SUPPLEMENT TO "INFERENCE UNDER RANDOM LIMIT BOOTSTRAP MEASURES"

GIUSEPPE CAVALIERE
Department of Economics, University of Bologna and Department of Economics, University of Exeter

ILIYAN GEORGIEV
Department of Economics, University of Bologna and Institut d'Anàlisi Econòmica—CSIC

S.1. INTRODUCTION

THIS SUPPLEMENT TO CAVALIERE AND GEORGIEV (2020), CG hereafter, has four main sections. In Section S.2, we present the proofs of the results on weak convergence in distribution formulated in Appendix A of CG. In Sections S.3 and S.4, some derivations pertaining to the applications in, respectively, Sections 2 and 4 of CG are given. Finally, Section S.5 provides a description of the Monte Carlo simulation design used in Section 2 of CG. For notation, see CG. Unless differently specified, all references are to sections, equations and results in CG.

S.2. WEAK CONVERGENCE IN DISTRIBUTION: PROOFS

Throughout this supplement, references to extended Skorokhod coupling are based on Corollary 5.12 of Kallenberg (1997). The exposition could sometimes be shortened by explicitly considering the random measures of interest as random elements of a Polish space of measures. To avoid an extra level of abstraction, we do not adopt this perspective.

PROOF OF LEMMA A.1: The proof of part (a) is a straightforward modification of step 1 in the proof of Theorem 2.1 in Crimaldi and Pratelli (2005), where $X'_n = X_n$ is considered. For part (b), let $X'_n = \phi_n(X_n)$ $(n \in \mathbb{N})$ for some measurable functions ϕ_n. Without loss of generality, we can consider that $\mathcal{S}_X = \mathcal{S}'_X$, for otherwise we could identify X_n and X with some random elements (X_n, Y_n) and (Y, X) of $\mathcal{S}_X \times \mathcal{S}'_X$ for arbitrary degenerate random elements Y_n ($\mathcal{S}'_X$-valued) and Y ($\mathcal{S}_X$-valued) defined on the probability spaces of respectively X_n and X. Then, by extended Skorokhod coupling, consider a single probability space supporting $(\tilde{Z}_n, \tilde{X}_n, \tilde{Z}'_n) \stackrel{d}{=} (Z_n, X_n, Z'_n)$ and $(\tilde{Z}, \tilde{X}, \tilde{Z}') \stackrel{d}{=} (Z, X, Z')$ with the respective $\tilde{X}'_n := \phi_n(\tilde{X}_n)$ such that $(\tilde{Z}_n, \tilde{X}'_n, \tilde{Z}'_n) \stackrel{\text{a.s.}}{\to} (\tilde{Z}, \tilde{X}, \tilde{Z}')$. Then also $\tilde{Z}_n|\tilde{X}_n \stackrel{w}{\to}_w \tilde{Z}|\tilde{X}$ because weak convergence in distribution is a property of the distributions of $(\tilde{Z}_n, \tilde{X}_n)$ and $(\tilde{Z}, \tilde{X})$. From part (a), it follows that $\tilde{Z}_n|\tilde{X}_n \stackrel{w}{\to}_p \tilde{Z}|\tilde{X}$ such that $E\{h(\tilde{Z}_n)|\tilde{X}_n\} \stackrel{p}{\to} E\{h(\tilde{Z})|\tilde{X}\}$ for every $h \in \mathcal{C}_b(\mathcal{S}_Z)$. The convergence $(E\{h(\tilde{Z}_n)|\tilde{X}_n\}, \tilde{Z}_n, \tilde{X}'_n, \tilde{Z}'_n) \stackrel{p}{\to} (E\{h(\tilde{Z})|\tilde{X}\}, \tilde{Z}, \tilde{X}, \tilde{Z}')$ for such h implies (A.3) on a general probability space. *Q.E.D.*

Giuseppe Cavaliere: giuseppe.cavaliere@unibo.it
Iliyan Georgiev: i.georgiev@unibo.it

REMARK S.1: We establish here the "natural" fact that the convergence

$$\left((Z'_n|X'_n), (Z'_n|X'_n), Z'''_n\right) \overset{w}{\to}_w \left((Z'|X'), (Z'|X'), Z'''\right)$$

is equivalent to $((Z'_n|X'_n), Z'''_n) \overset{w}{\to}_w ((Z'|X'), Z''')$ under separability of the space $\mathcal{S}'''_Z$ where Z'''_n, Z''' take values. In fact, in this case, (A.5) with $Z'_n = Z''_n$ and $X'_n = X''_n$ is equivalent to

$$\left(E\{h(Z'_n)|X'_n\}, Z'''_n\right) \overset{w}{\to} \left(E\{h(Z')|X'\}, Z'''\right) \tag{S.1}$$

for all $h \in \mathcal{C}_b(\mathcal{S}'_Z)$, since both (A.5) and (S.1) are equivalent to $uE\{h(Z'_n)|X'_n\} + v\mathbb{I}_{\{Z'''_n \in A\}} \overset{w}{\to} uE\{h(Z')|X'\} + v\mathbb{I}_{\{Z''' \in A\}}$ for all such h, all continuity sets A of the distribution of Z''' and all $(u, v) \in \mathbb{R}^2$, by Theorem 3.1 of Billingsley (1968) and the Cramér–Wold theorem.   □

PROOF OF LEMMA A.2(a): Let (A.4)–(A.5) hold. Then $Z'_n \overset{w}{\to} Z'$ such that the sequence of probability measures P'_n induced by Z'_n is tight. The sequence of conditional measures $Z'_n|X'_n$ has the tight sequence P'_n as its sequence of average measures. As a result, there exists a countable set of continuous and bounded real functions, say $\{h'_i\}_{i \in \mathbb{N}}$, such that the convergence $E\{h'_i(Z'_n)|X'_n\} \overset{a.s.}{\to} E\{h'_i(Z')|X'\}$, were it to hold for all $i \in \mathbb{N}$, would imply $E\{h'(Z'_n)|X'_n\} \overset{a.s.}{\to} E\{h'(Z')|X'\}$ for all continuous and bounded real h' with the domain of h'_i (by Theorem 2.2 of Berti, Pratelli, and Rigo (2006)). Similarly, there exists a sequence of continuous and bounded real functions $\{h''_i\}_{i \in \mathbb{N}}$, such that the convergence $E\{h''_i(Z''_n)|X''_n\} \overset{a.s.}{\to} E\{h''_i(Z'')|X''\}$ for all $i \in \mathbb{N}$ would imply $E\{h''(Z''_n)|X''_n\} \overset{a.s.}{\to} E\{h''(Z'')|X''\}$ for all continuous and bounded real h'' with the domain of h''_i.

Consider the measurable functions H_n with values in $\mathcal{S}'''_Z \times \mathbb{R}^\infty$ defined by

$$H_n\left(X'_n, X''_n, Z'''_n\right) = \left(Z'''_n, \phi'_{n1}\left(X'_n\right), \phi''_{n1}\left(X''_n\right), \phi'_{n2}\left(X'_n\right), \phi''_{n2}\left(X''_n\right), \dots\right)$$

such that a version $\phi'_{ni}(X'_n)$ of $E\{h'_i(Z'_n)|X'_n\}$ and a version $\phi''_{ni}(X''_n)$ of $E\{h''_i(Z''_n)|X''_n\}$ appear respectively at positions $2i$ and $2i + 1$, and the analogous

$$H\left(X', X'', Z'''\right) = \left(Z''', \phi'_1\left(X'\right), \phi''_1\left(X''\right), \phi'_2\left(X'\right), \phi''_2\left(X''\right), \dots\right),$$

where $\phi'_i(X')$ and $\phi''_i(X'')$ are versions of respectively $E\{h'_i(Z')|X'\}$ and $E\{h''_i(Z'')|X''\}$ ($i \in \mathbb{N}$). By separability and Theorem 3.1 of Billingsley (1968), $H_n(X'_n, X''_n, Z'''_n) \overset{w}{\to} H(X', X'', Z''')$ in $\mathcal{S}'''_Z \times \mathbb{R}^\infty$ would follow if

