

A An Introduction to the Wild Bootstrap

Bootstrap methods aim to evaluate the accuracy of the sample estimates - they are particularly useful when dealing with complicated distributions, or when the assumptions of a parametric procedure are in doubt. Bootstrap methods randomize the dataset used for the sample estimate calculation, so that a new dataset with a similar statistical properties is obtained, e.g. one popular method is resampling. In the wild bootstrap method the observations in the dataset are multiplied by appropriate random numbers. To present the idea behind the wild bootstrap we will discuss a toy example similar to that in [27], and then relate it to the wild bootstrap method used in this article.

Consider a stationary autoregressive-moving-average random process $\{X_i\}_{i \in \mathbb{Z}}$ with zero mean. The normalized sample mean of the process X_t has normal distribution

$$\frac{\sum_{i=1}^N X_i}{\sqrt{n}} \xrightarrow{d} N(0, \sigma_\infty^2), \quad (9)$$

where $\sigma_\infty^2 = \sum_{j=-\infty}^{\infty} \text{cov}(X_0, X_j)$. The variance σ_∞^2 is not easy to estimate (the naive approach of approximating different covariances separately and summing them has several drawbacks, e.g. how many empirical covariances should be calculated?). Using the wild bootstrap method we will construct processes that mimic behaviour of the X_t process and then use them to approximate the distribution of the normalized sample mean, $\frac{\sum_{i=1}^N X_i}{\sqrt{n}}$. The bootstrap process used to randomize the sample meets the following criteria:

- $\{W_{t,n}\}_{1 \leq t \leq n}$ is a row-wise strictly stationary triangular array independent of all X_t , such that $EW_{t,n} = 0$ and $\sup_n E|W_{t,n}^{2+\sigma}| < \infty$ for some $\sigma > 0$.
- The autocovariance of the process is given by $EW_{s,n}W_{t,n} = \rho(|s - t|/l_n)$ for some function ρ , such that $\lim_{u \rightarrow 0} \rho(u) = 1$.
- The sequence $\{l_n\}$ is taken such that $\lim_{n \rightarrow \infty} l_n = \infty$.

A process that fulfils those criteria, given also in the main article, is

$$W_{t,n} = e^{-1/l_n} W_{t-1,n} + \sqrt{1 - e^{-2/l_n}} \epsilon_t \quad (10)$$

We need to show that the distribution of the normalized sample mean of the process $Y_t^n = W_t^n X_t$, where $|t| \leq n$, mimics the distribution $N(0, \sigma_\infty^2)$. It suffices to calculate the expected value and correlations:

$$EY_t^n = EW_t^n X_t = 0, \quad (11)$$

$$\text{cov}(Y_0^n, Y_t^n) = \text{cov}(X_0, X_t) \text{cov}(Y_0^n, Y_t^n) = \text{cov}(X_0, X_t) \rho(|t|/l_n) \rightarrow \text{cov}(X_0, X_t) \quad (12)$$

The asymptotic auto-covariance structure of the process Y_t is the same as the auto-covariance structure of the process X_t . Therefore

$$\frac{\sum_{i=1}^N Y_i}{\sqrt{n}} \xrightarrow{d} N(0, \sigma_\infty^2). \quad (13)$$

This mechanism is used in [20]. Recall that, under some assumptions, a normalized V-statistic can be written as

$$\sum_{k=0}^{\infty} \lambda_k \left(\frac{\sum_{i=1}^n \phi_k(X_i)}{\sqrt{n}} \right)^2 \stackrel{P}{=} \frac{1}{n} \sum_{1 \leq i, j \leq n} h(X_i, X_j)$$

where λ_k are eigenvalues and ϕ_k are eigenfunction of the kernel h , respectively. Since $E\phi_k(X_i) = 0$ (degeneracy condition) one may replace

$$\frac{\sum_{i=1}^n \phi_k(X_i)}{\sqrt{n}}$$

with a bootstrapped version

$$\frac{\sum_{i=1}^n W_t^n \phi_k(X_i)}{\sqrt{n}},$$

and conclude, as in the toy example, that the limiting distribution of the single component of the sum $\sum_k \lambda_k \dots$ remains the same. One of the main contributions of [20] is in showing that the distribution of the whole sum $\sum_k \lambda_k \left(\frac{\sum_{i=1}^n W_i^n \phi_k(X_i)}{\sqrt{n}} \right)^2$ with the components bootstrapped converges in a particular sense (in probability in Prokhorov metric) to the distribution of the normalized V-statistic, $\frac{1}{n} \sum_{1 \leq i, j \leq n} h(X_i, X_j)$.

B Relation between β, ϕ and τ mixing

Strong mixing coefficients. A process is called absolutely regular (β -mixing) if $\beta(m) \rightarrow 0$, where

$$\beta(m) = \frac{1}{2} \sup_n \sup_{i=1}^I \sum_{j=1}^J |P(A_i \cap B_j) - P(A_i)P(B_j)|.$$

The second supremum in the $\beta(m)$ definition is taken over all pairs of finite partitions $\{A_1, \dots, A_I\}$ and $\{B_1, \dots, B_J\}$ of the sample space such that $A_i \in \mathcal{A}_1^n$ and $B_j \in \mathcal{A}_{n+m}^\infty$, and \mathcal{A}_b^c is a sigma field spanned by a subsequence, $\mathcal{A}_b^c = \sigma(Z_b, Z_{b+1}, \dots, Z_c)$. A process is called uniform mixing (ϕ -mixing) if $\phi(m) \rightarrow 0$, where

$$\phi(m) = \sup_n \sup_{A \in \mathcal{A}_1^n} \sup_{B \in \mathcal{A}_{n+m}^\infty} |P(B|A) - P(B)|.$$

The process is called strongly mixing (α -mixing) if $\alpha(m) \rightarrow 0$, where

$$\alpha(m) = \sup_n \sup_{A \in \mathcal{A}_1^n} \sup_{B \in \mathcal{A}_{n+m}^\infty} |P(B \cap A) - P(B)P(A)|.$$

By [7] we have $\alpha(m) \leq \beta(m) \leq \phi(m)$.

Weak mixing coefficients. The process is called $\tilde{\alpha}$ -mixing if $\tilde{\alpha}(m) \rightarrow 0$, where

$$\begin{aligned} \tilde{\alpha}(m) &= \sup_{l \in \mathbb{N}} \frac{1}{l} \sup_{m \leq i_1 \leq \dots \leq i_l} \tilde{\alpha}(\mathcal{F}_0, (Z_{i_1}, \dots, Z_{i_l})) \xrightarrow{r \rightarrow \infty} 0, \text{ where} \\ \tilde{\alpha}(\mathcal{M}, X) &= \sup_{g \in \Lambda} \|E(g(X)|\mathcal{M}) - Eg(X)\|_1 \end{aligned}$$

and Λ is the set of all one-Lipschitz continuous real-valued functions on the domain of X . The other weak mixing coefficient, already introduced, is τ -mixing. [9, Remark 2.4] show that $\tilde{\alpha}(m) \leq 2\alpha(m)$. [10, Proposition 2] relates τ -mixing and $\tilde{\alpha}$ mixing, as follows: if Q_x is the generalized inverse of the tail function

$$Q_x(u) = \inf_{t \in \mathbb{R}} \{P(|X| > t) \leq u\},$$

then

$$\tau(\mathcal{M}, X) \leq 2 \int_0^{\tilde{\alpha}(\mathcal{M}, X)} Q_x(u) du.$$

