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 $\phi_n$. Without loss of generality, we can consider that $S_X = S'_X$, for otherwise we could identify $X_n$ and $X$ with some random elements $(X_n, Y_n)$ and $(Y, X)$ of $S_X \times S_X$ for arbitrary degenerate random elements $Y_n$ ($S'_X$-valued) and $Y$ ($S_X$-valued) defined on the probability spaces of respectively $X_n$ and $X$. Then, by extended Skorokhod coupling, consider a single probability space supporting $(\tilde{Z}_n, \tilde{X}_n, \tilde{Z}_n') = (Z_n, X_n, Z_n)$ and $(\tilde{Z}, \tilde{X}, \tilde{Z}') = (Z, X, Z')$ with the respective $\tilde{X}_n := \phi_n(\tilde{X}_n)$ such that $(\tilde{Z}_n, \tilde{X}_n, \tilde{Z}_n') \overset{a.s.}{\rightarrow} (\tilde{Z}, \tilde{X}, \tilde{Z}')$. Then also $\tilde{Z}_n|\tilde{X}_n \overset{w}{\rightarrow} \tilde{Z}|\tilde{X}$ 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 $\tilde{Z}_n|\tilde{X}_n \overset{\text{P}}{\rightarrow} \tilde{Z}|\tilde{X}$ such that $E[h(\tilde{Z}_n)|\tilde{X}_n] \overset{\text{P}}{\rightarrow} E[h(\tilde{Z})|\tilde{X}]$ for every $h \in \mathcal{C}_b(S_Z)$. The convergence $(E[h(\tilde{Z}_n)|\tilde{X}_n], \tilde{Z}_n, \tilde{X}_n', \tilde{Z}_n') \overset{\text{P}}{\rightarrow} (E[h(\tilde{Z})|\tilde{X}], \tilde{Z}, \tilde{X}, \tilde{Z}')$ for such $h$ implies (A.3) on a general probability space.

Q.E.D.
Remark S.1: We establish here the “natural” fact that the convergence
\[ ((Z_n'|X_n'), (Z_n'|X_n'), Z''_n') \xrightarrow{w} ((Z'|X'), (Z'|X'), Z'') \]
is equivalent to \(( (Z_n'|X_n'), Z''_n') \xrightarrow{w} ((Z'|X'), Z'') \) under separability of the space \( S''_Z \) 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
\[ (E[h(Z'_n)|X'_n], Z'_n') \xrightarrow{w} (E[h(Z')|X'], Z'') \]
for all \( h \in C_b(S''_Z) \), since both (A.5) and (S.1) are equivalent to \( uE[h(Z'_n)|X'_n] + v\|Z''_n\|_{a.e.} \xrightarrow{w} uE[h(Z')|X'] + v\|Z''\|_{a.e.} \) 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 \( Z'_n \xrightarrow{w} Z' \) 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 \( E[h'_i(Z'_n)|X'_n] \xrightarrow{a.s.} E[h'_i(Z')|X'] \) is equivalent to all continuous and bounded real functions \( h'_i \) 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 \( E[h''_i(Z''_n)|X''_n] \xrightarrow{a.s.} E[h''_i(Z'')|X''] \) for all \( i \in \mathbb{N} \) implies \( E[h''_i(Z''_n)|X''_n] \xrightarrow{a.s.} E[h''_i(Z'')|X''] \) for all continuous and bounded real \( h'' \) with the domain of \( h''_i \).

Consider the measurable functions \( H_n \) with values in \( S''_Z \times \mathbb{R}^\infty \) defined by
\[ H_n(X'_n, X''_n, Z''_n) = (Z''_n, \phi'_1(X'_n), \phi''_1(X''_n), \phi'_2(X'_n), \phi''_2(X''_n), \ldots) \]
such that a version \( \phi''_n(X'_n) \) of \( E[h'_i(Z'_n)|X'_n] \) and a version \( \phi''_n(X''_n) \) of \( E[h''_i(Z''_n)|X''_n] \) appear respectively at positions \( 2i \) and \( 2i + 1 \), and the analogous
\[ H(X', X'', Z'') = (Z'', \phi'_1(X'), \phi''_1(X''), \phi'_2(X'), \phi''_2(X''), \ldots), \]
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), \( H_n(X'_n, X''_n, Z''_n) \xrightarrow{w} H(X', X'', Z'') \) in \( S''_Z \times \mathbb{R}^\infty \) would follow if
\[ \left( \mathbb{I}_{[Z''_n \in A]}, \phi'_1(X'_n), \phi''_1(X''_n), \phi'_2(X'_n), \phi''_2(X''_n), \ldots \right) \xrightarrow{w} \left( \mathbb{I}_{[Z'' \in A]}, \phi'_1(X'), \phi''_1(X''), \phi'_2(X'), \phi''_2(X''), \ldots \right) \]
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 \( h' = \sum_{s=1}^m u_s h'_{s} \) and \( h'' = \sum_{s=1}^l v_s h''_{s} \), it holds that \( \sum_{s=1}^m u_s \phi'_{n,s}(X'_n) \sum_{s=1}^l v_s \phi''_{n,s}(X''_n) \) are versions of respectively \( E[h'(Z'_n)|X'_n] \) and \( E[h''(Z''_n)|X''_n] \), whereas \( \sum_{s=1}^m u_s \phi'_{s}(X') \sum_{s=1}^l v_s \phi''_{s}(X'') \) are versions of respectively.
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, $H_n(X_n', X_n'', Z_n''') \xrightarrow{w} H(X', X'', Z'')$ in $S_Z^2 \times \mathbb{R}^\infty$.

