UNIVERSITY of York

This is a repository copy of *Bootstrapping Structural Change Tests*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/146607/

Version: Published Version

Article:

Boldea, Otilia, Cornea-Madeira, Adriana orcid.org/0000-0002-0889-7145 and Hall, Alastair (2019) Bootstrapping Structural Change Tests. Journal of Econometrics. pp. 359-397. ISSN 0304-4076

https://doi.org/10.1016/j.jeconom.2019.05.019

Reuse

This article is distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivs (CC BY-NC-ND) licence. This licence only allows you to download this work and share it with others as long as you credit the authors, but you can't change the article in any way or use it commercially. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

White Rose university consortium
Universities of Leeds, Sheffield & York

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

ARTICLE IN PRESS
Journal of Econometrics xxx (xxxx) xxx

Contents lists available at ScienceDirect

Journal of Econometrics

journal homepage: www.elsevier.com/locate/jeconom

Bootstrapping structural change tests

Otilia Boldea [a,*], Adriana Cornea-Madeira [b], Alastair R. Hall [c]

[a] Department of Econometrics and OR, TiSEM, Tilburg University, Warandelaan 2, 5037AB, Tilburg, The Netherlands
[b] The York Management School, University of York, Freboys Lane, Heslington, York YO10 5GD, United Kingdom
[c] Department of Economics, School of Social Sciences, University of Manchester, Oxford Road, Manchester, M13 9PL, United Kingdom

ARTICLE INFO

Article history:
Received 7 July 2017
Received in revised form 30 September 2018
Accepted 17 May 2019
Available online xxxx

JEL classification:
C12
C13
C15
C22

Keywords:
Multiple break points
Instrumental variables estimation
Two-Stage Least Squares
Wild bootstrap
Heteroskedasticity

ABSTRACT

This paper demonstrates the asymptotic validity of methods based on the wild recursive and wild fixed bootstraps for testing hypotheses about discrete parameter change in linear models estimated via Two Stage Least Squares. The framework allows for the errors to exhibit conditional and/or unconditional heteroscedasticity, and for the reduced form to be unstable. Simulation evidence indicates the bootstrap tests yield reliable inferences in the sample sizes often encountered in macroeconomics. If the errors exhibit unconditional heteroscedasticity and/or the reduced form is unstable then the bootstrap methods are particularly attractive because the limiting distributions of the test statistics are not pivotal.

© 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1. Introduction

Linear models with endogenous regressors are commonly employed in time series econometric analysis.[1] In many cases, the parameters of these models are assumed constant throughout the sample. However, given the span of many economic time series data sets, this assumption may be questionable and a more appropriate specification may involve parameters that change value during the sample period. Such parameter changes could reflect legislative, institutional or technological changes, shifts in governmental and economic policy, political conflicts, or could be due to large macroeconomic shocks such as the oil shocks experienced over the past decades and the productivity slowdown. It is therefore important to test for parameter – or structural – change. Various tests for structural change have been proposed with one difference between them being in the type of structural change against which the tests are designed to have

* Corresponding author.
E-mail addresses: o.boldea@uvt.nl (O. Boldea), adriana.cornea-madeira@york.ac.uk (A. Cornea-Madeira), Alastair.Hall@manchester.ac.uk (A.R. Hall).

[1] For example, Brady (2008) examines consumption smoothing by regressing consumption growth on consumer credit, the latter being endogenous because it depends on liquidity constraints. Zhang et al. (2008), Kleibergen and Mavroeidis (2009), Hall et al. (2012) and Kim et al. (2014) investigate the New Keynesian Phillips curve, where inflation is driven by expected inflation and marginal costs, both endogenous since they are correlated with inflation surprises. Bunzel and Enders (2010) and Qian and Su (2014) estimate the forward-looking Taylor rule, a model where the Federal fund rate is set based on expected inflation and output, both endogenous as they depend either on forecast errors or on current macroeconomic shocks. All these studies test for structural change in their estimated equations as part of their analysis.

https://doi.org/10.1016/j.jeconom.2019.05.019
0304-4076/© 2019 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Please cite this article as: O. Boldea, A. Cornea-Madeira and A.R. Hall, Bootstrapping structural change tests. Journal of Econometrics (2019), https://doi.org/10.1016/j.jeconom.2019.05.019.

power. In this paper, we focus on the scenario in which the potential structural change consists of discrete changes in the parameter values at unknown points in the sample, known as break - (or change-) points. Within this framework, two types of hypotheses tests are of natural interest: tests of no parameter change against an alternative of change at a fixed number of break-points, and tests of whether the parameters change at ℓ break-points against an alternative that they change at $\ell + 1$ points. These hypotheses tests are of interest in their own right, and also because they can form the basis of a sequential testing strategy for estimating the number of parameter break-points, see Bai and Perron (1998).

Hall et al. (2012) propose various statistics for testing these hypotheses in linear models with endogenous regressors based on Two Stage Least Squares (2SLS).[2] Their tests are the natural extensions of the analogous tests for linear models with exogenous regressors estimated via Ordinary Least Squares (OLS) that are introduced in the seminal paper by Bai and Perron (1998). A critical issue in the implementation of these tests in a 2SLS setting is whether or not the reduced form (RF) for the endogenous regressors is stable. If it is then, under certain conditions, Hall et al.'s (2012) test statistics converge in distribution to the same distributions as their OLS counterparts and are pivotal, see Hall et al. (2012) and Perron and Yamamoto (2014). However, if the reduced form itself is unstable and/or there is unconditional heteroskedasticity, then these limiting distributions no longer apply (Hall et al., 2012), and are, in fact, no longer pivotal (Perron and Yamamoto, 2014). This is a severe drawback as in most cases of interest the reduced form is likely to be unstable. This problem has been circumvented in two ways. Hall et al. (2012) suggest a testing strategy based on dividing the sample into sub-samples over which the RF is stable but this is inefficient compared to inferences based on the whole sample, and can be infeasible if the sub-samples are small. Perron and Yamamoto (2015) propose using a variant of Hansen's (2000) fixed regressor bootstrap to calculate the critical values of the test. Their simulation evidence suggests the use of this bootstrap improves the reliability of inferences but they do not establish the asymptotic validity of the method.[3]

In this paper, we explore the use of bootstrap versions of 2SLS-based tests for parameter change in far greater detail than previous studies. We consider inferences based on two different types of bootstrap versions of the structural change tests, provide formal proofs of their asymptotic validity and report simulation results that demonstrate that the bootstrap tests provide reliable inferences in the finite sample sizes encountered in practice. More specifically, we consider the case where the right-hand side variables of the equation of interest contain endogenous regressors, contemporaneously exogenous variables, lagged values of both and lagged values of the dependent variable. This equation of interest is part of a system of equations that is completed by the reduced form for the endogenous regressors and equations for the contemporaneously exogenous variables. This system of equations is assumed to follow a Structural Vector Autoregressive (SVAR) model in which the parameters of the mean are subject to discrete shifts at a finite number of break-points in the sample. Both the number and location of the break-points are unknown to the researcher. These break-points define regimes over which the parameters are constant, and it is assumed that the implied reduced form VAR is stable within each such regime. The errors of the VAR are assumed to follow a vector martingale difference sequence (m.d.s.) that potentially exhibits both conditional and unconditional heteroskedasticity. Given this error structure, we explore methods for inference based on the wild bootstrap proposed by Liu (1988) because it has been found to replicate the conditional and unconditional heteroskedasticity of the errors in other contexts. In particular, we consider two versions of the wild bootstrap: the wild recursive bootstrap (which generates recursively the bootstrap observations) and the wild fixed-regressor bootstrap (which adds the wild bootstrap residuals to the estimated conditional mean, thus keeping all lagged regressors fixed). These bootstraps have been proposed by Gonçalves and Kilian (2004) to test the significance of parameters in autoregressions with (stationary) conditional heteroskedastic errors. Our primary focus is on bootstrap versions of sup-*Wald*-type statistics to test for structural changes in the parameters of the equation of interest (with endogenous variables) estimated by 2SLS, but our validity arguments also extend straightforwardly to analogous sup-F-type statistics.

While our primary focus is on models where the reduced form for the endogenous regressors is unstable, our results also cover the case where this reduced form is stable. In the latter case, the test statistics have a pivotal limiting distribution under conditions covered by our framework, specialized to errors that are unconditionally homoskedastic. For these situations, the bootstrap methods we propose are expected to provide a superior approximation to finite sample behavior compared to the limiting distribution because the bootstrap, by its nature, incorporates sample information. Thus bootstrap versions of the tests are attractive in this setting as well.

In the case where there are no endogenous regressors in the equation of interest, our framework reduces to a linear model estimated by OLS. For this set-up, Hansen (2000) proposes the wild fixed-design bootstrap to test for structural changes using a sup-F statistic. Very recently Georgiev et al. (2018) consider Hansen's (2000) bootstrap for versions of sup-F-type tests for parameter variation in predictive regressions with exogenous regressors. Both Hansen (2000) and Georgiev et al. (2018) establish the asymptotic validity of this bootstrap within the settings they consider.[4] There are some similarities and important differences between our framework (specialized to the no endogenous regressor case) and those in Hansen (2000) and Georgiev et al. (2018). We adopt similar assumptions about the error process to

[2] Perron and Yamamoto (2015) propose an alternative approach based on OLS.

[3] An alternative approach is to estimate the number and location of the breaks via an information criteria, see Hall et al. (2015). However, this approach has the drawback that inferences can be sensitive to the choice of penalty function.

[4] In fact, Georgiev et al. (2018) demonstrate that Hansen's (2000) proof of the asymptotic validity of the bootstrap needs an amendment when the predictive regressors are (near-) unit root processes.

Georgiev et al. (2018) and like both Hansen (2000) and Georgiev et al. consider fixed regressor bootstrap tests of a null of constant parameters versus an alternative of parameter change. Important differences include: Georgiev et al. (2018) allow for strongly persistent variables whereas our framework assumes the system is stable within (suitably defined) regimes; our analysis covers tests for additional breaks in the model, the use of the recursive bootstrap and also inferences based on sup-*Wald* tests. Thus our results for this case complement those of Hansen (2000) and Georgiev et al. (2018).[5]

Although the frameworks are different, Hansen (2000), Georgiev et al. (2018) and our own study all find their bootstrap versions of the structural change tests work well in finite samples. Interestingly, Chang and Perron (2018) find that bootstrap-based inferences about the location of breaks have similar advantages in finite samples.[6] Collectively, our paper and these other recent studies suggest the use of the bootstrap can yield reliable inferences in linear models with multiple break-points in the sample sizes encountered in practice.

An outline of the paper is as follows. Section 2 lays out the model, test statistics and their bootstrap versions. Section 3 details the assumptions and contains theoretical results establishing the asymptotic validity of the bootstrap methods. Section 4 contains simulation results that provide evidence on the finite sample performance of the bootstrap tests. Section 5 concludes. Appendix A contains all the tables for Section 4, with additional simulations relegated to a Supplementary Appendix.[7] Appendix B contains the proofs, with some background results relegated to the same Supplementary Appendix.

Notation: Matrices and vectors are denoted with bold symbols, and scalars are not. Define for a scalar N, the generalized vec operator $\mathbf{vect}_{s=1:N}(\boldsymbol{A}_s) = \mathbf{vect}(\boldsymbol{A}_1, \dots, \boldsymbol{A}_N)$, stacking in order the columns of the matrices $\boldsymbol{A}_s, s = 1, \dots, N$. Let $\mathbf{diag}_{s=1:N}(\boldsymbol{A}_s) = \mathbf{diag}(\boldsymbol{A}_1, \dots, \boldsymbol{A}_N)$ be the matrix that puts the blocks $\boldsymbol{A}_1, \dots, \boldsymbol{A}_N$ on the diagonal. If it is clear over which set **vect** and **diag** operations are taken, then the subscript $s = 1 : N$ is dropped on these operators. T denotes the number of time series observations. If N is the number of breaks in a quantity then $T_1, \dots, T_N$ are the ordered change-points. Also, $\boldsymbol{\tau}_N = (\tau_0, \mathbf{vect}_{s=1:N}(\tau_s)', \tau_{N+1})$ is a partition of the interval $[1, T]$ where each element is divided by T, such that $[T\tau_s] = T_{s,\boldsymbol{\tau}_N}$, for $s = 1, \dots, N$, and $\tau_0 = 0$ and $\tau_{N+1} = 1$. Define the regimes where parameters are assumed constant as $I_{s,\boldsymbol{\tau}_N} = [T_{s-1} + 1, T_s]$ for $s = 1, \dots, N + 1$. Below the breaks in the structural equation are denoted by $\boldsymbol{\tau}_N = \boldsymbol{\lambda}_m$, and those in the reduced form by $\boldsymbol{\tau}_N = \boldsymbol{\pi}_h$, where m an h are the number of breaks in each equation. A superscript zero on any quantity refers to the true quantity, which is a fixed number, vector or matrix. For any random vector or matrix $\boldsymbol{Z}$, denote by $\|\boldsymbol{Z}\|$ the Euclidean norm for vectors, or the Frobenius norm for matrices. Finally, $\mathbf{0}_a$ and $\mathbf{0}_{a\times b}$ denote, respectively, an $a \times 1$ vector and a $a \times b$ matrix of zeros, and 1_A denotes an indicator function that takes the value one if event A occurs. Let mI_a be the $a \times a$ identity matrix.

2. The model and test statistics with their bootstrap versions

This section is divided into three sub-sections. Section 2.1 outlines the model. Section 2.2 outlines the hypotheses of interest and the test statistics. Section 2.3 presents the bootstrap versions of the test statistics.

2.1. The model

Consider the case where the equation of interest takes the form

$$y_t = \underbrace{\boldsymbol{w}_t'}_{1\times(p_1+q_1)} \underbrace{\boldsymbol{\beta}_{(i)}^0}_{(p_1+q_1)\times 1} + u_t, \quad i = 1, \dots, m+1, \quad t \in I_{i,\boldsymbol{\lambda}_m^0}, \tag{1}$$

where $\boldsymbol{w}_t = \mathbf{vect}(\boldsymbol{x}_t, \boldsymbol{z}_{1,t})$, $\boldsymbol{z}_{1,t}$ includes the intercept, $\boldsymbol{r}_t$ and lagged values of y_t, $\boldsymbol{x}_t$, and $\boldsymbol{r}_t$, and $\boldsymbol{\beta}_{(i)}^0$ are the parameters in regime i. The key difference between $\boldsymbol{x}_t$ and $\boldsymbol{r}_t$ is that $\boldsymbol{x}_t$ represents the set of explanatory variables which are correlated with u_t, and $\boldsymbol{r}_t$ represents the set of explanatory variables that are uncorrelated with u_t. We therefore refer to $\boldsymbol{x}_t$ as the *endogenous* regressors and $\boldsymbol{r}_t$ as the *contemporaneously exogenous* regressors.[8] Eq. (1) can be re-written as:

$$y_t = \boldsymbol{x}_t'\boldsymbol{\beta}_{\boldsymbol{x},t}^0 + \boldsymbol{z}_{1,t}'\boldsymbol{\beta}_{\boldsymbol{z},t}^0 + u_t = \boldsymbol{w}_t'\boldsymbol{\beta}_t^0 + u_t,$$

where $\boldsymbol{\beta}_t^0 = \boldsymbol{\beta}_{(i)}^0$ if $t \in I_{i,\boldsymbol{\lambda}_m^0}$, $i = 1, \dots, m+1$ and similar notation holds for $\boldsymbol{\beta}_{\boldsymbol{x},t}$ and $\boldsymbol{\beta}_{\boldsymbol{z},t}$. For simplicity, we refer to (1) as the "structural equation" (SE).

[5] The wild fixed-regressor bootstrap is also included in the recent simulation study exploring the finite sample properties of inference methods about the location of the break-point in models estimated via OLS reported in Chang and Perron (2018).

[6] Chang and Perron (2018) report results from a comprehensive simulation study that investigates the finite sample properties of various methods for constructing confidence intervals for the break fractions in linear regression models with exogenous regressors. They consider variants of the intervals based on i.i.d., wild and sieve bootstraps.

[7] The Supplementary Appendix is available at https://sites.google.com/site/otiliaboldea/ and in the supplementary material archive of the Journal of Econometrics.

[8] This terminology is taken from Wooldridge (1994)[p.349] and reflects that fact $\boldsymbol{r}_t$ may be correlated with u_n for $t \neq n$.

The SE is assumed to be part of a system that is completed by the following equations for $\boldsymbol{x}_t$ and $\boldsymbol{r}_t$. The reduced form (RF) equation for the endogenous regressors $\boldsymbol{x}_t$ is a regression model with h breaks ($h+1$ regimes), that is:

$$\underbrace{\boldsymbol{x}_t'}_{1\times p_1} = \underbrace{\boldsymbol{z}_t'}_{1\times q}\underbrace{\boldsymbol{\Delta}^0_{(i)}}_{q\times p_1} + \underbrace{\boldsymbol{v}_t'}_{1\times p_1}, \qquad i=1,\ldots,h+1, \qquad t\in I_{i,\pi_h^0}. \tag{2}$$

The vector $\boldsymbol{z}_t$ includes the constant, $\boldsymbol{r}_t$ and lagged values of y_t, $\boldsymbol{x}_t$ and $\boldsymbol{r}_t$. It is assumed that the variables in $\boldsymbol{z}_{1,t}$ are a strict subset of those in $\boldsymbol{z}_t$ and therefore we write $\boldsymbol{z}_t = \mathbf{vect}(\boldsymbol{z}_{1,t}, \boldsymbol{z}_{2,t})$. Eq. (2) can also be rewritten as:

$$\boldsymbol{x}_t' = \boldsymbol{z}_t'\boldsymbol{\Delta}^0_t + \boldsymbol{v}_t',$$

where $\boldsymbol{\Delta}^0_t = \boldsymbol{\Delta}^0_{(i)}$ if $t\in I_{i,\pi_h^0}$, $i=1,\ldots,h+1$. The contemporaneously exogenous variables $\boldsymbol{r}_t$ are assumed to be generated as follows,

$$\underbrace{\boldsymbol{r}_t'}_{1\times p_2} = \boldsymbol{z}_{3,t}'\boldsymbol{\Phi}^0_{(i)} + \underbrace{\boldsymbol{\zeta}_t'}_{1\times p_2} \qquad i=1,\ldots,d+1, \quad t\in I_{i,\omega_d^0}, \tag{3}$$

where $\boldsymbol{z}_{3,t}$ includes the constant and lagged values of $\boldsymbol{r}_t$, y_t and $\boldsymbol{x}_t$.