While this definition can be hard to interpret, it can be simplified in the case $E|X|^p = M$ for some $p > 1$, since via Markov's inequality $P(|X| > t) \leq \frac{M}{t^p}$, and thus $\frac{M}{t^p} \leq u$ implies $P(|X| > t) \leq u$. Therefore $Q'(u) = \frac{M}{p u^{1/p}} \geq Q_x(u)$. As a result, under the assumption that the real valued random variable is p -integrable for some $p > 1$, we have the following inequality

$$\frac{\sqrt[p]{\tilde{\alpha}(\mathcal{M}, X)}}{M} \geq C\tau(\mathcal{M}, X)$$

C Proofs

The proofs are organized as follows

- Subsection C.1 introduces notation and states basic results concerning V -statistics.
- Subsection C.2 provides proofs of Lemmas 2 and 3 that constitute for most of the poof of the Theorem 1.

- Subsection C.2 provides proofs of Lemmas 1 and 5 from which Theorem 2 follows.
- Subsection C.4 provides proof of the Proposition 1.
- Finally, section C.5 lists all auxiliary lemmas.

C.1 Notation and basic facts.

The following Lemma introduces notation and basic facts about V -statistics. These can be found in the [26, Section 5.1.5] (page 178).

Lemma 1. [26, Section 5.1.5] Any core h can be written as a sum of canonical cores h_1, \dots, h_m and a constant h_0

$$\begin{aligned} h(z_1, \dots, z_m) &= h_m(z_1, \dots, z_m) + \sum_{1 \leq i_1 < \dots < i_{m-1} \leq m} h_{m-1}(z_{i_1}, \dots, z_{i_{m-1}}) \\ &+ \dots + \sum_{1 \leq i_1 < i_2 \leq m} h_2(z_{i_1}, z_{i_2}) + \sum_{1 \leq i \leq m} h_1(z_i) + h_0 \end{aligned}$$

We call h_1, \dots, h_m components of a core h . We do not call h_0 a component, its simply a constant. The components are defined in terms of auxiliary functions g_c

$$g_c(z_1, \dots, z_c) = E h(z_1, \dots, z_c, Z_{c+1}^*, \dots, Z_m^*)$$

for each $c = 0, \dots, m-1$ and we put $g_m = h$. We define components as follows

$$h_0 = g_0, \tag{14}$$

$$h_1(z_1) = g_1(z_1) - h_0, \tag{15}$$

$$h_2(z_1, z_2) = g_2(z_1, z_2) - h_1(z_1) - h_1(z_2) - h_0, \tag{16}$$

$$h_3(z_1, z_2, z_3) = g_3(z_1, z_2, z_3) - \sum_{1 \leq i < j \leq 3} h_2(z_i, z_j) - \sum_{1 \leq i \leq 3} h_1(z_i) - h_0, \tag{17}$$

$$\dots, \tag{18}$$

$$h_m(z_1, \dots, z_m) = g_m(z_1, \dots, z_m) - \sum_{1 \leq i_1 < \dots < i_{m-1} \leq m} h_{m-1}(z_{i_1}, \dots, z_{i_{m-1}}) \tag{19}$$

$$- \dots - \sum_{1 \leq i_1 < i_2 \leq m} h_2(z_{i_1}, z_{i_2}) - \sum_{1 \leq i \leq m} h_1(z_i) - h_0. \tag{20}$$

$$\tag{21}$$

[26, Section 5.1.5] shows that components h_c are symmetric (and therefore cores) and canonical. Finally a V -statistic of a core function h can be written as a sum of V -statistics with canonical cores

$$V(h) = V(h_m) + \binom{m}{1} V(h_{m-1}) + \dots + \binom{m}{m-2} V(h_2) + \binom{m}{m-1} V(h_1) + h_0. \tag{22}$$

C.2 Proof of Theorem 1

Lemma 2. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$. If core h is Lipschitz continuous, one-degenerate, bounded, and its h_2 component is a kernel then its normalized V statistic limiting distribution is proportional to its second component normalized V -statistic distribution. Shortly

$$\lim_{n \rightarrow \infty} \varphi(nV(h), \binom{m}{2} nV(h_2)) = 0 \tag{23}$$

where φ denotes Prokhorov metric.

Proof. Lemma 1 shows how to write core h as a sum of its components h_i ,

$$nV(h) = nV(h_m) + \binom{m}{1} nV(h_{m-1}) + \dots + \binom{m}{m-2} nV(h_2) + \binom{m}{m-1} nV(h_1) + h_0. \tag{24}$$

By Lemma 7 all components of h are bounded and Lipschitz continuous. Since h is one-degenerate, $h_0 = 0$ and component $h_1(z)$ is equal to zero everywhere

$$h_1(z) = Eh(z, Z_2^*, \dots, Z_m^*) = 0. \quad (25)$$

By Lemma 14, for $c \geq 3$, $nV(h_c)$ converges to zero in probability. Therefore the behaviour of $V(h)$ is determined by $\binom{m}{m-2}V(h_2) = \binom{m}{2}V(h_2)$. Convergence of $V(h_2)$ follows from the [20, Theorem 2.1]. \square

Lemma 3. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$. If core h is Lipschitz continuous, one-degenerate, bounded, and its h_2 component is a kernel then its normalized and bootstrapped V statistic limiting distribution is same its second component normalized and bootstrapped V -statistic distribution. Shortly

$$\lim_{n \rightarrow \infty} \varphi(nV_b(h), nV_b(h_2)) = 0, \quad (26)$$

in probability, where φ denotes Prokhorov metric.

Proof. We show that the proposition holds for V_{b1} and then we prove that $\varphi(nV_{b2}(h), nV_{b1}(h)) = 0$ converges to zero - the concept is to [20].

nV_{b1} convergence. We write core h as a sum of components h_i (h_0, h_1 are equal to zero and therefore omitted). By Lemma 1

$$nV_{b1}(h) = \frac{1}{n^{m-1}} \sum_{i \in N^m} \left[W_{i_1} W_{i_2} h_m(Z_{i_1}, \dots, Z_{i_m}) + \right. \quad (27)$$

$$\left. \sum_{1 \leq j_1 < \dots < j_{m-1} \leq m} W_{i_1} W_{i_2} h_{m-1}(Z_{i_{j_1}}, \dots, Z_{i_{j_{m-1}}}) + \dots + \sum_{1 \leq j_1 < j_2 \leq m} W_{i_1} W_{i_2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) \right]. \quad (28)$$

Consider a sum associated with h_2

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} \sum_{1 \leq j_1 < j_2 \leq m} W_{i_1} W_{i_2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}). \quad (29)$$

Fix j_1, j_2 . If $j_1 \neq 1, j_2 \neq 2$ then the sum

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} W_{i_2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) \stackrel{L.8}{=} \frac{1}{n^3} \sum_{i \in N^4} W_{i_1} W_{i_2} h_2(Z_{i_3}, Z_{i_4}) = \quad (30)$$

$$\left(\frac{1}{n} \sum_{i \in N^2} h_2(Z_{i_1}, Z_{i_2}) \right) \left(\frac{1}{n} \sum_{i=1}^n W_i \right)^2 \stackrel{L.16}{\rightarrow} 0 \text{ in probability.} \quad (31)$$

