Stochastic Equicontinuity in Nonlinear Time Series Models

Andreas Hagemann[*]

University of Notre Dame

June 10, 2013

In this paper I provide simple and easily verifiable conditions under which a strong form of stochastic equicontinuity holds in a wide variety of modern time series models. In contrast to most results currently available in the literature, my methods avoid mixing conditions. I discuss several applications in detail.

1. Introduction

Stochastic equicontinuity typically captures the key difficulty in weak convergence proofs of estimators with non-differentiable objective functions. Precise and elegant methods have been found to deal with cases where the data dependence structure can be described by mixing conditions; see Dedecker et al. (2007) for an excellent summary. Mixing assumptions are convenient in this context because they measure how events generated by time series observations—rather than the observations themselves—relate to one another and therefore also measure dependence of functions of such time series. The downside to these assumptions is that they can be hard to verify for a given application. Hansen (1996) describes alternatives and considers parametric classes of functions that behave like mixingales, but his results come at the expense of Lipschitz continuity conditions on these functions and rule out many applications of interest.

In this paper I give simple and easily verifiable conditions under which objective functions of econometric estimators are stochastically equicontinuous when the underlying process is a stationary time series of the form

$$\xi_i = \xi(\varepsilon_i, \varepsilon_{i-1}, \varepsilon_{i-2}, \dots). \tag{1.1}$$

Here $(\varepsilon_i)_{\in \mathbb{Z}}$ is a sequence of iid copies of a random variable ε and ξ is a measurable, possibly unknown function that transforms the input $(\varepsilon_i, \varepsilon_{i-1}, \dots)$ into the output ξ_i. The stochastic

equicontinuity problem does not have to be parametric and no continuity conditions are needed. The class (1.1) allows for the construction of dependence measures that are directly related to the stochastic process and includes a large number of commonly-used stationary time series models. The next section provides several specific examples.

In the following, $\|X\|_p$ denotes $(\mathbb{E}|X|^p)^{1/p}$ and $\mathbb{P}^*$ and $\mathbb{E}^*$ are outer probability and expectation, respectively (see van der Vaart, 1998, p. 258). Limits are as $n \to \infty$.

2. Stochastic Equicontinuity in Nonlinear Time Series Models

Let $\nu_n f := n^{-1/2} \sum_{i=1}^n \big(f(\xi_i) - \mathbb{E}f(\xi_0)\big)$ be the empirical process evaluated at some function f. Here f is a member of a class of real-valued functions $\mathcal{F}$. In econometric applications, $\mathcal{F}$ is typically a parametric class $\{f_\theta : \theta \in \Theta\}$, where Θ is a bounded subset of $\mathbb{R}^k$, although no parametric restriction on $\mathcal{F}$ is necessary in the following. Define a norm by $\rho(f) = \|f(\xi_0)\|_2$. An empirical process is said to be *stochastically equicontinuous* (see, e.g., Pollard, 1985, p. 139) on $\mathcal{F}$ if for all $\epsilon > 0$ and $\eta > 0$, there is a $\delta > 0$ such that

$$\limsup_{n\to\infty} \mathbb{P}^* \left(\sup_{f,g\in\mathcal{F}:\rho(f-g)<\delta} |\nu_n(f-g)| > \eta \right) < \epsilon. \tag{2.1}$$

As mentioned above, proving stochastic equicontinuity is often the key difficulty in weak convergence proofs. The next four examples illustrate typical applications.

Example 1 (Quantilograms). Linton and Whang (2007) measure the directional predictive ability of stationary time series $(X_i)_{i\in\mathbb{Z}}$ with the quantilogram, a normalized version of $\mathbb{E}(\alpha - 1\{X_0 < \theta_\alpha\})(\alpha - 1\{X_h < \theta_\alpha\})$ with $\alpha \in (0,1)$ and $h = 1, 2, \ldots$, where θ_α is the α-quantile of the marginal distribution of $(X_i)_{i\in\mathbb{Z}}$. Let $\xi_i = (X_{i-h}, X_i)^\top$ and $f_\theta(\xi_i) = (\alpha - 1\{X_{i-h} < \theta\})(\alpha - 1\{X_i < \theta\})$. Under the null hypothesis of no directional predictability, we have $\mathbb{E}f_{\theta_\alpha}(\xi_0) = 0$ for all $h = 1, 2, \ldots$. Let $\hat{\theta}_{n,\alpha}$ be the sample α-quantile and replace population moments by sample moments to obtain $(n-h)^{-1}\sum_{i=1+h}^n f_{\hat{\theta}_{n,\alpha}}(\xi_i)$, the sample version of $\mathbb{E}f_{\theta_\alpha}(\xi_0)$. Apart from a scaling factor, the asymptotic null distribution of the sample quantilogram can be determined through the decomposition

$$(n-h)^{-1/2} \sum_{i=1+h}^n f_{\hat{\theta}_{n,\alpha}}(\xi_i) = \sqrt{n-h}\, \mathbb{E}f_{\hat{\theta}_{n,\alpha}}(\xi_0) + \nu_{n-h} f_{\theta_\alpha} + \nu_{n-h}(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}).$$

