Invariance Principles for Dependent Processes Indexed by Besov Classes with an Application to a Hausman Test for Linearity

Guido M. Kuersteiner*
University of Maryland

Abstract

This paper considers functional central limit theorems for stationary absolutely regular mixing processes. Bounds for the entropy with bracketing are derived using recent results in Nickl and Pötscher (2007). More specifically, their bracketing metric entropy bounds are extended to a norm defined in Doukhan, Massart and Rio (1995, henceforth DMR) that depends both on the marginal distribution of the process and on the mixing coefficients. Using these bounds, and based on a result in DMR, it is shown that for the class of weighted Besov spaces polynomially decaying tail behavior of the function class is sufficient to obtain a functional central limit theorem under minimal conditions. A second class of functions that allow for a functional central limit theorem under minimal conditions are smooth functions defined on bounded sets. Similarly, a functional CLT for polynomially explosive tail behavior is obtained under additional moment conditions that are easy to check. An application to a Hausman specification test illustrates the theory.

Keywords: dependent process, empirical process, mixing, Besov classes, Hausman test

*University of Maryland, Department of Economics, Tydings Hall 3145, 7343 Prinkert Dr., College Park, MD 20742, USA. email: kuersteiner@econ.umd.edu; http://econweb.umd.edu/~kuersteiner/

1 Introduction

Central limit theorems for empirical processes defined on dependent data and indexed by smooth classes of functions are being considered. Doukhan, Massart and Rio (1994)**?** and Doukhan, Massart and Rio (1995)**?** (henceforth DMR) are landmark contributions in this literature. The key insight from those papers is that a specific norm that combines dependence properties and the marginal distribution of the process provides the appropriate measure to assess the complexity of the function class in terms of bracketing entropy. However, as pointed out by Rio (1998, 2013)**??** the results of DMR are not minimal in the sense of providing convergence under dependence assumptions equivalent to finite dimensional cases. In fact, for a β-mixing process with mixing coefficients β_m, central limit theorems can be established under the minimal condition that $\sum_{m=0}^{\infty} \beta_m < \infty$. Rio (1998, 2013)**?** shows that such minimal results are possible in some cases involving VC classes as well as certain Lipschitz type functions. In this paper the function classes for which such minimal results are possible are expanded to smooth classes of rapidly asymptoting functions as well as function classes defined on a bounded set. This is achieved by directly employing recent results of complexity measures for weighted Besov spaces in Haroske and Triebel (2005)**?** and Nickl and Pötscher (2007)**?**. In addition to these improvements over the existing literature the paper also gives a number of explicit results that relate dependence properties of the underlying process to smoothness properties of the indexing function class.

Separate results then need to be employed to arrive at explicit central limit theorems. This is particularly relevant for dependent data where there is a potentially complex interaction between the properties of the function class, dependence of the process and properties of the marginal distribution of the process. An additional requirement, especially in econometric applications, is that function spaces be defined on unbounded sets, typically $\mathbb{R}^d$. This further limits applicability of many results available in the iid literature.

Andrews (1991)**?** has given similar results under related conditions but essentially under the assumption of function classes restricted to a bounded domain. Nickl (2007)**?** mentions the possibility of obtaining explicit empirical process central limit theorems for the dependent case using the approach pursued here but does not give such results. A useful by-product of obtaining empirical central limit theorems for specific function classes are stochastic equicontinuity results for these function classes. This fact is exploited in the part of the paper that develops a Hausman specification test for linearity of the conditional mean.

Empirical central limit theorems have a long history in probability and have found wide applications in statistics. Early results are due to Dudley (1978, 1984)**??** and Pollard (1982)**?**. General results for iid data using bracketing were obtained by Ossiander (1987)**?** and Pollard (1989)**?** and based on Vapnik-Cervonenkis (VC) classes by Pollard (1990)**?**. Early results for dependent processes include Berkes and Phillip (1977)**?** generalizing Donsker's theorem to strongly mixing stationary sequence. Uniform CLT's over function classes for dependent processes were studied in Doukhan, Leon and Portal (1987)**?**, Massart (1987)**?**, Andrews (1991)**?**, Andrews and Pollard (1994)**?** and Hansen (1996)**?**. Arcones and Yu (1994)**?** consider absolutely regular processes indexed by VC classes. A very influential paper is Doukhan, Massart and Rio (1995) which considers absolutely regular processes under a bracketing condition, extending results from Ossiander to the dependent case.

The paper is organized as follows. Section 2 presents definitions of smooth function classes and measures of dependence and presents the main results of the paper. Section 3 contains a detailed comparison with other related results in the literature. An application to the problem of testing for a linear conditional mean using a Hausman test is given in Section 4. Proofs are collected in the appendix in Section A.

2 A Functional CLT for Dependent Processes

Let the sequence χ_t be (measurable) random variables defined on the probability space $(\Omega, \mathcal{A}, \mathbb{P})$. Assume that $\{\chi_t\}_{t=-\infty}^{\infty}$ is strictly stationary with values in the measurable space $\left(\mathbb{R}^d, \mathcal{B}^d\right)$ where $\mathcal{B}^d$ is the Borel σ-field on $\mathbb{R}^d$ and $d \in \mathbb{N}_+$. Let $\mathcal{A}^l = \sigma\left(\chi_t : t \leq l\right)$ be the sigma field generated by $, ...\chi_{l-1}, \chi_l$ and $\mathcal{D}^l = \sigma\left(\chi_t : t \geq l\right)$. Following DMR, p.379 the absolutely regular mixing coefficient β_m is defined as

$$2\beta_m = \sup \sum_{(i,j) \in I \times J} \left|\mathbb{P}\left(A_i \cap D_j\right) - \mathbb{P}\left(A_i\right)\mathbb{P}\left(D_j\right)\right|$$

where the supremum is taken over all finite partitions A_i and D_j of $\mathcal{A}^0$ and $\mathcal{D}^m$. The definition of β_m is due to Volkonski and Rozanov (1959)**?** who give an alternative equivalent formulation that is sometimes used in the literature (see for example Arcones and Yu, 1994).**?** Strong mixing is defined as

$$\alpha_m = \sup_{(A,D) \in \mathcal{A}^0 \times \mathcal{D}^m} \left|\mathbb{P}\left(D \cap A\right) - \mathbb{P}\left(A\right)\mathbb{P}\left(D\right)\right|,$$

and φ-mixing is based on

$$\varphi_m = \sup_{(A,D) \in \mathcal{A}^0 \times \mathcal{D}^m} \left|\mathbb{P}\left(D|A\right) - \mathbb{P}\left(D\right)\right|$$

and the relationship $2\alpha_m \leq \beta_m \leq \varphi_m \leq 1$ holds. The condition

$$\sum_{m=0}^{\infty} \beta_m < \infty \tag{1}$$

is frequently imposed in what follows.

Define the Euclidian norm for a real valued matrix or vector A as $\|A\|^2 = \operatorname{tr} AA'$. Let $\chi \subseteq \mathbb{R}^d$ be a non-empty Borel set. Define the sup-norm $\|f\|_\infty = \sup_{x\in\chi} |f(x)|$ for any measurable function $f : \mathcal{X} \to \mathbb{R}$. Similarly, for $r \geq 1$ let $\|f\|_{r,P} = \left(\int |f(x)|^r dP(x)\right)^{1/r}$ where P is the marginal distribution of χ_t and let $\mathcal{L}_r(P)$ be the set of functions with $\|f\|_{r,P} < \infty$. The following definitions are given in Rio (1993)**?** and DMR. For a nonincreasing function $h : \mathbb{R} \to \mathbb{R}$ define the inverse $h^{-1}(u) = \inf\{t : h(t) \leq u\}$. Let $Q_f(u)$ be the quantile function defined as the inverse of the tail probability $P(|f(\chi_t)| > t)$. Let $\lfloor t \rfloor$ be the largest integer smaller or equal to $t \in \mathbb{R}$ and define $\beta^{-1}(u) = \inf\left\{t : \beta_{\lfloor t \rfloor} \leq u\right\}$. Now define the norm

$$\|f\|_{2,\beta} = \sqrt{\int_0^1 \beta^{-1}(u)\,(Q_f(u))^2\, du} < \infty.$$

DMR, Lemma 1, show that if (1) holds, the set $\mathcal{L}_{2,\beta}(P)$ of functions with $\|f\|_{2,\beta} < \infty$ equipped with the norm $\|.\|_{2,\beta}$ is a normed subspace of $\mathcal{L}_2(P)$ and that $\|f\|_{2,P} \leq \|f\|_{2,\beta}$.

Consider the class of functions $\mathcal{F}$ with elements $f : \mathcal{X} \to \mathbb{R}$. For a sample $\{\chi_t\}_{t=1}^n$ define the empirical process

$$v_n(f) = n^{1/2} \sum_{t=1}^n \left(f(\chi_t) - E[f(\chi_t)]\right).$$

When (1) is satisfied, Rio (1993, Theorem 1.2) shows that for $f \in \mathcal{L}_{2,\beta}(P)$,

$$\sum_{t=-\infty}^{\infty} |\operatorname{Cov}(f(\chi_0), f(\chi_t))| \leq 4\|f\|_{2,\beta}^2$$

and for $\Gamma(f,f) = \sum_{t=-\infty}^{\infty} \operatorname{Cov}(f(\chi_0), f(\chi_t))$ it follows that

$$\lim_{n\to\infty} \operatorname{Var}(v_n(f)) = \Gamma(f,f) \leq 4\|f\|_{2,\beta}^2.$$

Following DMR and van der Vaart and Wellner (1996, p.83)**?** let $\mathcal{F}$ be a subset of a normed space $(V, \|\|_V)$ of functions $f : \chi \to \mathbb{R}$ with norm $\|\|_V$. For any pair of functions, $l, u \in \mathcal{F}$ and $\delta > 0$, the set $[l, u] \subset \mathcal{F}$ is a δ-bracket if $l \leq u$ with $\|l - u\|_V \leq \delta$ and for all $f \in [l, u]$ it follows that $l \leq f \leq u$. The bracketing number $N_{[]}(\delta, \mathcal{F}, \|\|_V)$ is the smallest number of δ-brackets needed to cover $\mathcal{F}$. The entropy with bracketing is the logarithm of $N_{[]}(\delta, \mathcal{F}, \|\|_V)$ denoted by $H_{[]}(\delta, \mathcal{F}, \|\|_V)$.

DMR show in Theorem 1 that if χ_t is a strictly stationary β-mixing sequence with (1) holding, marginal distribution P and $\mathcal{F} \subset \mathcal{L}_{2,\beta}(P)$ such that

$$\int_0^1 \sqrt{H_{[]}\left(\delta, \mathcal{F}, \|\|_{2,\beta}\right)} d\delta < +\infty \tag{2}$$

then the finite dimensional vector $v_n(f_1), ..., v_n(f_k)$ converges weakly,

$$(v_n(f_1), ..., v_n(f_k)) \rightarrow^d (v(f_1), ..., v(f_k)), \tag{3}$$

where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths and the asymptotic equicontinuity condition holds for every $\epsilon > 0$:

$$\lim_{\delta \to 0} \limsup_{n \to \infty} \mathbb{P}^* \left(\sup_{\|f-g\|_{2,\beta} \leq \delta,\ f,g \in \mathcal{F}} |v_n(f) - v_n(g)| > \epsilon \right) = 0, \tag{4}$$

where $\mathbb{P}^*$ is outer probability. The short hand notation $v_n(f) \rightsquigarrow v(f)$ is used when both (3) and (4) hold.