If $j_1 = 1$ and $j_2 \neq 2$, then the sum

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} W_{i_2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) \stackrel{L.8}{=} \frac{1}{n^2} \sum_{i \in N^3} W_{i_1} W_{i_3} h_2(Z_{i_1}, Z_{i_3}) = \quad (32)$$

$$\left(\frac{1}{n} \sum_{i \in N^2} W_{i_1} h_2(Z_{i_1}, Z_{i_2}) \right) \left(\frac{1}{n} \sum_{i=1}^n W_i \right) \stackrel{L.15,18}{\rightarrow} 0 \text{ in probability.} \quad (33)$$

The similar reasoning holds for $j_i = 2$ and $j_2 > 2$. The sum associated with h_c for $c > 2$

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} \sum_{1 \leq j_1 < \dots < j_c \leq m} W_{i_1} W_{i_2} h_c(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) \stackrel{L.13}{\rightarrow} 0 \text{ in probability.} \quad (34)$$

Therefore

$$\lim_{n \rightarrow \infty} \left(nV_b(h) - \sum_{i \in N^2} W_{i_1} W_{i_2} h_2(Z_{i_1}, Z_{i_2}) \right) \stackrel{P}{=} 0. \quad (35)$$

what proofs the proposition for V_{b1} .

nV_{b1} convergence. To prove that nV_{b2} converges to the same distribution as nV_{b1} we investigate the difference

$$V_{b1} - V_{b2} = \frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} W_{i_2} h(Z_{i_1}, \dots, Z_{i_m}) - \frac{1}{n^{m-1}} \sum_{i \in N^m} \tilde{W}_{i_1} \tilde{W}_{i_2} h(Z_{i_1}, \dots, Z_{i_m}) = \quad (36)$$

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} W_{i_2} h(\cdot) - \frac{1}{n^{m-1}} \sum_{i \in N^m} (W_{i_1} - \frac{1}{n} \sum_{j=1}^n W_j) (W_{i_2} - \frac{1}{n} \sum_{j=1}^n W_j) h(\cdot) = \quad (37)$$

$$- \left(\frac{2}{n} \sum_{j=1}^n W_j \right) \left(\frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} h(\cdot) \right) + \left(\frac{1}{n^{m-1}} \sum_{i \in N^m} h(\cdot) \right) \left(\frac{1}{n} \sum_{j=1}^n W_j \right)^2. \quad (38)$$

The second term

$$\left(\frac{1}{n^{m-1}} \sum_{i \in N^m} h(Z_{i_1}, \dots, Z_{i_m}) \right) \left(\frac{1}{n} \sum_{j=1}^n W_j \right)^2 \xrightarrow{L.16} 0 \text{ in probability.} \quad (39)$$

Therefore we only need to show that the first term converges to zero

$$\left(\frac{2}{n} \sum_{j=1}^n W_j \right) \left(\frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} h(Z_{i_1}, \dots, Z_{i_m}) \right). \quad (40)$$

Since $\frac{2}{n} \sum_{j=1}^n W_j$ converges in probability to zero (by Lemma 15) we only need to show that $\frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} h(Z_{i_1}, \dots, Z_{i_m})$ converges. Using decomposition from Lemma 1 we write

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} W_{i_1} h(Z_{i_1}, \dots, Z_{i_m}) = \frac{1}{n^{m-1}} \sum_{i \in N^m} \left[W_{i_1} h_m(Z_{i_1}, \dots, Z_{i_m}) + \quad (41)$$

$$\sum_{1 \leq j_1 < \dots < j_{m-1} \leq m} W_{i_1} h_{m-1}(Z_{i_{j_1}}, \dots, Z_{i_{j_{m-1}}}) + \dots + \sum_{1 \leq j_1 < j_2 \leq m} W_{i_1} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) \right]. \quad (42)$$

Term associated with h_2 can be written as

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} \sum_{1 \leq j_1 < j_2 \leq m} W_{i_1} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) = \quad (43)$$

$$= \begin{cases} n^{-1} \sum_{i \in N^2} W_{i_1} h_2(Z_{i_1}, Z_{i_2}) : j_1 = 1 \text{ or } j_1 = 2 \\ (n^{-1} \sum_{i \in N^2} h_2(Z_{i_1}, Z_{i_2})) \left(\frac{1}{n} \sum_{j=1}^n W_j \right) : \text{otherwise} \end{cases}. \quad (44)$$

In the first case ($j_1 = 1$ or $j_1 = 2$) Lemma 18 assures convergence. In the second case we use Lemma 16 to show convergence to zero. Other terms with h_c for $c > 2$

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} \sum_{1 \leq j_1 < \dots < j_c \leq m} W_{i_1} h_{m-1}(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) \xrightarrow{L.13} 0 \text{ in probability.} \quad (45)$$

□

C.3 Proof of Proposition 3

Lemma 4. $nV_{b2}(h)$ converges to some non-zero random variable with finite variance.

Proof. Using decomposition from the Lemma 1 we write core h as a sum of components h_c and h_0

$$nV_{b2}(h) = \frac{1}{n^{m-1}} \sum_{i \in N^m} \left[h_0 \tilde{W}_{i_1} \tilde{W}_{i_2} + \sum_{1 \leq j \leq m} \tilde{W}_{i_1} \tilde{W}_{i_2} h_1(Z_{i_j}) \right. \quad (46)$$

$$\left. \sum_{1 \leq j_1 < j_2 \leq m} \tilde{W}_{i_1} \tilde{W}_{i_2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) + \dots + \tilde{W}_{i_1} \tilde{W}_{i_2} h_m(Z_{i_1}, \dots, Z_{i_m}) \right]. \quad (47)$$

We examine terms of the above sum starting from the one with h_0 - it is equal to zero

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} h_0 \tilde{W}_{i_1} \tilde{W}_{i_2} \stackrel{L.8}{=} \frac{1}{n} h_0 \sum_{i \in N^2} \tilde{W}_{i_1} \tilde{W}_{i_2} = \frac{1}{n} h_0 \left(\sum_{i=1}^n \tilde{W}_i \right)^2 \stackrel{L.19}{=} 0. \quad (48)$$

Term with h_1 is zero as well, to see that fix j and consider

$$T_j = \frac{1}{n^{m-1}} \sum_{i \in N^m} \tilde{W}_{i_1} \tilde{W}_{i_2} h_1(Z_{i_j}). \quad (49)$$

If $j = 1$ then

$$T_1 \stackrel{L.8}{=} \frac{1}{n} \sum_{i \in N^2} \tilde{W}_{i_1} \tilde{W}_{i_2} h_1(Z_{i_1}) = \frac{1}{n} \left(\sum_{i=1}^n \tilde{W}_i h_1(Z_i) \right) \left(\sum_{i=1}^n \tilde{W}_i \right) \stackrel{L.19}{=} 0. \quad (50)$$