By extended Skorokhod coupling, $(X_n', X_n'', Z_n', Z_n''')$ and $(X', X'', Z', Z''')$ can be redefined, maintaining their distribution, on a new probability space where $H_n(X_n', X_n'', Z_n''') \xrightarrow{a.s.} H(X', X'', Z''')$ in $S_Z^2 \times \mathbb{R}^\infty$ (we subsume the $\sim$-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, $Z_n'' \xrightarrow{a.s.} Z''$, $E[h_i'(Z_n')|X_n'] \xrightarrow{a.s.} E[h_i'(Z')|X']$ 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 $Z_n''|X_n' \xrightarrow{w} Z'|X'$ and $Z_n''|X_n' \xrightarrow{w} Z''|X''$. Q.E.D.

**Proof of Lemma A.2(b):** Let $Z_n', Z_n'', Z', Z''$ ($n \in \mathbb{N}$) be i.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 $P(Z_n' \leq |X_n'|) \xrightarrow{a.s.} P(Z' \leq |X'|)$ in $\mathcal{D}(\mathbb{R})$ and $P(Z_n'' \leq |X_n''|) \xrightarrow{a.s.} P(Z'' \leq |X''|)$ in $\mathcal{D}(\mathbb{R})$. Since on this space also $Z_n'' \xrightarrow{a.s.} Z''$, (A.6) follows on a general probability space.

Conversely, let (A.6) hold. Notice that $P(Z_n' \leq |X_n'|), P(Z_n'' \leq |X_n''|), P(Z' \leq |X'|)$ and $P(Z'' \leq |X''|)$ as random elements of $\mathcal{D}(\mathbb{R})$ are measurable transformations of resp. $X_n', X_n'', X', X''$ 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, $(X_n', X_n'', Z_n', Z_n''')$ and $(X', X'', Z', Z''')$ can be redefined, maintaining their distribution, on a new probability space where $Z_n'' \xrightarrow{a.s.} Z''$, $P(Z_n' \leq |X_n'|) \xrightarrow{a.s.} P(Z' \leq |X'|)$ in $\mathcal{D}(\mathbb{R})$ and $P(Z_n'' \leq |X_n''|) \xrightarrow{a.s.} P(Z'' \leq |X''|)$ in $\mathcal{D}(\mathbb{R})$. By the proof of Kallenberg (2017, Theorem 4.20), on this space

$$
(E[h'(Z_n')|X_n'], E[h''(Z_n')|X_n'], Z_n'' \xrightarrow{a.s.} (E[h'(Z)|X'], E[h''(Z)|X''], Z''),
$$

for all $h', h'' \in \mathcal{C}_h(\mathbb{R})$ 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 $Z_n \xrightarrow{w} Z$ and $Z_n|Z \xrightarrow{w} Z|Z$ 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 $S_X \times S_Y$-valued, whereas $X_n$ and $X$ are respectively $S$-valued and $S_X$-valued ($n \in \mathbb{N}$), with all the mentioned spaces being Polish metric spaces. Then the convergence $(Z_n|Z', Z_n''|X_n) \xrightarrow{w} (Z'|Z', Z''|X)$ in the sense of (A.1) is equivalent to the convergence $(Z_n', (Z_n''|X_n)) \xrightarrow{w} (Z', (Z''|X))$ in the sense of (A.2). If additionally $S_Z^2 = \mathbb{R}^\infty$ and the conditional distribution $Z'|X$ is diffuse, then both convergence facts are equivalent to $(Z_n, P(Z_n'' \leq |X_n''|)) \xrightarrow{w} (Z', P(Z'' \leq |X''|))$ as random elements of $S_Z^2 \times \mathcal{D}(\mathbb{R})$. 