$$\left(\mathbb{I}_{\{Z'''_n \in A\}}, \phi'_{n1}\left(X'_n\right), \phi''_{n1}\left(X''_n\right), \phi'_{n2}\left(X'_n\right), \phi''_{n2}\left(X''_n\right), \dots\right)$$
$$\overset{w}{\to} \left(\mathbb{I}_{\{Z''' \in A\}}, \phi'_1\left(X'\right), \phi''_1\left(X''\right), \phi'_2\left(X'\right), \phi''_2\left(X''\right), \dots\right)$$

in $\mathbb{R}^\infty$ for every continuity set A of the distribution of Z'''. The previous is equivalent to weak convergence of the finite-dimensional distributions of the considered sequences for every such A (Billingsley (1968, p. 19)). Any linear combination of finitely many functions among $\{h'_i\}_{i \in \mathbb{N}}$ is a bounded and continuous real function, and so for $\{h''_i\}_{i \in \mathbb{N}}$, and for any such two linear combinations (say $h' = \sum_{s=1}^m u_s h'_{i_s}$ and $h'' = \sum_{s=1}^l v_s h''_{j_s}$), it holds that $\sum_{s=1}^m u_s \phi'_{ni_s}(X'_n)$ and $\sum_{s=1}^l v_s \phi''_{nj_s}(X''_n)$ are versions of respectively $E\{h'(Z'_n)|X'_n\}$ and $E\{h''(Z''_n)|X''_n\}$, whereas $\sum_{s=1}^m u_s \phi'_{i_s}(X')$ and $\sum_{s=1}^l v_s \phi''_{j_s}(X'')$ are versions of respectively

$E\{h'(Z')|X'\}$ and $E\{h''(Z'')|X''\}$; therefore,

$$\left(\mathbb{I}_{\{Z_n'''\in A\}}, \sum_{s=1}^{m} u_s \phi'_{ni_s}(X_n'), \sum_{s=1}^{l} v_s \phi''_{nj_s}(X_n'')\right) \overset{w}{\to} \left(\mathbb{I}_{\{Z'''\in A\}}, \sum_{s=1}^{m} u_s \phi'_{i_s}(X'), \sum_{s=1}^{l} v_s \phi''_{j_s}(X'')\right)$$

by (A.5) and Theorem 3.1 of Billingsley (1968). By the Cramér–Wold theorem, this implies that the finite-dimensional distributions of $H_n(X_n', X_n'', Z_n''')$ weakly converge to those of $H(Z', X', X'')$. As a result, $H_n(X_n', X_n'', Z_n''') \overset{w}{\to} H(X', X'', Z''')$ in $\mathcal{S}_Z''' \times \mathbb{R}^\infty$.

By extended Skorokhod coupling, $(X_n', X_n'', Z_n', Z_n'', Z_n''')$ and (X', X'', Z', Z'', Z''') can be redefined, maintaining their distribution, on a new probability space where $H_n(X_n', X_n'', Z_n''') \overset{\text{a.s.}}{\to} H(X', X'', Z''')$ in $\mathcal{S}_Z''' \times \mathbb{R}^\infty$ (we subsume the ∼-notation for the redefined variables). On the new probability space, the relevant components of H_n, H are still versions of the conditional expectations for the redefined variables, for conditional expectations are determined up to equivalence by the underlying joint distributions. As a result, $Z_n''' \overset{\text{a.s.}}{\to} Z'''$, $E\{h_i'(Z_n')|X_n'\} \overset{\text{a.s.}}{\to} E\{h_i'(Z')|X'\}$ for all $i \in \mathbb{N}$, and similarly for h_i''. By the choice of $\{h_i'\}_{i\in\mathbb{N}}$ and $\{h_i''\}_{i\in\mathbb{N}}$, on this space $Z_n'|X_n' \overset{w}{\to}_{\text{a.s.}} Z'|X'$ and $Z_n''|X_n'' \overset{w}{\to}_{\text{a.s.}} Z''|X''$.   *Q.E.D.*

PROOF OF LEMMA A.2(b): Let Z_n', Z_n'', Z', Z'' ($n \in \mathbb{N}$) be r.v.'s. By the proof of Kallenberg (2017, Theorem 4.20), on the Skorokhod-coupling space considered in the proof of part (a) it holds that $P(Z_n' \leq \cdot|X_n') \overset{\text{a.s.}}{\to} P(Z' \leq \cdot|X')$ in $\mathcal{D}(\mathbb{R})$ and $P(Z_n'' \leq \cdot|X_n'') \overset{\text{a.s.}}{\to} P(Z'' \leq \cdot|X'')$ in $\mathcal{D}(\mathbb{R})$. Since on this space also $Z_n''' \overset{\text{a.s.}}{\to} Z'''$, (A.6) follows on a general probability space.

Conversely, let (A.6) hold. Notice that $P(Z_n' \leq \cdot|X_n')$, $P(Z_n'' \leq \cdot|X_n'')$, $P(Z' \leq \cdot|X')$ and $P(Z'' \leq \cdot|X'')$ as random elements of $\mathcal{D}(\mathbb{R})$ are measurable transformations of resp. X_n', X_n'', X', and X'' that are determined up to indistinguishability by the joint distributions of respectively (Z_n', X_n'), (Z_n'', X_n''), (Z', X'), and (Z'', X''). By extended Skorokhod coupling, $(X_n', X_n'', Z_n', Z_n'', Z_n''')$ and (X', X'', Z', Z'', Z''') can be redefined, maintaining their distribution, on a new probability space where $Z_n''' \overset{\text{a.s.}}{\to} Z'''$, $P(Z_n' \leq \cdot|X_n') \overset{\text{a.s.}}{\to} P(Z' \leq \cdot|X')$ in $\mathcal{D}(\mathbb{R})$ and $P(Z_n'' \leq \cdot|X_n'') \overset{\text{a.s.}}{\to} P(Z'' \leq \cdot|X'')$ in $\mathcal{D}(\mathbb{R})$. By the proof of Kallenberg (2017, Theorem 4.20), on this space

$$\left(E\{h'(Z_n')|X_n'\}, E\{h''(Z_n'')|X_n''\}, Z_n'''\right) \overset{\text{a.s.}}{\to} \left(E\{h'(Z')|X'\}, E\{h''(Z'')|X''\}, Z'''\right)$$

for all h', $h'' \in \mathcal{C}_b(\mathbb{R})$ and, therefore, (A.5) holds on a general probability space.   *Q.E.D.*

The following corollary, in its simplest version, establishes the "natural" equivalence of $Z_n \overset{w}{\to} Z$ and $Z_n|Z_n \overset{w}{\to}_w Z|Z$ for random elements Z_n, Z of a Polish space.

COROLLARY S.1: *Let (Z_n, X_n) and (Z, X) be random elements such that $Z_n = (Z_n', Z_n'')$ and $Z = (Z', Z'')$ are $\mathcal{S}_Z' \times \mathcal{S}_Z''$-valued, whereas X_n and X are respectively $\mathcal{S}$-valued and $\mathcal{S}_X$-valued ($n \in \mathbb{N}$), with all the mentioned spaces being Polish metric spaces. Then the convergence $(Z_n'|Z_n', Z_n''|X_n) \overset{w}{\to}_w (Z'|Z', Z''|X)$ in the sense of (A.1) is equivalent to the convergence $(Z_n', (Z_n''|X_n)) \overset{w}{\to}_w (Z', (Z''|X))$ in the sense of (A.2). If additionally $\mathcal{S}_Z'' = \mathbb{R}$ and the conditional distribution $Z''|X$ is diffuse, then both convergence facts are equivalent to $(Z_n', P(Z_n'' \leq \cdot|X_n'')) \overset{w}{\to} (Z', P(Z'' \leq \cdot|X''))$ as random elements of $\mathcal{S}_Z' \times \mathcal{D}(\mathbb{R})$.*