If $j = 2$ the same reasoning holds and if $j > 2$

$$T_j \stackrel{L.8}{=} \frac{1}{n^2} \sum_{i \in N^3} \tilde{W}_{i_1} \tilde{W}_{i_2} h_1(Z_{i_3}) = \frac{1}{n} \left(\sum_{i=1}^n h_1(Z_i) \right) \left(\sum_{i=1}^n \tilde{W}_i \right)^2 \stackrel{L.19}{=} 0. \quad (51)$$

Term containing h_2

$$T_{j_1, j_2} = \frac{1}{n^{m-1}} \sum_{i \in N^m} \tilde{W}_{i_1} \tilde{W}_{i_2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) \quad (52)$$

is not zero. In the Lemma 20 we show that for $j_1 = 1$ and $j_2 = 2$ it converges to some non-zero variable. For $j_1 = 1$ and $j_2 > 2$ we have

$$T_{1, j_2} \stackrel{L.8}{=} \frac{1}{n^2} \sum_{i \in N^3} \tilde{W}_{i_1} \tilde{W}_{i_2} h_2(Z_{i_1}, Z_{i_{j_2}}) = \frac{1}{n^2} \left(\sum_{i \in N^2} \tilde{W}_{i_1} h_2(Z_{i_1}, Z_{i_2}) \right) \left(\sum_{i=1}^n \tilde{W}_i \right) \stackrel{L.19}{=} 0. \quad (53)$$

Exactly the same argument works for $T_{j_2, 1}$. If both $j_1 \neq 1$ and $j_2 \neq 2$ then

$$T_{j_1, j_2} \stackrel{L.8}{=} \frac{1}{n^3} \sum_{i \in N^4} \tilde{W}_{i_1} \tilde{W}_{i_2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) = \frac{1}{n^3} \left(\sum_{i \in N^2} h_2(Z_{i_{j_1}}, Z_{i_{j_2}}) \right) \left(\sum_{i=1}^n \tilde{W}_i \right)^2 \stackrel{L.19}{=} 0. \quad (54)$$

Terms containing h_c for $c > 2$

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} \sum_{1 \leq j_1 < \dots < j_c \leq m} \tilde{W}_{i_1} \tilde{W}_{i_2} h_{m-1}(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) \stackrel{L.13}{\rightarrow} 0 \quad (55)$$

converge to zero in probability. \square

Lemma 5. V_{b1} converges to zero in probability.

Proof. The expected value and variance of V_{b1} converge to 0, therefore V_{b1} converges to zero in probability. Indeed for an expected value we have

$$EV_{b1} = \frac{1}{n^m} \sum_{i \in N^m} E W_{i_1} W_{i_2} E h(Z_{i_1}, \dots, Z_{i_m}) = \frac{1}{n^m} \sum_{i \in N^m} \rho(|i_2 - i_1|/l_n) E h(\cdot) \leq \quad (56)$$

$$\frac{1}{n^m} \sum_{i \in N^m} \rho(|i_2 - i_1|/l_n) \|h\|_\infty = \|h\|_\infty \frac{1}{n^2} \sum_{i \in N^2} \rho(|i_2 - i_1|/l_n) \rightarrow 0. \quad (57)$$

Last convergence follows from the fact that $\sum_{i \in N^2} \rho(|i_2 - i_1|/l_n) \leq \sum_{r=1}^{n-1} n \rho(r/l_n) = O(n l_n)$. Similar reasoning shows convergence of EV_{b1}^2 . \square

C.4 Proof of Proposition 1

Proposition. Let k be bounded and Lipschitz continuous, and let $\{X_t\}$ and $\{Y_t\}$ both be τ -dependent with coefficients $\tau(i) = O(i^{-6-\epsilon})$, but independent of each other. Further, let $n_x = \rho_x n$ and $n_y = \rho_y n$ where $n = n_x + n_y$. Then, under the null hypothesis $P_x = P_y$, $\varphi\left(\rho_x \rho_y n \widehat{\text{MMD}}_k, \rho_x \rho_y n \widehat{\text{MMD}}_{\tilde{k},b}\right) \rightarrow 0$ in probability as $n \rightarrow \infty$, where φ is the Prokhorov metric.

Proof. Since $\widehat{\text{MMD}}_k$ is just the MMD between empirical measures using kernel k , it must be the same as the empirical MMD $\widehat{\text{MMD}}_{\tilde{k}}$ with centred kernel $\tilde{k}(x, x') = \langle k(\cdot, x) - Ek(\cdot, X), k(\cdot, x') - Ek(\cdot, X) \rangle_{\mathcal{H}_k}$ according to [24, Theorem 22]. Using the Mercer expansion, we can write

$$\begin{aligned} \rho_x \rho_y n \widehat{\text{MMD}}_k &= \rho_x \rho_y n \sum_{r=1}^{\infty} \lambda_r \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \Phi_r(x_i) - \frac{1}{n_y} \sum_{j=1}^{n_y} \Phi_r(y_j) \right)^2 \\ &= \sum_{r=1}^{\infty} \lambda_r \left(\sqrt{\frac{\rho_y}{n_x}} \sum_{i=1}^{n_x} \Phi_r(x_i) - \sqrt{\frac{\rho_x}{n_y}} \sum_{j=1}^{n_y} \Phi_r(y_j) \right)^2, \end{aligned}$$

where $\{\lambda_r\}$ and $\{\Phi_r\}$ are the eigenvalues and the eigenfunctions of the integral operator $f \mapsto \int f(x) k(\cdot, x) dP_x(x)$ on $L_2(P_x)$. Similarly as in [20, Theorem 2.1], the above converges in distribution to $\sum_{r=1}^{\infty} \lambda_r Z_r^2$, where $\{Z_r\}$ are marginally standard normal, jointly normal and given by $Z_r = \sqrt{\rho_x} A_r - \sqrt{\rho_y} B_r$. $\{A_r\}$ and $\{B_r\}$ are in turn also marginally standard normal and jointly normal, with a dependence structure induced by that of $\{X_t\}$ and $\{Y_t\}$ respectively. This suggests individually bootstrapping each of the terms $\Phi_r(x_i)$ and $\Phi_r(y_j)$, giving rise to

$$\begin{aligned} \widehat{\text{MMD}}_{\tilde{k},b} &= \sum_{r=1}^{\infty} \lambda_r \left(\frac{1}{n_x} \sum_{i=1}^{n_x} \Phi_r(x_i) \tilde{W}_i^{(x)} - \frac{1}{n_y} \sum_{j=1}^{n_y} \Phi_r(y_j) \tilde{W}_j^{(y)} \right)^2 \\ &= \frac{1}{n_x^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_x} \tilde{W}_i^{(x)} \tilde{W}_j^{(x)} \tilde{k}(x_i, x_j) - \frac{1}{n_x^2} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \tilde{W}_i^{(x)} \tilde{W}_j^{(y)} \tilde{k}(x_i, y_j) \\ &\quad - \frac{2}{n_x n_y} \sum_{i=1}^{n_x} \sum_{j=1}^{n_y} \tilde{W}_i^{(x)} \tilde{W}_j^{(y)} \tilde{k}(x_i, y_j). \end{aligned}$$