Proof: As in the proof of Lemma A.1, there is no loss of generality in assuming the equality $S_X = S$. First, let $(Z_n'\mid Z_n') \xrightarrow{w} (Z\mid Z, Z'\mid X)$ by Lemma A.2(a), consider a Skorokhod representation such that $\tilde{Z}_n \xrightarrow{w} \tilde{Z}$ and $\tilde{Z}_n' \xrightarrow{a.s.} \tilde{Z}'$. Then $h(\tilde{Z}_n') \xrightarrow{a.s.} h(\tilde{Z}')$ for every $h' \in C_b(S_Z')$. This implies that $\tilde{Z}_n \xrightarrow{L} \tilde{Z}$ by the proof of Proposition 4.3(i) of Crimaldi and Pratelli (2005). As further $E[h''(\tilde{Z}_n')\mid \tilde{Z}_n']$ for every $h'' \in C_b(S_Z')$, the convergence $(\tilde{Z}_n', E[h''(\tilde{Z}_n')\mid \tilde{Z}_n]) \xrightarrow{L} (\tilde{Z}', E[h'(\tilde{Z}'')\mid \tilde{X}])$ implies that $(Z_n', E[h''(Z_n')\mid \tilde{Z}_n]) \xrightarrow{w} (Z', E[h'(Z'')\mid \tilde{X}])$ for every such $h''$, which is (A.2).

Second, let $(Z_n', E[h''(Z_n')\mid X_n]) \xrightarrow{w} (Z', E[h''(Z'')\mid X])$ hold for every $h'' \in C_b(S_Z')$. Then $(h''(Z_n'), E[h''(Z_n')\mid X_n]) \xrightarrow{w} (h''(Z'), E[h''(Z'')\mid X])$ for every $h'' \in C_b(S_Z')$, 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. $P(Z'' \leq \cdot \mid X'')$ 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 $S_X = S_X$. By Lemma A.2(a), consider a Skorokhod representation such that $\tilde{Z}_n\mid X_n \xrightarrow{w} \tilde{Z}\mid \tilde{X}$. Then $h(\tilde{Z}_n\mid X_n) \xrightarrow{w} h(\tilde{Z})\mid \tilde{X}$ on the Skorokhod-representation space by Theorems 8(i) and 10 of Sweeting (1989). Therefore, $h(Z_n)\mid X_n \xrightarrow{w} h(Z)\mid X$ 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 $E[h(X_n'')\mid X_n'] \xrightarrow{w} E[h(X'')\mid X]$ for all continuous and bounded real functions $h$ with matching domain, where $(X_n', X''_n)$ are $X_n$-measurable, then the iterated expectations

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

satisfy the convergence

$$(E(z_n\mid X_n'), E(z_n\mid X_n), X_n', X''_n, Y_n) \xrightarrow{w} (E(z\mid X'), E(z\mid X', X''), X', X'', Y). \quad (S.2)$$

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 $X''_n\mid X_n' \xrightarrow{w} X''\mid X'$ imply $(X_n', X''_n)\mid X_n' \xrightarrow{w} (X', X'')\mid X'$; that is, for all $h \in C_b(S_X)$, it holds that $E[h(X_n', X''_n)\mid X_n'] \xrightarrow{w} E[h(X', X'')\mid X']$.

Let $\phi_n$ and $\phi$ be measurable real functions such that $\phi_n(X_n)$ and $\phi(X', X'')$ are versions respectively of the conditional expectations $E(z_n\mid X_n)$ and $E(z\mid X', X'')$. 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 $(\phi_n(X_n), X_n', X''_n, Y_n) \xrightarrow{a.s.} (\phi(X', X''), X', X'', Y)$ and $E[h(X_n', X''_n)\mid X_n'] \xrightarrow{p} E[h(X', X'')\mid X']$ for all $h \in C_b(S_X')$. Second, we show that on this space $E[\phi_n(X_n')\mid X_n'] \xrightarrow{p} E[\phi(X', X'')\mid X']$, which yields convergence (S.2) on a general probability space.

Step 1. Let the measurable function $\psi_n$ 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 $(\tilde{X}_n, \tilde{Y}_n) \overset{d}{=} (X_n, Y_n), (\tilde{X}', \tilde{X}'', \tilde{Y}) \overset{d}{=} (X', X'', Y)$ defined on this space such that $(\phi_n(\tilde{X}_n), \psi_n(\tilde{X}_n), \tilde{Y}_n) \overset{a.s.}{\rightarrow} (\phi(\tilde{X}', \tilde{X}''), \tilde{X}', \tilde{X}'', \tilde{Y})$. The proof of Theorem A.2 is thus concluded.