If the distribution of X_i is smooth, the delta method can be used to control the first term on the right and, under dependence conditions, an ordinary central limit theorem applies to the second term. Further, we have $\rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) \to_p 0$ whenever $\hat{\theta}_{n,\alpha} \to_p \theta_\alpha$ (see Example 5 below). Hence, we can take $\mathcal{F} = \{f_\theta : \theta \in \Theta\}$, where Θ is a compact neighborhood of θ_α, and as long as (2.1) holds, the third term on the right-hand side of the preceding display converges to zero in probability because in large samples

$$\mathbb{P}\left(\nu_{n-h}(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) > \eta, \rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) < \delta\right) \le \mathbb{P}^*\left(\sup_{f_\theta\in\mathcal{F}:\rho(f_\theta-f_{\theta_\alpha})<\delta} |\nu_{n-h}(f_\theta - f_{\theta_\alpha})| > \eta\right).$$

2

Example 2 (Robust M-estimators of location). Robust location estimators can often be defined implicitly as an M-estimator $\hat{\theta}_n$ that nearly solves $n^{-1}\sum_{i=1}^n f_\theta(\xi_i) = 0$ in the sense that $\sum_{i=1}^n f_{\hat{\theta}_n}(\xi_i) = o_p(\sqrt{n})$. Popular examples include the median with $f_\theta(x) = \operatorname{sign}(x-\theta)$ and Huber estimators with $f_\theta(x) = -\Delta 1\{x-\theta < -\Delta\} + (x-\theta)1\{|x-\theta| \le \Delta\} + \Delta 1\{x-\theta > \Delta\}$ for some $\Delta > 0$. Add and subtract in $\sum_{i=1}^n f_{\hat{\theta}_n}(\xi_i) = o_p(\sqrt{n})$ to see that stochastic equicontinuity implies $\sqrt{n}\,\mathbb{E}f_{\hat{\theta}_n}(\xi_0) + \nu_n f_{\theta_0} = o_p(1)$. The limiting behavior of $\sqrt{n}(\hat{\theta}_n - \theta_0)$ can then again be determined through the delta method and a central limit theorem.

Example 3 (Stochastic dominance). When comparing two stationary time series $(X_{i,1})_{i\in\mathbb{Z}}$ and $(X_{i,2})_{i\in\mathbb{Z}}$, $X_{i,1}$ is said to (weakly) stochastically dominate $X_{i,2}$ over Θ if $\mathbb{P}(X_{0,1} \le \theta) \le \mathbb{P}(X_{0,2} \le \theta)$ for all $\theta \in \Theta$. Let $\xi_i = (X_{i,1}, X_{i,2})^\top$ and $f_\theta(\xi_i) = 1\{X_{i,1} \le \theta\} - 1\{X_{i,2} \le \theta\}$. The null of stochastic dominance can be expressed as $\sup_{\theta\in\Theta} \mathbb{E}f_\theta(\xi_0) \le 0$. Linton et al. (2005) study weak convergence of the rescaled sample equivalent $\sup_{\theta\in\Theta} n^{-1/2}\sum_{i=1}^n f_\theta(\xi_i)$ to the supremum of a certain Gaussian process when the null hypothesis is satisfied with $\sup_{\theta\in\Theta} \mathbb{E}f_\theta(\xi_0) = 0$. This convergence follows from the continuous mapping theorem as long as (i) (Θ, ρ) is a totally bounded pseudometric space, (ii) $(\nu_n f_{\theta_1}, \ldots, \nu_n f_{\theta_k})^\top$ converges in distribution for every finite set of points $\theta_1, \ldots, \theta_k$ in Θ, and (iii) $\nu_n f_\theta$ is stochastically equicontinuous; see, e.g., van der Vaart (1998, p. 261). Condition (i) can be shown to hold if Θ is bounded and ξ_i has Lipschitz continuous marginal distribution functions. Condition (ii) can be verified with the help of a multivariate central limit theorem.

Example 4 (Censored quantile regression). Volgushev et al. (2012) develop Bahadur representations for quantile regression processes arising from a linear latent variable model with outcome T_i and covariate vector Z_i. Denote the random censoring time by C_i. Only Z_i, the random event time $\min\{T_i, C_i\}$, and the associated censoring indicator $1\{T_i \le C_i\}$ are observed. Let $\xi_i = (T_i, C_i, Z_i^\top)^\top$ and $f_\theta(\xi_i) = Z_i 1\{T_i \le C_i\}1\{T_i \le Z_i^\top\theta\}$. As Volgushev et al. point out in their Remark 3.2, a key condition for the validity of their Bahadur representations under dependence is stochastic equicontinuity of $\nu_n f_\theta$.