Besov spaces are now defined as in Nickl and Pötscher (2007, Remark 2). For Lebesgue measure λ let $\mathcal{L}_p\left(\mathbb{R}^d, \lambda\right)$ be the set of all functions $f : \mathbb{R}^d \to \mathbb{R}$ with $\|f\|_{p,\lambda} = \left(\int |f(x)|^p dx\right)^{1/p} < \infty$. Let $\alpha = (\alpha_1, ..., \alpha_d)$ be a multi index of non-negative integers α_i, with $|\alpha| = \sum_{i=1}^d \alpha_i$ and let D^α denote the partial differential operator $\partial^{|\alpha|} / \left((\partial x_1)^{\alpha_1} ... (\partial x_d)^{\alpha_d}\right)$ of order $|\alpha|$ in the sense of distributions - see Stein (1970, p. 121).**?** For a function $f : \mathbb{R}^d \to \mathbb{R}$ the difference operator Δ_z is defined as $\Delta_z f(.) = f(. + z) - f(.)$ and $\Delta_z^2 f(.) = \Delta_z(\Delta_z f(.))$ for $z \in \mathbb{R}^d$. Let $0 < s < \infty$ and set $s = [s]^- + \{s\}^+$ where $[s]^-$ is integer and $0 < \{s\}^+ \leq 1$. For example, when $s = 1$, $\{s\}^+ = 1$ and $[s]^- = 0$. Let $1 \leq p \leq \infty$ and $1 \leq q \leq \infty$. For $f \in \mathcal{L}_p\left(\mathbb{R}^d, \lambda\right)$ with $\|D^\alpha f\|_{p,\lambda} < \infty$ and for $0 \leq \alpha \leq [s]^-$ define

$$\|f\|_{s,p,q,\lambda}^* = \sum_{0 \leq \alpha \leq [s]^-} \|D^\alpha f\|_{p,\lambda} + \sum_{\alpha = [s]^-} \left(\int_{\mathbb{R}^d} |z|^{-\{s\}^+ q - d} \left\|\Delta_z^2 D^\alpha f\right\|_{p,\lambda}^q dz \right)^{1/q}$$

for $q < \infty$, and for $q = \infty$ define

$$\|f\|_{s,p,\infty,\lambda}^* = \sum_{0 \leq \alpha \leq [s]^-} \|D^\alpha f\|_{p,\lambda} + \sum_{\alpha = [s]^-} \sup_{0 \neq z \in \mathbb{R}^d} |z|^{-\{s\}^+} \left\|\Delta_z^2 D^\alpha f\right\|_{p,\lambda}.$$

Then the Besov space $\mathcal{B}_{pq}^s\left(\mathbb{R}^d\right)$ is defined as $\mathcal{B}_{pq}^s\left(\mathbb{R}^d\right) = \left\{ f \in \mathcal{L}_p\left(\mathbb{R}^d, \lambda\right) : \|f\|_{s,p,q,\lambda}^* < \infty \right\}$. An equivalent definition can be given in terms of Fourier transforms F acting on the space of complex tempered distributions on $\mathbb{R}^d$ (see Edmunds and Triebel, 1996, 2.2.1)**?**. Denote by F^{-1} the inverse of F. Let $\varphi_0(x)$ be a complex valued C^∞-function on $\mathbb{R}^d$ with $\varphi_0(x) = 1$

if $\|x\| \leq 1$ and $\varphi_0(x) = 0$ if $\|x\| \geq 3/2$. Define $\varphi_1(x) = \varphi_0(x/2) - \varphi_0(x)$ and $\varphi_k(x) = \varphi_1(2^{-k+1}x)$ for $k \in \mathbb{N}$. Let $0 \leq s < \infty$, $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, with $q = 1$ if $s = 0$. For $f \in \mathcal{L}_p(\mathbb{R}^d, \lambda)$ and $q < \infty$ define

$$\|f\|_{s,p,q,\lambda} = \left(\sum_{k=0}^{\infty} 2^{ksq} \left\| F^{-1}(\varphi_k F f) \right\|_{p,\lambda}^q \right)^{1/q}$$

and for $q = \infty$

$$\|f\|_{s,p,\infty,\lambda} = \sup_{0 \leq k < \infty} 2^{ks} \left\| F^{-1}(\varphi_k F f) \right\|_{p,\lambda}.$$

Then, it follows (see Nickl and Pötscher, 2007, p. 180) that

$$\mathcal{B}^s_{pq}\left(\mathbb{R}^d\right) = \left\{ f \in \mathcal{L}_p\left(\mathbb{R}^d, \lambda\right) : \|f\|_{s,p,q,\lambda} < \infty \right\}$$

and the norms $\|f\|^*_{s,p,q,\lambda}$ and $\|f\|_{s,p,q,\lambda}$ are equivalent on $\mathcal{B}^s_{pq}(\mathbb{R}^d)$. Define $\langle x \rangle = 1 + \|x\|^2$. Weighted Besov spaces are now defined as in Edmunds and Triebel (1996, 4.2) and Nickl and Pötscher (2007, p.181) for $\vartheta \in \mathbb{R}$ as

$$\mathcal{B}^s_{pq}\left(\mathbb{R}^d, \vartheta\right) = \left\{ f : \left\| f(.) \langle x \rangle^{\vartheta/2} \right\|_{s,p,q,\lambda} < \infty \right\}.$$

For $s > d/p$ or $s = d/p$ with $q = 1$ define

$$B^s_{pq}\left(\mathbb{R}^d, \vartheta\right) = \mathcal{B}^s_{pq}\left(\mathbb{R}^d, \vartheta\right) \cap \left\{ f : f(.) \langle x \rangle^{\vartheta/2} \in C\left(\mathbb{R}^d\right) \right\}$$

where $C(\mathbb{R}^d)$ is the vector space of bounded continuous real valued functions on $\mathbb{R}^d$ with the sup-norm $\|.\|_\infty$. Nickl and Pötscher (2007, Proposition 3) show that $f \in B^s_{pq}(\mathbb{R}^d)$ implies that f is bounded and if $p < \infty$ it also follows that $\lim_{\|x\| \to \infty} f(x) = 0$. These restrictions do not necessarily apply when $f \in B^s_{pq}(\mathbb{R}^d, \vartheta)$ and $\vartheta < 0$. This feature of weighted spaces is important for applications in econometrics, as will be demonstrated in Section 4.

The following result gives upper bounds for entropy with bracketing on the normed space $\mathcal{L}_{2,\beta}(P)$. It extends Theorem 1 of Nickl and Pötscher (2007) to the space $\mathcal{L}_{2,\beta}(P)$ which plays a crucial role in obtaining a functional CLT for dependent processes.

Theorem 1 *Assume that* $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\vartheta \in \mathbb{R}$ *and* $s > d/p$. *Further assume that* $\mathcal{F} \subset B^s_{pq}(\mathbb{R}^d, \vartheta)$ *is nonempty and bounded. If* $\vartheta > 0$ *then*

$$H_{[]}\left(\delta, \mathcal{F}, \|\|_{2,\beta}\right) \precsim \begin{cases} \delta^{-d/s} & \text{if } \vartheta > s - d/p \\ \delta^{-(\vartheta/d + 1/p)^{-1}} & \text{if } \vartheta < s - d/p \end{cases}.$$

If $\vartheta \leq 0$ and if for some $\gamma > 0$ it holds that

$$\left\| \langle \chi_t \rangle^{(\gamma-\vartheta)/2} \right\|_{2,\beta} < \infty$$

then it follows that

$$H_{[]}\left(\delta, \mathcal{F}, \|\|_{2,\beta}\right) \precsim \begin{cases} \delta^{-d/s} & \text{if } \gamma > s - d/p \\ \delta^{-(\gamma/d+1/p)^{-1}} & \text{if } \gamma < s - d/p \end{cases}.$$

The difference between Nickl and Pötscher (2007, Theorem 1) and Theorem 1 is that bracketing is with respect to the norm $\|.\|_{2,\beta}$ rather than the conventional $\|.\|_{r,P}$ norm on $\mathcal{L}_r\left(\mathbb{R}^d, P\right)$. The result obtained here directly leads to a functional CLT based on the results of DMR. A corollary to Theorem 1 is obtained for the case when the function space $\mathcal{F}$ is restricted to a bounded domain $\mathfrak{X}$. Then the following result applies.

Corollary 2 *Let $\mathfrak{X} \subset \mathbb{R}^d$ and there exists a finite M with $\langle x \rangle \leq M$ for all $x \in \mathfrak{X}$. Assume that $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\vartheta \in \mathbb{R}$ and $s > d/p$. Further assume that $\mathcal{F} \subset B^s_{pq}(\mathfrak{X}, \vartheta)$ is nonempty and bounded. Then,*

$$H_{[]}\left(\delta, \mathcal{F}, \|\|_{2,\beta}\right) \precsim \begin{cases} \delta^{-d/s} & \text{if } \vartheta > s - d/p \\ \delta^{-(\vartheta/d+1/p)^{-1}} & \text{if } \vartheta < s - d/p \end{cases}.$$

The bounds on bracketing numbers obtained in Theorem 1 and Corollary 2 can now be applied to obtain a functional central limit theorem based on Theorem 1 of DMR. The proof is based on using the tail decay properties of weighted function spaces to establish that $\mathcal{F} \subset \mathcal{L}_{2,\beta}(P)$. This property is satisfied without further assumptions on the marginal distribution of χ_t if $\vartheta > 0$.

Theorem 3 *Let χ_t be a strictly stationary and β-mixing process. Assume that (1) holds. Assume that $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\vartheta \in \mathbb{R}$ and $s > d/p$. Further assume that $\mathcal{F} \subset B^s_{pq}\left(\mathbb{R}^d, \vartheta\right)$ is nonempty and bounded. Assume that one of the following conditions hold: (i) $\vartheta > 0$, $\vartheta > s - d/p$ and $s/d > 1/2$; (ii) $\vartheta > 0$, $\vartheta < s - d/p$ and $\vartheta/d + 1/p > 1/2$; (iii) $\vartheta \leq 0$ and for some $\gamma > 0$ it follows that $\left\| \langle \chi_t \rangle^{(\gamma-\vartheta)/2} \right\|_{2,\beta} < \infty$, $\gamma > s - d/p$ and $s/d > 1/2$; (iv) $\vartheta \leq 0$ and for some $\gamma > 0$ it follows that $\left\| \langle \chi_t \rangle^{(\gamma-\vartheta)/2} \right\|_{2,\beta} < \infty$, $\gamma < s - d/p$ and $\gamma/d+1/p > 1/2$. Then, $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

We note that the conditions $1/2 < s/d$ and $1/2 < \gamma/d + 1/p$ are the same as the conditions given in Corollary 5 of Nickl and Pötscher (2007) for the iid case. Here these conditions need to hold in conjunction with bounds on the β-mixing coefficients and the moment condition in (5).

Theorem 3 shows that an empirical process CLT can be obtained under the minimal Condition (1) if $\mathcal{F}$ is a space of functions that asymptote to zero rapidly enough, measured by the parameter $\vartheta > 0$. If the decay is rapid enough relative to smoothness as in case (i) then the functional CLT holds under the minimal condition $s/d > 1/2$. Even in case (ii) one still obtains a result with only Condition (1) imposed on the dependence of the process. An immediate corollary obtains for the case where χ_t takes values in a bounded set $\mathfrak{X} \subset \mathbb{R}^d$.

Corollary 4 *Let χ_t be strictly stationary and β-mixing. Assume that $P(\chi_t \in \mathfrak{X}) = 1$ for a bounded Borel set $\mathfrak{X} \subset \mathbb{R}^d$ and there exists a finite M with $\langle x \rangle \leq M$ for all $x \in \mathfrak{X}$. Assume that (1) holds. Assume that $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\vartheta \in \mathbb{R}$ and for $s, d < \infty$ and $s > d/p$. Further assume that $\mathcal{F} \subset B^s_{pq}(\mathfrak{X})$ is nonempty and bounded. Assume that $s/d > 1/2$. Then, $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

When the asymptotic behavior of f as $\|x\| \to \infty$ is proportional to $\langle \chi_t \rangle^{-\vartheta/2}$ and $\vartheta \leq 0$, then more restrictive conditions on the dependence need to be imposed. This happens implicitly through the condition

$$\left\| \langle \chi_t \rangle^{(\gamma-\vartheta)/2} \right\|_{2,\beta} < \infty \tag{5}$$

which must hold for some $\gamma > 0$. The advantage of this condition is that it only involves the marginal distribution of χ_t and not the properties of the functional class, other than through the parameter ϑ. Results in DMR can be used to give simple sufficient conditions for 5. Under additional assumptions about the order of β_m and moment restrictions on the marginal distribution of $\|\chi_t\|^2$ the following result can be given for the case when $\vartheta \leq 0$, i.e. when $\lim_{\|x\|} f(x) \to 0$ does necessarily not hold.