PROOF: As in the proof of Lemma A.1, there is no loss of generality in assuming the equality $\mathcal{S}_X = \mathcal{S}$. First, let $(Z_n'|Z_n', Z_n''|X_n) \xrightarrow{w} (Z'|Z', Z''|X)$. By Lemma A.2(a), consider a Skorokhod representation such that $\tilde{Z}_n'|\tilde{Z}_n' \xrightarrow{w}_{\text{a.s.}} \tilde{Z}'|\tilde{Z}'$ and $\tilde{Z}_n''|\tilde{X}_n \xrightarrow{w}_{\text{a.s.}} \tilde{Z}''|\tilde{X}$. Then $h'(\tilde{Z}_n') \xrightarrow{\text{a.s.}} h'(\tilde{Z}')$ for every $h' \in \mathcal{C}_b(\mathcal{S}_Z')$. This implies that $\tilde{Z}_n' \xrightarrow{p} \tilde{Z}'$ by the proof of Proposition 4.3(i) of Crimaldi and Pratelli (2005). As further $E\{h''(\tilde{Z}_n'')|\tilde{X}_n\} \xrightarrow{\text{a.s.}} E\{h''(\tilde{Z}'')|\tilde{X}\}$ for every $h'' \in \mathcal{C}_b(\mathcal{S}_Z'')$, the convergence $(\tilde{Z}_n', E\{h''(\tilde{Z}_n'')|\tilde{X}_n\}) \xrightarrow{p} (\tilde{Z}', E\{h''(\tilde{Z}'')|\tilde{X}\})$ implies that $(Z_n', E\{h''(Z_n'')|X_n\}) \xrightarrow{w} (Z', E\{h''(Z'')|X\})$ for every such h'', which is (A.2). Second, let $(Z_n', E\{h''(Z_n'')|X_n\}) \xrightarrow{w} (Z', E\{h''(Z'')|X\})$ hold for every $h'' \in \mathcal{C}_b(\mathcal{S}_Z'')$. Then $(h'(Z_n'), E\{h''(Z_n'')|X_n\}) \xrightarrow{w} (h'(Z'), E\{h''(Z'')|X\})$ for every $h' \in \mathcal{C}_b(\mathcal{S}_Z')$, by the CMT. The latter statement is equivalent to (A.1) with $Z_n' = X_n'$. Finally, equivalence to the convergence involving the random c.d.f. $P(Z'' \leq \cdot|X'')$ follows from Lemma A.2(b); see also Remark S.1. *Q.E.D.*

PROOF OF THEOREM A.1: As in the proof of Lemma A.1, without loss of generality, we can assume that $\mathcal{S}_X = \mathcal{S}_X'$. By Lemma A.2(a), consider a Skorokhod representation such that $\tilde{Z}_n|\tilde{X}_n \xrightarrow{w}_{\text{a.s.}} \tilde{Z}|\tilde{X}$. Then $h(\tilde{Z}_n)|\tilde{X}_n \xrightarrow{w}_{\text{a.s.}} h(\tilde{Z})|\tilde{X}$ on the Skorokhod-representation space by Theorems 8(i) and 10 of Sweeting (1989). Therefore, $h(Z_n)|X_n \xrightarrow{w}_w h(Z)|X$ on a general probability space. *Q.E.D.*

PROOF OF THEOREM A.2: In terms of conditional expectations, the first part of the theorem asserts that if (A.7) holds and $E\{h(X_n'')|X_n'\} \xrightarrow{w} E\{h(X'')|X'\}$ for all continuous and bounded real functions h with matching domain, where (X_n', X_n'') are X_n-measurable, then the iterated expectations

$$E(z_n|X_n') = E\{E(z_n|X_n)|X_n'\} \quad \text{and} \quad E(z|X') = E\{E(z|X', X'')|X'\}$$

satisfy the convergence

$$\left(E(z_n|X_n'), E(z_n|X_n), X_n', X_n'', Y_n\right) \xrightarrow{w} \left(E(z|X'), E(z|X', X''), X', X'', Y\right). \tag{S.2}$$

We set up the proof in these terms.

By Theorem 2.1 of Crimaldi and Pratelli (2005), $(X_n', X_n'') \xrightarrow{w} (X', X'')$ and $X_n''|X_n' \xrightarrow{w}_w X''|X'$ imply $(X_n', X_n'')|X_n' \xrightarrow{w}_w (X', X'')|X'$; that is, for all $h \in \mathcal{C}_b(\mathcal{S}_X)$, it holds that $E\{h(X_n', X_n'')|X_n'\} \xrightarrow{w} E\{h(X', X'')|X'\}$.

Let ϕ_n and ϕ be measurable real functions such that $\phi_n(X_n)$ and $\phi(X', X'')$ are versions respectively of the conditional expectations $E(z_n|X_n)$ and $E(z|X', X'')$. We proceed in two steps. First, we argue that we can redefine (X_n, Y_n) and (X', X'', Y), maintaining their distribution, on a new probability space where $(\phi_n(X_n), X_n', X_n'', Y_n) \xrightarrow{\text{a.s.}} (\phi(X', X''), X', X'', Y)$ and $E\{h(X_n', X_n'')|X_n'\} \xrightarrow{p} E\{h(X', X'')|X'\}$ for all $h \in \mathcal{C}_b(\mathcal{S}_X')$. Second, we show that on this space $E\{\phi_n(X_n)|X_n'\} \xrightarrow{p} E\{\phi(X', X'')|X'\}$, which yields convergence (S.2) on a general probability space.

STEP 1. Let the measurable function ψ_n be such that $(X_n', X_n'') = \psi_n(X_n)$, thus $(\phi_n(X_n), \psi_n(X_n), Y_n) \xrightarrow{w} (\phi(X', X''), X', X'', Y)$. By extended Skorokhod coupling, there exist a probability space and random elements $(\tilde{X}_n, \tilde{Y}_n) \xlongequal{d} (X_n, Y_n)$, $(\tilde{X}', \tilde{X}'', \tilde{Y}) \xlongequal{d} (X', X'', Y)$ defined on this space such that $(\phi_n(\tilde{X}_n), \psi_n(\tilde{X}_n), \tilde{Y}_n) \xrightarrow{\text{a.s.}} (\phi(\tilde{X}', \tilde{X}''), \tilde{X}',$

$\tilde{X}''$, $\tilde{Y}$). On this space, it also holds that $E\{h(\tilde{X}'_n, \tilde{X}''_n)|\tilde{X}'_n\} \overset{w}{\to} E\{h(\tilde{X}', \tilde{X}'')|\tilde{X}'\}$ for all $h \in \mathcal{C}_b(\mathcal{S}'_X)$ and $(\tilde{X}'_n, \tilde{X}''_n) := \psi_n(\tilde{X}_n)$, as a consequence of the distributional equalities $(\tilde{X}'_n, \tilde{X}''_n) \overset{d}{=} (X'_n, X''_n)$ and $(\tilde{X}', \tilde{X}'') \overset{d}{=} (X', X'')$. Moreover, this convergence can be strengthened to $E\{h(\tilde{X}'_n, \tilde{X}''_n)|\tilde{X}'_n\} \overset{p}{\to} E\{h(\tilde{X}', \tilde{X}'')|\tilde{X}'\}$ for all $h \in \mathcal{C}_b(\mathcal{S}'_X)$ by Lemma A.1(a). The next step of the proof takes place on this special probability space (we subsume the $\sim$-notation).