Stochastic equicontinuity cannot hold without restrictions on the complexity of the set $\mathcal{F}$; see, e.g., Andrews (1994, pp. 2252–2253). Here, complexity of $\mathcal{F}$ is measured via its bracketing number $N = N(\delta, \mathcal{F})$, the smallest number for which there are functions $f_1, \ldots, f_N \in \mathcal{F}$ and functions $b_1, \ldots, b_N$ (not necessarily in $\mathcal{F}$) such that $\rho(b_k) \le \delta$ and $|f - f_k| \le b_k$ for all $1 \le k \le N$. In addition, some restrictions are required on the memory of the time series. For processes of the form (1.1), the memory is most easily controlled by comparing ξ_i to a slightly perturbed version of itself (see Wu, 2005). Let $(\varepsilon_i^*)_{i\in\mathbb{Z}}$ be an iid copy of $(\varepsilon_i)_{i\in\mathbb{Z}}$, so that the difference between ξ_i and $\xi_i' := \xi(\varepsilon_i, \ldots, \varepsilon_1, \varepsilon_0^*, \varepsilon_{-1}^*, \ldots)$ are the inputs prior to period 1. Assume the following:

Assumption A. *Let $\mathcal{F}$ be a uniformly bounded class of real-valued functions with bracketing numbers $N(\delta, \mathcal{F}) < \infty$. Then there exists some $\alpha \in (0,1)$ and $p > 0$ such that*
   (i) $\sup_{f\in\mathcal{F}} \|f(\xi_n) - f(\xi_n')\|_p = O(\alpha^n)$ *and*
   (ii) $\max_{1\le k\le N(\delta,\mathcal{F})} \|b_k(\xi_n) - b_k(\xi_n')\|_p = O(\alpha^n)$ *for any given $\delta > 0$.*

Remarks. (i) Assumption A is a short-range dependence condition. Proposition 1 below presents a device to establish this condition and shows that Assumption A often imposes

only mild restrictions on the dependence structure. At the end of this section, I provide a detailed discussion of how to verify this assumption for Examples 1–4.

(ii) Because $\mathcal{F}$ is assumed to be uniformly bounded, the bounding functions b_k can be chosen to be bounded as well. Hence, in view of Lemma 2 of Wu and Min (2005), the exact choice of p is irrelevant, for if Assumption A holds for some p, then it holds for all $p > 0$.

Assumption A and a complexity requirement on $\mathcal{F}$ given by a bracketing integral imply a strong form of stochastic equicontinuity. The following theorem (the proof of which is found in the Appendix) is similar to Andrews and Pollard's (1994) Theorem 2.2 with their mixing condition replaced by Assumption A. It implies (2.1) via the Markov inequality.

Theorem 1. *Suppose that Assumption A holds and $\int_0^1 x^{-\gamma/(2+\gamma)} N(x, \mathcal{F})^{1/Q} \, dx < \infty$ for some $\gamma > 0$ and an even integer $Q \geq 2$. Then for every $\epsilon > 0$, there is a $\delta > 0$ such that*

$$\limsup_{n \to \infty} \mathbb{E}^* \left(\sup_{f,g \in \mathcal{F}: \rho(f-g) < \delta} |\nu_n(f-g)| \right)^Q < \epsilon.$$

Remark. A useful feature of this theorem is that the constants γ and Q are not connected to the dependence measures as in Andrews and Pollard (1994). In contrast to their result, γ and Q can therefore be chosen to be as small and large, respectively, as desired to make the bracketing integral converge without restricting the set of time series under consideration.

As a referee points out, Assumption A is not primitive and therefore the application at hand determines how difficult it is to verify this assumption. Suitable sufficient conditions can be obtained, e.g., if $\|f(\xi_n) - f(\xi_n')\|_p$ and $\|b_k(\xi_n) - b_k(\xi_n')\|_p$ can be uniformly bounded above by constant multiples of $\|\xi_n - \xi_n'\|_q$ for some $q > 0$. Since the expectation operator is a smoothing operator, these bounds can hold even if f and b_k are not Lipschitz or, more generally, Hölder continuous. Assumption A is then satisfied as long as the geometric moment contraction (GMC) property of Wu and Min (2005) holds, i.e., there is some $\beta \in (0, 1)$ and $q > 0$ such that $\|\xi_n - \xi_n'\|_q = O(\beta^n)$. Time series models with the GMC property include, among many others, stationary (causal) ARMA, ARCH, GARCH, ARMA-ARCH, ARMA-GARCH, asymmetric GARCH, generalized random coefficient autoregressive, and quantile autoregressive models; see Shao and Wu (2007) and Shao (2011) for proofs and more examples.

The problem in Examples 1–4 and in a variety of other applications is the appearance of one or more indicator functions that cause kinks or discontinuities in the objective function. The following result (a generalization of Proposition 3.1 of Hagemann, 2011) combines the GMC property and smoothness conditions on the distribution of the underlying stochastic process to generate the kinds of bounds needed for Assumption A when indicator functions are present. I discuss the result in the examples below. Here and in the remainder of the paper, if ξ_i is vector-valued and has a subvector X_i, then the corresponding subvector of the perturbed version ξ_i' is denoted by X_i'.