Eqs. (1), (2) and (3) imply $\tilde{\boldsymbol{z}}_t = \mathbf{vect}(y_t, \boldsymbol{x}_t, \boldsymbol{r}_t)$ evolves over time via a SVAR process whose parameters are subject to discrete shifts at unknown points in the sample. To present the reduced form VAR version of the model, define $n = \dim(\tilde{\boldsymbol{z}}_t)$ and let $\boldsymbol{\tau}_N$ denote the partition of the sample such that all three equations have constant parameters within the associated regimes.[9] We can then write Eqs. (1), (2) and (3) as:

$$\tilde{\boldsymbol{z}}_t = \boldsymbol{c}_{\tilde{\boldsymbol{z}},s} + \sum_{i=1}^{p}\boldsymbol{C}_{i,s}\tilde{\boldsymbol{z}}_{t-i} + \boldsymbol{e}_t, \qquad [\tau_{s-1}T]+1 \le t \le [\tau_s T],\ s=1,2,\ldots,N+1, \tag{4}$$

where $\boldsymbol{e}_t = \boldsymbol{A}_s^{-1}\boldsymbol{\epsilon}_t$,

$$\boldsymbol{A}_s = \begin{bmatrix} 1 & -\boldsymbol{\beta}^{0\prime}_{\boldsymbol{x},s} & -\boldsymbol{\beta}^{0\prime}_{\boldsymbol{r},s} \\ \mathbf{0}_{p_1} & \boldsymbol{I}_{p_1} & \boldsymbol{\Delta}^{0\prime}_{\boldsymbol{r},s} \\ \mathbf{0}_{p_2} & \mathbf{0}_{p_2\times p_1} & \boldsymbol{I}_{p_2} \end{bmatrix}, \tag{5}$$

$\boldsymbol{\beta}^{0\prime}_{\boldsymbol{r},s}$ denotes the sub-vector of $\boldsymbol{\beta}^{0\prime}_s$ that contain the coefficients on $\boldsymbol{r}_t$ in (1) ($\boldsymbol{\beta}^{0\prime}_{\boldsymbol{r},s}$ and $\boldsymbol{\beta}^{0\prime}_s$ are the values of $\boldsymbol{\beta}^{0\prime}_{\boldsymbol{r},t}$ and $\boldsymbol{\beta}^{0\prime}_t$ for $[\tau_{s-1}T]+1 \le t \le [\tau_s T]$); $\boldsymbol{\Delta}^{0\prime}_{\boldsymbol{r},s}$ denotes the sub-matrix of $\boldsymbol{\Delta}^{0\prime}_s$ that contains the coefficients on $\boldsymbol{r}_t$ in (2) ($\boldsymbol{\Delta}^{0\prime}_{\boldsymbol{r},s}$ and $\boldsymbol{\Delta}^{0\prime}_s$ are the values of $\boldsymbol{\Delta}^{0\prime}_{\boldsymbol{r},t}$ and $\boldsymbol{\Delta}^{0\prime}_t$ for $[\tau_{s-1}T]+1 \le t \le [\tau_s T]$), and $\boldsymbol{\epsilon}_t = \mathbf{vect}(u_t, \boldsymbol{v}_t, \boldsymbol{\zeta}_t)$. For ease of notation, we assume the order of the VAR is the same in each regime, but our results easily extend to the case where the order varies by regime.

2.2. Testing parameter variation

As stated in the introduction, this paper focuses on the issue of testing for structural change in the SE. Within the model described above, there are two types of test that are of particular interest. The first tests the null hypothesis of no parameter change against the alternative of a fixed number of parameter changes in the sample that is, a test of $H_0: m=0$ versus $H_1: m=k$. The second tests the null of a fixed number of parameter changes against the alternative that there is one more, that is, it tests $H_0: m=\ell$ versus $H_1: m=\ell+1$. We consider appropriate test statistics for each of these scenarios in turn below.

As the tests are based on the Wald principle, calculation of the test statistics here requires 2SLS estimation of the SE under H_1. On the first stage, the RF is estimated via least squares methods. If the number and location of the breaks in the RF are known then this estimation is straightforward. However, in general, neither the number nor the location of the breaks are known and so they must be estimated. For our purposes here, it is important that both h and $\boldsymbol{\pi}_h^0$ are consistently estimated and that $\hat{\boldsymbol{\pi}}_h$, the estimator of $\boldsymbol{\pi}_h^0$, converges sufficiently fast (see Lemma 7 in Appendix B). These properties can be achieved by estimating the RF either as a system or equation by equation, and using a sequential testing strategy to estimate h; see, respectively Qu and Perron (2007) and Bai and Perron (1998). Provided the significance levels of the tests shrink to zero slowly enough, $\hat{h}$ approaches h with probability one as the sample size T grows; *e.g.* see Bai and Perron (1998) [Proposition 8]. The same consistency result holds if we estimate h via the information criteria; *e.g.* see Hall et al. (2013). For this reason, in the rest of the theoretical analysis, we treat h as known. However, we explore the potential sensitivity of the finite sample performance of the tests for structural change in the SE to the estimation of h in our simulation study. Let $\hat{\boldsymbol{\Delta}}_{(j)}$ be the estimator of $\boldsymbol{\Delta}^0_{(j)}$, $\hat{\boldsymbol{\Delta}}_t = \sum_{j=1}^{h+1}\hat{\boldsymbol{\Delta}}_{(j)}1_{t\in\hat{I}_j^*}$, where $\hat{I}_j^* = \{[\hat{\pi}_{j-1}T]+1, [\hat{\pi}_{j-1}T]+2,\ldots,[\hat{\pi}_j T]\}$, and $\hat{\boldsymbol{x}}_t = \hat{\boldsymbol{\Delta}}_t'\boldsymbol{z}_t$ that is, $\hat{\boldsymbol{x}}_t$ is the predicted value for $\boldsymbol{x}_t$ from the estimated RF.

Case (i): $H_0: m=0$ versus $H_1: m=k$

9 For example, suppose $m=1$, $h=2$ and $d=1$ with $\boldsymbol{\lambda}^0_m = [0, 0.5, 1]'$, $\boldsymbol{\pi}^0_h = [0, 0.3, 0.5, 1]'$ and $\boldsymbol{\omega}^0_d = [0, 0.7, 1]'$, then $N=3$ and $\boldsymbol{\tau}_N = [0, 0.3, 0.5, 0.7, 1]'$.

Under H_1, the second stage estimation involves estimation via OLS of the model,

$$y_t = \hat{\boldsymbol{w}}_t' \boldsymbol{\beta}_{(i)} + \text{error}, \quad i = 1, \ldots, k+1, \quad t \in I_{i,\boldsymbol{\lambda}_k}, \tag{6}$$

for all possible k-partitions $\boldsymbol{\lambda}_k$. Let $\hat{\boldsymbol{\beta}}_{(i)}$ denote the OLS estimator of $\boldsymbol{\beta}_{(i)}$ in (6), $\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k} \equiv \mathbf{vect}_{i=1:k+1}(\hat{\boldsymbol{\beta}}_{(i)}) = \mathbf{vect}_{i=1:k+1}(\hat{\boldsymbol{\beta}}_{i,\boldsymbol{\lambda}_k})$ denote the OLS estimator of $\mathbf{vect}_{i=1:k+1}(\boldsymbol{\beta}_{(i)}) \equiv \mathbf{vect}_{i=1:k+1}(\boldsymbol{\beta}_{i,\boldsymbol{\lambda}_k})$ in (6) (that is, $\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k}$ is the OLS estimator of $\mathbf{vect}_{i=1:k+1}(\boldsymbol{\beta}_{(i)})$ based on partition $\boldsymbol{\lambda}_k$). To present the sup-*Wald* test, we define $\boldsymbol{R}_k = \tilde{\boldsymbol{R}}_k \otimes \boldsymbol{I}_p$ where $\tilde{\boldsymbol{R}}_k$ is the $k \times (k+1)$ matrix whose $(i,j)^{th}$ element, $\tilde{R}_k(i,j)$, is given by: $\tilde{R}_k(i,i) = 1$, $\tilde{R}_k(i,i+1) = -1$, $\tilde{R}_k(i,j) = 0$ for $i = 1, 2, \ldots, k$, and $j \neq i, j \neq i+1$. Also let $\boldsymbol{\Lambda}_{\epsilon,k} = \{\boldsymbol{\lambda}_k : |\lambda_{i+1} - \lambda_i| \geq \epsilon, \lambda_1 \geq \epsilon, \lambda_k \leq 1 - \epsilon\}$. With this notation, the test statistic is:

$$\sup\text{-}Wald_T = \sup_{\boldsymbol{\lambda}_k \in \boldsymbol{\Lambda}_{\epsilon,k}} Wald_{T\boldsymbol{\lambda}_k}, \tag{7}$$

$$Wald_{T\boldsymbol{\lambda}_k} = T\, \hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k}' \boldsymbol{R}_k' \left(\boldsymbol{R}_k \hat{\boldsymbol{V}}_{\boldsymbol{\lambda}_k} \boldsymbol{R}_k'\right)^{-1} \boldsymbol{R}_k \hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k}, \tag{8}$$

where:

$$\hat{\boldsymbol{V}}_{\boldsymbol{\lambda}_k} = \mathbf{diag}_{i=1:k+1}(\hat{\boldsymbol{V}}_{(i)}), \qquad \hat{\boldsymbol{V}}_{(i)} = \hat{\boldsymbol{Q}}_{(i)}^{-1} \hat{\boldsymbol{M}}_{(i)} \hat{\boldsymbol{Q}}_{(i)}^{-1}, \qquad \hat{\boldsymbol{Q}}_{(i)} = T^{-1} \sum_{t \in I_{i,\boldsymbol{\lambda}_k}} \hat{\boldsymbol{w}}_t \hat{\boldsymbol{w}}_t', \tag{9}$$

$$\hat{\boldsymbol{M}}_{(i)} \xrightarrow{p} \lim_{T\to\infty} \mathrm{Var}\left(T^{-1/2} \sum_{t \in I_{i,\boldsymbol{\lambda}_k}} \boldsymbol{\Upsilon}_t^{0\prime} \boldsymbol{z}_t \left(u_t + \boldsymbol{v}_t' \boldsymbol{\beta}_{\boldsymbol{x}}^0\right)\right), \tag{10}$$

$\boldsymbol{\beta}_{\boldsymbol{x}}^0$ is the true value of $\boldsymbol{\beta}_{\boldsymbol{x},(i)}^0$ for $i = 1, 2, \ldots, m+1$ under H_0, and $\boldsymbol{\Upsilon}_t^0 = (\boldsymbol{\Delta}_t^0, \boldsymbol{\Pi})$ and $\boldsymbol{\Pi}' = \left(\boldsymbol{I}_{q_1}, \mathbf{0}_{q_1 \times (q-q_1)}\right)$.

As mentioned in the introduction, our framework assumes the errors are a m.d.s. that potentially exhibits heteroskedasticity, and so the natural choice of $\hat{\boldsymbol{M}}_{(i)}$ is the Eicker–White estimator, see Eicker (1967) and White (1980). This can be constructed using the estimator of $\boldsymbol{\beta}_{\boldsymbol{x},(i)}$ in (6) under either H_0 or H_1, where $\boldsymbol{\beta}_{\boldsymbol{x},(i)}$ are the elements of $\boldsymbol{\beta}_{(i)}$ containing the coefficients on $\hat{\boldsymbol{x}}_t$. For the purposes of the theory presented below, it does not matter which is used because the null hypothesis is assumed to be true. However, the power properties may be sensitive to this choice. In our simulation study reported below, we use the Eicker–White estimator based on $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}$, the estimator of $\boldsymbol{\beta}_{\boldsymbol{x},(i)}$ under H_1, that is,

$$\hat{\boldsymbol{M}}_{(i)} = \widehat{EW}\left[\, \hat{\boldsymbol{\Upsilon}}_t' \boldsymbol{z}_t (\hat{u}_t + \hat{\boldsymbol{v}}_t' \hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)});\ I_{i,\boldsymbol{\lambda}_k} \right],$$

where $\hat{u}_t = y_t - \boldsymbol{w}_t' \hat{\boldsymbol{\beta}}_{(i)}$ for $t \in I_{i,\boldsymbol{\lambda}_k}$, $\hat{\boldsymbol{v}}_t = \boldsymbol{x}_t - \hat{\boldsymbol{\Delta}}_t' \boldsymbol{z}_t$, $\hat{\boldsymbol{\Upsilon}}_t = [\hat{\boldsymbol{\Delta}}_t, \boldsymbol{\Pi}]$, $\hat{\boldsymbol{\Delta}}_t$ is defined before (6), $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}$ are the first p_1 elements of $\hat{\boldsymbol{\beta}}_{(i)}$, and for any vector $\boldsymbol{a}_t$ and $I \subseteq \{1, 2, \ldots, T\}$, $\widehat{EW}\,[\boldsymbol{a}_t;\ I\,] = T^{-1} \sum_{t \in I} \boldsymbol{a}_t \boldsymbol{a}_t'$.

Case (ii): $H_0: m = \ell$ versus $H_1: m = \ell + 1$

Following the same approach used by Bai and Perron (1998) for OLS based inferences, suitable tests statistics can be constructed as follows. The model with ℓ breaks is estimated via a global minimization of the sum of squared residuals associated with the second stage of the 2SLS estimation of the SE. For each of the $\ell+1$ regimes of this estimated model, the sup-*Wald* statistic for testing no breaks versus one break is calculated. Inference about $H_0: m = \ell$ versus $H_1: m = \ell + 1$ is based on the largest of these $\ell + 1$ sup-*Wald* statistics.

More formally, let the estimated SE break fractions for the ℓ-break model be $\hat{\boldsymbol{\lambda}}_\ell$ and the associated break points be denoted $\{\hat{T}_i\}_{i=1}^{\ell}$ where $\hat{T}_i = [T\hat{\lambda}_i]$. Let $\hat{I}_i = I_{i,\hat{\boldsymbol{\lambda}}_\ell}$, the set of observations in the ith regime of the ℓ-break model and partition this set as $\hat{I}_i = \hat{I}_i^{(1)}(\varpi_i) \cup \hat{I}_i^{(2)}(\varpi_i)$ where $\hat{I}_i^{(1)}(\varpi_i) = \{t : [\hat{\lambda}_{i-1}T] + 1, [\hat{\lambda}_{i-1}T] + 2, \ldots, [\varpi_i T]\}$ and $\hat{I}_i^{(2)}(\varpi_i) = \{t : [\varpi_i T] + 1, [\varpi_i T] + 2, \ldots, [\hat{\lambda}_i T]\}$. Consider the estimation of the model

$$y_t = \hat{\boldsymbol{w}}_t' \boldsymbol{\beta}_{(j)} + \text{error}, \quad j = 1, 2, \quad t \in \hat{I}_i^{(j)}(\varpi_i), \tag{11}$$

for all possible choices of ϖ_i (where for notational brevity we suppress the dependence of $\boldsymbol{\beta}_{(j)}$ on i). Let $\hat{\boldsymbol{\beta}}(\varpi_i) = \mathbf{vect}\left(\hat{\boldsymbol{\beta}}_{(1)}(\varpi_i), \hat{\boldsymbol{\beta}}_{(2)}(\varpi_i)\right)$ be the OLS estimators of $\mathbf{vect}(\boldsymbol{\beta}_{(1)}, \boldsymbol{\beta}_{(2)})$ from (11). Also let $\mathcal{N}_i(\hat{\boldsymbol{\lambda}}_\ell) = [\hat{\lambda}_{i-1}+\epsilon, \hat{\lambda}_i - \epsilon]$. The sup-*Wald* statistic for testing $H_0: m = \ell$ versus $H_1: m = \ell + 1$ is given by

$$\sup\text{-}Wald_T(\ell+1 \mid \ell) = \max_{i=1,2,\ldots,\ell+1} \left\{ \sup_{\varpi_i \in \mathcal{N}_i(\hat{\boldsymbol{\lambda}}_\ell)} T \hat{\boldsymbol{\beta}}(\varpi_i)' \boldsymbol{R}_1' [\boldsymbol{R}_1 \hat{\boldsymbol{V}}(\varpi_i) \boldsymbol{R}_1']^{-1} \boldsymbol{R}_1 \hat{\boldsymbol{\beta}}(\varpi_i) \right\} \tag{12}$$

where[10]

$$\hat{\boldsymbol{V}}(\varpi_i) = \mathbf{diag}\left(\hat{\boldsymbol{V}}_1(\varpi_i), \hat{\boldsymbol{V}}_2(\varpi_i)\right), \qquad \hat{\boldsymbol{V}}_j(\varpi_i) = \{\hat{\boldsymbol{Q}}_j(\varpi_i)\}^{-1} \hat{\boldsymbol{M}}_j(\varpi_i) \{\hat{\boldsymbol{Q}}_j(\varpi_i)\}^{-1},$$

10 The comment above (after (10)) about the calculation of $\hat{\boldsymbol{M}}_{(i)}$ applies equally to $\hat{\boldsymbol{M}}_j(\varpi_i)$.

$$\hat{\boldsymbol{Q}}_j(\varpi_i) = T^{-1} \sum_{t \in \hat{I}_i^{(j)}} \hat{\boldsymbol{w}}_t \hat{\boldsymbol{w}}_t', \qquad \hat{\boldsymbol{M}}_j(\varpi_i) = \widehat{EW}\left[\ \hat{\boldsymbol{\Upsilon}}_t' \boldsymbol{z}_t(\hat{u}_t + \hat{\boldsymbol{v}}_t' \hat{\boldsymbol{\beta}}_{\boldsymbol{x},(j)}(\varpi_i));\ \hat{I}_i^{(j)}(\varpi_i)\right],$$

where $\hat{u}_t = y_t - \boldsymbol{w}_t' \hat{\boldsymbol{\beta}}_{(j)}(\varpi_i)$ for $t \in \hat{I}_i^{(j)}(\varpi_i)$, $j = 1, 2$, $\hat{\boldsymbol{v}}_t = \boldsymbol{x}_t - \hat{\boldsymbol{\Delta}}_t' \boldsymbol{z}_t$, $\hat{\boldsymbol{\Upsilon}}_t = [\hat{\boldsymbol{\Delta}}_t, \boldsymbol{\Pi}]$, $\hat{\boldsymbol{\Delta}}_t$ is defined before (6), and $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(j)}(\varpi_i)$ are the first p_1 elements of $\hat{\boldsymbol{\beta}}_{(j)}(\varpi_i)$.

2.3. Bootstrap versions of the test statistics

In this section, we introduce the bootstrap analogues of the test statistics presented in the previous section. As noted above, our framework assumes the error vector $\boldsymbol{\epsilon}_t$ to be a m.d.s that potentially exhibits conditional and unconditional heteroskedasticity, and so we use the wild bootstrap proposed by Liu (1988) because it has been found to replicate the conditional and unconditional heteroskedasticity of the errors in other contexts.[11] We consider both the wild recursive (WR) bootstrap and the wild fixed regressor (WF) bootstrap. These procedures differ in their treatment of the right-hand side variables in the bootstrap samples as described below.

