holds in general for the following reason: observe that any one of the norms mentioned, when restricted to $\mathcal{S}_J(r)$, is a continuous function of the coefficients $\hat{\gamma}^{(r)}_{lJ}(\theta)$ because $\mathcal{S}_J(r)$ is finite-dimensional. The same is true if $p_{k,J,r}(\theta)$ is replaced by $p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)$ or $p_{k,J,r}(\theta) - p(\theta)$, in the latter case provided the respective norm of $p(\theta)$ is finite. [The argument is the same, except that $\mathcal{S}_J(r)$ is to be replaced by the linear span of $\mathcal{S}_J(r) \cup \{p(\theta)\}$ for establishing the latter claim.] Analogous statements obviously also hold for p_{n,j,r_*} for every $j \geq 1$, $r \geq 1$. (ii) The reasoning just given in fact establishes that the above mentioned norms of $p_{k,J,r}(\theta)$ and $p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)$ are continuous functions of θ, provided the coefficients $\hat{\gamma}^{(r)}_{lJ}(\theta)$ (and $E\hat{\gamma}^{(r)}_{lJ}(\theta)$) are continuous in θ (which is, e.g., the case if $r \geq 2$ and Assumption R(i) holds); consequently, suprema over θ of the above mentioned norms of $p_{k,J,r}(\theta)$ and $p_{k,J,r}(\theta) - Ep_{k,J,r}(\theta)$ are then measurable. [We note that this argument does not apply to suprema of norms of $p_{k,J,r}(\theta) - p(\theta)$, because $p(\theta)$ may not vary in a finite-dimensional space when θ varies.]*

C Appendix: Moment Bounds for Empirical Processes

The following moment inequalities can be deduced from a general theorem in Giné and Koltchinskii (2006) and a refinement with explicit constants in Giné and Nickl (2009a).

Proposition 12 *Let Z_i, $i \in \mathbb{N}$, be i.i.d. random variables with values in a measurable space $(S, \mathcal{A})$ and common law R. Let $\mathcal{F}$ be a countable R-centered class of real valued measurable functions from $(S, \mathcal{A})$ to $\mathbb{R}$. Assume that $\mathcal{F}$ is uniformly bounded by a finite positive constant U and let further σ, $0 < \sigma \leq U$, be some constant satisfying $\sup_{f\in\mathcal{F}} Ef^2(Z_i) \leq \sigma^2$.*

a. Assume that the $\mathcal{L}^2(Q)$-covering numbers satisfy

$$\sup_Q \log N(\mathcal{F}, \mathcal{L}^2(Q), \tau) \leq v \log\left(\frac{AU}{\tau}\right), \quad 0 < \tau \leq 2U,$$

for some $A > e$ and $v \geq 2$ (the supremum extending over all probability measures Q on S). Then, for every $b_0 > 0$ satisfying

$$n\sigma^2 \geq b_0 v U^2 \log(5AU/\sigma) \quad \textit{for all } n \in \mathbb{N}, \tag{56}$$

there exists a finite positive constant $b_1(v, b_0)$, that depends only on v and b_0, such that for every $n \in \mathbb{N}$

$$E\left\|\sum_{i=1}^n f(Z_i)\right\|_{\mathcal{F}}^2 \leq b_1(v, b_0) n\sigma^2 \log\frac{AU}{\sigma}$$

holds.

b. Assume that the $\mathcal{L}^2(Q)$-covering numbers satisfy

$$\sup_Q \log N(\mathcal{F}, \mathcal{L}^2(Q), \tau) \leq \left(\frac{A'U}{\tau}\right)^w, \quad 0 < \tau \leq 2U,$$

for some $0 < A' < \infty$ and $0 < w < 2$. Then, for all $n \in \mathbb{N}$ and some positive constant b_2, that depends only on A', w, we have

$$E\left\|\sum_{i=1}^n f(Z_i)\right\|_{\mathcal{F}} \leq b_2 \sqrt{n} U.$$

Proof. Since the results depend only on the distribution of $\|\sum_{i=1}^n f(Z_i)\|_{\mathcal{F}}$, we may assume w.l.o.g. that – as in Giné and Koltchinskii (2006) – the random variables are realized as coordinate projections on the infinite product space of $(S, \mathcal{A})$. The second claim of the proposition then follows directly from Theorem 3.1 in Giné and Koltchinskii (2006) applied to the class $\mathcal{F}' = \{f/U : f \in \mathcal{F}\}$ with envelope $F = 1$ and $H(x) = (A'x)^w$ for $x \geq 1/2$ and $H(x) = 0$ for $0 \leq x < 1/2$. The first claim is proved as follows: By Proposition 3.1 in Giné, Latała and Zinn (2000) (applied to $\mathcal{F} \cup (-\mathcal{F})$ and observing that σ^2 in that reference is bounded by $n\sigma^2$ in our notation) we have

$$E\left\|\sum_{i=1}^n f(Z_i)\right\|_{\mathcal{F}}^2 \leq K^2\left[\left(E\left\|\sum_{i=1}^n f(Z_i)\right\|_{\mathcal{F}}\right)^2 + 2n\sigma^2 + 4U^2\right],$$

where K is a universal constant. We then bound the first term on the right-hand side by using Proposition 3 in Giné and Nickl (2009a) and simplify the resulting bound using (56), $A > e$, and $U/\sigma \geq 1$ to arrive at the result. ■

References

Adams, R. A. & J. J. F. Fournier (2003): *Sobolev Spaces.* 2nd edition, Elsevier.