Proposition 1. *Suppose $(\xi_i)_{i \in \mathbb{Z}} = (U_i, V_i^\top, W_i^\top)_{i \in \mathbb{Z}}^\top$ has the form (1.1) and ξ takes values in $\mathcal{U} \times \mathcal{V} \times \mathcal{W} \subseteq \mathbb{R} \times \mathbb{R}^l \times \mathbb{R}^m$. Assume that uniformly in $w \in \mathcal{W}$, $\mathbb{P}(U_i \leq x \mid W_i = w)$ is Lipschitz in x on an open interval containing $\mathcal{X} = \{v^\top \lambda \mid v \in \mathcal{V}, \lambda \in \Lambda \subseteq \mathbb{R}^l\}$. Suppose $(U_i, V_i^\top)_{i \in \mathbb{Z}}^\top$*

has the GMC property. If either (i) $\mathcal{V}$ *is a singleton, or* (ii) *there is a measurable function* $g\colon \mathcal{W} \to \mathcal{V}$ *with* $g(W_i) = V_i$ *and* Λ *is bounded, or* (iii) U_i *and* V_i *are independent conditional on* W_i *and* Λ *is bounded, then* $\sup_{\lambda \in \Lambda} \|1\{U_n < V_n^\top \lambda\} - 1\{U_n' < V_n'^\top \lambda\}\|_p = O(\alpha^n)$ *and* $\sup_{\lambda \in \Lambda} \|1\{U_n \le V_n^\top \lambda\} - 1\{U_n' \le V_n'^\top \lambda\}\|_p = O(\alpha^n)$ *for some* $\alpha \in (0,1)$ *and all* $p > 0$.

Before concluding this section, the next four examples illustrate how to apply Theorem 1 and how to verify Assumption A in practice.

Example 5 (Quantilograms, continued). Suppose $F_X(\theta) := \mathbb{P}(X_0 \le \theta)$ is Lipschitz on an open interval containing Θ. Take a grid of points $\min \Theta := t_0 < t_1 < \cdots < t_N =: \max \Theta$ and let $b_k(\xi_i) = 1\{X_{i-h} < t_k\} - 1\{X_{i-h} < t_{k-1}\} + 1\{X_i < t_k\} - 1\{X_i < t_{k-1}\}$. Given a $\theta \in \Theta$, we can then find an index k such that $|f_\theta - f_{t_k}| \le b_k$, where I used the fact that $|\alpha - 1\{\cdot\}| \le \max\{\alpha, 1 - \alpha\} < 1$. Moreover, by stationarity $\rho(b_k) \le 2\|1\{X_0 < t_k\} - 1\{X_0 < t_{k-1}\}\|_2 \le 2\sqrt{F_X(t_k) - F_X(t_{k-1})}$, which is bounded above by a constant multiple of $\sqrt{t_k - t_{k-1}}$ due to Lipschitz continuity. Hence, if $\rho(b_k) \le \delta$ for all $1 \le k \le N$, we can choose bracketing numbers with respect to ρ of order $N(\delta, \mathcal{F}) = O(\delta^{-2})$ as $\delta \to 0$ (see Andrews and Pollard, 1994; van der Vaart, 1998, pp. 270–272) and the bracketing integral converges, e.g., for $\gamma = 1$ and $Q = 4$. By the same calculations as for $\rho(b_k)$, all $\theta, \theta' \in \Theta$ satisfy $\rho(f_\theta - f_{\theta'}) = O(|\theta - \theta'|^{1/2})$ as $\theta \to \theta'$ and therefore $\rho(f_{\hat{\theta}_{n,\alpha}} - f_{\theta_\alpha}) \to_p 0$ if $\hat{\theta}_{n,\alpha} \to_p \theta_\alpha$. In addition, suppose the GMC property holds. Because $\|f_\theta(\xi_n) - f_\theta(\xi_n')\|_p \le 2^{1/p}\|1\{X_n < \theta\} - 1\{X_n' < \theta\}\|_p + 2^{1/p}\|1\{X_{n-h} < \theta\} - 1\{X_{n-h}' < \theta\}\|_p$ uniformly in θ by Loève's c_r inequality, apply Proposition 1(i) twice with $U_i = X_i$ and $U_i = X_{i-h}$ to see that Assumption A(i) is also satisfied. In both cases we can take $\Lambda = \Theta$, $V_i \equiv 1$, and $W_i \equiv 0$ (say). The same reasoning applies to b_k.

Example 6 (Robust M-estimators of location, continued). Nearly identical arguments as in the preceding example yield stochastic equicontinuity for the median. For the Huber estimator, take the grid from before and note that we can find a k such that $|f_\theta - f_{t_k}| \le \min\{t_k - t_{k-1}, 2\Delta\} =: b_k$. A routine argument (Andrews and Pollard, 1994; van der Vaart, 1998, Example 19.7, pp. 270–271) yields bracketing numbers of order $N(\delta, \mathcal{F}) = O(\delta^{-1})$ as $\delta \to 0$; the bracketing integral is finite, e.g., for $\gamma = 1$ and $Q = 2$. Assumption A(i) can be verified via the bound $\sup_{\theta \in \Theta} \|f_\theta(\xi_n) - f_\theta(\xi_n')\|_p \le \|\xi_n - \xi_n'\|_p$ and (ii) holds trivially.