Generation of the bootstrap samples:
Let $\tilde{\boldsymbol{z}}_t^b = \mathbf{vect}(y_t^b, \boldsymbol{x}_t^b, \boldsymbol{r}_t)$ where y_t^b and $\boldsymbol{x}_t^b$ denote the bootstrap values of y_t and $\boldsymbol{x}_t$. Note that because $\boldsymbol{r}_t$ is contemporaneously exogenous its sample value is used in the bootstrap samples. In all cases below, the bootstrap residuals are obtained as $u_t^b = \hat{u}_t \nu_t$ and $\boldsymbol{v}_t^b = \hat{\boldsymbol{v}}_t \nu_t$, where $\hat{u}_t$ and $\hat{\boldsymbol{v}}_t$ are the (non-centered) residuals under the null hypothesis and ν_t is a random variable that is discussed further in Section 3 (Assumption 10).
For the WR bootstrap, $\{y_t^{b\prime}\}$ and $\{\boldsymbol{x}_t^{b\prime}\}$ are generated recursively as follows:

$$\boldsymbol{x}_t^{b\prime} = \boldsymbol{z}_t^{b\prime} \hat{\boldsymbol{\Delta}}_t + \boldsymbol{v}_t^{b\prime}, \tag{13}$$

$$y_t^b = \boldsymbol{x}_t^{b\prime} \hat{\boldsymbol{\beta}}_{\boldsymbol{x},t} + \boldsymbol{z}_{1,t}^{b\prime} \hat{\boldsymbol{\beta}}_{\boldsymbol{z},t} + u_t^b, \tag{14}$$

where the vector $\boldsymbol{z}_t^b$ contains a constant, $\boldsymbol{r}_t$ and lags of y_t^b, $\boldsymbol{x}_t^b$ and $\boldsymbol{r}_t$; $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},t}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{z},t}$ are the sample estimates of $\boldsymbol{\beta}_{\boldsymbol{x},t}^0$ and $\boldsymbol{\beta}_{\boldsymbol{z},t}^0$ under H_0 of the test in question.
For the WF bootstrap, $\boldsymbol{z}_t$ is kept fixed and, following Gonçalves and Kilian (2004), the bootstrap samples are generated as follows:

$$\boldsymbol{x}_t^{b\prime} = \boldsymbol{z}_t' \hat{\boldsymbol{\Delta}}_t + \boldsymbol{v}_t^{b\prime}, \tag{15}$$

$$y_t^b = \boldsymbol{x}_t^{b\prime} \hat{\boldsymbol{\beta}}_{\boldsymbol{x},t} + \boldsymbol{z}_{1,t}' \hat{\boldsymbol{\beta}}_{\boldsymbol{z},t} + u_t^b, \tag{16}$$

where again $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},t}$ and $\hat{\boldsymbol{\beta}}_{\boldsymbol{z},t}$ are the sample estimates of $\boldsymbol{\beta}_{\boldsymbol{x},t}^0$ and $\boldsymbol{\beta}_{\boldsymbol{z},t}^0$ under H_0 of the test in question.

Case (i): $H_0: m = 0$ vs $H_1: m = k$
First consider the WR bootstrap. 2SLS estimation is implemented in the bootstrap samples as follows. On the first stage, the following model is estimated via OLS

$$\boldsymbol{x}_t^{b\prime} = \boldsymbol{z}_t^{b\prime} \boldsymbol{\Delta}_j + \text{error}, \qquad t \in \hat{I}_j^*, \qquad j = 1, 2, \ldots, h+1,$$

to obtain $\hat{\boldsymbol{\Delta}}_j^b = \left\{\sum_{t \in \hat{I}_j^*} \boldsymbol{z}_t^b \boldsymbol{z}_t^{b\prime}\right\}^{-1} \sum_{t \in \hat{I}_j^*} \boldsymbol{z}_t^b \boldsymbol{x}_t^{b\prime}$. Define $\hat{\boldsymbol{\Delta}}_t^b = \sum_{j=1}^{\hat{h}+1} 1_{t \in \hat{I}_j^*} \hat{\boldsymbol{\Delta}}_j^b$, $\hat{\boldsymbol{x}}_t^{b\prime} = \boldsymbol{z}_t^{b\prime} \hat{\boldsymbol{\Delta}}_t^b$, and $\hat{\boldsymbol{w}}_t^b = \mathbf{vect}(\hat{\boldsymbol{x}}_t^b, \boldsymbol{z}_{1,t}^b)$. For a given k-partition $\boldsymbol{\lambda}_k$, the second stage of the 2SLS in the bootstrap samples involves OLS estimation of

$$y_t^b = \hat{\boldsymbol{w}}_t^{b\prime} \boldsymbol{\beta}_{(i)} + \text{error}, \quad i = 1, \ldots, k+1, \quad t \in I_{i,\boldsymbol{\lambda}_k}, \tag{17}$$

and let $\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k}^b \equiv \mathbf{vect}_{i=1:k+1}(\hat{\boldsymbol{\beta}}_{(i)}^b) = \mathbf{vect}_{i=1:k+1}(\hat{\boldsymbol{\beta}}_{i,\boldsymbol{\lambda}_k}^b)$ be the resulting OLS estimator of $\mathbf{vect}_{i=1:k+1}(\boldsymbol{\beta}_{(i)}) \equiv \mathbf{vect}_{i=1:k+1}(\boldsymbol{\beta}_{i,\boldsymbol{\lambda}_k})$ based on partition $\boldsymbol{\lambda}_k$. The WR bootstrap version of the sup-$Wald_T$ statistic is:

$$\text{sup-}Wald_T^b = \sup_{\boldsymbol{\lambda}_k \in \boldsymbol{\Lambda}_{\epsilon,k}} Wald_{T\boldsymbol{\lambda}_k}^b, \tag{18}$$

$$Wald_{T\boldsymbol{\lambda}_k}^b = T\, \hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k}^{b\prime} \boldsymbol{R}_k' \left(\boldsymbol{R}_k \hat{\boldsymbol{V}}_{\boldsymbol{\lambda}_k}^b \boldsymbol{R}_k'\right)^{-1} \boldsymbol{R}_k \hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k}^b, \tag{19}$$

where:

$$\hat{\boldsymbol{V}}_{\boldsymbol{\lambda}_k}^b = \mathbf{diag}_{i=1:k+1}(\hat{\boldsymbol{V}}_{(i)}^b), \qquad \hat{\boldsymbol{V}}_{(i)}^b = (\hat{\boldsymbol{Q}}_{(i)}^b)^{-1} \hat{\boldsymbol{M}}_{(i)}^b (\hat{\boldsymbol{Q}}_{(i)}^b)^{-1}, \qquad \hat{\boldsymbol{Q}}_{(i)}^b = T^{-1} \sum_{t \in I_{i,\boldsymbol{\lambda}_k}} \hat{\boldsymbol{w}}_t^b \hat{\boldsymbol{w}}_t^{b\prime}, \tag{20}$$

$$\hat{\boldsymbol{M}}_{(i)}^b = \widehat{EW}\left[\ \hat{\boldsymbol{\Upsilon}}_t^{b\prime} \boldsymbol{z}_t^b \left(\hat{u}_t^b + \hat{\boldsymbol{v}}_t^{b\prime} \hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b\right);\ I_{i,\boldsymbol{\lambda}_k}\right], \tag{21}$$

[11] The wild bootstrap has been developed in Liu (1988) following suggestions in Wu (1986) and Beran (1986) in the context of static linear regression models with (unconditionally) heteroskedastic errors.

where $\hat{u}_t^b = y_t^b - \boldsymbol{w}_t^{b\prime}\hat{\boldsymbol{\beta}}_{(i)}^b$ for $t \in I_{i,\lambda_k}$, $\hat{\boldsymbol{v}}_t^b = \boldsymbol{x}_t^b - \hat{\boldsymbol{\Delta}}_t^{b\prime}\boldsymbol{z}_t^b$, $\hat{\boldsymbol{\Upsilon}}_t^b = (\hat{\boldsymbol{\Delta}}_t^b, \boldsymbol{\Pi})$, $\hat{\boldsymbol{\Delta}}_t^b$ is defined before (17), $\boldsymbol{w}_t^b = \textbf{vect}(\boldsymbol{x}_t^b, \boldsymbol{z}_{1,t}^b)$, and $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b$ are the first p_1 elements of $\hat{\boldsymbol{\beta}}_{(i)}^b$.

Now consider the WF bootstrap, for which y_t^b and $\boldsymbol{x}_t^b$ are generated via (15)–(16). The first stage of the 2SLS involves LS estimation of

$$\boldsymbol{x}_t^b = \boldsymbol{z}_t'\boldsymbol{\Delta}_j + \text{error}, \qquad t \in \hat{I}_j^*, \qquad j = 1, 2, \ldots, h+1,$$

to obtain $\hat{\boldsymbol{\Delta}}_j^b = \left\{\sum_{t\in\hat{I}_j^*}\boldsymbol{z}_t\boldsymbol{z}_t'\right\}^{-1}\sum_{t\in\hat{I}_j^*}\boldsymbol{z}_t\boldsymbol{x}_t^{b\prime}$. Now re-define $\hat{\boldsymbol{\Delta}}_t^b = \sum_{j=1}^{\hat{h}+1} 1_{t\in\hat{I}_j^*}\hat{\boldsymbol{\Delta}}_j^b$, $\hat{\boldsymbol{x}}_t^{b\prime} = \boldsymbol{z}_t'\hat{\boldsymbol{\Delta}}_t^b$, and $\hat{\boldsymbol{w}}_t^b = \textbf{vect}(\hat{\boldsymbol{x}}_t^b, \boldsymbol{z}_{1,t})$. For a given k-partitions $\boldsymbol{\lambda}_k$, the second stage of the 2SLS in the bootstrap samples involves OLS estimation of (17) and let $\hat{\boldsymbol{\beta}}_{\boldsymbol{\lambda}_k}^b = \textbf{vect}_{i=1:k+1}(\hat{\boldsymbol{\beta}}_{(i)}^b)$ be the resulting OLS estimator of $\textbf{vect}_{i=1:k+1}(\boldsymbol{\beta}_{(i)})$ based on partition $\boldsymbol{\lambda}_k$. The WF bootstrap sup-*Wald* statistic is defined as in (18) with $Wald_{T\boldsymbol{\lambda}_k}^b$ defined as in (19) only with $\hat{\boldsymbol{w}}_t^b$ and $\hat{\boldsymbol{\Delta}}_t^b$ redefined in the way described in this paragraph, and $\hat{\boldsymbol{M}}_{(i)}^b$ in (21) replaced by $\hat{\boldsymbol{M}}_{(i)}^b = \widehat{EW}\left[\hat{\boldsymbol{\Upsilon}}_t^{b\prime}\boldsymbol{z}_t(\hat{u}_t^b + \hat{\boldsymbol{v}}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b);\, I_{i,\lambda_k}\right]$, where $\hat{u}_t^b = y_t^b - \boldsymbol{w}_t^{b\prime}\hat{\boldsymbol{\beta}}_{(i)}^b$ for $t \in I_{i,\lambda_k}$, $\hat{\boldsymbol{v}}_t^b = \boldsymbol{x}_t^b - \hat{\boldsymbol{\Delta}}_t^{b\prime}\boldsymbol{z}_t$, $\hat{\boldsymbol{\Upsilon}}_t^b = (\hat{\boldsymbol{\Delta}}_t^b, \boldsymbol{\Pi})$, $\boldsymbol{w}_t^b = \textbf{vect}(\boldsymbol{x}_t^b, \boldsymbol{z}_{1,t})$, and $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b$ are the first p_1 elements of $\hat{\boldsymbol{\beta}}_{(i)}^b$.

Case (ii): $H_0: m = \ell$ versus $H_1: m = \ell + 1$

For each bootstrap the first stage of the 2SLS estimation and the construction of $\hat{\boldsymbol{w}}_t$ is the same as described under *Case (i)* above. Let $\hat{I}_i^{(j)}$ be defined as in the discussion of *Case (ii)* in Section 2.2, and consider

$$y_t^b = \hat{\boldsymbol{w}}_t^{b\prime}\boldsymbol{\beta}_{(j)} + \text{error}, \quad j = 1, 2 \quad t \in \hat{I}_i^{(j)}, \tag{22}$$

for all possible choices of ϖ_i (where, once again, we suppress the dependence of $\boldsymbol{\beta}_{(j)}$ on i). Let $\hat{\boldsymbol{\beta}}^b(\varpi_i) = \textbf{vect}\left(\hat{\boldsymbol{\beta}}_{(1)}^b(\varpi_i), \hat{\boldsymbol{\beta}}_{(2)}^b(\varpi_i)\right)$ be the OLS estimators of $\textbf{vect}(\boldsymbol{\beta}_{(1)}, \boldsymbol{\beta}_{(2)})$ from (22). The bootstrap version of sup-$Wald_T(\ell+1 \mid \ell)$ is given by

$$\text{sup-}Wald_T^b(\ell+1 \mid \ell) = \max_{i=1,2,\ldots,\ell+1}\left\{\sup_{\varpi_i\in\mathcal{N}(\hat{\lambda}_\ell)} T\hat{\boldsymbol{\beta}}^b(\varpi_i)'\boldsymbol{R}_1'[\boldsymbol{R}_1\hat{\boldsymbol{V}}^b(\varpi_i)\boldsymbol{R}_1']^{-1}\boldsymbol{R}_1\hat{\boldsymbol{\beta}}^b(\varpi_i)\right\} \tag{23}$$

where

$$\hat{\boldsymbol{V}}^b(\varpi_i) = \textbf{diag}\left(\hat{\boldsymbol{V}}_1^b(\varpi_i), \hat{\boldsymbol{V}}_2^b(\varpi_i)\right), \qquad \hat{\boldsymbol{V}}_j^b(\varpi_i) = \{\hat{\boldsymbol{Q}}_j^b(\varpi_i)\}^{-1}\,\hat{\boldsymbol{M}}_j^b(\varpi_i)\,\{\hat{\boldsymbol{Q}}_j^b(\varpi_i)\}^{-1},$$

$$\hat{\boldsymbol{Q}}_j^b(\varpi_i) = T^{-1}\sum_{t\in\hat{I}_i^{(j)}}\hat{\boldsymbol{w}}_t^b\hat{\boldsymbol{w}}_t^{b\prime},$$

and $\hat{\boldsymbol{M}}_j^b(\varpi_i) = \widehat{EW}\left[\hat{\boldsymbol{\Upsilon}}_t^{b\prime}\boldsymbol{z}_t^b(\hat{u}_t^b + \hat{\boldsymbol{v}}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(j)}^b(\varpi_i));\, \hat{I}_i^{(j)}(\varpi_i)\right]$ for the WR bootstrap, where $\hat{u}_t^b = y_t^b - \boldsymbol{w}_t^{b\prime}\hat{\boldsymbol{\beta}}_{(j)}^b(\varpi_i)$ for $t \in \hat{I}_i^{(j)}(\varpi_i)$, $\hat{\boldsymbol{v}}_t^b = \boldsymbol{x}_t^b - \hat{\boldsymbol{\Delta}}_t^{b\prime}\boldsymbol{z}_t^b$, $\hat{\boldsymbol{\Upsilon}}_t^b = (\hat{\boldsymbol{\Delta}}_t^b, \boldsymbol{\Pi})$, $\hat{\boldsymbol{\Delta}}_t^b$ is defined before (17), $\boldsymbol{w}_t^b = \textbf{vect}(\boldsymbol{x}_t^b, \boldsymbol{z}_{1,t}^b)$, and $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(j)}^b(\varpi_i)$ are the first p_1 elements of $\hat{\boldsymbol{\beta}}_{(j)}^b(\varpi_i)$; and $\hat{\boldsymbol{M}}_j^b(\varpi_i) = \widehat{EW}\left[\hat{\boldsymbol{\Upsilon}}_t^{b\prime}\boldsymbol{z}_t(\hat{u}_t^b + \hat{\boldsymbol{v}}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(j)}^b(\varpi_i));\, \hat{I}_i^{(j)}(\varpi_i)\right]$ for the WF bootstrap, where $\hat{u}_t^b = y_t^b - \boldsymbol{w}_t^{b\prime}\hat{\boldsymbol{\beta}}_{(j)}^b(\varpi_i)$ for $t \in \hat{I}_i^{(j)}(\varpi_i)$, $\hat{\boldsymbol{v}}_t^b = \boldsymbol{x}_t^b - \hat{\boldsymbol{\Delta}}_t^{b\prime}\boldsymbol{z}_t$, $\hat{\boldsymbol{\Upsilon}}_t^b = (\hat{\boldsymbol{\Delta}}_t^b, \boldsymbol{\Pi})$, $\hat{\boldsymbol{\Delta}}_t^b$ is defined in the last paragraph of *Case (i)* in this section, $\boldsymbol{w}_t^b = \textbf{vect}(\boldsymbol{x}_t^b, \boldsymbol{z}_{1,t})$, and $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(j)}^b(\varpi_i)$ are the first p_1 elements of $\hat{\boldsymbol{\beta}}_{(j)}^b(\varpi_i)$.

3. The asymptotic validity of the bootstrap tests

In this section, we establish the asymptotic validity of the bootstrap versions of the test statistics described above. To this end we impose the following conditions.

Assumption 1. If $m > 0$, $T_i^0 = [T\lambda_i^0]$, where $0 < \lambda_1^0 < \cdots < \lambda_m^0 < 1$.

Assumption 2. If $m > 0$, $\boldsymbol{\beta}_{(i+1)}^0 - \boldsymbol{\beta}_{(i)}^0$ is a non-zero vector of constants for $i = 1, \ldots, m$.

Assumption 3. If $h > 0$, then $T_i^* = [T\pi_i^0]$, where $0 < \pi_1^0 < \cdots < \pi_h^0 < 1$.

Assumption 4. If $h > 0$, $\boldsymbol{\Delta}_{(j+1)}^0 - \boldsymbol{\Delta}_{(j)}^0$ is a non-zero matrix of constants for $j = 1, \ldots, h$.

Assumption 5. If $k > 0$, then $0 < \omega_1^0 < \cdots < \omega_k^0 < 1$ and $\boldsymbol{\Phi}_{(i+1)}^0 - \boldsymbol{\Phi}_{(i)}^0$ is a non-zero matrix of constants for $i = 1, \ldots, k$.

Assumption 6. The first and second stage estimations in 2SLS are over respectively all partitions of $\boldsymbol{\pi}$ and $\boldsymbol{\lambda}$ such that $T_i - T_{i-1} > \max(q-1, \epsilon T)$ for some $\epsilon > 0$ and $\epsilon < \min_i(\lambda_{i+1}^0 - \lambda_i^0)$ and $\epsilon < \min_j(\pi_{j+1}^0 - \pi_j^0)$.

Assumption 7. (i) $p < \infty$; (ii) $\left|\boldsymbol{I}_n - \boldsymbol{C}_{1,s}a - \boldsymbol{C}_{2,s}a^2 - \ldots - \boldsymbol{C}_{p,s}a^p\right| \neq 0$, for all $s = 1, \ldots, N+1$, and all $|a| \leq 1$.

Assumption 8. $\text{rk}\left(\boldsymbol{\Upsilon}_t^0\right) = p_1 + q_1$.

Assumption 9. The innovations can be written as $\boldsymbol{\epsilon}_t = \boldsymbol{S}\boldsymbol{D}_t\boldsymbol{l}_t$, where:

(i) $\boldsymbol{S}$ is a $n \times n$ lower triangular non-stochastic matrix with real-valued diagonal elements $s_{ii} = 1$ and elements below the diagonal equal to s_{ij} (which are also zero for $i > p_1+1$, $j < p_1+1$), such that $\boldsymbol{S}\boldsymbol{S}'$ is positive definite; $\boldsymbol{D}_t = \textbf{diag}_{i=1:n}(d_{it})$, a non-stochastic matrix where $d_{it} = d_i(t/T) : [0, 1] \to D^n[0, 1]$, the space of cadlag strictly positive real-valued functions equipped with the Skorokhod topology;

(ii) $\boldsymbol{l}_t = \textbf{vect}(l_{u,t}, \boldsymbol{l}_{v,t}, \boldsymbol{l}_{\zeta,t})$ is a $n \times 1$ vector m.d.s. w.r.t to $\mathcal{F}_t = \{\boldsymbol{l}_t, \boldsymbol{l}_{t-1}, \ldots\}$ to which it is adapted, with conditional covariance matrix $\boldsymbol{\Sigma}_{t|t-1} = \text{E}(\boldsymbol{l}_t\boldsymbol{l}_t' \mid \mathcal{F}_{t-1}) = \textbf{diag}\left(\boldsymbol{\Sigma}_{t|t-1}^{(1)}, \boldsymbol{\Sigma}_{t|t-1}^{(2)}\right)$ and unconditional variance $\text{E}(\boldsymbol{l}_t\boldsymbol{l}_t') = \boldsymbol{I}_n$.