Now, since \tilde{k} is degenerate under the null distribution, the first two terms (after appropriate normalization) converge in distribution to $\rho_x \sum_{r=1}^{\infty} \lambda_r A_r^2$ and $\rho_y \sum_{r=1}^{\infty} \lambda_r B_r^2$ by [20, Theorem 3.1] as required. The last term follows the same reasoning - it suffices to check part (b) of [20, Theorem 3.1] (which is trivial as processes $\{X_t\}$ and $\{Y_t\}$ are assumed to be independent of each other) and apply the continuous mapping theorem to obtain convergence to $-2\sqrt{\rho_x \rho_y} \sum_{r=1}^{\infty} \lambda_r A_r B_r$ implying that $\widehat{\text{MMD}}_{\tilde{k},b}$ has the same limiting distribution as $\widehat{\text{MMD}}_k$. While we cannot compute \tilde{k} as it depends on the underlying probability measure P_x , it is readily checked that due to the empirical centering of processes $\{\tilde{W}_t^{(x)}\}$ and $\{\tilde{W}_t^{(y)}\}$, $\widehat{\text{MMD}}_{\tilde{k},b} = \widehat{\text{MMD}}_{k,b}$ holds and the claim follows. Note that the result fails to be valid for wild bootstrap processes that are not empirically centred. \square

C.5 Auxiliary results

The following section lists all auxiliary lemmas.

Proposition 4. [20, p.259, Equation 2.1] If process $\{Z_t, \mathcal{F}_t\}_{t \in \mathbb{N}}$ is τ -dependent and \mathcal{F} is rich enough (see [9, Lemma 5.3]), then there exists, for all $t < t_1 < \dots < t_l$, $l \in \mathbb{N}$, a random vector $(Z_{t_1}^*, \dots, Z_{t_l}^*)$ that is independent of \mathcal{F}_t , has the same distribution as $(Z_{t_1}, \dots, Z_{t_l})$ and

$$E\|(Z_{t_1}^*, \dots, Z_{t_l}^*) - (Z_{t_1}, \dots, Z_{t_l})\|_1 \leq l\tau(t_1 - t).$$

Lemma 6. Let $\{Z_i, \mathcal{F}_i\}$ be a τ -mixing sequence, $\{\delta_i\}$ a sequence of i.i.d random variables independent of filtration \mathcal{F} , $(i_1 \leq \dots \leq i_m)$ a non-decreasing sequence of positive integers, k a positive integer such that $1 < k < m$ and $(Z_{i_1}, \dots, Z_{i_m})$ some random vector. Further let $A = (Z_{i_1}, \dots, Z_{i_{k-1}})$, $B = Z_{i_k}$ and $C = (Z_{i_{k+1}}, \dots, Z_{i_m})$, $\mathcal{F}_A = \mathcal{F}_{k-1}$. Then, there exist the independent random variables B^* and C^* , both independent of \mathcal{F}_A , such that

$$E|B - B^*| = \tau(i_k - i_{k+1}) \text{ and } \frac{1}{m-k} E \|C - C^*\|_1 \leq \tau(i_{k+1} - i_k) \quad (58)$$

Proof. Let $\mathcal{F}_B = \mathcal{F}_k$. We first use [20, Equation 2.1]] (also [9, Lemma 5.3]) to construct C^* such that $\frac{1}{m-k} E \|C - C^*\|_1 \leq \tau(i_{k+1} - i_k)$. By construction C^* is independent of \mathcal{F}_B . Since $\mathcal{F}_A \subset \mathcal{F}_B$ and $\sigma(B) \subset \mathcal{F}_B$ and $C^* \perp \sigma(\delta_k)$, then $C^* \perp (\mathcal{F}_A \vee \sigma(B) \vee \sigma(\delta_k))$. Next by [9, Lemma 5.2] we construct B^* such that $E|B - B^*| = \tau(i_{k+1} - i_k)$, and B^* independent of \mathcal{F}_A , but $\mathcal{F}_A \vee \sigma(B) \vee \sigma(\delta_k)$ measurable. Since $\sigma(C^*) \perp (\mathcal{F}_A \vee \sigma(B) \vee \sigma(\delta_k))$ then C^* and B^* are independent. Finally both C^* and B^* are independent of \mathcal{F}_A \square

Lemma 7. If h is bounded and Lipschitz continuous core then its components are also bounded and Lipschitz continuous.

Proof. Note that

$$g_c(z_1, \dots, z_c) = E h(z_1, \dots, z_c, Z_{c+1}^*, \dots, Z_m^*) \leq E \|h\|_\infty. \quad (59)$$

To prove boundedness we use induction - we assume that components with low index are bounded and use the fact that sum of bounded functions is bounded to obtain the required results.

We prove Lipschitz continuity similarly, first by showing that $g_c(z_1, \dots, z_c)$ are Lipschitz continuous with the same coefficient as the core h and then we use the fact that sum of Lipschitz continuous functions is Lipschitz continuous. \square

Lemma 8. Let f be a symmetric function and let $j = \{j_1, \dots, j_q\}$ be a subset of $\{1, \dots, m\}$. Then

$$\sum_{i \in N^m} f(Z_{i_{j_1}}, \dots, Z_{i_{j_q}}) = n^{m-q} \sum_{i \in N^q} f(Z_{i_1}, \dots, Z_{i_q}) \quad (60)$$

Proof. Denote $r = \{r_1, \dots, r_{m-q}\} = \{1, \dots, m\} \setminus j$. Then

$$\begin{aligned} \sum_{i \in N^m} f(Z_{i_{j_1}}, \dots, Z_{i_{j_q}}) &= \sum_{(i_{j_1}, \dots, i_{j_q}) \in N^q} \sum_{(i_{r_1}, \dots, i_{r_{m-q}}) \in N^{m-q}} f(Z_{i_{j_1}}, \dots, Z_{i_{j_q}}) \\ &= \sum_{(i_{j_1}, \dots, i_{j_q}) \in N^q} \left(f(Z_{i_{j_1}}, \dots, Z_{i_{j_q}}) \sum_{(i_{r_1}, \dots, i_{r_{m-q}}) \in N^{m-q}} 1 \right) \\ &= n^{m-q} \sum_{(i_{j_1}, \dots, i_{j_q}) \in N^q} f(Z_{i_{j_1}}, \dots, Z_{i_{j_q}}) = n^{m-q} \sum_{i \in N^q} f(Z_{i_1}, \dots, Z_{i_q}). \end{aligned}$$

\square

Lemma 9. Let $\{Z_t\}$ be a τ -dependent stationary process with $\tau(r) = O(r^{-6-\epsilon})$. Let h be a bounded Lipschitz continuous function of $m \geq 3$ arguments (not necessarily symmetric) such that $\forall j \in \{1, \dots, m\}$,

$$\mathbb{E} h(z_1, \dots, z_{j-1}, Z_j, z_{j+1}, \dots, z_m) = 0. \quad (61)$$

Then,

$$\sum_{i \in [n]^m} |\mathbb{E} h(Z_{i_1}, \dots, Z_{i_m})| = O\left(n^{\lfloor \frac{m}{2} \rfloor} + n^{2\lfloor \frac{m}{2} \rfloor - 4 - \epsilon}\right).$$