Example 7 (Stochastic dominance, continued). Replace (X_{i-h}, X_i) in Example 5 with $(X_{i,1}, X_{i,2})$. A slight modification of the arguments presented there to account for the fact that the distribution functions of $X_{i,1}$ and $X_{i,2}$ are not identical yields stochastic equicontinuity of $\nu_n f_\theta$. Linton et al. (2005) also consider a more general case with $X_{i,j} = Y_{i,j} - Z_{i,j}^\top \eta_{0,j}$ for $j = 1, 2$. Here $(X_{i,1}, X_{i,2})$ is unobserved and $\eta_0 = (\eta_{0,1}^\top, \eta_{0,2}^\top)^\top$ has to be estimated by some $\hat{\eta}_n = (\hat{\eta}_{n,1}^\top, \hat{\eta}_{n,2}^\top)^\top$. Let $\eta = (\eta_1^\top, \eta_2^\top)^\top$ and $f_{\theta,\eta}(\xi_i) = 1\{Y_{i,1} \le Z_{i,1}^\top \eta_1 + \theta\} - 1\{Y_{i,2} \le Z_{i,2}^\top \eta_2 + \theta\}$. If $\sup_{\theta \in \Theta} \mathbb{E} f_{\theta,\eta_0} = 0$, the asymptotic null distribution of the test statistic $\sup_{\theta \in \Theta} n^{-1/2} \sum_{i=1}^n f_{\theta,\hat{\eta}_n}(\xi_i)$ can be derived from an application of the continuous mapping theorem in

$$n^{-1/2} \sum_{i=1}^n \left(f_{\theta,\hat{\eta}_n}(\xi_i) - \mathbb{E} f_{\theta,\eta_0}(\xi_0) \right) = \sqrt{n}\mathbb{E}(f_{\theta,\hat{\eta}_n} - f_{\theta,\eta_0})(\xi_0) + \nu_n f_{\theta,\eta_0} + \nu_n(f_{\theta,\hat{\eta}_n} - f_{\theta,\eta_0}).$$

Under appropriate conditions on $\hat{\eta}_n$, the behavior of the first term on the right follows from the functional delta method; see Lemma 3 and 4 of Linton et al. (2005). The asymptotic properties of the second term are the same as those of $\nu_n f_\theta$ from the simple case where $(X_{i,1}, X_{i,2})$ is observed directly.

Verifying stochastic equicontinuity of $\nu_n f_{\theta,\eta}$ to control the third term is more involved. Let $H_1 \times H_2$ be a bounded neighborhood of $(\eta_{0,1}^\top, \eta_{0,2}^\top)^\top$ and assume the conditional distribution functions of $X_{i,j}$ given $Z_{i,j}$, $j = 1, 2$, are Lipschitz continuous. The bounded set $\Theta \times H_1 \times H_2$ can then be covered by balls with suitably chosen radius $r = r(\delta)$ and centers $(t_k, e_{k,1}^\top, e_{k,2}^\top)^\top$, $1 \le k \le N$, so that the bracketing integral in Theorem 1 converges (Linton et al., 2005, Lemma 1). Here we take $\xi_i = (X_{i,1}, Z_{i,1}^\top, X_{i,2}, Z_{i,2}^\top)^\top$ and

$$
\begin{aligned}
b_k(\xi_i) = {}& 1\{X_{i,1} < Z_{i,1}^\top(e_{k,1} - \eta_{0,1}) + t_k + (|Z_{i,1}| + 1)r\} \\
& - 1\{X_{i,1} \le Z_{i,1}^\top(e_{k,1} - \eta_{0,1}) + t_k - (|Z_{i,1}| + 1)r\} \\
& + 1\{X_{i,2} < Z_{i,2}^\top(e_{k,2} - \eta_{0,2}) + t_k + (|Z_{i,2}| + 1)r\} \\
& - 1\{X_{i,2} \le Z_{i,2}^\top(e_{k,2} - \eta_{0,2}) + t_k - (|Z_{i,2}| + 1)r\},
\end{aligned}
$$

where $|\cdot|$ is Euclidean norm. For fixed k and r, define a function g such that the first term on the right-hand side of the preceding display equals $1\{X_{i,1} < g(Z_{i,1})\}$. Assume ξ_i has the GMC property. By the reverse triangle and Cauchy-Schwartz inequalities, $g(Z_i)$ has this property as well. Apply Proposition 1(ii) with $U_i = X_{i,1}$, $V_i = g(Z_{i,1})$, $W_i = Z_{i,1}$, and $\Lambda = \{1\}$ to establish $\|1\{X_{n,1} < g(Z_{n,1})\} - 1\{X'_{n,1} < g(Z'_{n,1})\}\|_p = O(\alpha^n)$. The same holds for the remaining indicator functions with appropriate choices of g. Conclude $\|b_k(\xi_n) - b_k(\xi'_n)\|_p = O(\alpha^n)$ for all $1 \le k \le N$, which verifies Assumption A(ii). Further, write $1\{Y_{i,j} \le Z_{i,j}^\top \eta_j + \theta\} = 1\{X_{i,j} \le Z_{i,j}^\top(\eta_j - \eta_{0,j}) + \theta\}$ so that for all $p > 0$ we can bound $\sup_{\theta,\eta} \|f_{\theta,\eta}(\xi_n) - f_{\theta,\eta}(\xi'_n)\|_p$, $\theta \in \Theta$ and $\eta \in H_1 \times H_2$, by a constant multiple of

$$
\begin{aligned}
& \sup_{\theta,\eta_1} \|1\{X_{n,1} \le Z_{n,1}^\top(\eta_1 - \eta_{0,1}) + \theta\} - 1\{X'_{n,1} \le Z'^\top_{n,1}(\eta_1 - \eta_{0,1}) + \theta\}\|_p \\
& + \sup_{\theta,\eta_2} \|1\{X_{n,2} \le Z_{n,2}^\top(\eta_2 - \eta_{0,2}) + \theta\} - 1\{X'_{n,2} \le Z'^\top_{n,2}(\eta_2 - \eta_{0,2}) + \theta\}\|_p.
\end{aligned}
$$