(iii) $\sup_t \text{E}\,\|\boldsymbol{l}_t\|^{4+\delta} < \infty$ for some $\delta > 0$; $\text{E}\,\|\boldsymbol{\xi}_0\|^4 < \infty$, where $\boldsymbol{\xi}_0 = \textbf{vect}(\tilde{\boldsymbol{z}}_0, \tilde{\boldsymbol{z}}_{-1}, \ldots, \tilde{\boldsymbol{z}}_{-p+1})$.

(iv) $\text{E}\left((\boldsymbol{l}_t\boldsymbol{l}_t') \otimes \boldsymbol{l}_{t-i}\right) = \boldsymbol{\rho}_i$ for all $i \geq 0$, with $\sup_{i\geq 0}\|\boldsymbol{\rho}_i\| < \infty$.

(v) $\text{E}\left((\boldsymbol{l}_t\boldsymbol{l}_t') \otimes (\boldsymbol{l}_{t-i}\boldsymbol{l}_{t-j}')\right) = \boldsymbol{\rho}_{i,j}$, for all $i, j \geq 0$ with $\sup_{i,j\geq 0}\|\boldsymbol{\rho}_{i,j}\| < \infty$.

Assumption 9′. Let $\boldsymbol{n}_t = \textbf{vect}(l_{u,t}, \boldsymbol{l}_{v,t})$. Then:

(i) Assumption 9(iv) holds with $\text{E}[(\boldsymbol{n}_t\boldsymbol{n}_t') \otimes \boldsymbol{n}_{t-i}] = \boldsymbol{0}_{(p_1+1)^2\times(p_1+1)}$ for all $i \geq 1$.

(ii) Assumption 9(v) holds with $\text{E}[(\boldsymbol{n}_t\boldsymbol{n}_t') \otimes (\boldsymbol{n}_{t-i}\boldsymbol{n}_{t-j}')] = \boldsymbol{0}_{(p_1+1)^2\times(p_1+1)^2}$ for all $i, j \geq 1$ and $i \neq j$.

(iii) Assumption 9(v) holds with $\text{E}[(\boldsymbol{n}_t\boldsymbol{n}_t') \otimes (\boldsymbol{n}_{t-i}\boldsymbol{l}_{\zeta,t-j}')] = \boldsymbol{0}_{(p_1+1)^2\times(p_1+1)p_2}$ for all $i \geq 1$ and $j \geq 0$.

(iv) $\sup_t \text{E}\,\|\boldsymbol{l}_t\|^8 < \infty$.

Assumption 10. (i) $\nu_t \overset{i.i.d.}{\sim} (0, 1)$ independent of the original data generated by (1), (2) and (3); (ii) $\text{E}^b\,|\nu_t|^{4+\delta^*} = \bar{c} < \infty$, for some $\delta^* > 0$, for all t, where E^b denotes the expectation under the bootstrap measure.

Before presenting our main theoretical results, we discuss certain aspects of the assumptions.

Remark 1. Assumptions 1–5 indicate that the breaks are "fixed" in the sense that the size of the associated shifts in the parameters between regimes is constant and does not change with the sample size.

Remark 2. It follows from Assumption 7 that $\tilde{\boldsymbol{z}}_t$ follows a finite order VAR in (4) that is stable within each regime.

Remark 3. Assumption 8 is the identification condition for estimation of the structural equation parameters; see Hall et al. (2012) for further discussion.

Remark 4. From Assumption 9 it follows that $\boldsymbol{\epsilon}_t$ is a vector m.d.s. relative to $\mathcal{F}_{t-1}$ with time varying conditional and unconditional variance given by $\text{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'|\mathcal{F}_{t-1}) = \boldsymbol{S}\boldsymbol{D}_t\boldsymbol{\Sigma}_{t|t-1}\boldsymbol{D}_t'\boldsymbol{S}'$ and $\text{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t') = \boldsymbol{S}\boldsymbol{D}_t\boldsymbol{D}_t'\boldsymbol{S}'$ respectively. The m.d.s. property implies that all the dynamic structure in the SE for $\boldsymbol{y}_t$ and RF for $\boldsymbol{x}_t$ is accounted for by the variables in $\boldsymbol{z}_{1,t}$ and $\boldsymbol{z}_t$ respectively. As noted by Boswijk et al. (2016) and Georgiev et al. (2018), Assumption 9 allows for $\boldsymbol{\epsilon}_t$ to exhibit conditional and unconditional heteroskedasticity of unknown and general form that can include single or multiple variance shifts, variances that follow a broken trend or follow a smooth transition model. When $\boldsymbol{D}_t = \boldsymbol{D}$, the unconditional variance is constant but we may have conditional heteroskedasticity. When $\boldsymbol{\Sigma}_{t|t-1} = \boldsymbol{I}_n$, the unconditional variance may still be time-varying. Note that Assumption 9(i)–(ii) imply that $\boldsymbol{x}_t$ is endogenous and $\boldsymbol{r}_t$ is contemporaneously exogenous in the SE. Assumption 9(iii) is a moment condition about $\boldsymbol{l}_t$ (similar to Assumption A(iv) in Gonçalves and Kilian (2004) and Assumption 2(iv) in Boswijk et al. (2016)) and a moment condition on the initial values of the VAR in (4). Assumption 9(iv) allows for leverage effects (the correlation between the conditional variance and $\boldsymbol{l}_{t-i}$ is nonzero, when $i \geq 1$). Assumption 9(v) allows for (asymmetric) volatility clustering (the conditional variance is correlated with cross-products $\boldsymbol{l}_{t-i}\boldsymbol{l}_{t-j}$, for $i, j \geq 1$).[12]

Remark 5. Assumption 9′ is only imposed in the case of the WR bootstrap. Assumption 9′(i)–(iii) is needed because the WR bootstrap sets to zero certain covariance terms in the distribution of the bootstrapped parameter estimates given the data. This happens because these moments depend on products of bootstrap errors at different lags and these terms have zero expectation under the bootstrap measure due to the fact that ν_t is mean zero and i.i.d. Assumption 9′(i) is a restriction on the leverage effects and Assumption 9′(ii) is a restriction of the asymmetric effects allowed in volatility clustering. Note that Assumption 9′(i) is only needed when we have an intercept in (4). Assumption 9′(iii) arises because the WR design bootstraps the lags of y_t and $\boldsymbol{x}_t$ in (4), but it does not bootstrap $\boldsymbol{r}_t$ and its lags. Therefore, certain fourth cross-moments involving both types of quantities are set to zero by the WR bootstrap, leading to the restriction on clustering effects in Assumption 9′(iii) (where $i = j$ is imposed for replicating certain variances, and $i \neq j$ is imposed for replicating certain

[12] The clustering is asymmetric if $\rho_{i,j} \neq 0$ when $i \neq j$.

covariances in the asymptotic distribution of the parameter estimates). Assumption 9′(iv) is needed in order to verify one of the conditions of the CLT for m.d.s., by ensuring the convergence of the WR bootstrap variance to the correct limiting variance. Assumption 9′(iv) is similar to Assumption A′(vi′) in Gonçalves and Kilian (2004) and Assumption 2′ in Boswijk et al. (2016). However, Assumption 9′(iv) can be replaced with Assumption 9(iii) if ν_t (in Assumption 10) used in the WR bootstrap follows the Rademacher two point distribution suggested in Liu (1988).[13]

Remark 6. There are several choices for the distribution of ν_t, the random variable used in construction of the bootstrap errors: Gonçalves and Kilian (2004) use the standard normal distribution, while Mammen (1993) suggested an asymmetric two-point distribution and Liu (1988) suggested the Rademacher two-point distribution. In this paper, we report simulation results for Liu's (1988) two-point distribution, which we found performed the best compared to the other distributions in simulations not reported here. This conclusion is similar to Davidson and Flachaire (2008) and Davidson and MacKinnon (2010).

The following theorems establish the asymptotic validity of the bootstrap versions of the *sup-Wald* tests.

Theorem 1. *If the WF bootstrap is used let Assumptions 1–10 hold and if the WR bootstrap is used let Assumptions 1–10 and 9′ hold. If y_t, $\boldsymbol{x}_t$ and $\boldsymbol{r}_t$ are generated by (1), (2)* and *(3) and $m = 0$ then it follows that*

$$\sup_{c\in\mathbb{R}} \left|P^b\left(\sup\text{-}Wald_T^b \le c\right) - P(\sup\text{-}Wald_T \le c)\right| \xrightarrow{p} 0$$

as $T \to \infty$, where P^b denotes the probability measure induced by the bootstrap.

Theorem 2. *If the WF bootstrap is used let Assumptions 1–10 hold and if the WR bootstrap is used let Assumptions 1–10 and 9′ hold. If y_t, $\boldsymbol{x}_t$ and $\boldsymbol{r}_t$ are generated by (1), (2)* and *(3) and $m = \ell$ then it follows that:*

$$\sup_{c\in\mathbb{R}} \left|P^b\left(\sup\text{-}Wald_T^b(\ell+1 \mid \ell) \le c\right) - P(\sup\text{-}Wald_T(\ell+1 \mid \ell) \le c)\right| \xrightarrow{p} 0$$

as $T \to \infty$, where P^b denotes the probability measure induced by the bootstrap.

Remark 7. The proof rests on showing the sample and bootstrap statistics have the same limiting distribution. Although this distribution is known to be non-pivotal if the RF is unstable (see Perron and Yamamoto, 2014), to our knowledge this distribution has not previously been presented in the literature. A formal characterization of this distribution is provided in the Supplementary Appendix.

Remark 8. Theorem 1–2 cover the case where the reduced form is stable and the errors are unconditionally homoskedastic. In this case, the sup-*Wald* tests are asymptotically pivotal and so the bootstrap is expected to provide a superior approximation to finite sample behavior compared to the limiting distribution because the bootstrap, by its nature, incorporates sample information. However, a formal proof is left to future research.

Remark 9. Hall et al. (2012) also propose testing the hypotheses described above using sup-F tests. While F-tests are designed for use in regression models with homoskedastic errors,[14] wild bootstrap versions of the tests can be used as a basis for inference when the errors exhibit heteroskedasticity. In the Supplementary Appendix, we present WR bootstrap and WF bootstrap versions of appropriate sup-F statistics for testing both $H_0 : m = 0$ versus $H_1 : m = k$ and $H_0 : m = \ell$ versus $H_1 : m = \ell + 1$, and show that these bootstrap versions of the sup-F tests are asymptotically valid under the same conditions as their sup-*Wald* counterparts. Simulation evidence indicated no systematic difference in the finite sample behavior of the sup-*Wald* and sup-F tests for a given null and bootstrap method, and so further details about this approach are relegated to the Supplementary Appendix.

Remark 10. In the special case where there are no endogenous regressors in the equation of interest then our framework reduces to one in which a linear regression model is estimated via OLS. For this set-up, the asymptotic validity of wild fixed bootstrap versions of sup-F test for parameter variation (our Case(i) above) has been established under different sets of conditions by Hansen (2000) and Georgiev et al. (2018). Hansen (2000) considers the case where the marginal distribution of the exogenous regressors changes during the sample. Georgiev et al. (2018) consider Hansen's (2000) bootstrap in the context of predictive regressions with strongly persistent exogenous regressors. Our results complement these earlier studies because we provide results for the wild recursive bootstrap and a theoretical justification for tests of ℓ breaks against $\ell + 1$ based on bootstrap methods.

13 See the proof of Lemma 10.

14 If the reduced form is stable then the limiting distribution of the *sup-F* statistics are only pivotal if the errors are homoskedastic.

4. Simulation results

In this section, we investigate the finite sample performance of the bootstrap versions of the sup-*Wald* and sup-*F* statistics. We consider a number of designs that involve stability or instability in the SE and/or the RF. In all the designs the variable x_t is endogenous and the SE is estimated by 2SLS. Recalling from above that h and m denote the true number of breaks in the RF and SE respectively, the four scenarios we consider are as follows.

- *Scenario: (h,m)=(0,0)*
  The DGP is as follows:

$$x_t = \alpha_x + \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 + \delta_{x_1}^0 x_{t-1} + \delta_{y_1}^0 y_{t-1} + v_t, \quad \text{for } t = 1, \ldots, T, \tag{24}$$
$$y_t = \alpha_y + x_t\beta_x^0 + \beta_{r_1}^0 r_{1,t} + \beta_{y_1}^0 y_{t-1} + u_t, \quad \text{for } t = 1, \ldots, T, \tag{25}$$

  where the parameters of the SE – see Eq. (25) – are $\alpha_y = 0.5$, $\beta_x^0 = 0.5$, $\beta_{r_1}^0 = 0.5$, $\beta_{y_1}^0 = 0.8$; the parameters of the RF in Eq. (24) are $\alpha_x = 0.5$, $\boldsymbol{\delta}_r^0 = (1.5, 1.5, 1.5, 1.5)'$ a 4×1 parameter vector, $\delta_{x_1}^0 = 0.5$, $\delta_{y_1}^0 = 0.2$; $\boldsymbol{r}_t = (r_{1,t}, \boldsymbol{r}_{2,t}')'$.
- *Scenario: (h,m)=(1,0)*
  The DGP is as follows:

$$x_t = \alpha_{x,(1)} + \boldsymbol{r}_t'\boldsymbol{\delta}_{r,(1)}^0 + \delta_{x_1,(1)}^0 x_{t-1} + \delta_{y_1,(1)}^0 y_{t-1} + v_t, \quad \text{for } t = 1, \ldots, [T/4], \tag{26}$$
$$= \alpha_{x,(2)} + \boldsymbol{r}_t'\boldsymbol{\delta}_{r,(2)}^0 + \delta_{x_1,(2)}^0 x_{t-1} + \delta_{y_1,(2)}^0 y_{t-1} + v_t, \quad \text{for } t = [T/4] + 1, \ldots, T, \tag{27}$$
$$y_t = \alpha_y + x_t\beta_x^0 + \beta_{r_1}^0 r_{1,t} + \beta_{y_1}^0 y_{t-1} + u_t, \quad \text{for } t = 1, \ldots, T, \tag{28}$$

  where the parameters of the SE – Eq. (28) – are the same as in scenario $(h, m) = (0, 0)$, and the RF parameters – Eqs. (26)–(27) – are: $\alpha_{x,(1)} = 0.1$, $\alpha_{x,(2)} = 0.5$, $\boldsymbol{\delta}_{r,(1)}^0 = (0.1, 0.1, 0.1, 0.1)'$, $\boldsymbol{\delta}_{r,(2)}^0 = (1.5, 1.5, 1.5, 1.5)'$, $\delta_{x_1,(1)}^0 = 0.1$, $\delta_{x_1,(2)}^0 = 0.5$, $\delta_{y_1,(1)}^0 = 0.1$, and $\delta_{y_1,(2)}^0 = 0.2$. In our simulation study, prior to testing the null hypothesis of zero breaks in the SE parameters from (28), we test sequentially for breaks in the RF parameters (assuming for a maximum of 2 breaks) by applying our bootstrap sup-*Wald* test. More exactly we tested the null hypothesis $H_0 : h = \ell$ against $H_1 : h = \ell + 1$, $\ell = 0, 1$ using the WR and WF bootstrap sup-*Wald* for OLS. If the bootstrap p-value (given by the fraction of bootstrap statistics more extreme than the sup-*Wald* based on the original sample) was larger than 5%, then we imposed the ℓ breaks (assumed under null $H_0 : h = \ell$) in the RF and estimated their locations which were subsequently accounted for in the estimation of the SE; see the first two columns of Table A.3.
- *Scenario: $(h, m) = (0,1)$*
  The DGP is as follows:

$$x_t = \alpha_x + \boldsymbol{r}_t'\boldsymbol{\delta}_r^0 + \delta_{x_1}^0 x_{t-1} + \delta_{y_1}^0 y_{t-1} + v_t, \quad \text{for } t = 1, \ldots, T, \tag{29}$$
$$y_t = \alpha_{y,(1)} + x_t\beta_{x,(1)}^0 + \beta_{r_1,(1)}^0 r_{1,t} + \beta_{y_1,(1)}^0 y_{t-1} + u_t, \quad \text{for } t = 1, \ldots, [3T/4], \tag{30}$$
$$= \alpha_{y,(2)} + x_t\beta_{x,(2)}^0 + \beta_{r_1,(2)}^0 r_{1,t} + \beta_{y_1,(2)}^0 y_{t-1} + u_t, \quad \text{for } t = [3T/4] + 1, \ldots, T, \tag{31}$$

  where the parameter values for the RF – Eq. (29) – are as in scenario $(h, m) = (0, 0)$, and the parameters on the SE – equations (30)–(31) – are: $\alpha_{y,(1)} = 0.5$, $\alpha_{y,(2)} = -0.5$; $\beta_{x,(1)}^0 = 0.5$, $\beta_{x,(2)}^0 = -0.5$; $\beta_{r_1,(1)}^0 = 0.5$, $\beta_{r_1,(2)}^0 = -0.5$, $\beta_{y_1,(1)}^0 = 0.8$, and $\beta_{y_1,(2)}^0 = 0.1$.
- *Scenario: (h,m)=(1,1)*
  The DGP is as follows:

$$x_t = \alpha_{x,(1)} + \boldsymbol{r}_t'\boldsymbol{\delta}_{r,(1)}^0 + \delta_{x_1,(1)}^0 x_{t-1} + \delta_{y_1,(1)}^0 y_{t-1} + v_t, \quad \text{for } t = 1, \ldots, [T/4], \tag{32}$$
$$= \alpha_{x,(2)} + \boldsymbol{r}_t'\boldsymbol{\delta}_{r,(2)}^0 + \delta_{x_1,(2)}^0 x_{t-1} + \delta_{y_1,(2)}^0 y_{t-1} + v_t, \quad \text{for } t = [T/4] + 1, \ldots, T, \tag{33}$$
$$y_t = \alpha_{y,(1)} + x_t\beta_{x,(1)}^0 + \beta_{r_1,(1)}^0 r_{1,t} + \beta_{y_1,(1)}^0 y_{t-1} + u_t, \quad \text{for } t = 1, \ldots, [3T/4], \tag{34}$$
$$= \alpha_{y,(2)} + x_t\beta_{x,(2)}^0 + \beta_{r_1,(2)}^0 r_{1,t} + \beta_{y_1,(2)}^0 y_{t-1} + u_t, \quad \text{for } t = [3T/4] + 1, \ldots, T, \tag{35}$$

  where the parameters of the RF – Eqs. (32)–(33) – are as in scenario $(h, m) = (1, 0)$ and the parameters in the SE – Eqs. (34)–(35) – are as in $(h, m) = (0, 1)$. In our simulation study, prior to testing the null hypothesis of zero breaks in the SE parameters from (28), we test sequentially for breaks in the RF parameters (assuming for a maximum of 2 breaks) by applying our bootstrap sup-*Wald* test as described in Scenario $(h, m) = (1, 0)$; see the first two columns of Table A.4. For Scenarios $(h, m) = (0, 0)$ and $(h, m) = (0, 1)$ the true number of breaks in RF $(h = 0)$ is imposed before testing for breaks in SE.

For the four scenarios above we consider the following choices for u_t, $\boldsymbol{v}_t$ and $\boldsymbol{r}_t$:

Case A: u_t and $v_t \overset{i.i.d.}{\sim} N(0, 1)$, $\text{Cov}(u_t, v_t) = 0.5$, $t = 1, \ldots, T$, $\boldsymbol{r}_t \overset{i.i.d.}{\sim} N(\boldsymbol{0}_{4\times 1}, \boldsymbol{I}_4)$.