Theorem 5 *Let χ_t be strictly stationary and β-mixing. Assume that for some $r > 1$, $\sum_{m=1}^{\infty} m^{1/(r-1)} \beta_m < \infty$ holds. Assume that $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\vartheta \in \mathbb{R}$, $\vartheta \leq 0$ and $s > d/p$. Further assume that $\mathcal{F} \subset B^s_{pq}\left(\mathbb{R}^d, \vartheta\right)$ is nonempty and bounded. Assume that for some $\gamma > 0$ such that $(r(\gamma - \vartheta)) > 1$ it holds that either (i) $\gamma > s - d/p$ and $s/d > 1/2$ or (ii) $\gamma < s - d/p$ and $\gamma/d + 1/p > 1/2$, and that $E\left[\|\chi_t\|^{2r(\gamma-\vartheta)}\right] < \infty$. Then, $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

The form of the last theorem is particularly useful when a comparison with other result in the literature is desired, since those results are often presented in terms of moment bounds and size restrictions on mixing coefficients.

More generally, the results show that in weighted Besov spaces control over tail behavior of the function class can be utilized to give sufficient conditions for a CLT that directly involve the marginal distribution of χ_t rather than that of $f(\chi_t)$. This is possible because the asymptotic behavior of $f(\chi_t)$ is controlled by terms that are functions of $\|\chi_t\|$. The next corollary gives explicit versions of this result for Sobolev, Hölder and Lipschitz classes of functions.

A special case of Besov spaces are Sobolev spaces. They are defined as follows (see Nickl and Pötscher, 2007, Section 3.3.2). Let $1 < p < \infty$, real $s \geq 0$ and

$$\mathcal{H}_p^s\left(\mathbb{R}^d\right) = \left\{f \in \mathcal{L}_p\left(\mathbb{R}^d, \lambda\right) : \|f\|_{s,p,\lambda} \equiv \left\|F^{-1}\left(\langle x \rangle^s Ff\right)\right\|_{p,\lambda} < \infty\right\}$$

where the norms are formulated in terms of the Fourier transform F. When $s \geq 0$ is integer, an equivalent norm on $\mathcal{H}_p^s\left(\mathbb{R}^d\right)$ is given by

$$\|f\| = \sum_{0 \leq |\alpha| \leq s} \|D^\alpha f\|_{p,\lambda}.$$

Similar as before define the Banach space $H_p^s\left(\mathbb{R}^d\right)$ of continuous functions for $s > d/p$ as

$$H_p^s\left(\mathbb{R}^d\right) = \mathcal{H}_p^s\left(\mathbb{R}^d\right) \cap \left\{f :\in C\left(\mathbb{R}^d\right)\right\}.$$

The space of weighted Sobolev functions is given by

$$H_p^s\left(\mathbb{R}^d, \vartheta\right) = \left\{f : f(.)\langle x \rangle^{\vartheta/2} \in H_p^s\left(\mathbb{R}^d\right)\right\}.$$

The following Corollary is a special case of Theorem 3. The proof follows in the same way as the proofs of similar corollaries in Nickl and Pötscher (2007) by arguing that bounded subsets of $H_p^s\left(\mathbb{R}^d, \vartheta\right)$ are also bounded subsets of $B_{p\infty}^s\left(\mathbb{R}^d, \vartheta\right)$.

Corollary 6 *Let χ_t be a strictly stationary and β-mixing process. Assume that (1) holds. Assume that $1 < p \leq \infty$, $\vartheta \in \mathbb{R}$ and $s > d/p$. Further assume that $\mathcal{F} \subset H_p^s\left(\mathbb{R}^d, \vartheta\right)$ is nonempty and bounded. Assume that one of the following conditions hold: (i) $\vartheta > 0$, $\vartheta > s - d/p$ and $s/d > 1/2$; (ii) $\vartheta > 0$, $\vartheta < s - d/p$ and $\vartheta/d + 1/p > 1/2$; (iii) $\vartheta \leq 0$ and for some $\gamma > 0$ it follows that $\left\|\langle \chi_t \rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta} < \infty$, $\gamma > s - d/p$ and $s/d > 1/2$; (iv) $\vartheta \leq 0$ and for some $\gamma > 0$ it follows that $\left\|\langle \chi_t \rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta} < \infty$, $\gamma < s - d/p$ and $\gamma/d + 1/p > 1/2$. Then, $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

The following Corollary again considers the special case where the domain of the function space is a bounded subset of $\mathbb{R}^d$.

Corollary 7 *Let χ_t be a strictly stationary and β-mixing process. Assume that $P(\chi_t \in \mathfrak{X}) = 1$ where $\mathfrak{X} \subset \mathbb{R}^d$ and there exists a finite M with $\langle x \rangle \leq M$ for all $x \in \mathfrak{X}$. Assume that (1) holds. Assume that $1 < p \leq \infty$, $\vartheta \in \mathbb{R}$ and $s, d < \infty$ with $s > d/p$. Further assume that $\mathcal{F} \subset H_p^s(\mathfrak{X},\vartheta)$ is nonempty and bounded. Assume that $s/d > 1/2$. Then, $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

For $s > 0$, s not integer, the Hölder space is the space $C^s(\mathbb{R}^d)$ of all $\lfloor s \rfloor$-times differentiable functions f with finite norm

$$\|f\|_{s,\infty} = \sum_{0 \leq |\alpha| \leq \lfloor s \rfloor} \|D^\alpha f\|_\infty + \sum_{|\alpha| = \lfloor s \rfloor} \sup_{x \neq y} \frac{|D^\alpha f(x) - D^\alpha f(y)|}{|x - y|^{s - \lfloor s \rfloor}}.$$

The weighted space $C^s(\mathbb{R}^d, \vartheta)$ is given by

$$C^s\left(\mathbb{R}^d, \vartheta\right) = \left\{ f : \left\| f(.) \langle x \rangle^{\vartheta/2} \right\|_{s,\infty} < \infty \right\}.$$

Related is the Zygmund space $\mathcal{C}^s(\mathbb{R}^d)$ for $s > 0$ defined in Triebel (1983, p.36) or Triebel (1992, p.4). Let

$$\|f\|_{s,\infty}^z = \sum_{0 \leq |\alpha| \leq [s]^-} \|D^\alpha f\|_\infty + \sum_{|\alpha| = [s]^-} \sup_{0 \neq z \in \mathbb{R}^d} |z|^{-\{s\}^+} \left\| \Delta_z^2 D^\alpha f \right\|_\infty .$$

By Triebel (1992, p.5), $\mathcal{C}^s(\mathbb{R}^d, \vartheta) = C^s(\mathbb{R}^d, \vartheta)$ when $s > 0$ and s is not integer. The space $C^s(\mathfrak{X})$ is considered by van der Vaart and Wellner (1996, p. 154) under the additional constraint that $\|f\|_{s,\infty} \leq M$ for some bounded constant M. As noted there, when $0 < s < 1$, $C^s(\mathfrak{X})$ contains the Lipschitz functions (see Adams and Fournier 2003, Theorem 1.34)**?**. The following corollaries specialize previous results to Hölder spaces.

Corollary 8 *Let χ_t be strictly stationary and β-mixing. Assume that (1) holds. Assume that $\vartheta \in \mathbb{R}$ and $s > d/2$. Further assume that $\mathcal{F} \subset C^s(\mathbb{R}^d, \vartheta)$ is nonempty and bounded. Assume that one of the following conditions hold: (i) $\vartheta > 0$, $\vartheta > s$ and $s/d > 1/2$; (ii) $\vartheta > 0$, $\vartheta < s$ and $\vartheta/d > 1/2$; (iii) $\vartheta \leq 0$ and for some $\gamma > 0$ it follows that $\left\| \langle \chi_t \rangle^{(\gamma - \vartheta)/2} \right\|_{2,\beta} < \infty$, $\gamma > s$ and $s/d > 1/2$; (iv) $\vartheta \leq 0$ and for some $\gamma > 0$ it follows that $\left\| \langle \chi_t \rangle^{(\gamma - \vartheta)/2} \right\|_{2,\beta} < \infty$, $\gamma < s$ and $\gamma/d > 1/2$. Then, $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

The proof follows again from noting that $\mathcal{F}$ is a bounded subset in $B^s_{\infty\infty}\left(\mathbb{R}^d,\vartheta\right)$, see Nickl and Pötscher (2007, p. 188). As before additional results for the cases of bounded support can be stated as follows.

Corollary 9 *Let χ_t be a strictly stationary and β-mixing. Assume that $P\left(\chi_t \in \mathfrak{X}\right) = 1$ where $\mathfrak{X} \subset\mathbb{R}^d$ and there exists a finite M with $\langle x\rangle \leq M$ for all $x \in \mathfrak{X}$. Assume that (1) holds. Assume that $\vartheta \in \mathbb{R}$, $s, d < \infty$ and $s > 0$. Further assume that $\mathcal{F} \subset C^s\left(\mathfrak{X},\vartheta\right)$ is nonempty and bounded. Assume that $s/d > 1/2$. Then, $v_n\left(f\right) \rightsquigarrow v\left(f\right)$ where $v\left(f\right)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

When $\vartheta \leq 0$ such that $\lim_{\|x\|} f\left(x\right) \rightarrow 0$ does not hold, a more specific result can be given for functions in $C^s\left(\mathbb{R}^d,\vartheta\right)$ as long as one is willing to impose additional conditions on the rate of decay of β_m. This is done in the following corollary.

Corollary 10 *Let χ_t be strictly stationary and β-mixing. Assume that for some $r > 1$, $\sum_{m=1}^{\infty} m^{1/(r-1)}\beta_m < \infty$ holds. Assume that $\vartheta \in \mathbb{R}$, $\vartheta \leq 0$ and $s > 0$. Further assume that $\mathcal{F} \subset C^s\left(\mathbb{R}^d,\vartheta\right)$ is nonempty and bounded. Assume that for some $\gamma > 0$ such that $\left(r\left(\gamma-\vartheta\right)\right) > 1$ it holds that either (i) $\gamma > s$ and $s/d > 1/2$ or (ii) $\gamma < s$ and $\gamma/d > 1/2$, and that $E\left[\|\chi_t\|^{2r(\gamma-\vartheta)}\right] < \infty$. Then, $v_n\left(f\right) \rightsquigarrow v\left(f\right)$ where $v\left(f\right)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.*

Corollary 10 should only be applied to cases where $\vartheta \leq 0$. As in previous results, when $\vartheta > 0$, the functional central limit theorem can be established under weaker assumptions.

3 Discussion and Comparison with the Literature

Andrews (1991)**?** considers the space $\mathcal{H}^s_p\left(\mathfrak{X}\right)$ where $\mathfrak{X}$ is a bounded subset of $\mathbb{R}^d$. He allows for heterogeneous near epoch dependent processes which include as special cases strong mixing stationary sequences. Since β-mixing considered here implies strong mixing the results of this paper are obtained under somewhat stronger assumptions as far as the mixing concept and stationarity requirements are concerned. On the other hand, no boundedness of $\mathfrak{X}$ is required. Andrews (1991, p.199) discusses some ways of relaxing the boundedness assumption regarding the support but does not provide a general treatment. Moreover, as pointed out by Nickl and Pötscher (2007, p. 179) it follows for $f \in H^s_p\left(\mathbb{R}^d\right)$, $\lim_{\|x\|} f\left(x\right) \rightarrow 0$ while this is not necessarily the case for $f \in H^s_p\left(\mathbb{R}^d,\vartheta\right)$ and $\vartheta < 0$.

Andrews (1991, Theorem 4 and Comment 1) obtains a functional central limit theorem for strong mixing processes of size -2, $f \in \mathcal{H}^s_2\left(\mathfrak{X}\right)$ and $s/d > 1/2$. Corollary 7 shows that, at least under the additional assumption of stationarity and β-mixing but only satisfying (1), this result can be obtained for all functions in $H_p^s(\mathfrak{X})$ with $s/d > 1/2$. Note that a β-mixing process that satisfies Condition (1) also is α-mixing with $\sum_{m=1}^{\infty} \alpha_m < \infty$ but is not necessarily α-mixing of size -2. In this sense, the conditions given here are complementary to Andrews (1991).