View the two suprema in the display as suprema over $\lambda = (\theta, \lambda_j^\top)^\top \in \Lambda = \Theta \times \{\eta_j - \eta_{0,j} : \eta_j \in H_j\}$ for $j = 1, 2$ and apply Proposition 1(ii) twice with $U_i = X_{i,j}$, $V_i = (1, Z_{i,j}^\top)^\top$, $W_i = Z_{i,j}$ to see that Assumption A(i) holds as well.

Example 8 (Censored quantile regression, continued). Following Volgushev et al. (2012, condition C1), assume there is some fixed $\Delta > 0$ with $|Z_i| \le \Delta$. Add and subtract to see that $\sup_{\theta \in \Theta} \|f_\theta(\xi_n) - f_\theta(\xi'_n)\|_p$ does not exceed a constant multiple of

$$
\|Z_n - Z'_n\|_p + \Delta \sup_{\theta \in \Theta} \|1\{T_n < Z_n^\top \theta\} - 1\{T'_n < Z'^\top_n \theta\}\|_p + \Delta \|1\{T_n < C_n\} - 1\{T'_n < C'_n\}\|_p.
$$

If $\xi_i = (T_i, C_i, Z_i^\top)^\top$ has the GMC property, then the first term is of size $O(\alpha^n)$ by Wu and Min's (2005) Lemma 2. Proposition 1(ii) can be used to establish the same order of

magnitude for the second term as long as the conditional distribution of T_i given Z_i satisfies the smoothness condition stated in the proposition. In addition, Volgushev et al. assume that T_i and C_i are independent conditional on Z_i. The size of the third term is then also $O(\alpha^n)$ by Proposition 1(iii). This verifies Assumption A(i). An argument similar to the one provided in Example 7 establishes Assumption A(ii).

References

Andrews, D. W. K. (1994). Empirical process methods in econometrics. In R. F. Engle and D. L. McFadden (Eds.), *Handbook of Econometrics*, Volume IV, Chapter 37, pp. 2248–2294. Elsevier.

Andrews, D. W. K. and D. Pollard (1994). An introduction to functional central limit theorems for dependent stochastic processes. *International Statistical Review 62*, 119–132.

Dedecker, J., P. Doukhan, G. Lang, J. R. León, S. Louhichi, and C. Prieur (2007). *Weak Dependence: With Examples and Applications*. Springer.

Hagemann, A. (2011). Robust spectral analysis. Unpublished manuscript, Department of Economics, University of Illinois, `arXiv:1111.1965v1`.

Hansen, B. E. (1996). Stochastic equicontinuity for unbounded heterogeneous arrays. *Econometric Theory 12*, 347–359.

Linton, O., E. Maasoumi, and Y.-J. Whang (2005). Consistent testing for stochastic dominance under general sampling schemes. *Review of Economic Studies 72*, 735–765.

Linton, O. and Y.-J. Whang (2007). The quantilogram: With an application to evaluating directional predictability. *Journal of Econometrics 141*, 250–282.

Pollard, D. (1985). *Convergence of Stochastic Processes*. Springer.

Shao, X. (2011). Testing for white noise under unknown dependence and its applications to goodness-of-fit for time series models. *Econometric Theory 27*, 312–343.

Shao, X. and W. B. Wu (2007). Asymptotic spectral theory for nonlinear time series. *Annals of Statistics 35*, 1773–1801.

van der Vaart, A. W. (1998). *Asymptotic Statistics*. Cambridge University Press.

Volgushev, S., J. Wagener, and H. Dette (2012). Censored quantile regression processes under dependence and penalization. Unpublished manuscript, Ruhr-Universität Bochum.

Wu, W. B. (2005). Nonlinear system theory: Another look at dependence. *Proceedings of the National Academy of Sciences 102*, 14150–14154.

Wu, W. B. and W. Min (2005). On linear processes with dependent observations. *Stochastic Processes and their Applications 115*, 939–958.