Case B: u_t and v_t are GARCH(1,1) processes *i.e.* $u_t = \tilde{u}_t/\sqrt{\text{Var}(\tilde{u}_t)}$ and $v_t = \tilde{v}_t/\sqrt{\text{Var}(\tilde{v}_t)}$ with $\tilde{u}_t = \sigma_{\tilde{u},t}\vartheta_{\tilde{u},t}$ and $\tilde{v}_t = \sigma_{\tilde{v},t}\vartheta_{\tilde{v},t}$, $\vartheta_{\tilde{u},t}$ and $\vartheta_{\tilde{v},t} \overset{i.i.d.}{\sim} N(0,1)$, $\text{Cov}(\vartheta_{\tilde{u},t}, \vartheta_{\tilde{v},t}) = 0.5$, $\sigma^2_{\tilde{u},t} = \gamma_0 + \gamma_1\tilde{u}^2_{t-1} + \gamma_2\sigma^2_{\tilde{u},t-1}$, $\sigma^2_{\tilde{v},t} = \gamma_0 + \gamma_1\tilde{v}^2_{t-1} + \gamma_2\sigma^2_{\tilde{v},t-1}$, where $\gamma_0 = 0.1$ and $\gamma_1 = \gamma_2 = 0.4$, $t = 1, \ldots, T$, $\boldsymbol{r}_t$ is as in *Case A*.

Case C: u_t and $v_t \overset{i.i.d.}{\sim} N(0,1)$, $\text{Cov}(u_t, v_t) = 0.5$, $t = 1, \ldots, [T/3]$; u_t and $v_t \overset{i.i.d.}{\sim} N(0,2)$, $\text{Cov}(u_t, v_t) = 0.5$, $t = [T/3]+1, \ldots, T$, $\boldsymbol{r}_t$ is as in *Case A*.

Case D: u_t and v_t are as in *Case C* and $\boldsymbol{r}_t \overset{i.i.d.}{\sim} N(\boldsymbol{0}_{4\times1}, \boldsymbol{I}_4)$ for $t = 1, \ldots, [3T/5]$, and $\boldsymbol{r}_t \overset{i.i.d.}{\sim} N(\boldsymbol{0}_{4\times1}, 1.5\boldsymbol{I}_4)$ for $t = [3T/5]+1, \ldots, T$.

In *Case A*, the errors u_t and v_t are homoskedastic and the contemporaneous exogenous regressors $\boldsymbol{r}_t$ are stable. In *Case B*, the errors are conditionally heteroskedastic. In *Case C* the errors have a contemporaneous upward shift in the unconditional variance, while in *Case D* there is also an upward shift in the variance of $\boldsymbol{r}_t$.

In our simulations we consider the behavior of the bootstrap tests both under their null and alternative hypotheses. For scenarios $(h,m) = (0,0)$ and $(h,m) = (1,0)$ we consider the behavior of the sup-$Wald_T$. For scenarios $(h,m) = (0,1)$ and $(h,m) = (1,1)$ we consider the performance of the sup-$Wald_T(2|1)$. In order to assess the power of our bootstrap tests we also consider the case when the null hypotheses are not true and there is an additional break in the SE parameters at $[T/2]$. Specifically, we consider in all the four scenarios described above the following:

$$y_t = (\alpha_{y,(i)} + g) + x_t(\beta^0_{x,(i)} + g) + (\beta^0_{r_1,(i)} + g)r_{1,t} + (\beta^0_{y_1,(i)} + g)y_{t-1} + u_t, \quad \text{for } t = [T/2]+1, \ldots, \tilde{T}, \tag{36}$$

with g a constant; $i = 1$ and $\tilde{T} = T$ for scenarios $(h,m) = (0,0)$ and $(h,m) = (1,0)$, and the equation for y_t for $t < [T/2]+1$ is the same as that given in the two scenarios $(h,m) = (0,0)$ and $(h,m) = (1,0)$; $i = 2$ and $\tilde{T} = [3T/4]$ for scenarios $(h,m) = (1,0)$ and $(h,m) = (1,1)$, and the equation for y_t for $t < [T/2]+1$ and $t > [3T/4]$ is the same as that given in the two scenarios $(h,m) = (1,0)$ and $(h,m) = (1,1)$. When $g = 0$, the null hypothesis is satisfied. We illustrate the behavior of the tests under the alternative hypothesis for the following values of g: $g = -0.007, -0.009$ for scenario $(h,m) = (0,0)$; $g = -0.05, -0.07$ for scenario $(h,m) = (1,0)$; $g = 0.3, 0.4$ for scenario $(h,m) = (0,1)$; $g = -0.5, 0.5$ for scenario $(h,m) = (1,1)$.

For scenarios $(h,m) = (1,0)$ and $(h,m) = (1,1)$ we have tested for the presence of max 2 breaks in the RF for x_t (in (26)–(27) and (32)–(33) respectively) prior to testing for breaks in the SE. More exactly we tested the null hypothesis $H_0 : h = \ell$ against $H_1 : h = \ell + 1$, $\ell = 0, 1$ using the WR and WF bootstrap sup-*Wald* for OLS. If the bootstrap p-value (given by the fraction of bootstrap statistics more extreme than the sup-*Wald* based on the original sample) was larger than 5%, then we imposed the ℓ breaks (assumed under null $H_0 : h = \ell$) in the RF and estimated their locations which were subsequently accounted for in the estimation of the SE.

We now describe other features of the calculations before discussing the results. For the WR and the WF bootstraps the auxiliary distribution (from Assumption 10) is the Rademacher distribution proposed by Liu (1988) which assigns 0.5 probability to the value $\nu_t = -1$ and 0.5 probability to $\nu_t = 1$, $t = 1, \ldots, T$. The same ν_t is used to obtain the bootstrap residuals $u^b_t = \hat{u}_t\nu_t$ and $v^b_t = \hat{v}_t\nu_t$ in order to preserve the contemporaneous correlation between the error terms. We consider $T = 120, 240, 480$ for the sample size and $B = 399$ for the number bootstrap replications. All results are calculated using $N^* = 1,000$ replications.

The reported rejection rates of the WR and WF bootstraps are calculated as: $\frac{1}{N^*}\sum_{j=1}^{N^*} 1_{t_j \geq t^b_{1-\alpha_1,j}}$, where $\alpha_1 = 0.10, 0.05, 0.01$ are the nominal values of the tests; t_j is the statistic (sup-*Wald*) computed from the original sample; $t^b_{1-\alpha_1,j}$ is $1-\alpha_1$ quantile of the bootstrap distribution calculated as $(1-\alpha_1)(B+1)$ bootstrap order statistic from the sample of bootstrap statistics in simulation $j = 1, \ldots, N^*$.

For the WR bootstrap, the bootstrap samples were generated recursively with start-up values for y^b_1 and x^b_1 being given by the first observations from the sample (x_1, y_1); see Davidson and MacKinnon (1993).

In all settings, the bootstrap samples are generated by imposing the null hypothesis. The value of ϵ, the trimming parameter in Assumption 6, is set to 0.15 which is a typical value used in the literature.

We now turn to our results. We present results for the sup-*Wald* test under both the null and alternative hypotheses in Tables A.1–A.4 of the paper. In Tables H.1–H.4 of Appendix H in the Supplementary Appendix we also present similar results for the sup-F test. The first two columns of these tables give the rejection rates of the tests under the null hypothesis, while columns 3–6 give the rejection rates of the tests under the alternative hypothesis.[15]

From the first two columns of Tables A.1–A.4, it can be seen that the WR bootstrap works better in general than the WF bootstrap. The latter has large size distortions for scenarios $(h,m) = (0,0)$, $(h,m) = (0,1)$ and $(h,m) = (1,0)$ whether the errors are conditionally homoskedastic, are conditionally heteroskedastic or have a break in the unconditional variance. For scenario $(h,m) = (1,1)$, the WF bootstrap is only slightly undersized or oversized. Regarding the behavior of the sup-*Wald* test under the alternative hypothesis, the main conclusion that emerges from columns 3–6 of Tables A.1–A.4 is that the power is influenced in small samples ($T = 120$) by the number of breaks in RF and SE, the distribution of the

[15] The rejection rates under the alternative are not level-adjusted, but since we have used the same sequence of random numbers for repetition j, $j = 1, \ldots, N^*$, in the experiments under both null and the alternative hypotheses, one can always subtract (or add) the positive (or negative) size discrepancy (relative to the nominal size) from the rejection rate under the alternative in order to obtain the level-adjusted power of the test; see Davidson and MacKinnon (1998).

errors u_t and v_t, the distribution of r_t, as well as the number of breaks in the variance of the errors and in the variance of $\boldsymbol{r}_t$. When there is a break in SE, we need a larger g in (36) to be able to see an increase in the power of the test, compared with scenarios with no break in SE ($g = 0.3, 0.4$ for scenario $(h, m) = (0, 1)$, and $g = -0.5, 0.5$ for scenario $(h, m) = (1, 1)$, while $g = -0.007, -0.009$ for scenario $(h, m) = (0, 0)$ and $g = -0.05, -0.07$ for scenario $(h, m) = (1, 0)$). This can be explained by the fact that the second break in the SE is tested over smaller samples than the first break in the SE. Moreover, the power is lower for the smallest sample ($T = 120$) when the error terms have an upward shift in the variance (*Case C* in Tables A.1–A.4) and the contemporaneous exogenous regressors also have an upward shift in their variance (*Case D*). However, for $T = 240, 480$ the power increases sharply in all cases.

In Tables A.3 and A.4 we have sequentially tested for the presence of max 2 breaks in the RF for x_t (in (26)–(27) and (32)–(33) respectively) using the WR/WR sup-*Wald* for OLS, and the resulting number of RF breaks was imposed in each simulation prior to estimating the RF and SE and computing the test statistics for 2SLS. The fraction of times that 0, 1, 2 breaks were detected in RF (out of 1,000 replications of the scenarios), is given in Tables H.7–H.8 from Appendix H of the Supplementary Appendix. To assess the impact of the pre-testing in RF (in the first two columns of Tables A.3 and A.4), we have obtained the rejection frequencies of the bootstrap tests when the number of breaks in the RF is held at the true number; see (the first two columns of) Tables H.5 and H.6 from Appendix H of the Supplementary Appendix. To complement our results, we have also considered a break in RF of smaller size than the one mentioned after (26)–(27) by taking $\boldsymbol{\delta}^0_{r,(1)} = (1, 1, 1, 1)'$ (and the rest of the parameter values are as mentioned after (26)–(27)); see Tables H.9 and H.10 from Appendix H of the Supplementary Appendix.

Looking at the results for the sup-*Wald* our results suggest that in the smaller samples ($T = 120, 240$) the recursive bootstrap is clearly to be preferred over the fixed regressor bootstrap. In the larger sample ($T = 480$), the case for the WR over the WF is more marginal as the latter yields only slightly oversized tests. This relative ranking of the two methods is intuitive from the perspective of Davidson's (2016) first "golden rule" of bootstrap, which states: "The bootstrap DGP [...] must belong to the model [...] that represents the null hypothesis". The fixed regressor bootstraps treat the lagged dependent variables in the RF and SE as fixed across bootstrap samples, and as such do not seem to replicate the true model that represents the null hypothesis. This would seem to point toward a recommendation to use the WR but it is important to note an important caveat to our results: our designs involve models for which both recursive and fixed bootstraps are valid. As discussed in Section 3, the fixed regressor bootstrap is asymptotically valid under weaker conditions than the recursive bootstrap. Therefore, while the recursive bootstrap works best in the settings considered here, there may be other settings of interest in which only the fixed bootstrap is valid and so would obviously be preferred.

5. Concluding remarks

In this paper, we analyze the use of bootstrap methods to test for parameter change in linear models estimated via Two Stage Least Squares (2SLS). Two types of test are considered: one where the null hypothesis is of no change and the alternative hypothesis involves discrete change at k unknown break-points in the sample; and a second test where the null hypothesis is that there is discrete parameter change at ℓ break-points in the sample against an alternative in which the parameters change at $\ell + 1$ break-points. In both cases, we consider inferences based on a sup-*Wald*-type statistic using either the wild recursive bootstrap or the wild fixed regressor bootstrap. We establish the asymptotic validity of these bootstrap tests under a set of general conditions that allow the errors to exhibit conditional and/or unconditional heteroskedasticity and the regressors to have breaks in their marginal distributions. While we focus on inferences based on sup-*Wald* statistics, our arguments are easily extended to establish the asymptotic validity of inferences based on bootstrap versions of the analogous tests based on sup-*F* statistics; see Appendix G from the Supplementary Appendix available online.

Our simulation results show that the wild recursive bootstrap is more reliable compared to the wild fixed regressor bootstrap, yielding sup-*Wald*-type tests with empirical size equal or close to the nominal size. The gains from using the wild recursive bootstrap are quite clear in the smaller sample sizes, but are more marginal in the largest sample size ($T = 480$) in our simulation study. This would seem to point toward a recommendation to use the wild recursive bootstrap but it is important to note that the wild fixed bootstrap is asymptotically valid under less restrictive conditions than the wild recursive bootstrap. Thus, while both bootstraps are valid in our simulation design, there may be other circumstances when the recursive bootstrap is invalid and the fixed bootstrap would be preferred. The powers of the bootstrap tests are affected in small samples by the characteristics of the error distribution, but in moderate sample sizes often encountered in macroeconomics, there is a very sharp increase in power.

Our analysis covers the cases where the first-stage estimation of 2SLS involves a model whose parameters are either constant or themselves subject to discrete parameter change. If the errors exhibit unconditional heteroscedasticity and/or the reduced form is unstable then the bootstrap methods are particularly attractive because the limiting distributions are non-pivotal. As a result, critical values have to be simulated on a case-by-case basis. In principle it may be possible to simulate these critical values directly from the limiting distributions presented in Appendix C from our Supplementary Appendix replacing unknown moments and parameters by their sample estimates but this would seem to require knowledge (or an estimate of) the function driving the unconditional heteroskedasticity. In contrast, the bootstrap approach is far more convenient because it involves simulations of the estimated data generation process using the residuals and so does not require knowledge of the form of heteroskedasticity. Furthermore, our results indicate that the bootstrap approach yields reliable inferences in the sample sizes often encountered in macroeconomics.

Table A.1
Scenario:(h,m) = *(0,0)* – rejection probabilities from testing $H_0 : m = 0$ vs. $H_1 : m = 1$ with bootstrap sup-*Wald* test.

	WR bootstrap Size $g = 0$			WF bootstrap Size $g = 0$			WR bootstrap Power $g = -0.007$			WF bootstrap Power $g = -0.007$			WR bootstrap Power $g = -0.009$			WF bootstrap Power $g = -0.009$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Case A																		
120	11.8	6.1	1.6	15.1	8.7	2.4	59.2	48.3	25	61.1	55.3	31.5	79.4	70.3	49.1	84.5	75	56.3
240	9.3	4	0.8	12.9	6.4	0.9	99.7	99.7	99.6	99.8	99.7	99.7	100	100	99.9	100	100	100
480	10.08	5.09	1.15	9.76	5.52	1.04	100	100	100	100	100	100	100	100	100	100	100	100
Case B																		
120	12	5.9	0.7	14.4	8.5	1.7	65.5	54.2	32.6	71.3	61.1	38.6	83.2	75.9	54.7	87.1	80.3	62.6
240	9.5	4.7	1.1	11.9	6.2	1.4	99.8	99.8	99.5	99.9	99.9	99.7	100	100	100	100	100	100
480	10.1	4.9	0.5	11.5	6.1	1.3	100	100	100	100	100	100	100	100	100	100	100	100
Case C																		
120	9.9	5.6	1.5	15.7	8.3	1.9	46.8	33.6	12.5	58.9	45	22.1	70	56.8	29.4	78.4	68.1	43
240	9.9	5.1	0.7	15.2	8.4	1.2	99.6	99.5	99	99.6	99.5	99	99.8	99.8	99.5	99.9	99.9	99.7
480	8.8	5.1	1.3	12	6.5	2.1	100	100	100	100	100	100	100	100	100	100	100	100
Case D																		
120	10.7	5.3	1	13.8	7.3	1.9	50.1	37.7	14.1	57.6	44.2	22.7	70.7	58.7	34.4	76.1	65.8	43.2
240	9.8	4.5	0.9	14.1	7.2	1.7	100	99.8	98.6	100	100	99.4	100	100	99.9	100	100	100
480	10.1	4.3	0.9	12.2	6	1.2	100	100	100	100	100	100	100	100	100	100	100	100

Notes. The first two columns refer to the case when $H_0 : m = 0$ is true ($g = 0$ in Eq. (36)). The next columns refer to the case when we test for $H_0 : m = 0$, but $H_1 : m = 1$ is true ($g = -0.007, -0.009$ in Eq. (36)). Under the null and the alternative hypotheses we impose $h = 0$ in the RF.

Table A.2
Scenario:(h,m) = *(0,1)* – rejection probabilities from testing $H_0 : m = 1$ vs. $H_1 : m = 2$ with bootstrap sup-*Wald* test.

	WR bootstrap Size $g = 0$			WF bootstrap Size $g = 0$			WR bootstrap Power $g = 0.3$			WF bootstrap Power $g = 0.3$			WR bootstrap Power $g = 0.4$			WF bootstrap Power $g = 0.4$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Case A																		
120	10.7	5	1.2	15.5	9.9	5.6	54.9	36.9	10.8	61.5	45.6	19.8	78.1	60.3	24.8	82.2	67.8	38.5
240	10.2	4.9	0.5	12.5	7.1	3.4	99.5	98.9	89.9	99.6	98.6	92.2	100	100	98.8	100	100	99.1
480	8	4.5	1	8.6	4.4	0.8	100	100	100	100	100	100	100	100	100	100	100	100
Case B																		
120	9.7	4.6	1	16	10.2	6.5	62.3	44.8	16.1	67.2	53.9	26.2	82.2	67.6	31.1	84.1	73.6	44.7
240	10.6	5.2	1.2	13.8	8.1	3	99.3	97.5	86.2	99.1	93	91.5	99.9	99.6	96.6	100	99.9	98.3
480	8	4.2	0.9	8.4	4.8	0.8	100	99.8	99.5	99.8	99.7	99.5	100	100	100	100	100	99.9
Case C																		
120	10.5	5.2	0.9	16.3	11	5.8	26.3	14.8	3.3	36.1	21.4	6.6	40.1	24.5	7.5	51.1	34.6	13
240	11	4.8	0.9	13.2	8.3	2.4	83.1	68.7	31.4	87.2	77.6	47.2	98.5	93.4	68.7	99	97	80.1
480	10.4	5.6	0.5	11.2	6.1	1.2	100	99.9	98.4	100	99.9	99.2	100	100	100	100	100	100
Case D																		
120	11.6	5.8	1.5	15.3	9.5	5.3	39.8	24.1	6.5	51.2	33.2	13.3	64.8	43.33	14	72.5	54.8	24.2
240	11.5	6	1	14.9	9.1	2.9	98.9	94.6	73	98.9	97.1	82.7	100	99.8	95.6	100	99.9	97.9
480	9.6	4	1.3	9.5	5.3	1.5	100	100	100	100	100	100	100	100	100	100	100	100

Notes. The first two columns refer to the case when $H_0 : m = 1$ is true ($g = 0$ in Eq. (36)). The next columns refer to the case when we test for $H_0 : m = 1$, but $H_1 : m = 2$ is true ($g = 0.3, 0.4$ in Eq. (36)). Under the null and the alternative hypotheses we impose $h = 0$ in the RF.