Altissimo, F. & A. Mele (2009): Simulated nonparametric estimation of dynamic models. *Review of Economic Studies*, forthcoming.

Beran, R. (1977): Minimum Hellinger distance estimates for parametric models. *Annals of Statistics* 5, 445-463.

Bickel, P. & Y. Ritov (2003): Nonparametric estimators that can be 'plugged-in'. *Annals of Statistics* 31, 1033-1053.

Carrasco, M., M. Chernov, J. P. Florens & E. Ghysels (2007): Efficient estimation of general dynamic models with a continuum of moment conditions. *Journal of Econometrics* 140, 529-573.

DeVore, R. A. & G. G. Lorentz (1993): *Constructive Approximation.* Springer-Verlag.

Donoho, D. L. & R. C. Liu (1988): The "automatic" robustness of minimum distance functionals. *Annals of Statistics* 16, 552-586.

Fermanian, J. D. & B. Salanié (2004): A nonparametric simulated maximum likelihood estimation method. *Econometric Theory* 20, 701-734.

Folland, G. (1999): *Real Analysis: Modern Techniques and Their Applications,* 2nd edition, Wiley.

Gach, F. (2009): *Efficiency in Indirect Inference.* PhD Thesis, University of Vienna.

Gallant, R. & G. Tauchen (1996): Which moments to match? *Econometric Theory* 12, 657-681.

Gallant R. & J. Long (1997): Estimating stochastic differential equations efficiently by minimum chi-squared. *Biometrika* 84, 125-141.

Giné, E. & V. Koltchinskii (2006): Concentration inequalities and asymptotic results for ratio type empirical processes. *Annals of Probability* 34, 1143-1216.

Giné, E., R. Latała & J. Zinn (2000): Exponential and moment inequalities for U-statistics. In: Giné, E., Mason, D. M., Wellner, J. A. (eds.): High-dimensional Probability II, Progress in Probability 47, 13-38.

Giné, E. & R. Nickl (2008): Uniform central limit theorems for kernel density estimators. *Probability Theory and Related Fields* 141, 333-387.

Giné, E. & R. Nickl (2009a): An exponential inequality for the distribution function of the kernel density estimator, with applications to adaptive estimation. *Probability Theory and Related Fields*, forthcoming.

Giné, E. & R. Nickl (2009b): Uniform limit theorems for wavelet density estimators. *Annals of Probability*, forthcoming.

Gourieroux, C., A. Monfort, & E. Renault (1993): Indirect inference. *Journal of Applied Econometrics* 8, 85-118.

Gourieroux, C. & A. Monfort (1996): *Simulation-based econometric methods.* Oxford University Press.

Huber, P. J. (1972): Robust statistics: A review. *Annals of Mathematical Statistics* 43, 1041-1067.

Jiang, W. & B. Turnbull (2004): The indirect method: Inference based on intermediate statistics – A synthesis and examples. *Statistical Science* 19, 239-263.

Lindsay (1994): Efficiency versus robustness: The case for minimum Hellinger distance and related methods. *Annals of Statistics* 22, 1081-1114.

Lorentz, G. G., v.Golitschek, M. & Y. Makovoz (1996): *Constructive Approximation: Advanced Problems.* Springer-Verlag.

Millar, P. W. (1981): Robust estimation via minimum distance methods. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete* 55, 73-89.

Nickl, R. (2007): Donsker-type theorems for nonparametric maximum likelihood estimators. *Probability Theory and Related Fields* 138, 411-449.

Pötscher, B. M. & I. R. Prucha (1997): *Dynamic Nonlinear Econometric Models: Asymptotic Theory.* Springer-Verlag.

Shadrin, A. Yu. (2001): The L_∞-norm of the L_2-spline projector is bounded independently of the knot sequence: a proof of de Boor's conjecture. *Acta Mathematica* 187, 59-137.

Smith, A. (1993): Estimating nonlinear time-series models using simulated vector autoregressions. *Journal of Applied Econometrics* 8, 63-84.

van der Vaart, A. W. & J. A. Wellner (1996): *Weak Convergence and Empirical Processes With Applications to Statistics.* Springer-Verlag.