STEP 2. As $\mathcal{C}_b(\mathcal{S}'_X)$ is dense in the real functions on $\mathcal{S}'_X$ that are integrable w.r.t. the probability measure induced by (X', X''), it follows that for every $\varepsilon \in (0, 1)$ there exists a $\phi_\varepsilon \in \mathcal{C}_b(\mathcal{S}'_X)$ such that $E|\phi_\varepsilon(X', X'') - \phi(X', X'')| < (\varepsilon/5)^2$. We decompose

$$
\begin{aligned}
\big| E\{\phi_n(X_n)|X'_n\} &- E\{\phi(X', X'')|X'\} \big| \\
&\leq E\big\{ \big| \phi_n(X_n) - \phi(X', X'') \big| |X'_n \big\} \\
&\quad + \big| E\{\phi(X', X'') - \phi_\varepsilon(X', X'')|X'_n\} \big| \\
&\quad + \big| E\{\phi_\varepsilon(X', X'') - \phi_\varepsilon(X'_n, X''_n)|X'_n\} \big| \\
&\quad + \big| E\{\phi_\varepsilon(X'_n, X''_n)|X'_n\} - E\{\phi_\varepsilon(X', X'')|X'\} \big| \\
&\quad + \big| E\{\phi_\varepsilon(X', X'') - \phi(X', X'')|X'\} \big|
\end{aligned}
$$

and label the addends on the right-hand side ρ_i, $i = 1, \ldots, 5$, in order of appearance. The term ρ_1 is $o_p(1)$ by Markov's inequality, because $|\phi_n(X_n) - \phi(X', X'')|$ is $o_p(1)$ and is uniformly integrable by the uniform integrability of z_n and Jensen's inequality. Again by Markov's inequality and the choice of ϕ_ε, it follows that $P(\rho_2 \geq \varepsilon/5) \leq \varepsilon/5$. Since $(X'_n, X''_n) \overset{a.s.}{\to} (X', X'')$ and ϕ_ε is continuous, it holds that $|\phi_\varepsilon(X', X'') - \phi_\varepsilon(X'_n, X''_n)| \overset{a.s.}{\to} 0$ and ρ_3 is $o_p(1)$ by Markov's inequality and the bounded convergence theorem. For $h = \phi_\varepsilon$, it holds that $E\{h(X'_n, X''_n)|X'_n\} \overset{p}{\to} E\{h(X', X'')|X'\}$ such that $\rho_4 = o_p(1)$. Finally, $P(\rho_5 \geq \varepsilon/5) \leq \varepsilon/5$ by Markov's inequality, like ρ_2. By combining these results, it follows that

$$
P\big(\big| E\{\phi_n(X_n)|X'_n\} - E\{\phi(X', X'')|X'\} \big| \geq \varepsilon \big) < \varepsilon
$$

for large enough n. This proves $E(z_n|X'_n) = E\{\phi_n(X_n)|X'_n\} \overset{p}{\to} E\{\phi(X', X'')|X'\} = E(z|X')$ on the special probability space, and since also $(E(z_n|X_n), X'_n, X''_n, Y_n) \overset{a.s.}{\to} (E(z|X'), X', X'', Y)$ on that space, (S.2) follows on the original probability spaces.

To prove the second part of the theorem, let $h', h'' \in \mathcal{C}_b(\mathcal{S}_Z)$ be arbitrary. By the arguments in Remark S.1, (A.8) implies (A.7) with $z_n = h'(Z_n)$, $Y_n = E\{h''(Z_n)|X_n\}$, $z = h'(Z)$, $Y = E\{h''(Z)|X', X''\}$. By the first part of the theorem and the arbitrariness of h', h'', (A.9) follows. *Q.E.D.*

PROOF OF LEMMA A.3: Let $\{f_j\}_{j \in \mathbb{N}}$ be a convergence-determining countable set of bounded Lipschitz functions $\mathcal{S}_X \times \mathcal{S}_Z \to \mathbb{R}$, such that $(X_n, Z^*_n) \overset{w}{\to} (X, Z)$ is implied by the convergence

$$
Ef_j(X_n, Z^*_n) \to Ef_j(X, Z) \quad \text{as } n \to \infty \text{ for all } j \in \mathbb{N}. \tag{S.3}
$$

The existence of such $\{f_j\}_{j \in \mathbb{N}}$ follows from the proof of Proposition 3.4.4 of Ethier and Kurtz (2005); see also Proposition 2.2 of Worm and Hille (2011). If $\{n_m\}$ is an arbitrary subsequence of the naturals, there exists a further subsequence $\{n_{m_k}\}$ such that

$X_{n_{m_k}} \overset{\text{a.s.}}{\to} X$ and the (random) conditional distribution of $Z^*_{n_{m_k}}$ given $D_{n_{m_k}}$ a.s. converges to the (random) conditional distribution of Z^* given X' (the latter by Corollary 2.4 of Berti, Pratelli, and Rigo (2006)). In particular, $E\{h(Z^*_{n_{m_k}})|D_{n_{m_k}}\} \overset{\text{a.s.}}{\to} E\{h(Z^*)|X'\}$ as $k \to \infty$ for every $h \in \mathcal{C}_b(\mathcal{S}_Z)$. If we show that $Ef_j(X_{n_{m_k}}, Z^*_{n_{m_k}}) \to Ef_j(X, Z)$ as $k \to \infty$ for all $j \in \mathbb{N}$, (S.3) will follow. Hence, without loss of generality we can take $X_n \overset{\text{a.s.}}{\to} X$ and $E\{h(Z^*_n)|D_n\} \overset{\text{a.s.}}{\to} E\{h(Z^*)|X'\}$ for every $h \in \mathcal{C}_b(\mathcal{S}_Z)$, and prove that, as a result, (S.3) holds.

Write $Z^*_n = \zeta^*_n(D_n, W^*_n)$ and define the measurable functions $\phi_{nj} : \mathcal{S}_X \times \mathcal{S}_D \to \mathbb{R}$ and $\phi_j : \mathcal{S}_X \times \Omega \to \mathbb{R}$ by

$$\phi_{nj}(x, d) := E_{P^b}\{f_j(x, \zeta^*_n(d, W^*_n))\} \quad \text{and} \quad \phi_j(x, \omega) := \int_{\mathcal{S}_Z} f_j(x, z)\nu(dz, X'(\omega)),$$

where ν is a regular conditional distribution of Z^* given X'. First, we show that there exists an event $A \in \mathcal{F}$ with $P(A) = 1$ such that

$$\phi_{nj}(x, D_n(\omega)) \to \phi_j(x, \omega) \quad \text{for all } j \in \mathbb{N}, x \in \mathcal{S}_X, \omega \in A. \tag{S.4}$$

Second, we conclude that $Ef_j(X, Z^*_n) \to Ef_j(X, Z)$ as $n \to \infty$ for all $j \in \mathbb{N}$, and then we obtain (S.3).

Let $\{x_i\}_{i \in \mathbb{N}}$ be a countable dense subset of $\mathcal{S}_X$. As $f_j(x_i, \cdot) \in \mathcal{C}_b(\mathcal{S}_Z)$, it holds that $E\{f_j(x_i, Z^*_n)|D_n\} \overset{\text{a.s.}}{\to} E\{f_j(x_i, Z^*)|X'\}$ (take $h = f_j(x_i, \cdot)$). Since $\phi_{nj}(x_i, D_n)$ and $\phi_j(x_i, \omega)$ are versions of $E\{f_j(x_i, Z^*_n)|D_n\}$ and $E\{f_j(x_i, Z^*)|X'\}$ respectively (see Example 10.1.9 of Dudley (2004, p. 341), for the former and Theorem 5.4 of Kallenberg (1997), for both or the latter), there exist sets $A_{ij} \in \mathcal{F}$ with $P(A_{ij}) = 1$ such that $\phi_{nj}(x_i, D_n(\omega)) \to \phi_j(x_i, \omega)$ for all $\omega \in A_{ij}$ and every $i, j \in \mathbb{N}$. Define $A := \cap_{i,j \in \mathbb{N}} A_{ij}$ with $P(A) = 1$. It then holds that

$$\phi_{nj}(x_i, D_n(\omega)) \to \phi_j(x_i, \omega) \quad \text{for all } i, j \in \mathbb{N}, \omega \in A.$$

Since, for every $x \in \mathcal{S}_X$ and $j \in \mathbb{N}$, $|f_j(x_i, \cdot) - f_j(x, \cdot)| \leq L_j\{\rho_X(x_i, x) \wedge 1\}$ can be made arbitrarily small by an appropriate choice of x_i (where ρ_X is the metric on $\mathcal{S}_X$ and L_j depend on the Lipschitz constants of f_j and on $\sup|f_j| < \infty$), from the definitions of ϕ_{nj} and ϕ_j it follows that $|\phi_{nj}(x_i, D_n(\omega)) - \phi_{nj}(x, D_n(\omega))|$ and $|\phi_j(x_i, \omega) - \phi_j(x, \omega)|$ for every fixed x, j can be made arbitrarily small uniformly over n, ω. From this fact and from the previous display, (S.4) follows for $A = \bigcap_{i,j \in \mathbb{N}} A_{ij}$.