Acknowledgments

We thank the following for very valuable comments on this work: Yacine Aït-Sahalia (the co-editor), Donald Andrews, Xu Cheng, Russell Davidson, Frank Diebold, Sílvia Gonçalves, Bruce Hansen, Jan Magnus, Ulrich Müller, Anders Rahbek, Nikolaus Schweizer, Terence Tao, three anonymous referees, and the participants at University of Pennsylvania Econometrics Seminar, 2011; the Workshop on Recent Developments in Econometric Analysis, University of Liverpool, 2012; Tilburg University seminar, 2012; University of York seminar 2012; the European Meeting of the Econometric Society, Malaga 2012 and Toulouse 2014; CFE-ERCIM Conference, Pisa 2014; Durham University Business School seminar, 2015; Inference Issues in Econometrics, Amsterdam 2017; IAAE Conference, Montreal, 2018. Boldea acknowledges the support

Table A.3
Scenario:(h,m) = (1,0) – rejection probabilities from testing $H_0 : m = 0$ vs. $H_1 : m = 1$ with bootstrap sup-*Wald* test; number of breaks in the RF was estimated and imposed in each simulation using a sequential strategy based on the WR/WF sup-*Wald* for OLS.

	WR bootstrap Size $g = 0$			WF bootstrap Size $g = 0$			WR bootstrap Power $g = -0.05$			WF bootstrap Power $g = -0.05$			WR bootstrap Power $g = -0.07$			WF bootstrap Power $g = -0.07$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Case A																		
120	10.2	3.7	0.9	15.3	7.1	1.3	52.3	42.8	22.5	60.5	50.6	29.3	67.9	58.3	38.1	74.9	66.3	45.6
240	10.8	5.7	0.8	14	6.7	1.2	94.6	91.3	84.5	95.1	92.1	86	98.1	96.7	91.7	98.7	97.1	92.7
480	10.9	5.2	0.9	12.5	6	0.8	99.9	99.8	99.3	100	99.8	99.5	100	100	99.8	100	100	99.7
Case B																		
120	10.1	4.8	1	13.3	7.8	1.5	54.5	44.9	28.1	63.1	51.5	33.7	68.9	59.6	43.4	77	68	48.9
240	10	5.4	1.2	12.2	6.8	1.5	94.5	92.1	83.7	95.8	93.3	85.9	97.9	96.7	92	98.8	97.5	93.3
480	11	5.4	0.7	12.8	5.9	1.2	100	100	100	100	99.7	99.2	100	99.9	99.8	100	99.9	99.8
Case C																		
120	9.6	4.3	0.9	15.6	7.5	1.7	39.6	28.7	11.3	50.8	37.4	18.4	54.9	43.8	22.9	66.6	53.8	33.5
240	11.8	6	0.6	15.6	8.6	1.4	88.8	83.5	71.8	91.3	87.3	76.2	94.3	92	83.5	96	93.7	86.7
480	10.8	5.9	1.1	12.6	7.1	1.4	99.9	99.6	98.6	99.9	99.5	98.5	99.9	99.9	99.8	100	99.9	99.4
Case D																		
120	10.2	4.8	1.2	14.8	6.7	1.6	40.9	29.9	12.9	49	37.3	16.8	56	45.1	24.3	64.5	52.3	31.9
240	10.6	5.7	0.9	14.2	7.5	1.8	89.6	85.2	73.2	91.1	87.2	76.6	94.9	92.4	85.3	95.7	93.7	86.9
480	11.6	6.2	0.9	13.5	7.4	1	99.4	99.1	98	99.5	99.3	98	100	99.8	98.9	99.8	99.8	99.1

Notes. The first two columns refer to the case when $H_0 : m = 0$ is true ($g = 0$ in Eq. (36)). The next columns refer to the case when we test for $H_0 : m = 0$, but $H_1 : m = 1$ is true ($g = -0.05, -0.07$ in Eq. (36)). Prior to testing $H_0 : m = 0$ vs $H_1 : m = 1$ (for all columns above), we tested sequentially for the presence of maximum two breaks in the RF (we used the WR/WF bootstrap sup-*Wald* for OLS to test $H_0 : h = \ell$ vs. $H_1 : \ell + 1$, $\ell = 0, 1$). If breaks are detected in the RF, the number of breaks and the estimated locations are imposed when estimating the SE.

Table A.4
Scenario:(h,m) = (1,1) – rejection probabilities from testing $H_0 : m = 1$ vs. $H_1 : m = 2$ with bootstrap sup-*Wald* test; number of breaks in the RF was estimated and imposed in each simulation using a sequential strategy based on the WR/WF sup-*Wald* for OLS.

	WR bootstrap Size $g = 0$			WF bootstrap Size $g = 0$			WR bootstrap Power $g = 0.5$			WF bootstrap Power $g = 0.5$			WR bootstrap Power $g = -0.5$			WF bootstrap Power $g = -0.5$		
T	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%	10%	5%	1%
Case A																		
120	8.8	4.7	0.7	8.7	4.5	0.8	52	40.7	16	57.4	45.4	22.7	85	71.9	32.1	88.2	74.8	37.5
240	10.4	5.7	0.7	10.4	5.2	0.8	99.8	99.4	97.6	99.6	99.4	97.4	100	100	99.8	100	100	99.7
480	9.7	4.2	0.7	10.2	4.6	0.8	100	100	100	100	99.8	99.1	100	100	100	100	100	100
Case B																		
120	8.9	3.7	0.9	8.7	3.4	0.9	50.1	40.1	18.6	54.8	45.4	24.4	81.8	70.9	38.3	85.5	73	39.9
240	10.8	4.7	0.8	10.6	5.3	0.9	98.8	98.3	96	99.2	98.7	95.8	99.6	99.5	98.1	98	99.6	98.3
480	9.9	4.1	0.9	10.9	5.4	0.9	100	100	99.8	100	99.8	99.6	100	100	100	100	100	100
Case C																		
120	9.1	3.5	1	9.4	4	0.4	30.7	17.5	3.1	38.3	25.9	8	45.1	25.4	6.9	49.7	31.7	8.8
240	10.3	5.2	1	10.2	5	1	98.6	96.8	86.2	99	97.7	88.4	99.3	98.5	86.9	100	99.8	90.8
480	11.3	4.8	1	12.1	5.3	0.6	100	100	99.9	100	100	99.2	99.9	99.9	99.7	100	100	99.8
Case D																		
120	10.1	4.4	1.6	8.5	3.8	0.6	36.8	23.4	6.3	42.1	30.4	12.4	69.3	52.6	16.4	76.8	59.4	25.4
240	10.9	4.9	0.8	11.8	5.2	0.8	99.2	98.6	94	99.6	98.9	94.5	99.5	99.4	98	99.9	99.9	98.6
480	10.2	5.3	1.4	11	5.6	1.2	100	100	100	100	100	98.1	100	100	100	100	100	100

Notes. The first two columns refer to the case when $H_0 : m = 1$ is true ($g = 0$ in Eq. (36)). The next columns refer to the case when we test for $H_0 : m = 1$, but $H_1 : m = 2$ is true ($g = -0.5, 0.5$ in Eq. (36)). Prior to testing $H_0 : m = 1$ vs $H_1 : m = 2$ (for all columns above), we tested sequentially for the presence of maximum two breaks in the RF (we used the WR/WF bootstrap sup-*Wald* for OLS to test $H_0 : h = \ell$ vs. $H_1 : \ell + 1$, $\ell = 0, 1$). If breaks are detected in the RF, the number of breaks and the estimated locations are imposed when estimating the SE.

of the NWO VENI Grant 451-11-001. Cornea-Madeira acknowledges the support of the British Academy's PDF/009/370. Hall acknowledges the support of ESRC grant RES-062-23-1351.

Appendix A. Tables

See Tables A.1–A.4

Please cite this article as: O. Boldea, A. Cornea-Madeira and A.R. Hall, Bootstrapping structural change tests. Journal of Econometrics (2019), https://doi.org/10.1016/j.jeconom.2019.05.019.

Appendix B. Proof of Theorems 1 and 2

For the purposes of our analysis, it is convenient to write the system in (4) as a VAR(1) model.[16] To this end, define:

$$\underbrace{\boldsymbol{\xi}_t}_{np\times 1} \equiv \begin{bmatrix} \tilde{\boldsymbol{z}}_t \\ \tilde{\boldsymbol{z}}_{t-1} \\ \vdots \\ \tilde{\boldsymbol{z}}_{t-p+1} \end{bmatrix}, \quad \underbrace{\boldsymbol{F}_s}_{np\times np} \equiv \begin{bmatrix} \boldsymbol{C}_{1,s} & \boldsymbol{C}_{2,s} & \boldsymbol{C}_{3,s} & \dots & \boldsymbol{C}_{p-1,s} & \boldsymbol{C}_{p,s} \\ \boldsymbol{I}_n & \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n} & \dots & \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n} \\ \boldsymbol{0}_{n\times n} & \boldsymbol{I}_n & \boldsymbol{0}_{n\times n} & \dots & \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n} \\ \vdots & \vdots & \vdots & \dots & \vdots & \vdots \\ \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n} & \boldsymbol{0}_{n\times n} & \dots & \boldsymbol{I}_n & \boldsymbol{0}_{n\times n} \end{bmatrix},$$

$$\underbrace{\boldsymbol{\eta}_t}_{np\times 1} \equiv \begin{bmatrix} \boldsymbol{e}_t \\ \boldsymbol{0}_n \\ \vdots \\ \boldsymbol{0}_n \end{bmatrix}, \quad \text{and} \quad \underbrace{\boldsymbol{\mu}_s}_{np\times 1} \equiv \begin{bmatrix} \boldsymbol{c}_{\tilde{z},s} \\ \boldsymbol{0}_n \\ \vdots \\ \boldsymbol{0}_n \end{bmatrix}.$$

Then Eq. (4) is the first n entries of:

$$\boldsymbol{\xi}_t = \boldsymbol{\mu}_s + \boldsymbol{F}_s \boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t, \tag{B.37}$$

where we have suppressed the dependence of $\boldsymbol{\xi}_t$ and $\boldsymbol{\eta}_t$ on s for notational convenience, $s = 1, \dots, N+1$ (there are $N+1$ stable regimes).

From Assumption 9 it follows that $\boldsymbol{\eta}_t$ is a vector m.d.s. relative to $\mathcal{F}_{t-1}$ with conditional covariance matrix

$$\mathrm{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}_j' \mid \mathcal{F}_{t-1}) = \begin{cases} \boldsymbol{\Omega}_{t|t-1}, & \text{for } t = j, \\ \boldsymbol{0}_{np\times np} & \text{otherwise}, \end{cases} \tag{B.38}$$

$$\underbrace{\boldsymbol{\Omega}_{t|t-1}}_{np\times np} \equiv \begin{bmatrix} \boldsymbol{A}_s^{-1} \overline{\boldsymbol{\Sigma}}_{t|t-1} \boldsymbol{A}_s^{-1'} & \boldsymbol{0}_{n\times n(p-1)} \\ \boldsymbol{0}'_{n\times n(p-1)} & \boldsymbol{0}_{n(p-1)\times n(p-1)} \end{bmatrix},$$

where $\overline{\boldsymbol{\Sigma}}_{t|t-1} = \boldsymbol{S}\boldsymbol{D}_t \boldsymbol{\Sigma}_{t|t-1} \boldsymbol{D}_t' \boldsymbol{S}'$, and time-varying unconditional covariance matrix

$$\underbrace{\boldsymbol{\Omega}_t}_{np\times np} \equiv \mathrm{E}(\boldsymbol{\eta}_t \boldsymbol{\eta}_t') = \begin{bmatrix} \boldsymbol{A}_s^{-1} \overline{\boldsymbol{\Sigma}}_t \boldsymbol{A}_s^{-1'} & \boldsymbol{0}_{n\times n(p-1)} \\ \boldsymbol{0}'_{n\times n(p-1)} & \boldsymbol{0}_{n(p-1)\times n(p-1)} \end{bmatrix}$$

where $\overline{\boldsymbol{\Sigma}}_t = \boldsymbol{S}\boldsymbol{D}_t\, \mathrm{E}(\boldsymbol{l}_t \boldsymbol{l}_t') \boldsymbol{D}_t' \boldsymbol{S}'$.

From (B.37), it follows that within each stable regime we have, for $t = [\tau_{s-1}T]+1, [\tau_{s-1}T]+2, \dots, [\tau_s T]$,

$$\boldsymbol{\xi}_t = \boldsymbol{F}_s^{t-[\tau_{s-1}T]} \boldsymbol{\xi}_{[\tau_{s-1}T]} + \tilde{\boldsymbol{\xi}}_t + \left(\sum_{l=0}^{t-[\tau_{s-1}T]-1} \boldsymbol{F}_s^l \right) \boldsymbol{\mu}_s, \tag{B.39}$$

where $\tilde{\boldsymbol{\xi}}_t = \sum_{l=0}^{t-[\tau_{s-1}T]-1} \boldsymbol{F}_s^l \boldsymbol{\eta}_{t-l}$, $\{\boldsymbol{\eta}_t\}$ is a m.d.s. sequence, and, from Assumption 7, all the eigenvalues of $\boldsymbol{F}_s$ have modulus less than one.

The following lemmas (Lemmas 1–11) are used in proofs; Lemmas 2, 4–8 are proven in Appendix C from the Supplementary Appendix, which also contains the asymptotic distributions of the sup-*Wald* test statistics. The rest of the lemmas are proven below.

Lemma 1. *If $\{\vartheta_t, \mathcal{F}_t\}$ is a mean-zero sequence of L^1-mixingale random variables with constants $\{c_t\}$ that satisfy $\overline{\lim}_{T\to\infty} T^{-1} \sum_{t=1}^T c_t < \infty$, and $\sup_t \mathrm{E}|\vartheta_t|^b < \infty$ for some $b > 1$, then $\sup_{s\in(0,1]} |T^{-1} \sum_{t=1}^{[Ts]} \vartheta_t| \xrightarrow{p} 0$.*

This follows from applying the LLN in Andrews (1988)[Theorem 1], modified to be a uniform LLN in the proof of Lemma A2 of Andrews (1993).

Lemma 2. *For $s = 1, \dots, N+1$, where N is the total number of breaks in the coefficients of the VAR(p) representation of $\tilde{\boldsymbol{z}}_t$, define the following functions: $\boldsymbol{F}(\tau) = \boldsymbol{F}_s$, $\boldsymbol{A}(\tau) = \boldsymbol{A}_s$, $\boldsymbol{\mu}(\tau) = \boldsymbol{\mu}_s$, $\boldsymbol{\Upsilon}(\tau) = \boldsymbol{\Upsilon}_s$ for $\tau_{s-1} < \tau \le \tau_s$. Also, define the function $\overline{\boldsymbol{\Sigma}}(\tau)$ on $\tau \in [0,1]$ as follows $\overline{\boldsymbol{\Sigma}}(0) = 0$, and $\overline{\boldsymbol{\Sigma}}(\tau) = \boldsymbol{\Sigma}_t$ for $\tau \in [(t-1)/T, t/T]$, $t = 1, \dots, T$. Let $\mathcal{S}$ and $\mathcal{S}_{\boldsymbol{r}}$ be the selection matrices such that $\boldsymbol{z}_t = \mathbf{vect}(1, \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t, \mathcal{S}\boldsymbol{\xi}_{t-1}) = \mathbf{vect}(1, \boldsymbol{r}_t, \mathcal{S}\boldsymbol{\xi}_{t-1})$, and*

$$\mathbb{Q}_{\boldsymbol{z}}(\tau) = \begin{bmatrix} 1 & \{\mathcal{S}_{\boldsymbol{r}}\mathbb{Q}_1(\tau)\}' & \{\mathcal{S}\mathbb{Q}_1(\tau)\}' \\ \mathcal{S}_{\boldsymbol{r}}\mathbb{Q}_1(\tau) & \mathcal{S}_{\boldsymbol{r}}\mathbb{Q}_2(\tau)\mathcal{S}_{\boldsymbol{r}}' & \mathcal{S}_{\boldsymbol{r}}(\boldsymbol{\mu}(\tau)\mathbb{Q}_1'(\tau) + \boldsymbol{F}(\tau)\mathbb{Q}_2(\tau))\mathcal{S}' \\ \mathcal{S}\mathbb{Q}_1(\tau) & (\mathcal{S}_{\boldsymbol{r}}(\boldsymbol{\mu}(\tau)\mathbb{Q}_1'(\tau) + \boldsymbol{F}(\tau)\mathbb{Q}_2(\tau))\mathcal{S}')' & \mathcal{S}\mathbb{Q}_2(\tau)\mathcal{S}' \end{bmatrix},$$

[16] For example, see Hamilton (1994)[p.259].

where $\mathbb{Q}_1(\tau) = \{\boldsymbol{I}_{np} - \boldsymbol{F}(\tau)\}\boldsymbol{\mu}(\tau)$ *and*

$$\mathbb{Q}_2(\tau) = \sum_{l=0}^{\infty} \boldsymbol{F}(\tau)^l \begin{bmatrix} \boldsymbol{A}(\tau)^{-1}\overline{\boldsymbol{\Sigma}}(\tau)\boldsymbol{A}(\tau)^{-1'} & \boldsymbol{0}_{n\times n(p-1)} \\ \boldsymbol{0}'_{n\times n(p-1)} & \boldsymbol{0}_{n(p-1)\times n(p-1)} \end{bmatrix} (\boldsymbol{F}(\tau)^l)' + \mathbb{Q}_1(\tau)\mathbb{Q}_1'(\tau).$$

Also, let $\mathbb{Q}_i = \int_{\lambda_{i-1}}^{\lambda_i} \boldsymbol{\Upsilon}'(\tau)\mathbb{Q}_{\boldsymbol{z}}(\tau)\boldsymbol{\Upsilon}(\tau)\mathrm{d}\tau$.

Under Assumptions 1–8,

$$\hat{\boldsymbol{Q}}_{(i)} = T^{-1}\sum_{t\in I_{i,\lambda_k}} \hat{\boldsymbol{\Upsilon}}_t'\boldsymbol{z}_t\boldsymbol{z}_t'\hat{\boldsymbol{\Upsilon}}_t \xrightarrow{p} \mathbb{Q}_i.$$

Lemma 3. *If* $(\boldsymbol{a}_t, \mathcal{F}_t)$ *is a* $o\times 1$ *vector of m.d.s. with* $\sup_t \mathrm{E}|a_{t,j}|^{2+\delta^*} < \infty$ *for some* $\delta^* > 0$ *and for all elements* $a_{t,j}$ *of the vector* $\boldsymbol{a}_t$, *if* $T^{-1}\sum_{t=1}^{[Tr]}[\mathrm{E}(\boldsymbol{a}_t\boldsymbol{a}_t'|\mathcal{F}_{t-1}) - \mathrm{E}(\boldsymbol{a}_t\boldsymbol{a}_t')] \xrightarrow{p} 0$ *uniformly in* r *and if* $T^{-1}\sum_{t=1}^{[Tr]}\mathrm{E}(\boldsymbol{a}_t\boldsymbol{a}_t') \to r\boldsymbol{I}_o$ *uniformly in* r, *then* $T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{a}_t \Rightarrow \boldsymbol{B}(r)$, *a* $o\times 1$ *vector of independent standard Brownian motions.*

Lemma 3 provides sufficient conditions so that Theorem 3 in Brown (1971) is satisfied.