Proof. The proof uses the same technique as [1, Lemma 3]. We will focus on ordered m -tuples $1 \leq i_1 \leq \dots \leq i_m \leq n$, and by considering all possible permutations of their indices, we obtain an upper bound

$$\sum_{i \in [n]^m} |Eh(Z_{i_1}, \dots, Z_{i_m})| \leq \sum_{1 \leq i_1 \leq \dots \leq i_m \leq n} \sum_{\pi \in S_m} |Eh(Z_{i_{\pi(1)}}, \dots, Z_{i_{\pi(m)}})|, \quad (62)$$

where (strict) inequality stems from the fact that the m -tuples i with some coinciding entries appear multiple times on the right. Now denote $s = \lfloor \frac{m}{2} \rfloor + 1$ and

$$j_1 = i_2 - i_1; \quad j_l = \min \{i_{2l} - i_{2l-1}, i_{2l-1} - i_{2l-2}\}, \quad l = 2, \dots, s-1; \quad j_s = i_m - i_{m-1}.$$

Let $w(i) = \max \{j_1, \dots, j_s\}$, i.e., $w(i)$ corresponds to the largest minimum gap between an individual entry in the ordered m -tuple i and its neighbours. For example, $w([1, 2, 5, 9, 9]) = 3$. Note that $w(i) = 0$ means that no entry in i appears exactly once. Let us assume that the maximum $w(i) = w > 0$ is obtained at j_r for some $r \in \{1, \dots, s\}$. Let us partition the vector $(Z_{i_1}, \dots, Z_{i_m})$ into three parts:

$$A = (Z_{i_1}, \dots, Z_{i_{2r-2}}), \quad B = Z_{i_{2r-1}}, \quad C = (Z_{i_{2r}}, \dots, Z_{i_m}).$$

Note that if $r = 1$, A is empty and if $r = s$ and m is odd, C is empty but this does not change our arguments below. Using Lemma 6, we can construct B^* and C^* that are independent of A and independent of each other and

$$E \|(A, B, C) - (A, B^*, C^*)\|_1 \leq m\tau(w). \quad (63)$$

Because B^* consists of a singleton and is independent of both A and C^* , (61) implies

$$Eh(A, B^*, C^*) = 0. \quad (64)$$

Thus, for $w(i) = w > 0$, we have that

$$\begin{aligned} |Eh(Z_{i_1}, \dots, Z_{i_m})| &\leq E|h(A, B, C) - h(A, B^*, C^*)| + |Eh(A, B^*, C^*)| \\ &\leq \text{Lip}(h)E \|(A, B, C) - (A, B^*, C^*)\|_1 + 0 \\ &\leq m\text{Lip}(h)\tau(w). \end{aligned}$$

Finally, if the entries within the ordered m -tuple i are permuted, L_1 -norm in (63) remains the same and (64) still holds, so also $|Eh(Z_{i_{\pi(1)}}, \dots, Z_{i_{\pi(m)}})| \leq m\text{Lip}(h)\tau(w) \forall \pi \in S_m$ and

$$\sum_{\pi \in S_m} |Eh(Z_{i_{\pi(1)}}, \dots, Z_{i_{\pi(m)}})| \leq m!m\text{Lip}(h)\tau(w).$$

Let us upper bound the number of ordered m -tuples i with $w(i) = w$. i_1 can take n different values, but since $i_2 \leq i_1 + w$, i_2 can take at most $w + 1$ different values. For $2 \leq l \leq s-1$, since $\min \{i_{2l} - i_{2l-1}, i_{2l-1} - i_{2l-2}\} \leq w$, we can either let i_{2l-1} take up to n different values and let i_{2l} take up to $w + 1$ different values (if $i_{2l} - i_{2l-1} \leq i_{2l-1} - i_{2l-2}$) or let i_{2l-1} take up to $w + 1$ different values and let i_{2l} take up to n different values (if $i_{2l} - i_{2l-1} > i_{2l-1} - i_{2l-2}$), upper bounding the total number of choices for $[i_{2l-1}, i_{2l}]$ by $2n(w + 1)$. Finally, the last term i_m can always have at most $w + 1$ different values. This brings the total number of m -tuples with $w(i) = w$ to at most $2^{s-2}n^{s-1}(w + 1)^s$. Thus, the number of m -tuples with $w(i) = 0$ is $O(n^{s-1})$ and since h is bounded, we have

$$\begin{aligned} &\sum_{1 \leq i_1 \leq \dots \leq i_m \leq n} \sum_{\pi \in S_m} |Eh(Z_{i_{\pi(1)}}, \dots, Z_{i_{\pi(m)}})| \\ &\leq O(n^{s-1}) + \sum_{w=1}^{n-1} \sum_{\substack{i \in [n]^m: \\ w(i)=w}} \sum_{\pi \in S_m} |Eh(Z_{i_{\pi(1)}}, \dots, Z_{i_{\pi(m)}})| \\ &\leq O(n^{s-1}) + 2^{s-2}m!m\text{Lip}(h)n^{s-1} \sum_{w=1}^{n-1} (w + 1)^s \tau(w) \\ &\leq O(n^{s-1}) + Cn^{s-1} \sum_{w=1}^{n-1} w^{s-6-\epsilon} \\ &\leq O(n^{s-1}) + O(n^{2s-6-\epsilon}), \end{aligned}$$

which proves the claim. We have used $\tau(w) = O(w^{-6-\epsilon})$ and collated the constants into C . \square

Lemma 10. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$. If h is a canonical and Lipschitz continuous core of three or more arguments, then

$$\lim_{n \rightarrow \infty} \frac{1}{n^{m-1}} \sum_{i \in N^m} |Eh(Z_{i_1}, \dots, Z_{i_m})| = 0. \quad (65)$$

Proof. We use Lemma 9 to obtain a bound

$$\sum_{i \in N^m} |Eh(Z_{i_1}, \dots, Z_{i_m})| = O\left(n^{\lfloor \frac{m}{2} \rfloor} + n^{2\lfloor \frac{m}{2} \rfloor - 4 - \epsilon}\right). \quad (66)$$

The right hand side of the above equation divided by n^{m-1} converges to zero. \square

Lemma 11. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$. If h is a canonical and Lipschitz continuous core of three or more arguments, then

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2m-2}} \sum_{i \in N^{2m}} E|h(Z_{i_1}, \dots, Z_{i_m})h(Z_{i_{m+1}}, \dots, Z_{i_{2m}})| = 0. \quad (67)$$

Proof. Let $g(z_{i_1}, \dots, z_{i_m}, z_{i_{m+1}}, \dots, z_{i_{2m}}) = h(z_{i_1}, \dots, z_{i_m})h(z_{i_{m+1}}, \dots, z_{i_{2m}})$. Since g meets assumptions of the Lemma 9 and for $m \geq 3$