Q.E.D.
\( \tilde{X}^\prime, \tilde{Y} \). On this space, it also holds that \( E[h(\tilde{X}_n, \tilde{X}_n)|\tilde{X}_n] \xrightarrow{w} E[h(\tilde{X}, \tilde{X}^\prime)|\tilde{X}^\prime] \) for all \( h \in C_b(S_X) \) and \( (\tilde{X}_n, \tilde{X}_n^\prime) := \psi_n(\tilde{X}_n) \), as a consequence of the distributional equalities \( (\tilde{X}_n, \tilde{X}_n^\prime) \overset{d}{=} (X_n, X_n^\prime) \) and \( (\tilde{X}, \tilde{X}^\prime) \overset{d}{=} (X, X^\prime) \). Moreover, this convergence can be strengthened to \( E[h(\tilde{X}_n, \tilde{X}_n)|\tilde{X}_n] \xrightarrow{p} E[h(\tilde{X}, \tilde{X}^\prime)|\tilde{X}^\prime] \) for all \( h \in C_b(S_X) \) 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 \( C_b(S_X) \) is dense in the real functions on \( S_X \) that are integrable w.r.t. the probability measure induced by \( (X^\prime, X^\prime) \), it follows that for every \( \varepsilon \in (0, 1) \) there exists a \( \phi_\varepsilon \in C_b(S_X) \) such that \( E|\phi_\varepsilon(X^\prime, X^\prime)| - \phi(X^\prime, X^\prime)| < (\varepsilon/5)^2 \). We decompose

\[
E[\phi_n(X_n)|X_n^\prime] - E[\phi(X^\prime, X^\prime)|X_n^\prime] \\
\leq E[|\phi_n(X_n) - \phi(X^\prime, X^\prime)||X_n^\prime] \\
+ E[|\phi(X^\prime, X^\prime) - \phi_\varepsilon(X^\prime, X^\prime)||X_n^\prime] \\
+ E[|\phi_\varepsilon(X^\prime, X^\prime) - \phi_\varepsilon(X^\prime, X^\prime)||X_n^\prime] \\
+ E[|\phi_\varepsilon(X^\prime, X^\prime) - \phi(X^\prime, X^\prime)||X_n^\prime] \\
+ E[|\phi_\varepsilon(X^\prime, X^\prime) - \phi(X^\prime, X^\prime)||X_n^\prime]
\]

and label the addends on the right-hand side \( \rho_i, i = 1, \ldots, 5 \), in order of appearance. The term \( \rho_1 \) is \( o_p(1) \) by Markov’s inequality, because \( |\phi_n(X_n) - \phi(X^\prime, X^\prime)| \) 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 \( \phi_\varepsilon \), it follows that \( P(\rho_2 \geq \varepsilon/5) \leq \varepsilon/5 \). Since \( (X_n^\prime, X_n^\prime) \overset{a.s.}{\rightarrow} (X^\prime, X^\prime) \) and \( \phi_\varepsilon \) is continuous, it holds that \( |\phi_\varepsilon(X^\prime, X^\prime) - \phi_\varepsilon(X_n^\prime, X_n^\prime)| \overset{a.s.}{\rightarrow} 0 \) and \( \rho_3 \) is \( o_p(1) \) by Markov’s inequality and the bounded convergence theorem. For \( h = \phi_\varepsilon \), it holds that \( E[h(X_n, X_n^\prime)|X_n^\prime] \xrightarrow{p} E[h(X^\prime, X^\prime)|X^\prime] \) such that \( \rho_4 = o_p(1) \). Finally, \( P(\rho_5 \geq \varepsilon/5) \leq \varepsilon/5 \) by Markov’s inequality, like \( \rho_2 \). By combining these results, it follows that

\[
P(\left| E[\phi_n(X_n)|X_n^\prime] - E[\phi(X^\prime, X^\prime)|X_n^\prime] \right| \geq \varepsilon) < \varepsilon
\]

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

To prove the second part of the theorem, let \( h^\prime, h^\prime'' \in C_b(S_Z) \) be arbitrary. By the arguments in Remark S.1, (A.8) implies (A.7) with \( z_n = h^\prime(Z_n), Y_n = E[h^\prime''(Z_n)|X_n], \) \( z = h^\prime(Z), Y = E[h^\prime''(Z)|X^\prime, X^\prime] \). By the first part of the theorem and the arbitrariness of \( h^\prime, h^\prime'' \), (A.9) follows.

**PROOF OF LEMMA A.3:** Let \( \{f_j\}_{j \in \mathbb{N}} \) be a convergence-determining countable set of bounded Lipschitz functions \( S_X \times S_Z \rightarrow \mathbb{R} \), such that \( (X_n, Z_n^\prime) \overset{w}{\rightarrow} (X, Z) \) is implied by the convergence

\[
Ef_j(X_n, Z_n^\prime) \rightarrow Ef_j(X, Z) \quad \text{as } n \rightarrow \infty \quad \text{for all } j \in \mathbb{N}.
\]

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
$X_{n_{mk}} \xrightarrow{a.s.} X$ and the (random) conditional distribution of $Z_{n_{mk}^*}$ given $D_{n_{mk}^*}$ 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, $E[h(Z_{n_{mk}^*})|D_{n_{mk}^*}] \xrightarrow{a.s.} E[h(Z^*)|X']$ as $k \to \infty$ for every $h \in C_b(S_Z)$. If we show that $Ef_j(X_{n_{mk}^*}, Z_{n_{mk}^*}^*) \to Ef_j(X, Z)$ as $k \to \infty$ for all $j \in \mathbb{N}$, (S.3) will follow. Hence, without loss of generality we can take $X_n \xrightarrow{a.s.} X$ and $E[h(Z_n)|D_n] \xrightarrow{a.s.} E[h(Z^*)|X']$ for every $h \in C_b(S_Z)$, and prove that, as a result, (S.3) holds.