Lemma 4. *Under Assumption 9,*[17]

(i) $T^{-1}\sum_{t=1}^{[Tr]}\mathrm{E}(\boldsymbol{l}_t\boldsymbol{l}_t'|\mathcal{F}_{t-1}) \xrightarrow{p} r\boldsymbol{I}_n$ *uniformly in* r.

(ii) $T^{-1}\sum_{t=1}^{[Tr]}\mathrm{E}((\boldsymbol{l}_t\boldsymbol{l}_t')\otimes\boldsymbol{l}_{t-i}|\mathcal{F}_{t-1}) \xrightarrow{p} r\boldsymbol{\rho}_i$ *uniformly in* r, *for all* $i\geq 0$.

(iii) $T^{-1}\sum_{t=1}^{[Tr]}\mathrm{E}((\boldsymbol{l}_t\boldsymbol{l}_t')\otimes(\boldsymbol{l}_{t-i}\boldsymbol{l}_{t-j})|\mathcal{F}_{t-1}) \xrightarrow{p} r\boldsymbol{\rho}_{i,j}$ *uniformly in* r, *for all* $i,j\geq 0$.

For the following lemmas and the rest of the proofs, we need additional notation. Define $\tilde{\mathcal{S}}_1 = [\boldsymbol{I}_{p_1+1} \quad \boldsymbol{0}_{(p_1+1)\times p_2}]$ and $\tilde{\mathcal{S}}_2 = [\boldsymbol{0}_{p_2\times(p_1+1)} \quad \boldsymbol{I}_{p_2}]$. Also, define the following vectors of Brownian motions: $\boldsymbol{B}_0(r)$, a $n\times 1$ vector with variance $r\boldsymbol{I}_n$, $\boldsymbol{B}_l(r)$, a $n^2\times 1$ vector with variance $r\boldsymbol{\rho}_{l,l}$ for each $l\geq 1$, $\boldsymbol{B}_\zeta(r) = \mathbf{vect}(\boldsymbol{B}_{u\zeta}(r), \boldsymbol{B}_{v\zeta}(r))$ with $\boldsymbol{B}_{u\zeta}(r)$ of dimension $p_2\times 1$ and $\boldsymbol{B}_{v\zeta}(r)$ of dimension $p_1p_2\times 1$, where the variance of $\boldsymbol{B}_\zeta(r)$ is $r(\tilde{\mathcal{S}}_1\otimes\tilde{\mathcal{S}}_2)\boldsymbol{\rho}_{0,0}(\tilde{\mathcal{S}}_1\otimes\tilde{\mathcal{S}}_2)' = r\boldsymbol{\rho}_{\xi,0,0} = r\begin{bmatrix} \boldsymbol{\rho}_{u,\xi,0,0} & \boldsymbol{\rho}_{u,v,\xi,0,0} \\ \boldsymbol{\rho}'_{u,v,\xi,0,0} & \boldsymbol{\rho}_{v,\xi,0,0} \end{bmatrix}$, where $\boldsymbol{\rho}_{u,\xi,0,0}$ is of dimension $p_2\times p_2$. The covariances of these processes are: $\mathrm{Cov}(\boldsymbol{B}_l(r_1), \boldsymbol{B}_\kappa(r_2)) = \min(r_1, r_2)\boldsymbol{\rho}_{l,\kappa}$ for all $l,\kappa\geq 1$, $l\neq\kappa$, and $\mathrm{Cov}(\boldsymbol{B}_\zeta(r_1), \boldsymbol{B}_l(r_2)) = \min(r_1, r_2)(\tilde{\mathcal{S}}_1\otimes\tilde{\mathcal{S}}_2)\boldsymbol{\rho}_{0,l}$ for all $l\geq 1$ and $\mathrm{Cov}(\boldsymbol{B}_\zeta(r_1), \boldsymbol{B}_0(r_2)) = \min(r_1, r_2)(\tilde{\mathcal{S}}_1\otimes\tilde{\mathcal{S}}_2)\boldsymbol{\rho}_0 = \min(r_1, r_2)\boldsymbol{\rho}_{\xi,0} = \min(r_1, r_2)\mathbf{vect}(\boldsymbol{\rho}_{u,0}, \boldsymbol{\rho}_{v,0})$, where $\boldsymbol{\rho}_{0,l}$ and $\boldsymbol{\rho}_0$ are given in Assumption 9(v) and (iv) respectively, and $\boldsymbol{\rho}_{u,0}$ is of dimension $p_2\times n$. Moreover, P^b denotes the probability measure induced by the bootstrap conditional on the original sample; E^b, Var^b denote expectation and variance with respect to the bootstrap data, conditional on the original sample. As in Gine and Zinn (1990), Hansen (1996), for any bootstrapped quantity $a_T^b(\lambda)$, we write $a_T^b(\lambda) \xrightarrow{p^b} 0$ or $a_T^b(\lambda) = o_p^b(1)$ in probability uniformly in λ when $\lim_{T\to\infty} P[P^b(|a_T^b(\lambda)| > \delta) > \epsilon] = 0$ for any $\delta > 0$, $\epsilon > 0$ that does not depend on λ. We write $a_T^b(\lambda) \overset{d_p^b}{\Rightarrow} a(\lambda)$ in probability uniformly in λ for any distribution $a(\lambda)$, when weak convergence under the bootstrap probability measure P^b occurs in a set with probability converging to one, uniformly in λ.

Lemma 5. *For fixed* n^*, *under Assumption 9,*

$$T^{-1/2}\sum_{t=1}^{[Tr]}\mathbf{vect}(\boldsymbol{l}_t, \boldsymbol{l}_t\otimes\boldsymbol{l}_{t-1}, \ldots, \boldsymbol{l}_t\otimes\boldsymbol{l}_{t-n^*}, l_{u,t}\boldsymbol{l}_{\zeta,t}, \boldsymbol{l}_{v,t}\otimes\boldsymbol{l}_{\zeta,t}) \Rightarrow \mathbf{vect}(\boldsymbol{B}_0(r), \boldsymbol{B}_1(r), \ldots \boldsymbol{B}_{n^*}(r), \boldsymbol{B}_\zeta(r)),$$

where if $t-l<0$ *for any* $l>0$, *the rest of the elements of this sum are artificially set to zero.*

Now define for $b = 1, 2$ and any $n^b\times 1$ vectors $\boldsymbol{a}$, $\boldsymbol{a}_\# = \mathbf{vect}(\boldsymbol{a}, \boldsymbol{0}_{n^b(p^b-1)})$, and for any $n^b\times n^b$ matrices $\boldsymbol{A}$, let $\boldsymbol{A}_\# = \mathbf{diag}(\boldsymbol{A}, \boldsymbol{0}_{n^b(p^b-1)\times n^b(p^b-1)})$, except for $\boldsymbol{\beta}_{\boldsymbol{x},s,\#}$, which is $\boldsymbol{\beta}_{\boldsymbol{x},s,\#} = \mathbf{vect}(0, \boldsymbol{\beta}^0_{\boldsymbol{x},(s)}, \boldsymbol{0}_{p_2+n(p-1)})$ and the subscript s indicates the value of $\boldsymbol{\beta}^0_{\boldsymbol{x},(i)}$ in the stable regime $\tilde{I}_s = [[\tau_{s-1}T]+1, [\tau_sT]]$. If $m = 0$, then $\boldsymbol{\beta}^0_{\boldsymbol{x},(s)} = \boldsymbol{\beta}^0_{\boldsymbol{x}}$, and $\boldsymbol{\beta}_{\boldsymbol{x},\#} = \mathbf{vect}(0, \boldsymbol{\beta}^0_{\boldsymbol{x}}, \boldsymbol{0}_{p_2+n(p-1)})$. Let $\mathcal{S}_u = \mathbf{vect}(1, \boldsymbol{0}_{n-1}, \boldsymbol{0}_{n(p-1)})$ and $\mathcal{S}_\dagger = \mathcal{S}_u$ or $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},s,\#}$, where the value $\mathcal{S}_\dagger$ takes is clarified in each context where the distinction between the two values is necessary. Let $\boldsymbol{S}$, defined in Assumption 9, and $\boldsymbol{D}(\tau)$, the function such that $\boldsymbol{D}(\tau) = \boldsymbol{D}_t$ for $\tau\in[\frac{t}{T}, \frac{t+1}{T})$, be partitioned as follows:

$$\boldsymbol{S} = \begin{bmatrix} 1 & \boldsymbol{0}_{1\times p_1} & \boldsymbol{0}_{1\times p_2} \\ \boldsymbol{s}_{p_1} & \boldsymbol{S}_{p_1} & \boldsymbol{0}_{p_1\times p_2} \\ \boldsymbol{0}_{p_2\times 1} & \boldsymbol{0}_{p_2\times p_1} & \boldsymbol{S}_{p_2} \end{bmatrix}, \quad \boldsymbol{D}(\tau) = \begin{bmatrix} d_u(\tau) & \boldsymbol{0}_{1\times p_1} & \boldsymbol{0}_{1\times p_2} \\ \boldsymbol{0}_{p_1\times 1} & \boldsymbol{D}_v(\tau) & \boldsymbol{0}_{p_1\times p_2} \\ \boldsymbol{0}_{p_2\times 1} & \boldsymbol{0}_{p_2\times p_1} & \boldsymbol{D}_\zeta(\tau) \end{bmatrix}, \tag{B.40}$$

where $\boldsymbol{s}_{p_1}$ is of dimension $p_1\times 1$, $\boldsymbol{S}_1$ and $\boldsymbol{D}_v(\tau)$ are of dimension $p_1\times p_1$, and $\boldsymbol{S}_{p_2}$ and $\boldsymbol{D}_\zeta(\tau)$ are of dimension $p_2\times p_2$. For any interval $[[\tau_{s-1}T]+1, [\tau_sT]]$ where the coefficients of the VAR representation in (4) are stable, let:

$$\mathbb{M}_1(\tau_{s-1}, \tau_s) = \mathcal{S}_\dagger'\boldsymbol{S}_\#\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}_\#(\tau)\mathrm{d}\boldsymbol{B}_{0,\#}(\tau)$$

[17] All elements with negative subscripts in (*ii*)-(*iii*) are set to zero.

$$\mathbb{M}_{2,1}(\tau_{s-1},\tau_s)=\sum_{l=0}^{\infty}((\mathcal{S}_{\dagger}'\boldsymbol{S}_{\#})\otimes(\mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_s^l))\left(\left[\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}_{\#}(\tau)\mathrm{d}\boldsymbol{B}_{0,\#}(\tau)\right]\otimes\boldsymbol{\mu}_s\right)$$

$$\mathbb{M}_{2,2}(\tau_{s-1},\tau_s)=\sum_{l=0}^{\infty}((\mathcal{S}_{\dagger}'\boldsymbol{S}_{\#})\otimes(\mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_s^{l+1}\boldsymbol{A}_{s,\#}^{-1}\boldsymbol{S}_{\#}))\int_{\tau_{s-1}}^{\tau_s}(\boldsymbol{D}_{\#}(\tau)\otimes\boldsymbol{D}_{\#}(\tau))\,\mathrm{d}\boldsymbol{B}_{l+1,\#}(\tau)$$

$$\mathbb{M}_{2,3}^{(1)}(\tau_{s-1},\tau_s)=\boldsymbol{S}_{p_2}\int_{\tau_{s-1}}^{\tau_s}d_u(\tau)\boldsymbol{D}_{\zeta}(\tau)\mathrm{d}\boldsymbol{B}_{u\zeta}(\tau)$$

$$\mathbb{M}_{2,3}^{(2)}(\tau_{s-1},\tau_s)=\boldsymbol{\beta}_{\boldsymbol{x},(s)}^{0'}\boldsymbol{S}_{p_1}\boldsymbol{S}_{p_2}\int_{\tau_{s-1}}^{\tau_s}d_u(\tau)\boldsymbol{D}_{\zeta}(\tau)\mathrm{d}\boldsymbol{B}_{u\zeta}(\tau)+((\boldsymbol{\beta}_{\boldsymbol{x},(s)}^{0'}\boldsymbol{S}_{p_1})\otimes\boldsymbol{S}_{p_2})\int_{\tau_{s-1}}^{\tau_s}(\boldsymbol{D}_{\boldsymbol{v}}(\tau)\otimes\boldsymbol{D}_{\zeta}(\tau))\mathrm{d}\boldsymbol{B}_{\boldsymbol{v}\zeta}(\tau)$$

$$\mathbb{M}_2(\tau_{s-1},\tau_s)=\mathbb{M}_{2,1}(\tau_{s-1},\tau_s)+\mathbb{M}_{2,2}(\tau_{s-1},\tau_s)+\mathbb{M}_{2,3}^{(j)}(\tau_{s-1},\tau_s),\ \text{where } j=1 \text{ if } \mathcal{S}_{\dagger}=\mathcal{S}_u \text{ and } j=2 \text{ otherwise}$$

$$\mathbb{M}_3(\tau_{s-1},\tau_s)=\mathbb{M}_{3,1}(\tau_{s-1},\tau_s)+\mathbb{M}_{3,2}(\tau_{s-1},\tau_s)$$

$$\mathbb{M}_{3,1}(\tau_{s-1},\tau_s)=\sum_{l=0}^{\infty}\left((\mathcal{S}_{\dagger}'\boldsymbol{S}_{\#})\otimes(\mathcal{S}\boldsymbol{F}_s^l)\right)\left(\left[\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}_{\#}(\tau)\mathrm{d}\boldsymbol{B}_{0,\#}(\tau)\right]\otimes\boldsymbol{\mu}_s\right)$$

$$\mathbb{M}_{3,2}(\tau_{s-1},\tau_s)=\sum_{l=0}^{\infty}((\mathcal{S}_{\dagger}'\boldsymbol{S}_{\#})\otimes(\mathcal{S}\boldsymbol{F}_s^l\boldsymbol{A}_{s,\#}^{-1}\boldsymbol{S}_{\#}))\int_{\tau_{s-1}}^{\tau_s}(\boldsymbol{D}_{\#}(\tau)\otimes\boldsymbol{D}_{\#}(\tau))\,\mathrm{d}\boldsymbol{B}_{l+1,\#}(\tau)$$

$$\mathbb{M}(\tau_{s-1},\tau_s)=\mathbf{vect}(\mathbb{M}_1(\tau_{s-1},\tau_s),\mathbb{M}_2(\tau_{s-1},\tau_s),\mathbb{M}_3(\tau_{s-1},\tau_s)),$$

where $\mathcal{S}_{\boldsymbol{r}}$ was defined in Lemma 2.

Lemma 6. *Let the interval I_i contain N_i breaks from the total set of N breaks. Then, under Assumptions 1–9,*

$$T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t u_t\Rightarrow\widetilde{\mathbb{M}}_i=\begin{cases}\mathbb{M}(\lambda_{i-1},\tau_s)+\sum_{j=1}^{N_i}\mathbb{M}(\tau_{s+j-1},\tau_{s+j})+\mathbb{M}(\tau_{s+N_i},\lambda_i) & \text{if } N_i\geq 2\\ \mathbb{M}(\lambda_{i-1},\tau_s)+\mathbb{M}(\tau_s,\lambda_i) & \text{if } N_i=1\\ \mathbb{M}(\lambda_{i-1},\lambda_i) & \text{if } N_i=0.,\end{cases}$$

with $\mathcal{S}_{\dagger}=\mathcal{S}_u$. Similarly, $T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t\boldsymbol{v}_t'\boldsymbol{\beta}_{\boldsymbol{x},(i)}^0\Rightarrow\widetilde{\mathbb{M}}_i$ but with $\mathcal{S}_{\dagger}=\boldsymbol{\beta}_{\boldsymbol{x},i,\#}=\mathbf{vect}(0,\boldsymbol{\beta}_{\boldsymbol{x},(i)}^0,\mathbf{0}_{p_2+n(p-1)})$. If $m=0$, then $\mathcal{S}_{\dagger}=\boldsymbol{\beta}_{\boldsymbol{x},\#}$.

Lemma 7. *Under Assumptions 1–9,*
(i) if $h>0$, then $T(\hat{\pi}_i-\pi_i^0)=O_p(1)$, $i=1,\ldots,h+1$;
(ii) $T^{1/2}(\hat{\boldsymbol{\Delta}}_{(i)}-\boldsymbol{\Delta}_{(i)}^0)=O_p(1)$ for $i=1,\ldots,h+1$;
(iii) if $m>0$, $T(\hat{\lambda}_i-\lambda_i^0)=O_p(1)$, $i=1,\ldots,m+1$.

Lemma 8. *Under Assumption 9, uniformly in r,*
(i) $T^{-1}\sum_{t=1}^{[Tr]}\{\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t'-\mathrm{E}(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\}\xrightarrow{p}0$,
(ii) $T^{-1}\sum_{t=1}^{[Tr]}\{(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\otimes\boldsymbol{\epsilon}_{t-i}-\mathrm{E}[(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\otimes\boldsymbol{\epsilon}_{t-i}]\}\xrightarrow{p}0$ for all $i\geq 0$,
(iii) $T^{-1}\sum_{t=1}^{[Tr]}\{(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\otimes(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-j}')-\mathrm{E}[(\boldsymbol{\epsilon}_t\boldsymbol{\epsilon}_t')\otimes(\boldsymbol{\epsilon}_{t-i}\boldsymbol{\epsilon}_{t-j}')]\}\xrightarrow{p}0$ for all $i,j\geq 0$
(iv) Parts (i)–(iii) hold with $\boldsymbol{l}_t,\boldsymbol{l}_{t-i},\boldsymbol{l}_{t-j}$ replacing $\boldsymbol{\epsilon}_t,\boldsymbol{\epsilon}_{t-i},\boldsymbol{\epsilon}_{t-j}$, uniformly in r.[18]

Lemma 9. *Let $\hat{\boldsymbol{Q}}_{(i)}^b=T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}_t'\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}\hat{\boldsymbol{\Upsilon}}_t$. Then, under Assumptions 1–9, $\hat{\boldsymbol{Q}}_{(i)}^b=\mathbb{Q}_i+o_p^b(1)$, where*

$$\mathbb{Q}_i=\int_{\lambda_{i-1}}^{\lambda_i}\boldsymbol{\Upsilon}(\tau)'\mathbb{Q}_{\boldsymbol{z}}(\tau)\boldsymbol{\Upsilon}(\tau)\mathrm{d}\tau. \quad \text{(B.41)}$$

Proof of Lemma 9. For the WF bootstraps, $\boldsymbol{z}_t^b=\boldsymbol{z}_t$, and therefore Lemma 9 holds by Lemma 2. Consider the WR bootstrap, first for $I_i=\tilde{I}_s$. Define $\tilde{\boldsymbol{z}}_t^b=(y_t^b,\boldsymbol{x}_t^{b'},\boldsymbol{r}_t)'$, and:

$$\hat{\boldsymbol{Q}}_{(i)}^b=T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}_t'\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}\hat{\boldsymbol{\Upsilon}}_t=\hat{\boldsymbol{\Upsilon}}_s'\begin{bmatrix}\Delta\tau_s & \mathcal{A}_1^{b'}\mathcal{S}_{\boldsymbol{r}}' & \mathcal{A}_2^{b'}\mathcal{S}'\\ \mathcal{S}_{\boldsymbol{r}}\mathcal{A}_1^b & \mathcal{S}_{\boldsymbol{r}}\mathcal{B}_1^b\mathcal{S}_{\boldsymbol{r}}' & \mathcal{S}_{\boldsymbol{r}}\mathcal{B}_2^b\mathcal{S}'\\ \mathcal{S}\mathcal{A}_2^b & \mathcal{S}\mathcal{B}_2^{b'}\mathcal{S}_{\boldsymbol{r}}' & \mathcal{S}\mathcal{B}_3^b\mathcal{S}'\end{bmatrix}\hat{\boldsymbol{\Upsilon}}_s,$$

where

$$\mathcal{A}_1^b\ =\ T^{-1}\sum_{t\in I_i}\boldsymbol{\xi}_t^b,\quad \mathcal{A}_2^b\ =\ T^{-1}\sum_{t\in I_i}\boldsymbol{\xi}_{t-1}^b$$

[18] All elements with negative subscripts in (ii)-(iv) are set to zero.