For arbitrary $j, n \in \mathbb{N}$, (S.4) ensures that $\phi_{nj}(X(\omega), D_n(\omega)) \to \phi_j(X(\omega), \omega)$ for all $\omega \in A$, and thus, a.s. Since $\phi_{nj}(X, D_n)$ is a version of $E\{f_j(X, Z^*_n)|D_n, X\}$ (by the product structure of the probability space; see again Example 10.1.9 of Dudley (2004)) and $\phi_j(x, \omega)$ is a version of $E\{f_j(X, Z^*)|X'\}$, it follows that

$$E\{f_j(X, Z^*_n)|D_n, X\} \overset{\text{a.s.}}{\to} \phi_j(X, \omega) = E\{f_j(x, Z^*)|X'\}\big|_{x=X} \overset{(1)}{=} E\{f_j(x, Z)|X'\}\big|_{x=X}$$

$$\overset{(2)}{=} E\{f_j(x, Z)|X\}\big|_{x=X} \overset{(3)}{=} E\{f_j(X, Z)|X\} \quad \text{a.s.},$$

where equalities (1) and (2) follow from the a.s. equality of the conditional distributions $Z^*|X'$, $Z|X'$ and $Z|X$, and equality (3) holds because for X-measurable $\mathcal{S}_X$-valued ξ's, $E\{f_j(x, Z)|X\}|_{x=\xi} = E\{f_j(\xi, Z)|X\}$ a.s. By the bounded convergence theorem, $Ef_j(X, Z^*_n) \to Ef_j(X, Z)$.

Next, $|Ef_j(X_n, Z_n^*) - Ef_j(X, Z_n^*)| \le L_j E\{\rho_X(X_n, X) \wedge 1\} \to 0$ for every $j \in \mathbb{N}$, again by the bounded convergence theorem, as $X_n \overset{\text{a.s.}}{\to} X$. Thus, $Ef_j(X_n, Z_n^*) = Ef_j(X, Z_n^*) + o(1) \to Ef_j(X, Z)$ and (S.3) is proved. This establishes the convergence $(X_n, Z_n^*) \overset{w}{\to} (X, Z)$.

Finally, $(X_n, Z_n^*) \overset{w}{\to} (X, Z)$ and $Z_n^*|D_n \overset{w}{\to}_p Z|X$, where X_n are D_n-measurable, imply that $(\{(X_n, Z_n^*)|D_n\}, X_n) \overset{w}{\to}_w (\{(X, Z)|X\}, X)$, by a straightforward modification of the proof of Theorem 2.1 in Crimaldi and Pratelli (2005). By Theorem A.2 (see also Remark A.1), the latter convergence implies that

$$\big((X_n, Z_n^*)|D_n, (X_n, Z_n^*)|X_n\big) \overset{w}{\to}_w \big((X, Z)|X, (X, Z)|X\big). \tag{S.5}$$

In their turn, $X_n \overset{p}{\to} X$ and $(X_n, Z_n^*)|X_n \overset{w}{\to}_w (X, Z)|X$ imply, by Corollary 4.4 of Crimaldi and Pratelli (2005), that $(X_n, Z_n^*)|X_n \overset{w}{\to}_p (X, Z)|X$, which jointly with (S.5) yields the convergence $(X_n, Z_n^*)|D_n \overset{w}{\to}_p (X, Z)|X$. *Q.E.D.*

S.3. PROOFS OF THE RESULTS IN SECTION 2

PROOF OF EQUATION (2.8): Let $\mathring{\varepsilon}_t := \varepsilon_t - E(\varepsilon_t|\eta_t)$, $t \in \mathbb{N}$. Then $(\mathring{\varepsilon}_t, E(\varepsilon_t|\eta_t), \eta_t)'$ is an i.i.d. sequence with diagonal covariance matrix $\text{diag}(\omega_{\varepsilon|\eta}, 1 - \omega_{\varepsilon|\eta}, 1)$, $\omega_{\varepsilon|\eta} := E\{\text{Var}(\varepsilon_t|\eta_t)\} \in (0, 1)$, and it is a standard fact that

$$n^{-1/2}\left(\sum_{t=1}^{\lfloor n\cdot \rfloor} \mathring{\varepsilon}_t, \sum_{t=1}^{\lfloor n\cdot \rfloor} E(\varepsilon_t|\eta_t), \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t\right) \overset{w}{\to} \big(\omega_{\varepsilon|\eta}^{1/2} B_{y1}, (1 - \omega_{\varepsilon|\eta})^{1/2} B_{y2}, B_\eta\big) \tag{S.6}$$

in $\mathcal{D}_3$, where (B_{y1}, B_{y2}, B_η) is a standard Brownian motion in $\mathbb{R}^3$. Further, by the conditional invariance principle of Rubshtein (1996),

$$n^{-1/2}\sum_{t=1}^{\lfloor n\cdot \rfloor} \mathring{\varepsilon}_t \Bigg| \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t \overset{w}{\to}_p \omega_{\varepsilon|\eta}^{1/2} B_{y1} = \omega_{\varepsilon|\eta}^{1/2} B_{y1}|(B_{y2}, B_\eta) \quad \text{a.s.} \tag{S.7}$$

as a convergence of random measures on $\mathcal{D}$. Since $\sigma(\sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t) = \sigma(\sum_{t=1}^{\lfloor n\cdot \rfloor} E(\varepsilon_t|\eta_t), \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t) = \sigma(X_n)$, the convergence

$$n^{-1/2}\left(\sum_{t=1}^{\lfloor n\cdot \rfloor} \mathring{\varepsilon}_t, \sum_{t=1}^{\lfloor n\cdot \rfloor} E(\varepsilon_t|\eta_t), \sum_{t=1}^{\lfloor n\cdot \rfloor} \eta_t\right)\Bigg| X_n \overset{w}{\to}_w \big(\omega_{\varepsilon|\eta}^{1/2} B_{y1}, (1 - \omega_{\varepsilon|\eta})^{1/2} B_{y2}, B_\eta\big)|(B_{y2}, B_\eta)$$

follows from (S.6) and (S.7) by Theorem 2.1 of Crimaldi and Pratelli (2005), for random measures on $\mathcal{D}_3$. Notice that $(n^{-1}\sum_{t=1}^n \eta_t \mathring{\varepsilon}_t, n^{-1}\sum_{t=1}^n \eta_t E(\varepsilon_t|\eta_t)) \overset{p}{\to} 0$ and the convergence is preserved upon conditioning on X_n. Then, by using conditional convergence to stochastic integrals (Theorem 3 of Georgiev, Harvey, Leybourne, and Taylor (2019)), it

further follows that

$$
\left(n^{-2} M_n, n^{-1} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_t \mathring{\varepsilon}_t, n^{-1} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_t E(\varepsilon_t | \eta_t) \right) \bigg| X_n
$$

$$
= \left(n^{-2} M_n, n^{-1} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_{t-1} \mathring{\varepsilon}_t + o_p(1), n^{-1} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_{t-1} E(\varepsilon_t | \eta_t) + o_p(1) \right) \bigg| X_n
$$

$$
\overset{w}{\to}_w \left(M, \omega_{\varepsilon|\eta}^{1/2} M^{1/2} \xi_1, (1 - \omega_{\varepsilon|\eta})^{1/2} M^{1/2} \xi_2 \right) | (B_{y2}, B_\eta)
$$