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2m-2}} (n^m + n^{2m-4-\epsilon}) = 0, \quad (68)$$

the Lemma follows from the lemma 9. \square

Lemma 12. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$. Let h be a canonical and Lipschitz continuous core of c arguments, $3 \leq c \leq m$, and $1 \leq j_1 < \dots < j_c \leq m$ be a sequence of c integers. If Q_{i_1, \dots, i_m} is a random variable independent of $(Z_{i_1}, \dots, Z_{i_m})$ such that $\sup_{i \in N^m} E|Q_i| \leq \infty$ and $\sup_{i \in N^m} \sup_{o \in N^m} E|Q_i Q_o| \leq \infty$ then

$$\lim_{n \rightarrow \infty} \frac{1}{n^{m-1}} E \sum_{i \in N^m} Q_i h(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) \stackrel{P}{=} 0. \quad (69)$$

$$\lim_{n \rightarrow \infty} \frac{1}{n^{2m-2}} E \sum_{i \in N^m} \sum_{o \in N^m} Q_i Q_o h(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) h(Z_{o_{j_1}}, \dots, Z_{o_{j_c}}) \stackrel{P}{=} 0. \quad (70)$$

$$(71)$$

For the first limit notice that

$$\frac{1}{n^{m-1}} E \sum_{i \in N^m} Q_i h(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) \leq \frac{1}{n^{m-1}} \sum_{i \in N^m} |EQ_i| |Eh(Z_{i_{j_1}}, \dots, Z_{i_{j_c}})| \stackrel{8}{\leq} \quad (72)$$

$$\sup_{i \in N^m} |EQ_i| \frac{1}{n^{c-1}} \sum_{i \in N^c} |Eh(Z_{i_1}, \dots, Z_{i_c})| \xrightarrow{10} 0 \text{ in probability.} \quad (73)$$

Similar reasoning, which uses Lemma 11 instead of 10, shows convergence of the second limit.

Lemma 13. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$. Let h be a canonical and Lipschitz continuous core of c arguments, $3 \leq c \leq m$. If Q_{i_1, \dots, i_m} is a random variable independent of $(Z_{i_1}, \dots, Z_{i_m})$ such that $\sup_{i \in N^m} E|Q_i| \leq \infty$ and $\sup_{i \in N^m} \sup_{o \in N^m} E|Q_i Q_o| \leq \infty$ then

$$\lim_{n \rightarrow \infty} \frac{1}{n^{m-1}} \sum_{i \in N^m} \sum_{1 \leq j_1 < \dots < j_c < m} EQ_i h_c(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) \stackrel{P}{=} 0. \quad (74)$$

Proof. For each sequence such that $1 \leq j_1 < \dots < j_c < m$ we apply Lemma 12 and conclude that the random sum

$$\frac{1}{n^{m-1}} \sum_{i \in N^m} EQ_{i_{j_1}, \dots, i_{j_c}} h_c(Z_{i_{j_1}}, \dots, Z_{i_{j_c}}) \quad (75)$$

converges to zero in a probability - from this the proposition follows. \square

Lemma 14. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$. If h is a one-degenerate and Lipschitz continuous core of three or more arguments

$$\lim_{n \rightarrow \infty} nV(h_c) = 0, \quad (76)$$

for $2 < c \leq m$.

Proof. For each c satisfying $2 < c \leq m$, h_c is canonical and Lipschitz continuous. It suffices to put $Q = 1$ and use Lemma 13. \square

Lemma 15. If W_i is a bootstrap process then

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n W_i \stackrel{P}{=} 0. \quad (77)$$

Proof. By the definition of W_i , $E(\sum_{i=1}^n W_i)^2 = O(nl_n)$, $\lim_{n \rightarrow \infty} \frac{l_n}{n} = 0$ and $E \sum_{i=1}^n W_i = 0$. Therefore $\frac{1}{n} \sum_{i=1}^n W_i$ converges to zero in probability. \square

Lemma 16. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$ and W_i is a bootstrap process. Let f be a one-degenerate, Lipschitz continuous, bounded core of at least m arguments, $m \geq 2$. Further assume that f_2 is a kernel. Then for a positive integer p

$$\lim_{n \rightarrow \infty} nV(f) \left(\frac{1}{n} \sum_{i=1}^n W_i \right)^p \stackrel{P}{=} 0 \quad (78)$$

Proof. By the Lemma 15 $\frac{1}{n} \sum_{i=1}^n W_i$ converges to zero in probability. By Theorem 1 $\frac{1}{n^{m-1}} \sum_{i \in N^m} f(Z_{i_m}, \dots, Z_{i_m})$ converges to some random variable. \square

Lemma 17. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$ and W_i is a bootstrap process. Let $x = (w, z)$ and suppose $f(x_1, x_2) = g(w_1)g(w_2)h(z_1, z_2)$ where g is Lipschitz continuous, $E|g(W_1)|^k \leq \infty$ for any finite k and h is symmetric, Lipschitz continuous, degenerate and bounded. Then

$$nV(f) = \frac{1}{n} \sum_{i,j} f(X_i, X_j) = \frac{1}{n} \sum_{i,j} g(W_i)g(W_j)h(Z_i, Z_j) \quad (79)$$

converges in law to some random variable.

Proof. We use [19, Theorem 2.1] to show that $nV(f)$ converges. We check the assumptions A1 - A3

Assumption A1. Point (i) requires that the process (W_n, Z_n) is a strictly stationary sequence of \mathbb{R}^d -values integrable random variables - this follows from the assumptions of this Lemma. For the point (ii) we put $\delta = \frac{1}{3}$ and check condition

$$\sum_{r=1}^{\infty} r\tau(r)^\delta \leq \sum_{r=1}^{\infty} rr^{-6\frac{1}{3}} = \sum_{r=1}^{\infty} r^{-2} < \infty. \quad (80)$$

Assumption A2. Point (i) requires that the function f is symmetric, measurable and degenerate. Symmetry and measurability are obvious and so we check degeneracy condition

$$Eg(W_1)g(w)h(Z_1, z) = Eg(W_1)g(w)Eh(Z_1, z) = 0. \quad (81)$$

Point (ii) requires that for $\nu > (2 - \delta)/(1 - \delta) = 2.5$ (since we have chosen $\delta = \frac{1}{3}$).

$$\sup_{k \in \mathbb{N}} E|f(X_1, X_k)|^\nu < \infty \text{ and } \sup_{k \in \mathbb{N}} E|f(X_1, X_k^*)|^\nu < \infty \quad (82)$$

Both requirements are met since h is bounded and the process $E|g(W_i)|^k \leq \infty$ for any finite k .

Assumption A3. Function f is Lipschitz continuous - this is met since both g and h are Lipschitz continuous. \square

Lemma 18. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$ and W_i is a bootstrap process. If h is a Lipschitz continuous, degenerate and bounded core of two arguments then

$$\frac{1}{n} \sum_{i \in N^2} W_{i_1} h(Z_{i_1}, Z_{i_2}) \quad (83)$$

converges in distribution to some random variable.

Proof.

$$\frac{1}{n} \sum_{i \in N^2} W_{i_1} h(Z_{i_1}, Z_{i_2}) = \frac{1}{4} (V_+ - V_-) \text{ where,} \quad (84)$$

$$V_- = n^{-1} \sum_{i \in N^2} (W_{i_1} - 1) h(Z_{i_1}, Z_{i_2}) (W_{i_2} - 1), \quad (85)$$

$$V_+ = n^{-1} \sum_{i \in N^2} (W_{i_1} + 1) h(Z_{i_1}, Z_{i_2}) (W_{i_2} + 1), \quad (86)$$

are normalized V statistics that converge. To see that we use Lemma 17 with $g_+(x) = x + 1$ and $g_-(x) = x - 1$ respectively. The only non-trivial assumption is that $E|g_+(W_i)|^k < \infty$ and $E|g_-(W_i)|^k < \infty$ - this follows from $E|W_i|^k$. \square

Lemma 19. If $\{W_i\}$ is a bootstrap process then

$$\sum_{i=1}^n \tilde{W}_i = \sum_{i=1}^n \left(W_i - \frac{1}{n} \sum_{j=1}^n W_j \right) = 0. \quad (87)$$

Lemma 20. Assume that the stationary process Z_t is τ -dependent with a coefficient $\tau(i) = i^{-6-\epsilon}$ for some $\epsilon > 0$ and W_i is a bootstrap process. If f is canonical, Lipschitz continuous, bounded core then a random variable

$$\frac{1}{n} \sum_{1 \leq i, j \leq n} \tilde{W}_i \tilde{W}_j f(Z_i, Z_j) \quad (88)$$

converges in law.