Write $Z_n^* = \xi_n^*(D_n, W_n^*)$ and define the measurable functions $\phi_{nj} : S_X \times S_D \to \mathbb{R}$ and $\phi_j : S_X \times \Omega \to \mathbb{R}$ by

$$ \phi_{nj}(x, d) := E_{D_n}(f_j(x, \xi_n^*(d, W_n^*)) \quad \text{and} \quad \phi_j(x, \omega) := \int_{S_Z} f_j(x, z) \nu(dz, X'(\omega)), $$

where $\nu$ 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

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

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 $S_X$. As $f_j(x_i, \cdot) \in C_b(S_Z)$, it holds that $E[f_j(x_i, Z^*)|D_n] \xrightarrow{a.s.} E[f_j(x_i, Z^*)|X']$ (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 $x \in S_X$ 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 $\rho_X$ is the metric on $S_X$ and $L_j$ depend on the Lipschitz constants of $f_j$ and on sup $|f_j| < \infty$, from the definitions of $\phi_{nj}$ and $\phi_j$ it follows that $|\phi_{nj}(x_i, D_n(\omega)) - \phi_j(x_i, D_n(\omega))|$ and $|\phi_j(x_i, \omega) - \phi_j(x_i, \omega)|$ for every fixed $x, j$ can be made arbitrarily small uniformly over $n, \omega$. From this fact and from the previous display, (S.4) follows for $A = \cap_{i,j \in \mathbb{N}} A_{ij}$.

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

$$ E[f_j(X, Z_n^*)|D_n, X] \xrightarrow{a.s.} \phi_j(X, \omega) = E[f_j(x, Z^*)|X']|_{x=X} \overset{(1)}{=} E[f_j(x, Z)|X']|_{x=X} = E[f_j(X, Z)|X'] \overset{(2)}{=} E[f_j(x, Z)|X']|_{x=X} = E[f_j(X, Z)|X'] \quad \text{a.s.,} $$

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 $S_X$-valued $\xi$'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, \(|Ef_j(X_n, Z_n^*) - Ef_j(X, Z^*)| \leq L_jE\{\rho_X(X_n, X) \wedge 1\} \rightarrow 0\) for every \(j \in \mathbb{N}\), again by the bounded convergence theorem, as \(X_n \stackrel{a.s.}{\rightarrow} X\). Thus, \(Ef_j(X_n, Z_n^*) = Ef_j(X, Z^*) + o(1) \rightarrow Ef_j(X, Z)\) and \((S.3)\) is proved. This establishes the convergence \((X_n, Z_n^*) \overset{w}{\rightarrow} (X, Z)\).

Finally, \((X_n, Z_n^*) \overset{w}{\rightarrow} (X, Z)\) and \(Z_n^* \mid D_n \overset{w}{\rightarrow}_p Z \mid X\), where \(X_n\) are \(D_n\)-measurable, imply that \(((X_n, Z_n^*) \mid D_n, \{X_n \overset{w}{\rightarrow} \} ((X, Z) \mid X), X)\), 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

\[
((X_n, Z_n^*) \mid D_n, (X_n, Z_n^*) \mid X_n) \overset{w}{\rightarrow} ((X, Z) \mid X, (X, Z) \mid X). \quad (S.5)
\]

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

\[Q.E.D.\]

\section*{S.3. Proofs of the Results in Section 2}

\textbf{Proof of Equation (2.8):} Let \(\hat{\varepsilon}_t := \varepsilon_t - E(\varepsilon_t \mid \eta_t), \ t \in \mathbb{N}\). Then \((\hat{\varepsilon}_t, E(\varepsilon_t \mid \eta_t), \eta_t)')\) is an i.i.d. sequence with diagonal covariance matrix \(\text{diag}(\omega_{\varepsilon_t \mid \eta_t}, 1 - \omega_{\varepsilon_t \mid \eta_t}, 1), \omega_{\varepsilon_t \mid \eta_t} := E[\text{Var}(\varepsilon_t \mid \eta_t)] \in (0, 1)\), and it is a standard fact that

\[
n^{-1/2} \left( \sum_{t=1}^{n} \hat{\varepsilon}_t, \sum_{t=1}^{n} E(\varepsilon_t \mid \eta_t), \sum_{t=1}^{n} \eta_t \right) \overset{w}{\rightarrow} (\omega_{\varepsilon_t \mid \eta_t} B_{Y_1}, (1 - \omega_{\varepsilon_t \mid \eta_t})^{1/2} B_{Y_2}, B_\eta) \quad (S.6)
\]

in \(\mathcal{D}_3\), where \((B_{Y_1}, B_{Y_2}, B_\eta)\) is a standard Brownian motion in \(\mathbb{R}^3\). Further, by the conditional invariance principle of Rubshtein (1996),

\[
n^{-1/2} \sum_{t=1}^{n} \hat{\varepsilon}_t \mid \sum_{t=1}^{n} \eta_t \overset{w}{\rightarrow}_p \omega_{\varepsilon_t \mid \eta_t}^{1/2} B_{Y_1} = \omega_{\varepsilon_t \mid \eta_t}^{1/2} B_{Y_1} \mid (B_{Y_2}, B_\eta) \overset{a.s.}{\rightarrow} (S.7)
\]