Andrews (1991, Comment 3) also considers the case of strong mixing processes of size -2 and Lipschitz function classes. More specifically, when $\mathfrak{X}$ is a bounded interval on $\mathbb{R}$, a functional central limit theorem holds for functions f such that $|f(x) - f(y)| \leq K |x - y|^s$ with $s \in (1/2, 1]$. By Adams and Fournier (2003, Theorem 1.34) the function class $C^s(\mathfrak{X})$ with $s \in (1/2, 1)$ contains the Lipschitz functions with $s \in (1/2, 1)$. Then, Corollary 9 can be used to establish a functional central limit theorem for Lipschitz functions and for stationary β-mixing processes that satisfy Condition (1). Note that when $s \in (1/2, 1)$ and $\mathfrak{X}$ is a bounded interval, it follows that for $d = 1$ the condition $s/d > 1/2$ is satisfied.

Rio (2013, Theorem 8.1) considers the generalized Lipschitz spaces $Lip^*\left(s, p, \mathbb{R}^d\right)$ defined in Meyer (1992)**?**. Rio (2013, Proposition 8.1) gives an equivalent norm $\|f\|_{ond}$ for functions $f \in Lip^*\left(s, p, \mathbb{R}^d\right)$. Meyer (1992, Proposition 7, p. 200) shows that every $f \in Lip^*\left(s, p, \mathbb{R}^d\right)$ is in $\mathcal{B}_{p\infty}^s\left(\mathbb{R}^d\right)$. Rio (2013, Theorem 8.1) shows that for every strongly mixing and stationary sequence with $\sum_{m=1}^{\infty} \alpha_m < \infty$, $f \in Lip^*\left(s, p, \mathbb{R}^d\right)$ with $p \in [1, 2]$, $s > d/p$ and $\|f\|_{ond} \leq a$ for some constant $a < \infty$, the empirical process $v_n(f)$ satisfies a stochastic equicontinuity condition and thus a functional central limit theorem.

The results given here complement the ones in Rio (2013). If a process is strictly stationary and β-mixing with Condition (1) and $f \in B_{pq}^s\left(\mathbb{R}^d, \vartheta\right)$ with $\vartheta > s - d/p$ then Theorem 3(i) establishes a functional CLT under the conditions that $s > d/p$ and $s/d > 1/2$. In particular, if $p = \infty$,then the FCLT holds under the minimal condition that $\vartheta > s > 0$ and $s/d > 1/2$. This case is not covered by the results in Rio (2013). To see this note that $B_{p_1\infty}^s\left(\mathbb{R}^d\right) \subset B_{p_2\infty}^{s+d/p_1-d/p_2}\left(\mathbb{R}^d\right)$ for $p_1 \leq p_2 \leq \infty$ by Triebel (1983, 2.7.1) indicating that the class $B_{p\infty}^s\left(\mathbb{R}^d\right)$ for $p > 2$, which is covered by Theorem 3, is a larger class than the one considered by Rio (2013). Further, from Haroske and Triebel (1994, 2005)**??** it follows for $\vartheta > 0, \vartheta/d < 1$, $s_1 - s_2 > 0$ and $p_1(1 - \vartheta/d) < p_2$ that $B_{p_1\infty}^{s_1}\left(\mathbb{R}^d, \vartheta\right)$ is embedded in $B_{p_2\infty}^{s_2}\left(\mathbb{R}^d\right)$. For example, when $d = 1$ the constraints $s > 1/2$, $s > 1/p$, $\vartheta < 1$ and $p_1(1 - \vartheta) < 2$ must hold for $B_{p_1\infty}^{s_1}\left(\mathbb{R}^d, \vartheta\right)$ to be embedded in $B_{p_2\infty}^{s_2}\left(\mathbb{R}^d\right)$. Thus, for Rio's results to encompass Theorem 3 one needs $p < 2/(1 - \vartheta)$. The results of Rio (2013) then cover the spaces $B_{p\infty}^s\left(\mathbb{R}^d, \vartheta\right)$ for values of $\vartheta < 1$ and values of $p \leq \infty$. However, as ϑ approaches 0, the largest value p can take approaches 2 while such a constraint does not apply to Theorem 3. On the other hand, Rio (2013) covers cases with $\vartheta = 0$ and $p \leq 2$ which can only be handled by Theorem 3 under additional moment restrictions and stronger assumptions on the β-mixing coefficients.

When $p = 2$, then $s/d > 1/2$ and $\vartheta > s - d/2$ lead to a FCLT by means of Theorem 3. This case essentially corresponds to Rio (2013) when $s - d/2$ is close to zero. By Triebel (1983, 2.7.1)**?** it follows that $B^{s_1}_{pq}\left(\mathbb{R}^d, \vartheta\right)$ is continuously embedded in $B^{s_0}_{pq}\left(\mathbb{R}^d, \vartheta\right)$ for $s_1 \geq s_0$. Thus, to apply Theorem 3 one can always choose s small enough such that $s - d/2$ is arbitrarily small and therefore ϑ can be chosen small. If $\vartheta < s - d/p$ then Theorem 3(ii) holds under the condition that $\vartheta/d > 1/2 + 1/p$ such that the CLT holds for p sufficiently large and $s/d > 1/2$.

These arguments indicate that the results in Rio (2013) are slightly sharper for the case when $p \in [1, 2]$ because of the requirement in Theorem 3 that $\vartheta > s - d/p$. In addition, by Triebel (1983, 2.3.2, Proposition 2), $B^{s}_{pq_0}\left(\mathbb{R}^d, \vartheta\right)$ is continuously embedded in $B^{s}_{pq_1}\left(\mathbb{R}^d, \vartheta\right)$ for $q_0 \leq q_1 \leq \infty$ and $p > 0$ such that $B^{s}_{pq}\left(\mathbb{R}^d\right)$ is continuously embedded in $Lip^*\left(s, p, \mathbb{R}^d\right)$. This implies that the results in Rio cover the spaces $B^{s}_{pq}\left(\mathbb{R}^d\right)$ for $p \in [1, 2]$ and $q \leq \infty$.

In summary, the results in Theorem 3 are very similar to Rio (2013) when $p \leq 2$ and the tail behavior of the function class is controlled by a polynomial. However, the results are achieved with simpler proofs. Because of the embedding result in Triebel (1983, 2.7.1), additional function classes are covered by Theorem 3 that are not contained in Rio (2013) when $p > 2$. Theorem 3 also covers cases when $\vartheta \leq 0$ and $p \leq \infty$ that are not covered by Rio (2013). However, in these situations somewhat stronger assumptions than (1) need to be imposed on dependence. Here the case $\vartheta = 0$ and $p = \infty$ may be of particular interest since the tail behavior of $f(x)$ no longer necessarily satisfies $\lim_{\|x\|} f(x) \to 0$ (see Proposition 3 of NP). This is one example of a case not covered by the results in Rio (2013).

The results in DMR are stated in general terms and form the basis for what is derived here. Nevertheless, on p.403-405 DMR provide a number of different approaches with which high level assumptions can be replaced with more primitive conditions. These methods do not lead to the sharpest possible results in terms of conditions imposed on β_m for the classes of functions considered by Rio (2013) and results given here for functions whose tail decay is well controlled by a polynomial or are restricted to a bounded domain. In particular, Theorem 3 shows that $\vartheta > 0$, i.e. when tail behavior is controlled by polynomials, the functional CLT can be obtained without requiring the additional moment bound in (5). As a result, neither the marginal distribution of χ_t nor the dependence of the process need further restrictions beyond Condition (1). On the other hand, the results in DMR lead to similar conditions as the ones given in Theorem 5 for spaces where $\vartheta \leq 0$. The following result illustrates this. By exploiting condition (2.11) in DMR and applying Theorem 1 in Nickl and Pötscher (2007) one obtains the following.

Theorem 11 *Let $1 \leq p \leq \infty$, $1 \leq q \leq \infty$, $\vartheta \in \mathbb{R}$ and $s - d/p > 0$. For $1 < r < \infty$ let χ_t be a strictly stationary, absolutely regular process such that $\sum_{m=1}^{\infty} m^{1/(r-1)}\beta_m < \infty$. For some $\gamma > 0$ it follows that*

$$\left\| \langle \chi_t \rangle^{(\gamma-\vartheta)/2} \right\|_{2r,P} < \infty \tag{6}$$

Let $\mathcal{F}$ be a bounded subset of $B^s_{pq}\left(\mathbb{R}^d, \vartheta\right)$. Further one of the conditions holds: i) $\gamma > s-d/p$ and $1/2 < s/d$ ii) $\gamma < s - d/p$ and $1/2 < \gamma/d + 1/p$. Then $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process with covariance function Γ and a.s. uniformly continuous sample paths.

The conditions of Theorem 11 are the same as given in Theorem 5 for the case when $\vartheta \leq 0$. However, the limitation of Theorem 11 over Theorems 3 and 5 is that it does not deliver a functional central limit theorem under the minimal condition (1) when $\vartheta > 0$.

4 Application: A Hausman Test for Linearity

In this section the problem of testing for linearity in the conditional expectation $E[y|x]$ of a process $\chi_t = (y_t, x_t)$ is considered. The purpose of the section is to illustrate how the central limit theory developed in this paper can be used to obtain limiting results for fairly general classes of processes and conditional mean functions. Because lineartity and unbounded domains are important ingredients to this application, the theory for weighted function spaces is particularly relevant for this application. Minimal dependence conditions in (1) could be obtained under the additional assumption that the domain of χ_t is bounded. This is an immediate consequence from results in earlier sections and is only noted in passing.

The insights underlying the Hausman test are ingenious and have found applications to a large number of testing problems in econometrics. In the particular case considered here the idea is to estimate the conditional mean by a linear regression of y_t on x_t. This estimate will not be consistent if the conditional expectation is non-linear, but is efficient, at least under additional regularity conditions, if the expectation is linear. An alternative estimator adds sieve basis functions to the linear regression. This estimator is consistent for the parameter of the linear term even if the conditional expectation is non-linear. Thus, under the null of linearity, both estimators should converge to the same parameter, however, with one of them having a smaller variance. Under the alternative only the second estimator is consistent for the linear term in a series expansion of $E[y|x]$ while the first estimator will be asymptotically biased. The Hausman test exploits these differences in asymptotic behavior by looking at the difference between the two estimators. Under the null, they have a well defined limiting distribution, while under alternatives the difference diverges, thus lending power to the test.

The estimation problem considered here is semi-parametric in nature. The distribution of the test statistic depends on the non-parametric sieve estimator employed in the second estimator. The influence function of the test statistic defines an empirical process that can be used to obtain the limiting distribution for the test statistic under the null and local alternatives. This is now formalized.