with $M := \int B_\eta^2$, $\xi_1 := (\int B_\eta^2)^{-1/2} \int B_\eta \, dB_{y1}$, $\xi_2 := (\int B_\eta^2)^{-1/2} \int B_\eta \, dB_{y2}$ jointly independent and $\xi_i \sim N(0, 1)$, $i = 1, 2$. Then, by Theorem A.1, τ_n of (2.6) satisfies

$$
\left(\tau_n, n^{-2} M_n, n^{-1} \sum_{t=1}^{n} x_t E(\varepsilon_t | \eta_t) \right) \bigg| X_n \overset{w}{\to}_w \left(\tau, M, (1 - \omega_{\varepsilon|\eta})^{1/2} M^{1/2} \xi_2 \right) | (B_{y2}, B_\eta) \quad \text{(S.8)}
$$

with $\tau := M^{-1/2} (\omega_{\varepsilon|\eta}^{1/2} \xi_1 + (1 - \omega_{\varepsilon|\eta})^{1/2} \xi_2)$. This yields (2.8). The bootstrap, instead of estimating consistently the limiting conditional distribution of τ_n given X_n, estimates the random distribution obtained by averaging this limit over ξ_2. As a result, conditionally on X_n the bootstrap p-value is not asymptotically uniformly distributed:

$$
p_n^* | X_n = \Phi\left(\hat{\omega}_\varepsilon^{-1/2} M_n^{1/2} (\hat{\beta} - \beta) \right) | X_n \overset{w}{\to}_w \Phi\left(\omega_{\varepsilon|\eta}^{1/2} \xi_1 + (1 - \omega_{\varepsilon|\eta})^{1/2} \xi_2 \right) | \xi_2, \quad \text{(S.9)}
$$

which is not the c.d.f. of a $U(0, 1)$ r.v.   *Q.E.D.*

S.4. PROOFS OF THE RESULTS IN SECTION 4

PROOF OF EQUATION (4.4): By extended Skorokhod coupling (Corollary 5.12 of Kallenberg (1997)), we can regard the data and U as defined on a single probability space such that $n^{-\alpha/2} x_{\lfloor n \cdot \rfloor} \overset{a.s.}{\to} U(\cdot)$ in $\mathcal{D}$. Then, by a product-space construction, we can extend this space to define also an i.i.d. standard Gaussian sequence $\{\varepsilon_t^*\}$ independent of the data and (by Lemma 5.9 of Kallenberg (1997)), a random element (W, b) of $\mathcal{D} \times \mathbb{R}$ such that $(W, b) | U$ has the conditional distribution specified in the text. Consider outcomes (say ω) in the factor-space of $n^{-\alpha/2} x_{\lfloor n \cdot \rfloor}$ such that $(n^{-\alpha-1} M_n(\omega), n^{-\alpha/2-1} \xi_n(\omega)) \to (M(\omega), \xi(\omega))$, $n^{-(\alpha+1)/2} \sup |x_{\lfloor n \cdot \rfloor}(\omega)| \to 0$ and $M(\omega) > 0$; such outcomes have probability one. Upon fixing such an outcome, the sequence $(n^{1/2} W_n^*, M_n^{1/2} \hat{\beta}^*)$, with randomness originating from $\{\varepsilon_t^*\}$, is tight in $\mathcal{D} \times \mathbb{R}$ because $n^{1/2} W_n^*$ and $M_n^{1/2} \hat{\beta}^*$ are tight in $\mathcal{D}$ and $\mathbb{R}$, respectively, and its finite-dimensional distributions converge, by the multivariate Lyapunov CLT (Bentkus (2005)), to those of (W^ω, b^ω), where W^ω and b^ω are respectively a standard Brownian bridge and a standard Gaussian r.v. with $\text{Cov}(W^\omega(u), b^\omega) = M(\omega)^{-1/2} \xi(\omega) \psi(u)$, $u \in [0, 1]$. It follows by disintegration (Theorem 5.4 of Kallenberg (1997)) that $(n^{1/2} W_n^*, M_n^{1/2} \hat{\beta}^*) | x_{\lfloor n \cdot \rfloor} \overset{w}{\to}_{a.s.} (W, b) | U$, with the limit conditional distribution a.s. equal to $(W, b) | (M, \xi)$, and further that

$$
\left(n^{1/2} W_n^*, M_n^{1/2} \hat{\beta}^*, n^{-\alpha-1} M_n, n^{-\alpha/2-1} \xi_n \right) | x_{\lfloor n \cdot \rfloor} \overset{w}{\to}_{a.s.} (W, b, M, \xi) | (M, \xi)
$$

by Lemma A.3, since $(n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n)$ are $x_{\lfloor n\cdot\rfloor}$-measurable. Still further, by a CMT for a.s. weak convergence (Theorem 10 of Sweeting (1989)),

$$\left(n^{1/2}W_n^*, n^{(\alpha+1)/2}\hat{\beta}^*, n^{-\alpha/2-1}\xi_n\right)|x_{\lfloor n\cdot\rfloor} \xrightarrow{w}_{\text{a.s.}} (W, M^{-1/2}b, \xi)|(M, \xi)$$

on the special probability space. This implies (4.4) on a general probability space. *Q.E.D.*

PROOF OF EQUATION (4.6): Under H_0, by extended Skorokhod coupling (Corollary 5.12 of Kallenberg (1997)), we regard the data and (τ, U) as defined on a single probability space such that $(\tau_n, n^{-\alpha/2}x_{\lfloor n\cdot\rfloor}) \xrightarrow{\text{a.s.}} (\tau, U)$ in $\mathbb{R} \times \mathcal{D}$, and which is extended to support the independent bootstrap sequence $\{\varepsilon_t^*\}$ and (W, b) such that $(W, b)|U$ has the conditional distribution specified in the text. We have by the same argument as for equation (4.5) that, on this space, $\tau_n^* \xrightarrow{w^*}_p \tau|(M, \xi)$ so that $F_n^*(\cdot) := P(\tau_n^* \le \cdot|D_n) \xrightarrow{p} F(\cdot) := P(\tau \le \cdot|M, \xi)$ in $\mathcal{D}(\mathbb{R})$, because F is sample-path continuous (e.g., by Proposition 3.2 of Linde (1989) applied conditionally on M, ξ). As further $\tau_n \xrightarrow{\text{a.s.}} \tau$ on this space, we can collect the previous convergence facts into $(\tau_n, F_n^*) \xrightarrow{p} (\tau, F)$, which proves that on a general probability space equation (4.6) holds.

For use in the discussion of conditional validity, consider again the Skorokhod probability space. On this space, (i) $(\tau_n, n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n) \xrightarrow{\text{a.s.}} (\tau, M, \xi)$, as implied by the a.s. convergence of τ_n and $n^{-\alpha/2}x_{\lfloor n\cdot\rfloor}$, and (ii) $\tau_n^* \xrightarrow{w^*}_p \tau|(M, \xi)$, jointly imply, by Lemma A.3, that $(\tau_n, \tau_n^*, n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n) \xrightarrow{w} (\tau, \tau^*, M, \xi)$ with the conditional distributions $\tau|(M, \xi)$ and $\tau^*|(M, \xi)$ equal a.s. The latter convergence remains valid on general probability spaces. *Q.E.D.*

DETAILS OF REMARK 4.2: By extended Skorokhod coupling (Corollary 5.12 of Kallenberg (1997)), consider a Skorokhod representation of D_n (resp., X_n) and (τ, τ^*, X, X') such that $(\tau_n, \tau_n^*, \phi_n(X_n), \psi_n(D_n)) \xrightarrow{\text{a.s.}} (\tau, \tau^*, X, X')$. Then, by Lemma A.1(a), on the Skorokhod-representation space it holds that $\tau_n|X_n \xrightarrow{w}_p \tau|X$ and $\tau_n^*|D_n \xrightarrow{w}_p \tau^*|X'$, such that on a general probability space $(\tau_n|X_n, \tau_n^*|D_n) \xrightarrow{w}_w (\tau|X, \tau^*|X')$. If the conditional distributions $\tau^*|X'$ and $\tau|X'$ are a.s. equal, (3.4) follows.