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

\[
n^{-1/2} \left( \sum_{t=1}^{n} \hat{\varepsilon}_t, \sum_{t=1}^{n} E(\varepsilon_t \mid \eta_t), \sum_{t=1}^{n} \eta_t \right) \overset{w}{\rightarrow} (\omega_{\varepsilon_t \mid \eta_t}^{1/2} B_{Y_1}, (1 - \omega_{\varepsilon_t \mid \eta_t})^{1/2} B_{Y_2}, B_\eta) \mid (B_{Y_2}, B_\eta)
\]

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 \((n^{-1} \sum_{t=1}^{n} \eta_t \hat{\varepsilon}_t, n^{-1} \sum_{t=1}^{n} \eta_t E(\varepsilon_t \mid \eta_t)) \overset{p}{\rightarrow} 0\) 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
\[
\left( n^{-2}M_n, n^{-1} \sum_{i=1}^{[n]} x_i \tilde{\epsilon}_i, n^{-1} \sum_{i=1}^{[n]} x_i E(\epsilon_i | \eta_i) \right)|X_n \\
= \left( n^{-2}M_n, n^{-1} \sum_{i=1}^{[n]} x_{i-1} \tilde{\epsilon}_i + o_p(1), n^{-1} \sum_{i=1}^{[n]} x_{i-1} E(\epsilon_i | \eta_i) + o_p(1) \right)|X_n \\
\xrightarrow{w} (M, \omega_{\varepsilon | \eta}^{1/2} \tilde{\xi}_1, (1 - \omega_{\varepsilon | \eta})^{1/2} M^{1/2} \xi_2)|(B_{y2}, B_\eta)
\]
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, \( \tau_n \) of (2.6) satisfies
\[
\left( \tau_n, n^{-2}M_n, n^{-1} \sum_{i=1}^{n} x_i E(\epsilon_i | \eta_i) \right)|X_n \xrightarrow{w} (\tau, M, (1 - \omega_{\varepsilon | \eta})^{1/2} M^{1/2} \xi_2)|(B_{y2}, B_\eta) \quad (S.8)
\]
with \( \tau := M^{-1/2}(\omega_{\varepsilon | \eta}^{1/2} \tilde{\xi}_1 + (1 - \omega_{\varepsilon | \eta})^{1/2} \xi_2) \). This yields (2.8). The bootstrap, instead of estimating consistently the limiting conditional distribution of \( \tau_n \) given \( X_n \), estimates the random distribution obtained by averaging this limit over \( \xi_2 \). As a result, conditionally on \( X_n \) the bootstrap \( p \)-value is not asymptotically uniformly distributed:
\[
p_n^*|X_n = \Phi(\omega_{\varepsilon}^{-1/2} M_n^{1/2} (\hat{\beta} - \beta))|X_n \xrightarrow{w} \Phi(\omega_{\varepsilon | \eta}^{1/2} \tilde{\xi}_1 + (1 - \omega_{\varepsilon | \eta})^{1/2} \xi_2)|\xi_2, \quad (S.9)
\]
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 \( n^{-\alpha/2} x_{[n]} \xrightarrow{a.s.} U(\cdot) \) in \( \mathcal{D} \). Then, by a product-space construction, we can extend this space to define also an i.i.d. standard Gaussian sequence \( \{\epsilon^*_i\} \) 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 \( \omega \)) in the factor-space of \( n^{-\alpha/2} x_{[n]} \) such that \((n^{-\alpha-1} M_n(\omega), n^{-\alpha/2-1} \xi_n(\omega)) \rightarrow (M(\omega), \xi(\omega)), n^{-(\alpha+1)/2} \sup |x_{[n]}(\omega)| \rightarrow 0 \) 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 \( \omega \) randomizing originating from \( \{\epsilon^*_i\} \), 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 \((W^\omega, b^\omega)\), where \( W^\omega \) and \( b^\omega \) are respectively a standard Brownian bridge and a standard Gaussian r.v. with \( \text{Cov}(W^\omega(u), b^\omega(u)) = M(\omega)^{-1/2} \xi(\omega) \psi(u), u \in [0, 1] \). It follows by disintegration (Theorem 5.4 of Kallenberg (1997)) that \((n^{1/2} W_n^*, M_n^{1/2} \hat{\beta}^*)|x_{[n]} \xrightarrow{a.s.} (W, b)|U, with the limit conditional distribution a.s. equal to \((W, b)|(M, \xi)\), and further that
\[
(n^{1/2} W_n^*, M_n^{1/2} \hat{\beta}^*, n^{-\alpha-1} M_n, n^{-\alpha/2-1} \xi_n)|x_{[n]} \xrightarrow{a.s.} (W, b, M, \xi)|(M, \xi)
\]
by Lemma A.3, since \((n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n)\) are \(x_{[n]}\)-measurable. Still further, by a CMT for a.s. weak convergence (Theorem 10 of Sweeting (1989)),