Let $\chi_t = (y_t, x_t) \in \mathbb{R}^2$ be a strictly stationary β-mixing process. Consider testing the hypothesis that $E[y_t|x_t] = \psi_0 + \psi_1 x_t$ is linear against the alternative that $E[y_t|x_t]$ is a general nonlinear function $g(x_t)$ of x_t. The problem of testing for non-linearites thus can be cast in a framework where a general model of the form

$$E[y_t|x_t] = \psi_0 + \psi_1 x_t + h(x_t)$$

with $h(x_t) = g(x_t) - \psi_1 x_t$ is estimated. A linear regression estimator for ψ_1 is generally inconsistent if $h(x_t) \neq 0$. A Hausman test is then based on the squared difference for two estimators of ψ_1. Under the null, ψ_1 is simply estimated as a regression of y_t on a constant and x_t. Under the alternative, ψ_1 is the coefficient of the linear term in a series regression of y_t on x_t. Define $P^\kappa(z) = (p_{1\kappa}(z), ..., p_{\kappa\kappa}(z))'$, where $p_{1\kappa}(z) = z$ for all κ, $\mu_P^\kappa = E[P^\kappa(z_t)]$ and $\tilde{P}^\kappa(z) = P^\kappa(z) - \mu_P^\kappa$. Define $P = [P^\kappa(x_1), ..., P^\kappa(x_n)]'$, $MP = \left[P^\kappa(x_1)^\kappa - \bar{P}^\kappa, ..., P^\kappa(x_n) - \bar{P}^\kappa\right]'$ where $M = I_n - n^{-1}\mathbf{1}_n\mathbf{1}_n'$ with $\mathbf{1}_n$ a vector of length one composed of the element one, $\bar{P}^\kappa = n^{-1}\sum_{t=1}^n P^\kappa(x_t)$. The series estimator for $E[y|x]$ is $\hat{g}(x) = \hat{\psi}_{0,\kappa} + P^\kappa(x)\hat{\psi}_\kappa$ where $\hat{\psi}_\kappa = (P'MP)^{-1} P'My$ and where $\hat{\psi}_{1\kappa}$ is the first component of $\hat{\psi}_\kappa$. The estimator for the constant is given by $\hat{\psi}_{0,j\kappa} = \bar{y} - \bar{P}^\kappa\hat{\psi}_\kappa$ with $\bar{y} = n^{-1}\sum_{t=1}^n y_t$. Partition $P = [P_1, P_2]$ where $P_1 = [x_1, ..., x_n]'$ and $P_2 = \left[(p_{2\kappa}(x_1), ..., p_{\kappa\kappa}(x_1))', ..., (p_{2\kappa}(x_n), ..., p_{\kappa\kappa}(x_n))'\right]'$. Then,

$$P'MP = \begin{bmatrix} P_1'MP_1 & P_1'MP_2 \\ P_2'MP_1 & P_2'MP_2 \end{bmatrix} := \begin{bmatrix} \hat{\Delta}_{11} & \hat{\Delta}_{12} \\ \hat{\Delta}_{21} & \hat{\Delta}_{22} \end{bmatrix}.$$

Using the partitioned inverse formula and focusing on the first component one obtains

$$\hat{\psi}_{1,\kappa} = \frac{P_1'\left(I - MP_2\hat{\Delta}_{22}^{-1}P_2'\right)My}{\left(\hat{\Delta}_{11} - \hat{\Delta}_{12}\hat{\Delta}_{22}^{-1}\hat{\Delta}_{21}\right)} \tag{7}$$

whereas the linear regression estimator is given by

$$\hat{\psi}_1 = \frac{P_1' M y}{\hat{\Delta}_{11}} = \frac{\sum_{t=1}^{n} (x_t - \bar{x}) y_t}{\hat{\Delta}_{11}}. \tag{8}$$

Let $\theta_\kappa = (\psi_1, \psi_{1,\kappa})$ and $\hat{\theta}_\kappa = \left(\hat{\psi}_1, \hat{\psi}_{1,\kappa}\right)$ where $\hat{\psi}_{1,\kappa}$ and $\hat{\psi}_1$ are given by (7) and (8) respectively. A Hausman test of linearity then compares the two estimators by forming the test statistic

$$\left(\hat{\psi}_1 - \hat{\psi}_{1,\kappa}\right)^2 / \widehat{\text{Var}} \left(\hat{\psi}_1 - \hat{\psi}_{1,\kappa}\right).$$

An alternative estimator for θ_κ is based on a Z-estimator[1] using a plug in non-parametric estimate $\hat{h}_\kappa (x_t) = M P_2 \hat{\Delta}_{22}^{-1} P_2 M y$. For this purpose define the moment function

$$\hat{m} \left(\chi_t, \theta_\kappa, \hat{h}_\kappa\right) = \begin{bmatrix} (y_t - \bar{y} - \psi_1 (x_t - \bar{x})) (x_t - \bar{x}) \\ \left(y_t - \bar{y} - \psi_{1,\kappa} (x_t - \bar{x}) - \hat{h}_\kappa (x_t)\right) (x_t - \bar{x}) \end{bmatrix} \tag{9}$$

and let

$$m_n (\theta_\kappa) = n^{-1} \sum_{t=1}^{n} \hat{m} \left(\chi_t, \theta_\kappa, \hat{h}_\kappa\right). \tag{10}$$

The Z-estimator $\tilde{\theta}_\kappa$ is obtained by solving $m_n \left(\tilde{\theta}_\kappa\right) = 0$. The limiting distribution of $\tilde{\psi}_1 - \tilde{\psi}_{1,\kappa}$ depends on non-parametric sieve estimators used in the construction of $\tilde{\psi}_{1,\kappa}$ and differs from the regression based estimators because the second component $\tilde{\psi}_{1,\kappa}$ is not estimated efficiently. The joint limiting distribution for $\tilde{\psi}_1$ and $\tilde{\psi}_{1,\kappa}$ can be analyzed in the framework of Newey (1994).**?** The following condition defines the sieve bases used for the non-parametric estimate $\hat{h}$.

Condition 1 *The functions $p_{j\kappa} (x_t) \in \mathcal{L}_2 (P)$ for all $j \leq \kappa$ and all κ. Define the closed span*

$$\mathcal{G} = \overline{\text{sp}} \{p_{j\kappa} (x_t), 1 \leq j \leq \kappa, \kappa \in \mathbb{N}_+\}$$

*as the smallest closed subset of $\mathcal{L}_2 (P)$ that contains all $p_{j\kappa} (x_t)$ (see Brockwell and Davis, 1991, p.54**?**). Further define the closed span*

$$\mathcal{G}_1 = \overline{\text{sp}} \{p_{j\kappa} (x_t), 2 \leq j \leq \kappa, \kappa \in \mathbb{N}_+\}$$

as the subspace that excludes the linear component $p_{1\kappa} (x_t)$.

[1] This terminology appreas for example in van der Vaart (1998, p 41).**?**

The limiting distribution and thus the test statistic is analyzed for the following data-generating mechanism under local alternatives. Let

$$y_t = \psi_0 + \psi_1 x_t + \frac{h_0(x_t)}{\sqrt{n}} + u_t$$

where $u_t = y_t - E[y_t|x_t]$ such that $E[u_t|x_t] = 0$ and $h_0(x) = h_0 \in B^s_{\infty\infty}(\mathbb{R}, \vartheta)$ for some $s > 0$ and some $\vartheta \in \mathbb{R}$. Let $E[x_t] = \mu_x$ and $\tilde{b}(h) = [b(h), 0]'$ with $b(h) = E[(x_t - \mu_x) h(x_t)]$. Under the null of a linear conditional mean the function $h_0 = 0$. Let $Q = E[\partial m(\chi_t, \theta, h)/\partial\theta]$. Denote $h_{0,n} = h_0/\sqrt{n}$ and let $m(\chi_t, \theta, h)$ be a population analog of $\hat{m}(\chi_t, \theta, h)$ defined in (9) where in $m(.)$ the empirical means $\bar{y}$ and $\bar{x}$ are replaced with μ_y and μ_x. Under regularity conditions it follows from arguments similar to Newey (1994) that for h fixed,

$$\sqrt{n}\left(\tilde{\theta}_\kappa - \theta\right) = Q^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\left(m(\chi_t, \theta, h_{0,n}) + \gamma(\chi_t)\right)\right) + o_p(1).$$

The correction term $\gamma(\chi_t)$ accounts for non-parametric estimation of the nuisance parameter h and can be derived using the methods developed in Newey (1994). It is given by

$$\gamma(\chi_t) = \begin{bmatrix} 0 \\ -\delta(x_t) \end{bmatrix}\left(y_t - \psi_0 - \psi_1 x_t - \frac{h_0(x_t)}{\sqrt{n}}\right)$$

where $\delta(x_t) = E[x_t|\mathcal{G}_1]$ is the $\mathcal{L}_2(P)$ projection of x_t onto $\mathcal{G}_1$. Define the empirical process

$$v_n(h) = n^{-1/2}\sum_{t=1}^{n}\left(m(\chi_t, \theta, h_{0,n}) + \gamma_h(\chi_t) - E[m(\chi_t, \theta, h_{0,n})]\right) \tag{11}$$

The central limit theorems developed in the first part of the paper play a dual role in analyzing the limiting properties of $\hat{\theta}_\kappa$. On the one hand, stochastic equicontinuity properties of the empirical process (11) can be used to verify regularity conditions in Newey (1994). On the other hand, the functional central limit theorem allows for a stochastic process representation of the limiting distribution of $\hat{\theta}_\kappa$ over the class of local alternatives.

Condition 2 *Let χ_t be a strictly stationary and β-mixing process. Assume that (1) holds. Assume that for some $\vartheta \in \mathbb{R}$, $\mathcal{F} \subset B^s_{\infty\infty}(\mathbb{R}^d, \vartheta_h)$ is nonempty and bounded, $0 \in \mathcal{F}$ and $h \in \mathcal{F}$. Assume that one of the following conditions hold: (i) $\vartheta \leq -2$ and for some $\gamma > 0$ it follows that $\left\|\langle\chi_t\rangle^{(\gamma-\vartheta-1)/2}\right\|_{2,\beta} < \infty$, $\gamma > s$ and $s > 1/2$; (iii) $\vartheta \leq -2$ and for some $\gamma > 0$ it follows that $\left\|\langle\chi_t\rangle^{(\gamma-\vartheta-1)/2}\right\|_{2,\beta} < \infty$, $\gamma < s$ and $\gamma > 1/2$.*

Condition (2) directly leads to the following lemma, which is a direct consequence of Theorem 3. Let $v_t = u_t \left([x_t - \mu_x, x_t - \mu_x - \delta(x_t)]' \right)$ and $\Gamma(h) = \sum_{j=-\infty}^{\infty} E\left[v_t v_{t-j}'\right]$.

Lemma 12 *Assume that Condition 2 holds. Let $v_n(h)$ be defined in (11). Then, $v_n(h) \rightsquigarrow v(h)$ where $v(h)$ is a Gaussian process with covariance function $\Gamma(h)$ and a.s. uniformly continuous sample paths.*

The following high level regularity conditions are similar to conditions imposed in Newey (1994)**?**. Since this section is mostly meant to highlight the usefulness of the functional central limit theory discussed in this paper the regularity conditions are high level with regard to the semiparametric estimators used here. Full development of these estimators is beyond the scope of this paper.

Condition 3 *i) Let $u_t = y_t - E[y_t|x_t]$. Then, $E\left[u_t^2|x_t\right] = \sigma_t^2(x_t)$ and $E\left[\sigma_t^2(x_t)x_t^2\right] < \infty$.*
ii)Let $\hat{h}$ be a series estimator of $h_{0,n}$. Then, there exists a sequence κ_n such that $\kappa_n \to \infty$ as $n \to \infty$ and $\sqrt{n}\left\|\hat{h} - h_{0,n}\right\|_{2,\beta}^2 = o_p(1)$.
iii) $1/\sqrt{n}\sum_{t=1}^n \left((x_t - \mu_x)\left(\hat{h}_t - h_{0,n}\right) - \gamma(x_t)\right) = o_p(1)$.

The next lemma establishes the limiting process for the empirical moment function $m_n(\theta_\kappa)$.

Lemma 13 *Assume that Conditions 2 and 3 hold. Let $m_n(\theta_\kappa)$ be defined in (10). Then,*

$$\sqrt{n} m_n(\theta_\kappa) = v_n(h) + \tilde{b}(h) + o_p(1)$$

and

$$\sqrt{n} m_n(\theta_\kappa) \rightsquigarrow v(h) + \tilde{b}(h)$$

where $v(h)$ is a Gaussian process with covariance function $\Gamma(h)$ and a.s. uniformly continuous sample paths.

The following condition is needed to derive an asymptotic limiting distribution of the two estimators for ψ_1.Note that the conditions here are much simpler than related conditions in Newey (1994) because the estimators considered here exist in closed form.