Proof.

$$\frac{1}{n} \sum_{1 \leq i, j \leq n} \tilde{W}_i \tilde{W}_j f(Z_i, Z_j) = \frac{1}{n} \sum_{1 \leq i, j \leq n} \left(W_i - \sum_{a=1}^n W_a \right) \left(W_i - \sum_{b=1}^n W_b \right) f(Z_i, Z_j) = \quad (89)$$

$$\frac{1}{n} \sum_{1 \leq i, j \leq n} W_i W_j f(Z_i, Z_j) - \left(\frac{2}{n} \sum_{1 \leq i, j \leq n} f(Z_i, Z_j) \right) \left(\frac{1}{n} \sum_{b=1}^n W_b \right) + \quad (90)$$

$$\left(\frac{1}{n} \sum_{1 \leq i, j \leq n} f(Z_i, Z_j) \right) \left(\frac{1}{n} \sum_{b=1}^n W_b \right)^2. \quad (91)$$

Last two terms converge to zero since by Lemma 19 $\left(\frac{1}{n} \sum_{b=1}^n W_b \right)$ converges to zero and by the Lemma 17 (with $g = 1$) $\frac{1}{n} \sum_{1 \leq i, j \leq n} f(Z_i, Z_j)$ converges in law. The first term converges by the Lemma 17. \square

D Lag-HSIC with $M \rightarrow \infty$

We here consider a multiple lags test described in Section 4.2 where the number of lags $M = M_n$ being considered goes to infinity with the sample size n . Thus, we will be testing if there exists a dependency between X_t and Y_{t+m} for $-M_n \leq m \leq M_n$ where $\{M_n\}$ is an increasing sequence of

positive numbers such that $M_n = o(n^r)$ for some $0 < r \leq 1$, but $\lim_{n \rightarrow \infty} M_n = \infty$. We denote $q_n = 1 - \frac{\alpha}{2M_n+1}$. As before, the shifted time series will be denoted $Z_t^m = (X_t, Y_{t+m})$ and $S_{m,n} = nV(h, Z^m)$ and $F_{b,n}$ is the empirical cumulative distribution function obtained from $nV_b(h, Z)$. We also let F_n and F denote respectively the finite-sample and the limiting distribution under the null hypothesis of $S_{0,n} = nV(h, Z)$ (or, equivalently, of any $S_{m,n}$ since the null hypothesis holds).

Let us assume that we have computed the empirical q_n -quantile based on the bootstrapped samples, denoted by $t_{b,q_n} = F_{b,n}^{-1}(q_n)$. The null hypothesis is then be rejected if the event $\mathcal{A}_n = \{\max_{-M_n \leq k \leq M_n} S_{m,n} > t_{b,q_n}\}$ occurs. By definition, since F is continuous, $F_n(x) \rightarrow F(x)$, $\forall x$. In addition, our Theorem 1 implies that $F_{b,n}(x) \rightarrow F(x)$ in probability. Thus, $|F_{b,n}(x) - F_n(x)| \rightarrow 0$ in probability as well. However, in order to guarantee that $|q_n - F_n(t_{b,q_n})| \rightarrow 0$, which we require for the Type I error control, we require a stronger assumption of uniform convergence, that $\|F_{b,n} - F_n\|_\infty \leq \frac{C}{n^r}$, for some $C < \infty$. Then, by continuity and sub-additivity of probability, the asymptotic Type I error is given by

$$\begin{aligned} \lim_{n \rightarrow \infty} P_{\mathbf{H}_0}(\mathcal{A}_n) &\leq \lim_{n \rightarrow \infty} \sum_{-M_n \leq m \leq M_n} P_{\mathbf{H}_0}(S_{m,n} > t_{b,q_n}) = \\ \lim_{n \rightarrow \infty} (2M_n + 1) (1 - F_n(t_{b,q_n})) &\leq \lim_{n \rightarrow \infty} (2M_n + 1) \left(1 - \left(1 - \frac{\alpha}{2M_n + 1} \right) + \frac{C}{n^r} \right) = \alpha, \end{aligned} \quad (92)$$

as long as $M_n = o(n^r)$. Intuitively, we require that the number of tests being performed increases at a slower rate than the rate of distributional convergence of the bootstrapped statistics.

On the other hand, under the alternative, there exists some m for which $n^{-1}S_{m,n}$ converges to some positive constant. In this case however, we do not have a handle on the asymptotic distribution F of $S_{m,n} = nV(h, Z^m)$: cumulative distribution function obtained from sampling $nV_{b2}(h)$ converges to G (possibly different from F) with a finite variance, while the behaviour of $nV_{b1}(h)$ is unspecified. We can however show that for any such cumulative distribution function G , the Type II error still converges to zero since

$$P_{\mathbf{H}_1}(\mathcal{A}_n) \geq P_{\mathbf{H}_1}(S_{m,n} > G^{-1}(q_n)) = P_{\mathbf{H}_1}(n^{-1}S_{m,n} > n^{-1}G^{-1}(q_n)) \rightarrow 1,$$

which follows from Lemma 21 below that shows that $n^{-1}G^{-1}(q_n)$ converges to zero.

Lemma 21. *If $X \sim G$ is a random variable such that $EX^2 < \infty$, $q_n = 1 - \frac{\alpha}{2M_n+1}$ and $M_n = o(n)$ then $n^{-1}G^{-1}(q_n) \rightarrow 0$.*

Proof. First observe that by Markov inequality $P(X \geq t) \leq \frac{EX^2}{t}$ and therefore $G(t) > g(t) = 1 - \frac{EX^2}{t}$. Therefore, on the interval $(EX, 1)$, $G^{-1}(x) < g^{-1}(x) = \frac{EX^2}{1-x}$. As a result

$$n^{-1}G^{-1}(q_n) \leq n^{-1}g^{-1}(q_n) = n^{-1} \frac{EX^2}{1 - (1 - \frac{\alpha}{2M_n+1})} = \frac{(2M_n+1)EX^2}{\alpha n} \xrightarrow{n \rightarrow \infty} 0. \quad (93)$$

□

E Various comments

E.1 Time complexity

The original HSIC and MMD tests for i.i.d. data, the computational cost of the wild bootstrap approach scales quadratically in the number of samples, and linearly in the number of bootstrap iterations (in the i.i.d. case, these were permutations of the data). The main alternative approaches are the lagged bootstrap of [8], which has the same scaling with data and number of bootstraps, and the spectrogram approach of [4] (note, however, that both these alternative approaches apply only to the independence testing case). The cost of [4] is comparable to our approach, however the statistical power of [4] was much weaker on the data we examined.