\[(n^{1/2}W_n, n^{(\alpha+1)/2}\beta_n, n^{-\alpha/2-1}\xi_n)|x_{[n]}| \xrightarrow{w} (W, M^{-1/2}b, \xi)|(M, \xi)\]
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 \((\tau, U)\) as defined on a single probability space such that \((\tau_n, n^{-\alpha/2}x_{[n]}) \xrightarrow{a.s.} (\tau, U)\) in \(\mathbb{R} \times \mathcal{D}\), and which is extended to support the independent bootstrap sequence \(\{\xi_n^*\}\) 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, \(\tau_n^* \xrightarrow{w} \tau(M, \xi)\) so that \(F_n^*(\cdot):=P(\tau_n^* \leq \cdot |D_n) \xrightarrow{p} F(\cdot):=P(\tau \leq \cdot |M, \xi)\) in \(\mathcal{D}(\mathbb{R})\), because \(F\) is sample-path continuous (e.g., by Proposition 3.2 of Linde (1989) applied conditionally on \(M, \xi\)). As further \(\tau_n \xrightarrow{a.s.} \tau\) 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) \((\tau_n, n^{-\alpha-1}M_n, n^{-\alpha/2-1}\xi_n) \xrightarrow{a.s.} (\tau, M, \xi)\), as implied by the a.s. convergence of \(\tau_n\) and \(n^{-\alpha/2}x_{[n]}\), and (ii) \(\tau_n^* \xrightarrow{w} \tau(M, \xi)\), 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 \((\tau, \tau^*, X, X')\) such that \((\tau_n, \tau_n^*, \phi_n(X_n), \psi_n(D_n)) \xrightarrow{a.s.} (\tau, \tau^*, X, X')\). Then, by Lemma A.1(a), on the Skorokhod-representation space it holds that \(\tau_n|X_n \xrightarrow{w} \tau|X\) and \(\tau_n^*|D_n \xrightarrow{w} \tau^*|X'|\), such that on a general probability space \((\tau_n|X_n, \tau_n^*|D_n) \xrightarrow{w} (\tau|X, \tau^*|X')\). If the conditional distributions \(\tau^*|X'\) and \(\tau|X'\) are a.s. equal, (3.4) follows.

**Proof of Theorem 4.1:** Let \((M_n, \tilde{V}_n) := (n^{-1}\sum_{i=1}^{[n]} x_{nt}x_{nt}', n^{-1}\sigma^{-2}\sum_{i=1}^{[n]} x_{nt}x_{nt}'\xi_n^2, a_n, 1)\) 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 \(\{w_{nt}^*\}_{t\in\mathbb{N}}\) can be regarded (upon packing with zeroes) as a random element of the Polish space \(\mathbb{R}^{k+2}\). 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 \((M_n, \tilde{V}_n) \xrightarrow{a.s.} (M, V)\).

Consider \(N_n^* := n^{-1/2}\sigma^{-1}\sum_{t=1}^{[n]} x_{nt}Y_{nt}'\). 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

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

$$(M_n, \tilde{V}_n, N^*_n)|D_n = (M_n, \tilde{V}_n, N^*_n)|(M_n, \tilde{V}_n) \xrightarrow{u} (M, V, N)|(M, V) \quad \text{(S.11)}$$

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 $\sup_{r \in [\zeta, \tau]} \|F_{\{nr\}}^* - \tilde{F}_n(r)\| = o_p(1)$ with

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

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

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

and since $\max_{r \in [\zeta, \tau]} \tilde{F}_n(r) \xrightarrow{u} F |(M, V)$ by (S.11) and a CMT for weak convergence in probability (Theorem 10 of Sweeting (1989)), with $F := \sup_{r \in [\zeta, \tau]} \{ \tilde{N}(r)M(1)^{-1}\tilde{N}(r) \}$

being the limit from (4.8), also $\max_{r \in [\zeta, \tau]} \tilde{F}_n(r) \xrightarrow{u} F |(M, V)$. Finally, as $F^* := \max_{r \in [\zeta, \tau]} F_{\{nr\}}^* = \max_{r \in [\zeta, \tau]} \tilde{F}_n(r) + o_p(1)$ and “($\cdot \xrightarrow{p} 0$)” becomes “($\cdot \xrightarrow{u} 0$)” upon conditioning on the data, we conclude that $F^* \xrightarrow{u} F |(M, V)$ on the special probability space. Then $F^* \xrightarrow{w} F |(M, V)$ in general.