Condition 4 *Let $\left(\hat{\delta}_\kappa(x_1), ..., \hat{\delta}(x_n)\right)' = MP_2\hat{\Delta}_{22}^{-1}\hat{\Delta}_{21}$ be the empirical projection of $x_t - \bar{x}$ on $\mathcal{G}_1$.*
i) For κ_n as specified in Condition 3 it follows that

$$\hat{Q} = n^{-1}\sum_{t=1}^n \begin{bmatrix} (x_t - \bar{x})^2 & 0 \\ 0 & (x_t - \bar{x})^2 \end{bmatrix} \to_p Q$$

where Q is a fixed, positive definite matrix that does not depend on h.
ii)Assume that $n^{-1}\left(\hat{\Delta}_{11}-\hat{\Delta}_{12}\hat{\Delta}_{22}^{-1}\hat{\Delta}_{21}\right)=\Delta_{11}-\Delta_{12}\Delta_{22}^{-1}\Delta_{21}+o_p(1)$ and $n^{-1}\hat{\Delta}_{11}=\Delta_{11}+o_p(1)$ where $\Delta_{11}=\operatorname{Var}(x_t)>0$ and $\Delta_{11}-\Delta_{12}\Delta_{22}^{-1}\Delta_{21}=\operatorname{Var}(x_t-\delta(x_t))>0$. Define

$$\Lambda=\left[\begin{array}{cc}\Delta_{11} & \Delta_{11}-\Delta_{12}\Delta_{22}^{-1}\Delta_{21}\\ \Delta_{11}-\Delta_{12}\Delta_{22}^{-1}\Delta_{21} & \Delta_{11}-\Delta_{12}\Delta_{22}^{-1}\Delta_{21}\end{array}\right].$$

iii) Let $\hat{\gamma}(x_t)=\left[0,-\hat{\delta}(x_t)\right]'\left(y_t-\bar{y}-\psi_1(x_t-\bar{x})-\hat{h}(x_t)\right)$. Then, it follows that

$$\sup_{h\in\mathcal{F}}\left\|n^{-1/2}\sum_{t=1}^{n}\hat{m}_t(\chi_t,\theta_\kappa,h)-m_t(\chi_t,\theta_\kappa,h)\right\|=o_p(1).$$

iv) Assume that $n^{-1}\sum_{t=1}^{n}(x_t-\bar{x})h_0(x_t)=b(h)+o_p(1)$.

We are now in a position to state the asymptotic limiting distribution of the estimators for ψ_1 under the null of $h=0$ and local alternatives. This distribution forms the basis for finding critical values for the Hausman test statistic for non-linearity.

Lemma 14 *Assume that Conditions 2, 3 and 4 hold. Then, it follows that for h_0 fixed,*

$$\sqrt{n}\left(\tilde{\theta}_{\kappa_n}-\theta_0\right)\to_d Q^{-1}\left(v(h)+\tilde{b}(h)\right)$$

where $\tilde{b}(0)=0$ and $Q^{-1}v(h)\sim N\left(0,Q^{-1}\Gamma(h)Q^{-1}\right)$ for h fixed. If in addition, $E\left[u_t|\mathcal{A}^{t-1}\right]=0$ and $E\left[u_t^2|x_t\right]=\sigma^2$ where σ^2 is constant and $\sigma^2>0$, then it follows that $Q^{-1}v(h)\sim N\left(0,\sigma^2Q^{-1}\Lambda Q^{-1}\right)$ where Δ is defined in Condition 4(ii).

To form the Hausman statistic assume that $\hat{\Gamma}$ is a consistent estimator of Γ and $\hat{Q}$ is consistent for Q by Condition 4. Let $e=(1,-1)'$. A generalized Hausman statistic to test the null hypothesis of a linear conditional mean then is given as

$$\tilde{H}_1=\frac{n\left(\tilde{\psi}_1-\tilde{\psi}_{1,\kappa}\right)^2}{e'\hat{Q}^{-1}\hat{\Gamma}\hat{Q}^{-1}e} \tag{12}$$

If the additional conditions imposed on u_t in Lemma 14 hold then test statistic can be simplified to

$$\tilde{H}_2=\frac{\left(\tilde{\psi}_1-\tilde{\psi}_{1,\kappa}\right)^2}{\hat{\sigma}^2\left(\hat{\Delta}_{12}\hat{\Delta}_{22}^{-1}\hat{\Delta}_{21}/\hat{\Delta}_{11}^2\right)}. \tag{13}$$

The limiting distributions of the two Hausman statistics are summarized in the following Theorem.

Theorem 15 *Assume that Conditions 2, 3 and 4 hold. Then,* $\hat{H}_1$ *defined in (12) converges (pointwise for* h_0 *fixed) to a non-central* χ^2 *process*

$$\hat{H}_1 \to_d \chi_1^2\left(\tilde{\lambda}_1\right)$$

where for fixed h, $\chi_1^2(\lambda_1)$ *is a non-central chi-square distribution with one degree of freedom and non-centrality parameter* λ_1 *and*

$$\tilde{\lambda}_1 = \frac{b(h)}{\Delta_{11}\sqrt{e'Q^{-1}\Gamma Q^{-1}e}}.$$

If in addition, $E\left[u_t|\mathcal{A}^{t-1}\right] = 0$ *and* $E\left[u_t^2|x_t\right] = \sigma^2$ *where* σ^2 *is constant and* $\sigma^2 > 0$, *then it follows that*

$$\hat{H}_1 \to_d \chi_1^2\left(\tilde{\lambda}_2\right),\ \hat{H}_2 \to_d \chi_2^2\left(\tilde{\lambda}_2\right)$$

where the non-centrality parameter λ_2 *is given by*

$$\tilde{\lambda}_2 = \frac{b(h)}{\Delta_{11}\sqrt{\sigma^2\left(\Delta_{12}\Delta_{22}^{-1}\Delta_{21}/\Delta_{11}^2\right)}}.$$

Theorem 15 establishes that under the null hypothesis of a linear conditional mean of y_t the limiting distribution of $\hat{H}_1$ and, under additional conditions, of $\hat{H}_2$ are asymptotically χ_1^2. For a significance level α, let c_α be the critical value of the central χ_1^2 distribution, i.e. $\alpha = \Pr\left(\chi_1^2 > c_\alpha\right)$. The null hypothesis of a linear conditional mean then is rejected if $\hat{H}_1 > c_\alpha$ or $\hat{H}_2 > c_\alpha$.

The analysis in Theorem 15 also shows how the power of the test against local alternatives depends on the efficiency gain of ψ_1 over $\psi_{1\kappa}$ under the null distribution. The asymptotic power function of the test is given by $\Pr\left(\chi_1^2(\lambda_1) > c_\alpha\right)$ as h ranges over the set of permissible alternatives.

Now assume that the martingale and homoskedasticity restrictions on u_t are satisfied. An alternative version of the test H_2 then is based on the OLS estimators $\hat{\theta}_\kappa$. By similar arguments as in the proof of Lemma 14 it can be shown that the asymptotic variance of $\hat{\psi}_1 - \hat{\psi}_{1,\kappa}$ is $\sigma^2\left(\left(\Delta_{11} - \Delta_{12}\Delta_{22}^{-1}\Delta_{21}\right)^{-1} - \Delta_{11}^{-1}\right)$. This implies that the concentration parameter for the regression based statistic is given by

$$\hat{\lambda}_2 = \frac{b(h)}{\Delta_{11}\sqrt{\sigma^2\left(\frac{\Delta_{12}\Delta_{22}^{-1}\Delta_{21}}{\Delta_{11}\left(\Delta_{11}-\Delta_{12}\Delta_{22}^{-1}\Delta_{21}\right)}\right)}}.$$

This result implies that the regression based test, while having the same limiting distribution under the null of linearity, is less powerful against local alternatives than the plug-in Z-estimator based test.

5 Conclusion

The paper combines recent results on bracketing numbers for weighted Besov spaces with a functional central limit theorem for strictly stationary β-mixing processes. It is shown that by specializing the bracketing results to a particular Hilbert space of relevance to the dependent limit theory, functional central limit theorems for dependent processes indexed by Besov classes can be obtained directly. These insights lead to some new results in function spaces with polynomially decaying functions over unbounded domains and smooth functions over bounded domains.

It is shown how the limit theory can be used to simplify some proofs in the analysis of semiparametric estimators and tests. An example for a Hausman test for linearity is considered in detail. More specifically, the central limit theorem implies a stochastic equicontinuity property that helps shorten arguments needed to establish the limiting behavior of the test. The central limit theory also allows to represent the limiting distribution over a class of local alternatives under general conditions. Finally, a comparison of two versions of the test when stronger conditions on the model are imposed is provided. It is shown that a test based on a less efficient plug in estimator is preferred over a regression based version of the test.

A detailed analysis of non-parametric estimation in weighted Besov spaces is beyond the scope of the paper and left for future research. As such, a number of the conditions imposed in Section 4 are high level.

A Proofs

Proof of Theorem 1. The proof follows the argument in Nickl and Pötscher (2007, p.184). Let $N(\delta, \mathcal{F}, \|.\|_\infty)$ be the minimal covering number of $\mathcal{F}$. with respect to $\|.\|_\infty$ and $H(\delta, \mathcal{F}, \|.\|_\infty) = \log N(\delta, \mathcal{F}, \|.\|_\infty)$ the metric entropy for $\mathcal{F}$. From Nickl and Pötscher (2007, p.184, Eq.3) it follows that for all $\vartheta \in \mathbb{R}$ and all $\gamma > 0$

$$H\left(\delta, \mathcal{F}, \left\|\langle x\rangle^{(\vartheta-\gamma)/2}\right\|_\infty\right) \precsim \begin{cases} \delta^{-d/s} & \text{if } \gamma > s - d/p \\ \delta^{-(\gamma/d+1/p)^{-1}} & \text{if } \gamma < s - d/p \end{cases} \tag{14}$$

Let B_i be closed balls in $C\left(\mathbb{R}^d, \langle x\rangle^{(\vartheta-\gamma)/2}\right) = \left\{f : f(.)\langle x\rangle^{(\vartheta-\gamma)/2} \in C\left(\mathbb{R}^d\right)\right\}$ with radius δ (relative to the norm $\left\|(.)\langle x\rangle^{(\vartheta-\gamma)/2}\right\|_\infty$) covering $\mathcal{F}$. Note that the number of such balls is $N\left(\delta, \mathcal{F}, \left\|(.)\langle x\rangle^{(\vartheta-\gamma)/2}\right\|_\infty\right)$. Let f_i be the center of B_i. Then each B_i contains the functions f such that

$$\sup_{x\in\mathbb{R}^d} |f(x) - f_i(x)|\langle x\rangle^{(\vartheta-\gamma)/2} \leq \delta.$$

The brackets

$$\left[f_i(x) - \delta\langle x\rangle^{(\gamma-\vartheta)/2}, f_i(x) + \delta\langle x\rangle^{(\gamma-\vartheta)/2}\right]$$

are contained in B_i and cover $\mathcal{F}$. The $\mathcal{L}_{2,\beta}(P)$ norm of these brackets is

$$\left\|2\delta\langle x\rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta}.$$

First consider the case when $\vartheta > 0$. In that case one can choose $\gamma = \vartheta$. Then, $\left\|2\delta\langle x\rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta} = \|2\delta\|_{2,\beta}$. Now note that for the constant function δ

$$Q_\delta(u) = \inf(t : P(|\delta| > t) \leq u) = \delta$$

such that

$$\|2\delta\|_{2,\beta}^2 = \sum_{m=0}^\infty \int_0^{\beta_m} (Q_{2\delta}(u))^2 du = (2\delta)^2 \sum_{m=0}^\infty \beta_m < \infty$$

by Condition (1). One obtains from Nickl and Pötscher (2007, p.184, eq. 4) that

$$H_{[]}\left(2\delta\sum_{m=0}^\infty \beta_m, \mathcal{F}, \|\|_{2,\beta}\right) \leq H(\delta, \mathcal{F}, \|\|_\infty)$$

such that the result follows immediately from (14).