PROOF OF THEOREM 4.1: Let $(M_n, \tilde{V}_n) := (n^{-1}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x'_{nt}, n^{-1}\sigma^{-2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x'_{nt}\tilde{e}_{nt}^2)$. As $\tilde{V}_n = n^{-1}\sigma^{-2}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}x'_{nt}\varepsilon_{nt}^2 + o_p(1)$ under H_0 and Assumption $\mathcal{H}$, it further holds that $(M_n, \tilde{V}_n) \xrightarrow{w} (M, V)$ in $\mathcal{D}_{m\times m} \times \mathcal{D}_{m\times m}$. The data $D_n := \{x_{nt}, y_{nt}\}_{t=1}^n$ and the bootstrap multipliers $\{w_t^*\}_{t\in\mathbb{N}}$ can be regarded (upon padding with zeroes) as a random element of the Polish space $(\mathbb{R}^\infty)^{k+2}$. Therefore, by Corollary 5.12 of Kallenberg (1997), there exists a special probability space where (M, V), and for every $n \in \mathbb{N}$, also the original and the bootstrap data can be redefined, maintaining their distribution (we also maintain the notation), such that $(M_n, \tilde{V}_n) \xrightarrow{\text{a.s.}} (M, V)$.

Consider $N_n^* := n^{-1/2}\sigma^{-1}\sum_{t=1}^{\lfloor n\cdot\rfloor} x_{nt}y_t^*$. As, conditionally on the data, N_n^* is a zero-mean Gaussian process with independent increments and variance function $\tilde{V}_n$, the argument for Theorem 5 of Hansen (2000) yields the conditional convergence

$$N_n^*|D_n \xrightarrow{w}_{\text{a.s.}} N|(M, V) \tag{S.10}$$

on the special probability space. Notice that the conditional distributions $N_n^*|D_n$ and $N_n^*|(M_n, \tilde{V}_n)$ are equal a.s. Then the marginal convergence $(M_n, \tilde{V}_n) \overset{\text{a.s.}}{\to} (M, V)$ in $(\mathcal{D}_{m \times m})^2$ and (S.10) jointly imply, by Lemma A.3, that

$$\left(M_n, \tilde{V}_n, N_n^*\right)|D_n = \left(M_n, \tilde{V}_n, N_n^*\right)|(M_n, \tilde{V}_n) \overset{w}{\to}_p (M, V, N)|(M, V) \qquad \text{(S.11)}$$

as a convergence of random measures on $(\mathcal{D}_{m \times m})^2 \times \mathcal{D}_m$, where the first equality is an a.s. equality of conditional distributions.

The proof is completed as in Theorems 5 and 6 of Hansen (2000), by using the expansion $\sup_{r \in [\underline{r}, \overline{r}]} |F_{\lfloor nr \rfloor}^* - \tilde{F}_n(r)| = o_p(1)$ with

$$\tilde{F}_n(r) = \left\| \left(M_n(r) - M_n(r)M_n(1)^{-1}M_n(r)\right)^{-1/2} \left(N^*(r) - M_n(r)M_n(1)^{-1}N_n^*(1)\right) \right\|^2$$

and where convergence is w.r.t. the joint measure over the original and the bootstrap data. As $\tilde{F}_n(r)$ depends on the original data only through $M_n, \tilde{V}_n$, it follows that

$$P^*\left(\max_{r \in [\underline{r}, \overline{r}]} \tilde{F}_n(r) \leq \cdot\right) = P\left(\max_{r \in [\underline{r}, \overline{r}]} \tilde{F}_n(r) \leq \cdot \big| M_n, \tilde{V}_n\right),$$

and since $\{\max_{r \in [\underline{r}, \overline{r}]} \tilde{F}_n(r)\}|(M_n, \tilde{V}_n) \overset{w}{\to}_p \mathcal{F}|(M, V)$ by (S.11) and a CMT for weak convergence in probability (Theorem 10 of Sweeting (1989)), with

$$\mathcal{F} := \sup_{r \in [\underline{r}, \overline{r}]} \left\{ \tilde{N}(r)' \tilde{M}(1)^{-1} \tilde{N}(r) \right\}$$

being the limit from (4.8), also $\max_{r \in [\underline{r}, \overline{r}]} \tilde{F}_n(r) \overset{w^*}{\to}_p \mathcal{F}|(M, V)$. Finally, as $\mathcal{F}_n^* := \max_{r \in [\underline{r}, \overline{r}]} F_{\lfloor nr \rfloor}^* = \max_{r \in [\underline{r}, \overline{r}]} \tilde{F}_n(r) + o_p(1)$ and "$(\cdot) \overset{p}{\to} 0$" becomes "$(\cdot) \overset{w^*}{\to}_p 0$" upon conditioning on the data, we conclude that $\mathcal{F}_n^* \overset{w^*}{\to}_p \mathcal{F}|(M, V)$ on the special probability space. Then $\mathcal{F}_n^* \overset{w^*}{\to}_w \mathcal{F}|(M, V)$ in general.   Q.E.D.

PROOF OF THEOREM 4.2: Additionally to the notation introduced in the proof of Theorem 4.1, let $V_n := n^{-1} \sigma^{-2} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_{nt} x_{nt}' \varepsilon_{nt}^2$ and $X_n := \{x_{nt}\}_{t=1}^n$. Under Assumption $\mathcal{H}$, by Corollary 5.12 of Kallenberg (1997), consider a single probability space where, for every $n \in \mathbb{N}$, the original and the bootstrap data are redefined together with (M, V, N), maintaining their distribution (we also maintain the notation), such that

$$\left(M_n, V_n, \tilde{V}_n, \frac{1}{n^{1/2}\sigma} \sum_{t=1}^{\lfloor n \cdot \rfloor} x_{nt}\varepsilon_{nt}, \mathcal{F}_n\right) \overset{\text{a.s.}}{\to} (M, V, V, N, \mathcal{F}) \qquad \text{(S.12)}$$

in $(\mathcal{D}_{m \times m})^3 \times \mathcal{D}_m \times \mathbb{R}$, with $\mathcal{F} := \sup_{r \in [\underline{r}, \overline{r}]} \{\tilde{N}(r)' \tilde{M}(r)^{-1} \tilde{N}(r)\}$ of equation (4.8). On this space also $\mathcal{F}_n^* \overset{w^*}{\to}_p \mathcal{F}|(M, V)$ holds, by the proof of Theorem 4.1, or equivalently, $P^*(\mathcal{F}_n^* \leq \cdot) \overset{p}{\to} P(\mathcal{F} \leq \cdot | M, V)$ in $\mathcal{D}(\mathbb{R})$, given that sample-path continuity of the conditional c.d.f. $P(\mathcal{F} \leq \cdot | M, V)$ is guaranteed by Proposition 3.2 of Linde (1989) applied conditionally on M, V. We see that $(\mathcal{F}_n, P^*(\mathcal{F}_n^* \leq \cdot)) \overset{p}{\to} (\mathcal{F}, P(\mathcal{F} \leq \cdot | M, V))$ on the special probability space. This implies that $(\mathcal{F}_n, P^*(\mathcal{F}_n^* \leq \cdot)) \overset{w}{\to} (\mathcal{F}, P(\mathcal{F} \leq \cdot | M, V))$ on general probability spaces. Theorem 3.1 becomes applicable and the conclusion of Theorem 4.1 about unconditional validity of the bootstrap follows.

Let now Assumption $\mathcal{C}$ hold. Let the original and the bootstrap data be redefined on another probability space where, by Lemma A.2(a), (S.12) holds (and thus, $\mathcal{F}_n^* \xrightarrow{w^*}_p \mathcal{F}|(M, V)$ by the proof of Theorem 4.1) and, additionally, the convergence in Assumption $\mathcal{C}$ holds as an a.s. convergence of random probability measures:

$$\left(M_n, V_n, \frac{1}{n^{1/2}\sigma} \sum_{t=1}^{\lfloor n\cdot \rfloor} x_{nt}\varepsilon_{nt} \right) \Bigg| X_n \xrightarrow{w}_{\text{a.s.}} (M, V, N)|(M, V).$$

By expanding $F_{\lfloor nr \rfloor}$ like $F_{\lfloor nr \rfloor}^*$ in the proof of Theorem 4.1 and applying the CMT of Sweeting (1989, Theorem 10), we can conclude that $\mathcal{F}_n|X_n \xrightarrow{w}_p \mathcal{F}|(M, V)$. Recalling that also $\mathcal{F}_n^* \xrightarrow{w^*}_p \mathcal{F}|(M, V)$, it follows that on a general probability space $(\mathcal{F}_n|X_n, \mathcal{F}_n^*|D_n) \xrightarrow{w}_w (\mathcal{F}|(M, V), \mathcal{F}|(M, V))$.