PROOF OF THEOREM 4.2: Additionally to the notation introduced in the proof of Theorem 4.1, let $V_n := n^{-1} \sigma^{-2} \sum_{t=1}^{[nt]} x_{nt}x_{nt}' \epsilon_{nt}^2$ and $X_n := [x_{nt}]_{t=1}^n$. Under Assumption $H$, 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

$$(M_n, V_n, \tilde{V}_n, \frac{1}{n^{1/2}} \sigma \sum_{t=1}^{[nt]} x_{nt} \epsilon_{nt}, \mathcal{F}_n) \xrightarrow{a.s.} (M, V, V, N, \mathcal{F}) \quad \text{(S.12)}$$

in $(\mathcal{D}_{m \times m})^3 \times \mathcal{D}_m \times \mathbb{R}$, with $\mathcal{F} := \sup_{r \in [\zeta, \tau]} \{ \tilde{N}(r)M(1)^{-1}\tilde{N}(r) \}$ of equation (4.8). On this space also $F^* \xrightarrow{w} F |(M, V)$ holds, by the proof of Theorem 4.1, or equivalently, $P^*(F^* \leq \cdot) \xrightarrow{p} P(\mathcal{F} \leq \cdot | M, V)$ in $D(\mathbb{R})$, 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 $(F_n, P^*(F^*_n \leq \cdot)) \xrightarrow{p} (F, P(\mathcal{F} \leq \cdot | M, V))$ on the special probability space. This implies that $(F_n, P^*(F^*_n \leq \cdot)) \xrightarrow{u} (F, P(\mathcal{F} \leq \cdot | M, V))$ 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 $C$ 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, $F_n^* \xrightarrow{w} F^*$ by the proof of Theorem 4.1) and, additionally, the convergence in Assumption $C$ holds as an a.s. convergence of random probability measures:

$$\left(M_n, V_n, \frac{1}{n^{1/2}} \sum_{i=1}^{[n^r]} x_{ni} \epsilon_{ni}\right) \leftarrow X_n \xrightarrow{w} (M, V, N) \mid (M, V).$$

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

As previously, the continuity requirement of Corollary 3.2(a) (with $\tau := F$ and $X = X' := (M, V')$) is satisfied by Proposition 3.2 of Linde (1989) applied conditionally. The bootstrap based on $F_n$ and $F_n^*$ 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 $\hat{\beta}$ based on $\tau_n := n^{1/2}(\hat{\beta} - \beta)$, with $\hat{\beta}$ denoting the OLS estimator of $\beta$; see Section 2.1. We consider the case where $x = \sum_{t=1}^\gamma \eta_t$ is a nonstationary $(I(1))$ process under the following three different distributional structures for $(\epsilon_t, \eta_t)$:

(i) $(\epsilon_t, \eta_t)$ is i.i.d. $N(0, I_2)$ such that if the true variance $\hat{\sigma}_x = 1$ was used in equation (2.2), the bootstrap would perform exact conditional inference (see Remark 2.1):

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

(iii) $\eta_t = \xi_t(1 + \delta \|x\|_{\ell^2} < 0)$, where $(\epsilon_t, \xi_t)$ 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 $\hat{\sigma}_x$ 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 $p_n^* \overset{d}{=} U(0, 1) + O_p(n^{-1/2})$. 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, $p_n^* \mid X_n \xrightarrow{w} U(0, 1)$. 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 $p_n^* \mid X_n$ 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 $X_{m}^{(m)} \sim X_n$ 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 \( \tau(m, v) \) and the associated bootstrap \( p \)-values, \( p^*(m, v) \), are used to estimate the conditional distribution of \( p^*_n \mid \{X_n = X^{(m)}_n\} \), for each \( m \), by the empirical c.d.f. \( N^{-1} \sum_{n=1}^{N} \mathbb{I}_{\{p^*_n \leq \cdot\}} \). We set \( M = 1000 \) and \( N = 100,000 \) throughout. Notice that for model (iii), once the regressor \( x_t \) (hence, \( \eta_t \)) is generated, simulation conditional on \( \{x_t\} \) requires drawing from the conditional distribution of \( \varepsilon_t \) given \( \eta_t \). An application of the Bayes rule yields that

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

and hence that the conditional distributions \( \varepsilon_t \mid \{x_t\} \), \( \varepsilon_t \mid \eta_t \), and \( (|\varepsilon_t|s_t) \mid \eta_t \) are a.s. equal, where \( |\varepsilon_t| \) is independent of \( \eta_t \) and \( s_t \mid \eta_t \) 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**


Co-editor Ulrich K. Müller handled this manuscript.

Manuscript received 31 July, 2018; final version accepted 26 June, 2020; available online 15 July, 2020.