Appendix

A. Proofs

Proof of Theorem 1. This follows from a simple modification of Andrews and Pollard's (1994) proof of their Theorem 2.1. The proof requires three steps: (i) Their "Proof of

inequality (3.2)," (ii) their "Proof of inequality (3.3)," and (iii) their "Comparison of pairs" argument. Replace their i with k and their $\tau(h_i)$ with $\tau(b_k)$; then apply Lemma A.1 below instead of Andrews and Pollard's (1994) Lemma 3.1 in the derivation of their inequality (3.5) to deduce $\| \max_{1 \leq k \leq N} | \nu_n \, b_k | \|_Q \leq C' N^{1/Q} \max\{n^{-1/2}, \max_{1 \leq k \leq N} \tau(b_k)\}$ and use this in (i) instead of their inequality (3.5). Another application of the lemma establishes the required analogue of their inequality (3.5) used in (ii). The same inequality can also be applied in (iii). The other arguments remain valid without changes. $\qquad\square$

Lemma A.1. *Let $\tau(f) := \rho(f)^{2/(2+\gamma)}$ for some $\gamma > 0$ and suppose that Assumption A holds. For all $n \in \mathbb{N}$, all $f, g \in \mathcal{F}$, and every even integer $Q \geq 2$ we have*

$$\mathbb{E}| \nu_n(f-g)|^Q \leq n^{-Q/2} C \Big(\big(\tau(f-g)^2 n \big) + \cdots + \big(\tau(f-g)^2 n \big)^{Q/2} \Big),$$

where C depends only on Q, γ, and α. The inequality remains valid when $f - g$ is replaced by b_k for any given $k \geq 1$.

Proof of Lemma A.1. Let $Z(i) := f(\xi_i) - \mathbb{E}f(\xi_0) - (g(\xi_i) - \mathbb{E}g(\xi_0))$. Assume without loss of generality that $|Z(i)| \leq 1$ for all $i \geq 1$; otherwise rescale and redefine C. Define $Z'(i) = f(\xi_i') - \mathbb{E}f(\xi_0) - (g(\xi_i') - \mathbb{E}g(\xi_0))$ and note that $\mathbb{E}Z(i) = \mathbb{E}Z'(i) = 0$ for all $i \in \mathbb{Z}$ and all $f, g \in \mathcal{F}$ because ξ_i and ξ_i' are identically distributed. For fixed $k \geq 2$, $d \geq 1$, and $1 \leq m < k$, consider integers $i_1 \leq \cdots \leq i_m \leq i_{m+1} \leq \cdots \leq i_k$ so that $i_{m+1} - i_m = d$. Since $Z(i)$ and $Z'(i)$ are stationary, repeatedly add and subtract to see that

$$\left| \mathbb{E}Z(i_1)Z(i_2) \cdots Z(i_k) - \mathbb{E}Z(i_1)Z(i_2) \cdots Z(i_m)\mathbb{E}Z(i_{m+1}) \cdots Z(i_k) \right|$$

$$= \left| \mathbb{E}Z(i_1 - i_m)Z(i_2 - i_m) \cdots Z(i_k - i_m) \right.$$

$$\left. - \mathbb{E}Z(i_1 - i_m)Z(i_2 - i_m) \cdots Z(0)\mathbb{E}Z(d) \cdots Z(i_k - i_m) \right|$$

$$\leq \left| \mathbb{E}Z(i_1 - i_m) \cdots Z(0)\big(Z(d) - Z'(d)\big)Z(i_{m+2} - i_m) \cdots Z(i_k - i_m) \right|$$

$$+ \sum_{j=2}^{k-m-1} \left| \mathbb{E}Z(i_1 - i_m) \cdots Z(0)Z'(d) \times \cdots \right.$$

$$\left. \times \big(Z(i_{m+j} - i_m) - Z'(i_{m+j} - i_m)\big) \cdots Z(i_k - i_m) \right|$$

$$+ \left| \mathbb{E}Z(i_1 - i_m) \cdots Z(0)Z'(d) \cdots Z'(i_k - i_m) \right.$$

$$\left. - \mathbb{E}Z(i_1 - i_m) \cdots Z(0)\mathbb{E}Z(d) \cdots Z(i_k - i_m) \right| \tag{A.1}$$

In particular, the last summand on the right-hand side is zero because $Z(i_1 - i_m) \cdots Z(0)$ and $Z'(d) \cdots Z'(i_k - i_m)$ are independent and $Z(d) \cdots Z(i_k - i_m)$ and $Z'(d) \cdots Z'(i_k - i_m)$ are identically distributed. For a large enough $M > 0$ and some $s > 1$, Assumption A(i) and distributional equivalence of $Z(d)$ and $Z'(d)$ imply $\|Z(d) - Z'(d)\|_s \leq \|f(\xi_d) - f(\xi_d')\|_s + \|g(\xi_d) - g(\xi_d')\|_s \leq 2\sup_{f \in \mathcal{F}} \|f(\xi_d) - f(\xi_d')\|_s \leq M\alpha^d$. Hölder's inequality then bounds the