When $\vartheta \leq 0$ the brackets have size

$$2\delta\left\|\langle x\rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta} < \infty$$

which is bounded by the conditions of the Theorem. It follows again by Nickl and Pötscher (2007, p.184, eq. 4) that

$$H_{[]}\left(2\delta\left\|\langle x\rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta},\mathcal{F},\|\|_{2,\beta}\right)\leq H\left(\delta,\mathcal{F},\left\|(.)\langle x\rangle^{(\vartheta-\gamma)/2}\right\|_{\infty}\right). \tag{15}$$

Then, (14) delivers the stated result. ∎

Proof of Corollary 2. From the proof of Theorem 1 the $\mathcal{L}_{2,\beta}(P)$ norm of the brackets is, for all $\gamma>0$ and all $\vartheta\in\mathbb{R}$,

$$\left\|2\delta\langle x\rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta}\leq 2\delta M^{(\gamma-\vartheta)/2}\sum_{m=0}^{\infty}\beta_m<\infty.$$

Therefore, the bound in (15) can be applied and the result again follows by (14). ∎

Proof of Theorem 3. The result follows from Theorem 1 in DMR once all of their conditions are verified. First show that $\mathcal{F}\in\mathcal{L}_{2,\beta}(P)$. Let $\mathcal{L}(\beta)$ be the class of integer valued random variables with distribution function $G_\beta(n)=1-\beta_n$ for any $n\in\mathbb{N}$ (see DMR, p. 423). For any $b\in\mathcal{L}(\beta)$ and some real number $K>0$ it follows that

$$\begin{aligned}E\left[bf^2(\chi_t)\right] &= E\left[b\langle\chi_t\rangle^{-\vartheta}\left(f(\chi_t)\langle\chi_t\rangle^{\vartheta/2}\right)^2\right] \\ &\leq \left(\sup_{x\in\mathbb{R}^d}\sup_{f\in\mathcal{F}}\left|f(x)\langle x\rangle^{\vartheta/2}\right|\right)^2 E\left[b\langle\chi_t\rangle^{-\vartheta}\right] \\ &\leq K^2E\left[b\langle\chi_t\rangle^{-\vartheta}\right]\end{aligned} \tag{16}$$

where the first inequality is obtained by applying Proposition 3 of Nickl and Pötscher (2007) and because $f(x)\langle x\rangle^{\vartheta/2}\in\mathcal{F}$ by assumption. For any $f\in\mathcal{F}$ it follows from DMR, Eq. (6.2) and

$$\begin{aligned}\|f\|_{2,\beta} &= \sup_{b\in\mathcal{L}(\beta)}\sqrt{E\left[bf^2(\chi_t)\right]} \\ &\leq K\sup_{b\in\mathcal{L}(\beta)}\sqrt{E\left[b\langle\chi_t\rangle^{-\vartheta}\right]}\end{aligned} \tag{17}$$

where the inequality uses (16). If $\vartheta\geq 0$ the inequality

$$\langle\chi_t\rangle^{-\vartheta}\leq 1$$

together with $b\geq 0$ leads to

$$\|f\|_{2,\beta}\leq K\sup_{b\in\mathcal{L}(\beta)}\sqrt{E[b]}=K\|1\|_{2,\beta}=K\sqrt{\sum_{m=0}^{\infty}\beta_m}. \tag{18}$$

When $\vartheta < 0$, (17) leads to

$$\|f\|_{2,\beta} \leq K \left\| \langle \chi_t \rangle^{-\vartheta} \right\|_{2,\beta}. \tag{19}$$

Since in this case,

$$\langle \chi_t \rangle^{-\vartheta} \geq 1$$

and for any $\gamma > 0$,

$$\langle \chi_t \rangle^{\gamma-\vartheta} \geq \langle \chi_t \rangle^{-\vartheta}$$

it follows from (19) that

$$\|f\|_{2,\beta} \leq K \left\| \langle \chi_t \rangle^{\gamma-\vartheta} \right\|_{2,\beta} < \infty \tag{20}$$

which is bounded by assumption. Thus, (18) and (20) show that $f \in \mathcal{F} \subset B^s_{pq}\left(\mathbb{R}^d, \vartheta\right)$ with either $\vartheta \geq 0$ or $\vartheta < 0$ and some $\gamma > 0$ such that $\left\| \langle x \rangle^{(\gamma-\vartheta)/2} \right\|_{2,\beta} < \infty$ implies that $\mathcal{F} \in \mathcal{L}_{2,\beta}(P)$.

It remains to be show that

$$\int_0^1 \sqrt{H_{[]}\left(\delta, \mathcal{F}, \|\|_{2,\beta}\right)} d\delta < +\infty. \tag{21}$$

For case (i) Theorem 1 implies that $H_{[]}\left(\delta, \mathcal{F}, \|\|_{2,\beta}\right) \precsim \delta^{-d/s}$ such that (21) holds for $d/2s < 1$. For case (ii) Theorem 1 implies that $H_{[]}\left(\delta, \mathcal{F}, \|\|_{2,\beta}\right) \precsim \delta^{-(\gamma/d+1/p)^{-1}}$ such that (21) holds for $1/2\left(\gamma/d + 1/p\right)^{-1} < 1$. Cases (iii) and (iv) follow in the same way. This establishes the result. ■

Proof of Corollary 4. For any $s > d/p$ fix ϑ such that $\vartheta > s - d/p$. By construction $0 < \vartheta < \infty$ and thus $f(.) \langle x \rangle^{\vartheta}$ is bounded for $x \in \mathfrak{X}$ and $f(.) \langle x \rangle^{\vartheta} \in B^s_{pq}(\mathfrak{X}, \vartheta)$. As in Nickl and Pötscher (2007, p.186), conclude that $\mathcal{F} \subseteq B^s_{pq}(\mathfrak{X}, \vartheta)$. The results of Theorem 3 can now be applied. In particular, using the bound in (16) leads to

$$\|f\|_{2,\beta} \leq K \sup_{b \in \mathcal{L}(\beta)} \sqrt{E\left[b \langle \chi_t \rangle^{-\vartheta}\right]} \leq K M^{-\vartheta/2} \sqrt{\sum_{m=0}^{\infty} \beta_m} < \infty.$$

The result now follows from the fact that 21 holds by the results in Corollary 2. ■

Proof of Theorem 5. From DMR Lemma 2, (S.1) and p. 404 it follows for $\phi(x) = x^r$ with $r > 1$ that

$$\sum_{m=1}^{\infty} m^{1/(r-1)} \beta_m < \infty \tag{22}$$

and

$$\left\| \langle \chi_t \rangle^{(\gamma-\vartheta)/2} \right\|_{2r,P} < \infty \tag{23}$$

is sufficient for $\left\|\langle\chi_t\rangle^{(\gamma-\vartheta)/2}\right\|_{2,\beta} < \infty$. Note that (23) holds since $r(\gamma-\vartheta) > 1$ and by Jensen's inequality

$$\left\|\langle\chi_t\rangle^{(\gamma-\vartheta)/2}\right\|_{2r,P}^{2r} = E\left[\langle\chi_t\rangle^{r(\gamma-\vartheta)}\right] \le 1 + E\left[\|\chi_t\|^{2r(\gamma-\vartheta)}\right] < \infty$$

where the expectation on the RHS is bounded by assumption. The result now follows from Theorem 3. ■

Proof of Theorem 11. The result follows from DMR (eq 2.11) and (eq. S.1). In particular, the condition

$$\int_0^1 \sqrt{H_{[]}\left(t,,\|.\|_{2p}\right)}\,dt < \infty \tag{24}$$

needs to hold. From Nickl and Pötscher (2007) it follows that under the stated conditions in (i),

$$H_{[]}\left(t,,\|.\|_{2p}\right) \precsim t^{-d/s}$$

such that (24) holds as long as $d/(2s) < 1$ or $1/2 < s/d$. Under conditions (ii) one obtains similarly that

$$H_{[]}\left(t,,\|.\|_{2p}\right) \precsim t^{-(\gamma/d+1/p)^{-1}}$$

such that (24) holds as long as $rp/(\gamma p + d) < 1$ or $1/2 < (\gamma/d + 1/p)$. ■

Proof of Lemma 12. It follows that

$$\begin{aligned}
v_n(h) &= n^{-1/2}\sum_{t=1}^{n}\left(m(\chi_t,\theta,h_{0,n}) + \gamma_h(\chi_t) - E[m(\chi_t,\theta,h_{0,n})]\right) \\
&= n^{-1/2}\sum_{t=1}^{n}\left(\begin{bmatrix}(u_t + h_{0,n}(x_t))(x_t-\mu_x) \\ u_t(x_t-\mu_x-\delta(x_t))\end{bmatrix} - \begin{bmatrix} n^{-1/2}E[(x_t-\mu_x)h_0(x_t)] \\ 0\end{bmatrix}\right) \\
&= n^{-1/2}\sum_{t=1}^{n}\begin{bmatrix}u_t(x_t-\mu_x) \\ u_t(x_t-\mu_x-\delta(x_t))\end{bmatrix} \qquad (25)\\
&\quad -n^{-1}\sum_{t=1}^{n}\begin{bmatrix}(x_t-\mu_x)h_0(x_t) - E[(x_t-\mu_x)h_0(x_t)] \\ 0\end{bmatrix}
\end{aligned}$$

where

$$n^{-1/2}\sum_{t=1}^{n}\begin{bmatrix}u_t(x_t-\mu_x) \\ u_t(x_t-\mu_x-\delta(x_t))\end{bmatrix} \rightsquigarrow v(h)$$

by Theorem 3 and the fact that $f(y,x) = (y - c - \psi x - h(x))(x-\mu_x) \in B^s_{\infty\infty}\left(\mathbb{R}^d, \min(\vartheta-1,-2)\right)$ if $h(x) \in B^s_{\infty\infty}\left(\mathbb{R}^d,\vartheta\right)$. It remains to be shown that the second term in (25) is $o_p(1)$. Since $(x_t-\mu_x)h_0(x_t) \in B^s_{\infty\infty}\left(\mathbb{R}^d,\vartheta-1\right)$ it follows by Nickl and Pötscher (2007, Theorem 1(2)), a strong law of large numbers for β-mixing processes and the arguments in the proof of Theorem 2.4.1. in van der Vaart and Wellner (1996, p. 122) that

$$\sup_{h\in\mathcal{F}}\left|n^{-1}\sum_{t=1}^{n}\left(x_t-\mu_x\right)h_0\left(x_t\right)-E\left[\left(x_t-\mu_x\right)h_0\left(x_t\right)\right]\right|=o_p\left(1\right).$$

■

Proof of Lemma 13. The proof closely follows arguments in Newey (1994, Sections 5 and 6), except for the fact that here $\|.\|_{2,\beta}$ norms rather than Sobolev norms are the natural norms to use. This is because stochastic equicontinuity of the empirical process determining the limiting distribution is directly tied to the $\|.\|_{2,\beta}$ norm. Let $m\left(\chi_t,\theta,\hat{h}\right)=\hat{m}_t\left(\theta\right)$ and $m\left(\chi_t,\theta,h_0\right)=m_t\left(\theta\right)$.Consider the expansion

$$\begin{aligned}\sqrt{n}m_n\left(\theta_\kappa\right) &= n^{-1/2}\sum_{t=1}^{n}\hat{m}_t\left(\theta_\kappa\right)=n^{-1/2}\sum_{t=1}^{n}\left(m_t\left(\theta_\kappa\right)+\gamma\left(\chi_t\right)\right)\\ &+n^{-1/2}\sum_{t=1}^{n}\left(\hat{m}_t\left(\theta_\kappa\right)-m_t\left(\theta_\kappa\right)-D\left(\chi_t,\hat{h}-h_0\right)\right) \quad (26)\\ &+n^{-1/2}\sum_{t=1}^{n}\left(D\left(\chi_t,\hat{h}-h_0\right)-\gamma\left(\chi_t\right)\right). \quad (27)\end{aligned}$$