As previously, the continuity requirement of Corollary 3.2(a) (with $\tau := \mathcal{F}$ and $X = X' := (M, V)$) is satisfied by Proposition 3.2 of Linde (1989) applied conditionally. The bootstrap based on $\mathcal{F}_n$ and $\mathcal{F}_n^*$ is then concluded to be valid conditionally on X_n. *Q.E.D.*

S.5. SIMULATION DESIGN

We provide here a description of the Monte Carlo [MC] simulation design used for the linear regression model of Section 2. Data $D_n := \{y_t, x_t\}_{t=1}^n$ are generated according to equation (2.1) and the object of interest is inference on β based on $\tau_n := n^{1/2}(\hat{\beta} - \beta)$, with $\hat{\beta}$ denoting the OLS estimator of β; see Section 2.1. We consider the case where $x_t = \sum_{s=1}^t \eta_s$ is a nonstationary ($I(1)$) process under the following three different distributional structures for (ε_t, η_t):

(i) (ε_t, η_t) is i.i.d. $N(0, I_2)$ such that if the true variance $\hat{\omega}_\varepsilon = 1$ was used in equation (2.2), the bootstrap would perform exact conditional inference (see Remark 2.1);

(ii) $\varepsilon_t = \zeta_t(1 + 0.3\varepsilon_{t-1}^2 + 0.3\eta_{t-1}^2)^{1/2}$ and $\eta_t = \xi_t(1 + 0.6\eta_{t-1}^2)^{1/2}$, where (ζ_t, ξ_t) is i.i.d. $N(0, I_2)$; this corresponds to a stationary and ergodic conditionally heteroskedastic process with non-Gaussian unconditional marginals;

(iii) $\eta_t = \xi_t(1 + \delta\mathbb{I}_{\{\varepsilon_t \leq 0\}})$, where (ε_t, ξ_t) is i.i.d. $N(0, I_2)$ and $\delta = 9$.

For model (ii), we initialize the process by setting the conditional variance equal to its unconditional expectation (results do not change if a burn-in method is employed instead). The bootstrap is implemented as in Section 2.1, with $\hat{\omega}_\varepsilon$ chosen as the OLS residual variance. For DGP (i) bootstrap inference is close to exact (see Remark 2.1) and it holds that the bootstrap p-value p_n^* satisfies $p_n^* \stackrel{d}{=} U(0, 1) + O_p(n^{-1/2})$. DGP (ii) satisfies the conditions discussed in Section 2.2.2 and bootstrap inference is valid conditionally on $X_n := \{x_t\}_{t=1}^n$; that is, $p_n^*|X_n \xrightarrow{w}_p U(0, 1)$. For DGP (iii), on the other hand, bootstrap inference is not valid conditionally on this X_n, but it is valid unconditionally; see Section 2.2.3. Hence, $p_n^* \xrightarrow{w} U(0, 1)$ while $p_n^*|X_n$ has a random limit distribution. Notice that since the bootstrap statistic is conditionally Gaussian, p-values can be obtained without resorting to simulation.

Standard MC experiments generating $D_n = (y_1, \ldots, y_n, X_n)$ at each MC iteration allow estimation of the unconditional distribution of the p-value p_n^*, rather than its distribution conditional on X_n (see, e.g., Hansen (2000, footnote 11)). To simulate the distribution of p_n^* conditional on X_n, we implement a double MC design where, for each $m = 1, \ldots, M$, we generate the regressors $X_n^{(m)} \sim X_n$ and then, for each $v = 1, \ldots, N$, we generate data $(y_1^{(m,v)}, \ldots, y_n^{(m,v)})$ from their distribution conditional on $X_n = X_n^{(m)}$. The

respective statistics $\tau_n^{(m,v)}$ and the associated bootstrap p-values, $p_n^{*(m,v)}$, are used to estimate the conditional distribution of $p_n^*|\{X_n = X_n^{(m)}\}$, for each m, by the empirical c.d.f. $N^{-1}\sum_{v=1}^N \mathbb{I}_{\{p_n^{*(m,v)} \leq \cdot\}}$. We set $M = 1000$ and $N = 100{,}000$ throughout. Notice that for model (iii), once the regressor x_t (hence, η_t) is generated, simulation conditional on $\{x_t\}$ requires drawing from the conditional distribution of ε_t given η_t. An application of the Bayes rule yields that

$$P(\varepsilon_t \leq 0 | \eta_t) = \frac{1}{1 + (1+\delta)e^{-\frac{\eta_t^2}{2}\frac{\delta(2+\delta)}{(1+\delta)^2}}} =: p_t,$$

and hence that the conditional distributions $\varepsilon_t|\{x_t\}$, $\varepsilon_t|\eta_t$, and $(|\varepsilon_t|s_t)|\eta_t$ are a.s. equal, where $|\varepsilon_t|$ is independent of η_t and $s_t|\eta_t$ is a random sign equal to -1 with probability p_t.

All computations have been performed using MATLAB R2019b. The code is available from the *Econometrica* website.

REFERENCES

BENTKUS, V. (2005): "A Lyapunov-Type Bound in R^d," *Theory of Probability and Its Applications*, 49 (2), 311–323. [8]

BERTI, P., L. PRATELLI, AND P. RIGO (2006): "Almost Sure Weak Convergence of Random Probability Measures," *Stochastics*, 78, 91–97. [2,6]

BILLINGSLEY, P. (1968): *Convergence of Probability Measures*. New York: Wiley. [2,3]

CAVALIERE, G., AND I. GEORGIEV (2020): "Inference Under Random Limit Bootstrap Measures," *Econometrica* (forthcoming). [1]

CRIMALDI, I., AND L. PRATELLI (2005): "Convergence Results for Conditional Expectations," *Bernoulli*, 11, 737–745. [1,4,7]

DUDLEY, R. M. (2004): *Real Analysis and Probability*. Cambridge: Cambridge University Press. [6]

ETHIER, S., AND T. KURTZ (2005): *Markov Processes. Characterization and Convergence*. New York: Wiley. [5]

GEORGIEV, I., D. HARVEY, S. LEYBOURNE, AND A. M. R. TAYLOR (2019): "A Bootstrap Stationarity Test for Predictive Regression Invalidity," *Journal of Business & Economic Statistics*, 37, 528–541. [7]

HANSEN, B. E. (2000): "Testing for Structural Change in Conditional Models," *Journal of Econometrics*, 97, 93–115. [9-11]

KALLENBERG, O. (1997): *Foundations of Modern Probability*. New York: Springer. [1,6,8-10]

——— (2017): *Random Measures: Theory and Applications*. New York: Springer. [3]

LINDE, W. (1989): "Gaussian Measure of Translated Balls in a Banach Space," *Theory of Probability and Its Applications*, 34, 349–359. [9-11]

RUBSHTEIN, B. (1996): "A Central Limit Theorem for Conditional Distributions," in *Convergence in Ergodic Theory and Probability*, ed. by V. Bergelson, P. March, and J. Rosenblatt. Berlin: De Gruyter. [7]

SWEETING, T. J. (1989): "On Conditional Weak Convergence," *Journal of Theoretical Probability*, 2, 461–474. [4, 9-11]

WORM, D., AND S. HILLE (2011): "Ergodic Decompositions Associated With Regular Markov Operators on Polish Spaces," *Ergodic Theory and Dynamical Systems*, 31, 571–597. [5]