8

first term on the right-hand side of the preceding display by

$$\|Z(i_1)\cdots Z(i_m)\|_p \|Z(i_{m+2})\cdots Z(i_k)\|_q M\alpha^d, \tag{A.2}$$

where the reciprocals of p, q, and s sum to 1. Proceeding similarly to Andrews and Pollard (1994), another application of the Hölder inequality yields $\|Z(i_1)\cdots Z(i_m)\|_p \leq \left(\prod_{j=1}^m \mathbb{E}|Z(i_j)|^{mp}\right)^{1/(mp)} \leq \tau(f-g)^{(2+\gamma)/p}$ whenever $mp \geq 2$ and similarly $\|Z(i_{m+2}) \times \cdots \times Z(i_k)\|_q \leq \tau(f-g)^{(2+\gamma)/q}$ whenever $(k-m-1)q \geq 2$. Suppose for now that $k \geq 3$. If $k > m+1$, take $s = (\gamma+Q)/\gamma$ and $mp = (k-m-1)q = (k-1)/(1-1/s)$. Decrease the resulting exponent of $\tau(f-g)$ from $Q(2+\gamma)/(Q+\gamma)$ to 2 so (A.2) is bounded by $M\alpha^d\tau(f-g)^2$. If $k \geq 2$ and $k = m+1$, the factor $\|Z(i_{m+2})\cdots Z(i_k)\|_q$ is not present in (A.2), but we can still choose $s = (\gamma+Q)/\gamma$ and $mp = (k-1)/(1-1/s)$ to obtain the same bound. Identical arguments also apply to each of the other summands in (A.1). Hence, we can find some $M' > 0$ so that

$$|\mathbb{E}Z(i_1)Z(i_2)\cdots Z(i_k)| \leq |\mathbb{E}Z(i_1)Z(i_2)\cdots Z(i_m)\mathbb{E}Z(i_{m+1})\cdots Z(i_k)| + M'\alpha^d\tau(f-g)^2.$$

Here M' in fact depends on k, but this does not disturb any of the subsequent steps.

Now replace (A.2) in Andrews and Pollard (1994) by the inequality in the preceding display. In particular, replace their $8\alpha(d)^{1/s}$ with $M'\alpha^d$ and their τ^2 with $\tau(f-g)^2$. The rest of their arguments now go through without changes. The desired result for b_k follows mutatis mutandis: Simply define $Z(i) = b_k(\xi_i)$, repeat the above steps, and invoke Assumption A(ii) in place of Assumption A(i). $\qquad\square$

Proof of Proposition 1. Suppose the conditions for (iii) hold. Add and subtract, then use $|1\{a < b\} - 1\{c < b\}| \leq 1\{|a-b| \leq |a-c|\}$ for all $a, b, c \in \mathbb{R}$ and the Cauchy-Schwartz inequality to see that $|1\{U_n < V_n^\top\lambda\} - 1\{U_n' < V_n'^\top\lambda\}|$ is at most

$$1\{|U_n - V_n^\top\lambda| \leq |U_n - U_n'|\} + 1\{|U_n' - V_n'^\top\lambda| \leq |\lambda||V_n - V_n'|\}, \tag{A.3}$$

where $|\cdot|$ is Euclidean norm.

Consider the second term in the preceding display. Denote the distribution function of U_i conditional on W_i by $F_{U|W}$. By assumption, we can always find a large enough n^* such that for all $n \geq n^*$ the set $\{x \pm |\lambda|\beta^{nq/(1+q)} : x \in \mathfrak{X}, \lambda \in \Lambda\}$ is contained in the interval on which $F_{U|W}$ is Lipschitz. Apply the Markov inequality, the GMC property, conditional independence, and continuity to write

$$\begin{aligned}
\|1\{|U_n' &- V_n'^\top\lambda| \leq |\lambda||V_n - V_n'|\}\|_p^p \\
&\leq \mathbb{P}(|U_n - V_n^\top\lambda| \leq |\lambda|\beta^{\frac{nq}{1+q}}) + \mathbb{E}|V_n - V_n'|^q\beta^{-nq^2/(1+q)} \\
&= \mathbb{E}\big(\mathbb{P}(|U_n - V_n^\top\lambda| \leq |\lambda|\beta^{\frac{nq}{1+q}} \mid V_n, W_n)\big) + O(\beta^{\frac{nq}{1+q}}) \\
&= \mathbb{E}\big(F_{U|W}(V_0^\top\lambda + |\lambda|\beta^{\frac{nq}{1+q}} \mid W_0) - F_{U|W}(V_0^\top\lambda - |\lambda|\beta^{\frac{nq}{1+q}} \mid W_0)\big) + O(\beta^{\frac{nq}{1+q}}) \\
&= O\big(\beta^{\frac{nq}{1+q}}(|\lambda| + 1)\big).
\end{aligned}$$

The first term in (A.3) satisfies $\|1\{|U_n - V_n^\top\lambda| \leq |U_n - U_n'|\}\|_p^p = O(\beta^{nq/(1+q)})$ for the same reasons. Take $\alpha = \beta^{q/(p+pq)}$ and combine these bounds via the Loève c_r inequality to establish the desired result.

Assertion (ii) also follows because the above arguments remain valid with $V_i = g(W_i)$. For (i), we have $V_i \equiv V_i'$ and hence $|1\{U_n < V_n^\top\lambda\} - 1\{U_n' < V_n'^\top\lambda\}| \leq 1\{|U_n - V_n^\top\lambda| \leq |U_n - U_n'|\}$. The second term in (A.3) is then no longer present, which removes the boundedness restriction on Λ. Finally, note that all of the above remains valid when strong inequalities are replaced by weak inequalities. $\square$