Let $A_{n,\varepsilon}=1\left\{\left\|n^{-1/2}\sum_{t=1}^{n}\left(\hat{m}_t\left(\theta_0\right)-m_t\left(\theta_0\right)+\gamma\left(\chi_t\right)\right)\right\|>\varepsilon\right\}$ and $B_{n,\varepsilon}=1\left\{\left\|\hat{h}-h_0\right\|_{2,\beta}\leq\varepsilon\right\}$. Then,

$$\begin{aligned}\lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}E\left[A_{n,\varepsilon}\right] &\leq \lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}E\left[A_{n,\varepsilon/2}\cap B_{n,\varepsilon/2}\right]\\ &+\lim_{\varepsilon\downarrow 0}\limsup_{n\to\infty}P\left(\left\|\hat{h}-h_0\right\|_{2,\beta}>\varepsilon/2\right)\end{aligned}$$

where the second term is zero by Condition 3(ii). Consequently, all subsequent arguments are restricted to the set $B_{n,\varepsilon}$. By the Markov inequality (26) and (27) are $o_p\left(1\right)$ if

$$\begin{aligned}&E\left\|n^{-1/2}\textstyle\sum_{t=1}^{n}\left(\hat{m}_t\left(\theta_0\right)-m_t\left(\theta_0\right)-D\left(\chi_t,\hat{h}-h_0\right)\right)\right\| \quad (28)\\ \leq\ &\sqrt{n}E\left\|\hat{m}_t\left(\theta_0\right)-m_t\left(\theta_0\right)-D\left(\chi_t,\hat{h}-h_0\right)\right\|\end{aligned}$$

tends to zero and

$$
\begin{aligned}
& \left\| n^{-1/2} \sum_{t=1}^{n} \left(D\left(\chi_t, \hat{h} - h_0\right) - \gamma\left(\chi_t\right)\right) \right\| \\
\leq & \left\| n^{-1/2} \sum_{t=1}^{n} \left(D\left(\chi_t, \hat{h} - h_0\right) - \int D\left(\chi, \hat{h} - h_0\right) dP \right) \right\| && (29) \\
& + \left\| \int D\left(\chi, \hat{h} - h_0\right) dP - n^{-1/2} \sum_{t=1}^{n} \gamma\left(\chi_t\right) \right\| && (30) \\
= & \; o_p(1).
\end{aligned}
$$

For (28) note that because $m(.)$ is linear in h one immediately obtains

$$
D(\chi, h - h_0) = \begin{bmatrix} 0 \\ (x - \mu_x)(h - h_0) \end{bmatrix} \tag{31}
$$

and

$$
\|(m(\chi, \theta, h) - m(\chi, \theta, h_0) - D(\chi, h - h_0))\| = 0
$$

such that the RHS of (28) is zero and consequently, the term in (26) is $o_p(1)$. Note that Newey (1994, Assumption 5.1) imposes that a second order approximation of $m(\chi, \theta, h_0)$ is well behaved. This includes requiring that $\sqrt{n}\left\|\hat{h} - h_0\right\|^2 = o_p(1)$. Here, such a restriction is not required because $m(\chi, \theta, h_0)$ is linear in h and thus second order terms are zero. For similar conditions see Andrews (1994, p.58).**?**

For (29) consider $D(\chi_t, h - h_0) = f(\chi_t)$ where only the second component is relevant. Thus focus on

$$
f(\chi_t) = (x_t - \mu_x) h(x_t) \tag{32}
$$

and where $f(\chi_t)$ is a class of functions indexed by $h \in \mathcal{F}_h \in B^s_{\infty\infty}(\mathbb{R}, \vartheta_h)$. It follows that $f \in \mathcal{F} \subset B^s_{\infty\infty}(\mathbb{R}, \vartheta_h - 1)$ as long as $h \in \mathcal{F}_h$. By Theorem 3 the empirical process

$$
v_n(f) := n^{-1/2} \sum_{t=1}^{n} \left(f(\chi_t) - \int f(\chi_t) \, dP \right)
$$

satisfies $v_n(f) \rightsquigarrow v(f)$ where $v(f)$ is a Gaussian process. Note that Theorem 3 is established by checking all the conditions for DMR, Theorem 1. That Theorem in turn is established by establishing stochastic equicontinity of the process $v_n(f)$. This shows that $v_n(f)$ is stochastically equicontinuous. Now, for $f_{0,t} = (x_t - \mu_x) h_0(x_t)$ and $f_t = (x_t - \mu_x) h(x_t)$ it follows by a routine argument that

$$
n^{-1/2} \sum_{t=1}^{n} \left(D(\chi_t, h - h_0) - \int D(\chi, h - h_0) \, dP_0 \right) = n^{-1/2} \sum_{t=1}^{n} \left(f_t - f_{0,t} - \int (f_t - f_{0,t}) \, dP \right)
$$

and

$$\begin{aligned}
&\Pr\left(\left\|n^{-1/2}\textstyle\sum_{t=1}^{n}\left(D\left(\chi_t,\hat{h}-h_0\right)-\int D\left(\chi,\hat{h}-h_0\right)dP\right)\right\|>\delta\right)\\
\leq\ &\Pr\left(\sup_{\|h-h_0\|_{2,\beta}\leq\epsilon}\left\|n^{-1/2}\sum_{t=1}^{n}\left(f_t-f_{0,t}-\int\left(f_t-f_{0,t}\right)dP\right)\right\|>\delta/2\right) && (33)\\
&+\Pr\left(\left\|\hat{h}-h_0\right\|_{2,\beta}>\delta/2\right) && (34)
\end{aligned}$$

where (33) tends to zero as $\delta\downarrow 0$ by the fact that $v_n(f)$ is stochastically equicontinuos and (34) tends to zero as $\delta\downarrow 0$ by Condition 3(ii). Together (33) and (34) establishes that (29) is $o_p(1)$.

To establish that (30) is $o_p(1)$ the conditions in Newey (1994, Assumption 5.3) are sufficient: there is a function $\gamma(\chi_t)$ such that

$$E\left[\gamma(\chi_t)\right]=0, \tag{35}$$

$$E\left[\|\gamma(\chi_t)\|^2\right]<\infty, \tag{36}$$

and for all $\left\|\hat{h}-h_0\right\|_{2,\beta}$ small enough,

$$n^{-1/2}\sum_{t=1}^{n}\left(\gamma(\chi_t)-\int D\left(\chi_t,\hat{h}-h_0\right)dP\right)\to^p 0. \tag{37}$$

For (35) note that formally differentiating $m(\chi,\theta,h)$ with respect to h leads to $D(\chi,h)$ in (31). Let τ index a path (see Newey, 1994, p.1352 for a definition). Let $\delta(x_t)=E\left[x_t|\mathcal{G}_1\right]$ be the projection of x_t onto $\mathcal{G}_1$ such that $E\left[D\left(\chi_t,\tilde{h}\right)\right]=E\left[E\left[(x_t-\mu_x)|\mathcal{G}_1\right]\tilde{h}\right]$ for all $\tilde{h}\in\mathcal{G}_1$. Let $g(x_t,\tau)$ be the projection of y_t on $\mathcal{G}$ for a path τ (see Newey, 1994, p. 1361). Since $\mathcal{G}_1\subset\mathcal{G}$ it follows by the Projection Theorem that $E_\tau\left[\delta(x_t)\,g(x_t,\tau)\right]=E_\tau\left[\delta(x_t)\,y_t\right]$. Then, it follows from Newey (1994, Eq. 4.5) that

$$\partial E\left[D(\chi,h(\tau))\right]/\partial\tau=E\left[\delta(x_t)\left(y_t-g(x_t)\right)S(\chi_t)\right]$$

where $S(\chi_t)$ is the score of a regular path (see Newey, 1994, Theorem 2.1). By Newey (1994, Theorem 4.1) the correction term $\gamma(\chi_t)$ is given by

$$\gamma(\chi_t)=\delta(x_t)\,u_t.$$

such that $E\left[\gamma(\chi_t)\right]=0$ follows immediately from $E\left[u_t|x_t\right]=0$.

For (36) note that

$$\left|\gamma(\chi_t)\right|\leq\left|E\left[x_t|\mathcal{G}_1\right]u_t\right|\leq E\left[|x_t|\,|\mathcal{G}_1\right]|u_t|$$

Then, by Jensen's inequality

$$\begin{aligned} E\left[\left|\gamma\left(\chi_t\right)\right|^2\right] &\leq E\left[\left(E\left[\left|x_t\right|\left|\mathcal{G}_1\right]\right)^2 u_t^2\right] \leq E\left[x_t^2 u_t^2\right] \\ &= E\left[\sigma_t^2\left(x_t\right) x_t^2\right] < \infty \end{aligned}$$

where $\sigma_t^2\left(x_t\right) = E\left[u_t^2|x_t\right]$ and $E\left[\sigma_t^2\left(x_t\right) x_t^2\right]$ is bounded by Condition (3)(i).

Finally, (37) is satisfied by Condition (3)(iii). This establishes that (26) and (27) are $o_p(1)$ and therefore that the first claim of the Lemma holds. The second part of the Lemma follows from Lemma (12). ■

Proof of Lemma 14. The estimator $\tilde{\theta}_\kappa$ solves

$$m_n\left(\tilde{\theta}_\kappa\right) = n^{-1}\sum_{t=1}^{n}\hat{m}\left(\chi_t, \tilde{\theta}_\kappa, \hat{h}_\kappa\right) = 0$$

which means that it can be expressed in closed form as as

$$\begin{bmatrix} \tilde{\psi}_1 \\ \tilde{\psi}_{1,\kappa} \end{bmatrix} = n^{-1}\hat{Q}^{-1}\begin{bmatrix} P_1' M y \\ P_1'\left(I - MP_2\hat{\Delta}_{22}^{-1}P_2'\right) My \end{bmatrix}.$$

Using the fact that

$$\begin{bmatrix} \psi_1 \\ \psi_{1,\kappa} \end{bmatrix} = n^{-1}\hat{Q}^{-1}\begin{bmatrix} \psi_1 P_1' M P_1 \\ \psi_{1,\kappa} P_1' M P_1 \end{bmatrix}$$

it follows that

$$\sqrt{n}\begin{bmatrix} \tilde{\psi}_1 - \psi_1 \\ \tilde{\psi}_{1,\kappa} - \psi_{1,\kappa} \end{bmatrix} = \hat{Q}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\begin{bmatrix} \left(x_t - \bar{x}\right)\left(y_t - \psi_1\left(x_t - \bar{x}\right)\right) \\ \left(x_t - \bar{x}\right)\left(y_t - \bar{y} - \psi_{1,\kappa}\left(x_t - \bar{x}\right) - \hat{h}\left(x_t\right)\right) \end{bmatrix} \tag{38}$$

$$= \hat{Q}^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n}\hat{m}_t\left(\chi_t, \theta_\kappa, \hat{h}\right) \tag{39}$$

By Condition 4(i) it follows that $\hat{Q}^{-1} - Q^{-1} = o_p(1)$. Then it follows by Condition 4(iii) and (iv) that

$$\sqrt{n}\left(\tilde{\theta} - \theta_\kappa\right) = Q^{-1}\frac{1}{\sqrt{n}}\sum_{t=1}^{n} m_t\left(\chi_t, \theta_\kappa, \hat{h}\right) + o_p(1).$$

The result then follows from Lemmas 12 and 13. ■

Proof of 15. It follows directly from Lemma 14 that for fixed h,

$$\tilde{H}_1^{1/2} := \frac{\sqrt{n}\left(\tilde{\psi}_1 - \tilde{\psi}_{1,\kappa}\right)}{\sqrt{e'\hat{Q}^{-1}\hat{\Gamma}\hat{Q}^{-1}e}} = e'\left(\sqrt{n}\left(\tilde{\theta} - \theta_\kappa\right)\right) \to_d \frac{Q^{-1}v(h)}{\sqrt{e'Q^{-1}\Gamma Q^{-1}e}} + \frac{e'Q^{-1}\tilde{b}(h)}{\sqrt{e'Q^{-1}\Gamma Q^{-1}e}}$$

where

$$\frac{e'Q^{-1}\tilde{b}\left(h\right)}{\sqrt{e'Q^{-1}\Gamma Q^{-1}e}} = \frac{b\left(h\right)}{\Delta_{11}\sqrt{e'Q^{-1}\Gamma Q^{-1}e}}$$

and

$$\frac{Q^{-1}v\left(h\right)}{\sqrt{e'Q^{-1}\Gamma Q^{-1}e}} \sim N\left(0,1\right).$$

The result follows now from the continuous mapping theorem and the fact that $\tilde{H}_1 = \left(\tilde{H}_1^{1/2}\right)^2$. The result for $\tilde{H}_2$ follows in the same way. ■