Please cite this article as: O. Boldea, A. Cornea-Madeira and A.R. Hall, Bootstrapping structural change tests. Journal of Econometrics (2019), https://doi.org/10.1016/j.jeconom.2019.05.019.

and

$$\mathcal{B}_1^b = T^{-1}\sum_{t\in I_i}\xi_t^b\xi_t^{b\prime},\quad \mathcal{B}_2^b = T^{-1}\sum_{t\in I_i}\xi_t^b\xi_{t-1}^{b\prime},\quad \mathcal{B}_3 = T^{-1}\sum_{t\in I_i}\xi_{t-1}^b\xi_{t-1}^{b\prime}.$$

Note that, because $\boldsymbol{r}_t$ is kept fixed, $\mathcal{S}_{\boldsymbol{r}}\mathcal{A}_1^b = T^{-1}\sum_{t\in I_i}\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t = \mathcal{S}_{\boldsymbol{r}}\mathcal{A}_1$, and $\mathcal{S}_{\boldsymbol{r}}\mathcal{B}_1^b\mathcal{S}_{\boldsymbol{r}} = \mathcal{S}_{\boldsymbol{r}}\mathcal{B}_1\mathcal{S}_{\boldsymbol{r}}'$, where $\mathcal{A}_1, \mathcal{A}_2$ are the sample counterparts of $\mathcal{A}_1^b, \mathcal{B}_1^b$ defined at the beginning of the proof of Lemma 2 in Appendix C from the Supplementary Appendix. By Lemma 2, the result in Lemma 9 holds automatically for these terms. We now analyze the rest of the terms. To that end, we first derive some preliminary results.

- **Preliminary results and bootstrap notation.** Note that in any stable subinterval $\tilde{I}_s$,

$$\tilde{\boldsymbol{z}}_t^b = \hat{\boldsymbol{c}}_{\tilde{z},s} + \sum_{i=1}^{p}\hat{\boldsymbol{C}}_{i,s}\tilde{\boldsymbol{z}}_{t-i}^b + \boldsymbol{e}_t^b, \qquad [\tau_{s-1}T]+1 \le t \le [\tau_s T],\ s = 1,2,\ldots,N+1, \tag{B.42}$$

where $\boldsymbol{e}_t^b = \hat{\boldsymbol{A}}_s^{-1}\boldsymbol{\epsilon}_t^b$, $\boldsymbol{\epsilon}_t^b = \mathbf{vect}(u_t^b, \boldsymbol{v}_t^b, \boldsymbol{\zeta}_t)$, of size $n\times 1$, and the elements of $\hat{\boldsymbol{A}}_s$, $\hat{\boldsymbol{c}}_{\tilde{z},s}$ and $\hat{\boldsymbol{C}}_{i,s}$ corresponding to the equation for $\boldsymbol{r}_t$ are the true parameters, not the estimated ones. Then,

$$\boldsymbol{\xi}_t^b = \hat{\boldsymbol{\mu}}_s^b + \hat{\boldsymbol{F}}_s\boldsymbol{\xi}_{t-1}^b + \boldsymbol{\eta}_t^b \tag{B.43}$$

$$= \hat{\boldsymbol{F}}_s^{t-[\tau_{s-1}T]}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b + \left(\sum_{l=0}^{t-[\tau_{s-1}T]-1}\hat{\boldsymbol{F}}_s^l\right)\hat{\boldsymbol{\mu}}_s + \sum_{l=0}^{t-[\tau_{s-1}T]-1}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l}^b, \tag{B.44}$$

where $\tilde{\boldsymbol{\xi}}_t^b = \mathbf{vect}(\tilde{\boldsymbol{z}}_t^b, \tilde{\boldsymbol{z}}_{t-1}^b, \ldots, \tilde{\boldsymbol{z}}_{t-p+1}^b)$, $\boldsymbol{\eta}_t^b = \hat{\boldsymbol{A}}_{s,\#}^{-1}\boldsymbol{\epsilon}_{t,\#}^b$, and $\hat{\boldsymbol{F}}_s$, $\hat{\boldsymbol{\mu}}_s$ are defined as $\boldsymbol{F}_s$, $\boldsymbol{\mu}_s$, but replacing the true coefficients that are estimated by 2SLS with those estimated counterparts. Also, let $\hat{\boldsymbol{\eta}}_t \equiv \hat{\boldsymbol{e}}_{t,\#} = \hat{\boldsymbol{A}}_{s,\#}^{-1}\hat{\boldsymbol{\epsilon}}_{t,\#}$, where $\hat{\boldsymbol{\epsilon}}_t = \mathbf{vect}(\hat{u}_t, \hat{\boldsymbol{v}}_t, \boldsymbol{\zeta}_t)$.

We now show two results that we repeatedly need in the proofs: $T^{-\alpha}\boldsymbol{\xi}_t^b = o_p^b(1)$ and $T^{-\alpha}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b\prime} = o_p^b(1)$ for any $\alpha > 0$.

For this purpose, we first show that $\mathrm{E}^b(T^{-\alpha}\boldsymbol{\eta}_t^b) = o_p^b(1)$ and that $\mathrm{Var}^b(T^{-\alpha}\boldsymbol{\eta}_t^b) = o_p^b(1)$. Then, by Markov's inequality, for any $C > 0$, $P^b(T^{-\alpha}\|\boldsymbol{\eta}_t^b - \mathrm{E}^b(\boldsymbol{\eta}_t^b)\| \ge C) \le C^{-2}T^{-2\alpha}\mathrm{Var}^b\|\boldsymbol{\eta}_t^b\| \overset{p}{\to} 0$, completing the proof.

Let $\mathcal{I} = \mathbf{vect}(\mathbf{0}_{p_1+1}, \boldsymbol{\iota}_{p_2}, \mathbf{0}_{p_2+n(p+1)})$ and $\mathcal{J} = [\mathbf{diag}(\boldsymbol{J}_{p_1+1}, \boldsymbol{J}_{p_2})]_\#$, where $\boldsymbol{\iota}_a$ is a $a\times 1$ vector of ones, and $\boldsymbol{J}_a = \boldsymbol{\iota}_a\boldsymbol{\iota}_a'$. Let $\tilde{\boldsymbol{\nu}}_t \equiv \mathbf{vect}(\nu_t\boldsymbol{\iota}_{p_1+1}, \boldsymbol{\iota}_{p_2})$ and $\boldsymbol{\nu}_t \equiv \tilde{\boldsymbol{\nu}}_{t,\#}$. Then $\mathrm{E}^b(\boldsymbol{\nu}_t) = \mathcal{I}$ and $\mathrm{E}^b(\boldsymbol{\nu}_t\boldsymbol{\nu}_t') = \mathcal{J}$.

Also, let $\boldsymbol{g}_t^b \equiv \boldsymbol{\epsilon}_{t,\#}^b = \hat{\boldsymbol{\epsilon}}_{t,\#} \odot \boldsymbol{\nu}_t$, where $\odot$ is the element-wise multiplication. Then $\boldsymbol{g}_t^b = \hat{\boldsymbol{A}}_{s,\#}\,\boldsymbol{\eta}_t^b$, and letting $\hat{\boldsymbol{g}}_t \equiv \hat{\boldsymbol{\epsilon}}_{t,\#}$, it follows that $\boldsymbol{g}_t^b = \hat{\boldsymbol{g}}_t \odot \boldsymbol{\nu}_t$. Further, let $\hat{\boldsymbol{g}}_{t,1} \equiv \mathbf{vect}(\hat{u}_t, \hat{\boldsymbol{v}}_t, \mathbf{0}_{p_2+n(p-1)})$ and $\boldsymbol{g}_{t,2} \equiv \mathbf{vect}(\mathbf{0}_{(p_1+1)}, \boldsymbol{\zeta}_t, \mathbf{0}_{n(p-1)})$. Also, note that $\boldsymbol{\eta}_t^b = \hat{\boldsymbol{A}}_{s,\#}^{-1}(\hat{\boldsymbol{g}}_t \odot \boldsymbol{\nu}_t)$. Then:

$$\mathrm{E}^b(\boldsymbol{\eta}_t^b) = \mathrm{E}^b(\hat{\boldsymbol{A}}_{s,\#}^{-1}(\hat{\boldsymbol{g}}_t \odot \boldsymbol{\nu}_t)) = \hat{\boldsymbol{A}}_{s,\#}^{-1}(\hat{\boldsymbol{g}}_t \odot \mathcal{I}) = \hat{\boldsymbol{A}}_{s,\#}^{-1}\mathbf{vect}(\mathbf{0}_{(p_1+1)}, \boldsymbol{\zeta}_t, \mathbf{0}_{n(p-1)}) = \hat{\boldsymbol{A}}_{s,\#}^{-1}\boldsymbol{g}_{t,2} \tag{B.45}$$

$$\mathrm{E}^b(\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b\prime}) = \mathrm{E}^b(\hat{\boldsymbol{A}}_{s,\#}^{-1}(\hat{\boldsymbol{g}}_t \odot \boldsymbol{\nu}_t)(\hat{\boldsymbol{g}}_t \odot \boldsymbol{\nu}_t)'\hat{\boldsymbol{A}}_{s,\#}^{\prime -1}) = \hat{\boldsymbol{A}}_{s,\#}^{-1}[(\hat{\boldsymbol{g}}_t\hat{\boldsymbol{g}}_t') \odot \mathcal{J}]\hat{\boldsymbol{A}}_{s,\#}^{-1}$$

$$= \hat{\boldsymbol{A}}_{s,\#}^{-1}\begin{bmatrix}\hat{u}_t^2 & \hat{u}_t\hat{\boldsymbol{v}}_t' & \mathbf{0}_{1\times p_2}\\ \hat{\boldsymbol{v}}_t\hat{u}_t & \hat{\boldsymbol{v}}_t\hat{\boldsymbol{v}}_t' & \mathbf{0}_{p_1\times p_2}\\ \mathbf{0}_{p_2} & \mathbf{0}_{p_2\times p_1} & \boldsymbol{\zeta}_t\boldsymbol{\zeta}_t'\end{bmatrix}_\#\hat{\boldsymbol{A}}_{s,\#}^{-1} \tag{B.46}$$

$$\mathrm{Var}^b(\boldsymbol{\eta}_t^b) = \mathrm{E}^b(\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b\prime}) - \mathrm{E}^b(\boldsymbol{\eta}_t^b)\mathrm{E}^b(\boldsymbol{\eta}_t^{b\prime}) = \hat{\boldsymbol{A}}_{s,\#}^{-1}\begin{bmatrix}\hat{u}_t^2 & \hat{u}_t\hat{\boldsymbol{v}}_t' & \mathbf{0}_{1\times p_2}\\ \hat{\boldsymbol{v}}_t\hat{u}_t & \hat{\boldsymbol{v}}_t\hat{\boldsymbol{v}}_t' & \mathbf{0}_{p_1\times p_2}\\ \mathbf{0}_{p_2} & \mathbf{0}_{p_2\times p_1} & \mathbf{0}_{p_2\times p_2}\end{bmatrix}_\#\hat{\boldsymbol{A}}_{s,\#}^{-1} = \hat{\boldsymbol{A}}_{s,\#}^{-1}\hat{\boldsymbol{g}}_{t,1}\hat{\boldsymbol{g}}_{t,1}'\hat{\boldsymbol{A}}_{s,\#}^{-1}. \tag{B.47}$$

By Lemmas 7, 8 and standard 2SLS theory, $\hat{\boldsymbol{g}}_{t,1} = \mathbf{vect}(\hat{u}_t, \hat{\boldsymbol{v}}_t, \mathbf{0}_{p_2+n(p-1)}) = O_p(1)$, and $\hat{\boldsymbol{A}}_{s,\#} = \boldsymbol{A}_{s,\#} + o_p(1)$, therefore $\hat{\boldsymbol{A}}_{s,\#}^{-1}\hat{\boldsymbol{g}}_{t,1} = O_p(1)$, so $\mathrm{E}^b[T^{-\alpha}\boldsymbol{\eta}_t^b] = o_p^b(1)$.

Next, we show that $T^{-\alpha}\boldsymbol{\xi}_t^b = o_p^b(1)$ by induction, for $\alpha > 0$. First, recall that $\boldsymbol{\xi}_0^b = \boldsymbol{\xi}_0$, and therefore, $T^{-\alpha}\boldsymbol{\xi}_1^b = T^{-\alpha}\hat{\boldsymbol{\mu}}_1 + \hat{\boldsymbol{F}}_1T^{-\alpha}\boldsymbol{\xi}_0 + T^{-\alpha}\boldsymbol{\eta}_1^b = o_p^b(1)$ because $\hat{\boldsymbol{\mu}}_s - \boldsymbol{\mu}_s = o_p(1)$, $\hat{\boldsymbol{F}}_s - \hat{\boldsymbol{F}}_s = o_p(1)$, and $T^{-\alpha}\boldsymbol{\eta}_t^b = o_p^b(1)$ and $\boldsymbol{\xi}_0 = O_p(1)$ by Assumption 9(iii). Now let $T^{-\alpha}\boldsymbol{\xi}_{t-1}^b = o_p^b(1)$; then for $t, t-1 \in \tilde{I}_s$, $T^{-\alpha}\boldsymbol{\xi}_t^b = T^{-\alpha}\hat{\boldsymbol{\mu}}_s + \hat{\boldsymbol{F}}_sT^{-\alpha}\boldsymbol{\xi}_{t-1}^b + T^{-\alpha}\boldsymbol{\eta}_t^b = o_p(1) + \hat{\boldsymbol{F}}_so_p^b(1) + o_p^b(1) = o_p^b(1)$.

Therefore, it follows that:

$$T^{-\alpha}\boldsymbol{\xi}_t^b = o_p^b(1). \tag{B.48}$$

Next, we show that $T^{-\alpha}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b\prime} = o_p^b(1)$, also by mathematical induction. Note that, from the results above,

$$T^{-\alpha}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b\prime} = T^{-\alpha}(\hat{\boldsymbol{\mu}}_s + \hat{\boldsymbol{F}}_s\boldsymbol{\xi}_{t-1}^b + \boldsymbol{\eta}_t^b)(\hat{\boldsymbol{\mu}}_s + \hat{\boldsymbol{F}}_s\boldsymbol{\xi}_{t-1}^b + \boldsymbol{\eta}_t^b)'$$

$$= T^{-\alpha}\hat{\boldsymbol{\mu}}_s\hat{\boldsymbol{\mu}}_s' + \hat{\boldsymbol{F}}_s(T^{-\alpha}\boldsymbol{\xi}_{t-1}^b\boldsymbol{\xi}_{t-1}^{b\prime})\hat{\boldsymbol{F}}_s' + T^{-\alpha}\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b\prime} + T^{-\alpha}\hat{\boldsymbol{\mu}}_s\boldsymbol{\xi}_{t-1}^{b\prime}\hat{\boldsymbol{F}}_s' + (T^{-\alpha}\hat{\boldsymbol{\mu}}_s\boldsymbol{\xi}_{t-1}^{b\prime}\hat{\boldsymbol{F}}_s')'$$

$$+T^{-\alpha}\hat{\boldsymbol{\mu}}_s\boldsymbol{\eta}_t^{b'}+(T^{-\alpha}\hat{\boldsymbol{\mu}}_s\boldsymbol{\eta}_t^{b'})'+\hat{\boldsymbol{F}}^s T^{-\alpha/2}\boldsymbol{\xi}_{t-1}^b T^{-\alpha/2}\boldsymbol{\eta}_t^{b'}+(\hat{\boldsymbol{F}}^s T^{-\alpha/2}\boldsymbol{\xi}_{t-1}^b T^{-\alpha/2}\boldsymbol{\eta}_t^{b'})'$$
$$=\hat{\boldsymbol{F}}_s(T^{-\alpha}\boldsymbol{\xi}_{t-1}^b\boldsymbol{\xi}_{t-1}^{b'})\hat{\boldsymbol{F}}_s'+T^{-\alpha}\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b'}+o_p^b(1). \tag{B.49}$$

Now consider $T^{-\alpha}\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b\prime}$. We have

$$\mathrm{E}^b(T^{-\alpha}\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b'})=\hat{\boldsymbol{A}}_{s,\#}^{-1}[(T^{-\alpha}\hat{\boldsymbol{g}}_t\hat{\boldsymbol{g}}_t')\odot\mathcal{J}]\hat{\boldsymbol{A}}_{s,\#}^{-1}=O_p(1)((T^{-\alpha}\boldsymbol{g}_t\boldsymbol{g}_t'+o_p(1))\odot\mathcal{J})O_p(1)=o_p(1),$$

where $T^{-\alpha}\boldsymbol{g}_t\boldsymbol{g}_t'=o_p(1)$ since by arguments before (C.7) in the proof of Lemma 2 in Appendix C from the Supplementary Appendix, $\sup_t \mathrm{E}\,\|\boldsymbol{g}_t\boldsymbol{g}_t'\|\le\sup_t(\mathrm{E}\,\|\boldsymbol{g}_t\|^2)^{1/2}\sup_t(\mathrm{E}\,\|\boldsymbol{g}_t\|^2)^{1/2}<c^*$ for some $c^*>0$ by Assumption 9, and by Markov's inequality, for any C, $P(T^{-\alpha}\|\boldsymbol{g}_t\boldsymbol{g}_t'\|\ge C)\le T^{-\alpha}C^{-1}\,\mathrm{E}\,\|\boldsymbol{g}_t\boldsymbol{g}_t'\|\le T^{-\alpha}C^{-1}c^*\to 0$. Hence, from Assumption 10 and by Markov's inequality, for any C, $P^b(T^{-\alpha}\|\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b'}\|\ge C)\le T^{-\alpha}C^{-1}\,\mathrm{E}^b\,\|\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b'}\|\overset{p}{\to}0$. It follows that $T^{-\alpha}\boldsymbol{\eta}_t^b\boldsymbol{\eta}_t^{b'}=o_p^b(1)$. Using this result in (B.49), by a similar mathematical induction argument as for $T^{-\alpha}\boldsymbol{\xi}_t^b=o_p^b(1)$, it follows that

$$T^{-\alpha}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b'}=o_p^b(1). \tag{B.50}$$

Besides (B.48) and (B.50), in the proof below we will assume $\left|\boldsymbol{I}_n-\hat{\boldsymbol{C}}_{1,s}a-\hat{\boldsymbol{C}}_{2,s}a^2-\ldots-\hat{\boldsymbol{C}}_{p,s}a^p\right|\ne 0$, for all $s=1,\ldots,N+1$, and all $|a|\le 1$; otherwise the estimated system is not stationary. Then we show in Appendix D of the Supplementary Appendix, that $\sum_{l=0}^{\infty}\|\boldsymbol{F}_s^l\|<\infty$, and similarly, it can be shown that $\sum_{l=0}^{\infty}\|\hat{\boldsymbol{F}}_s^l\|<\infty$ almost surely. Moreover, the results in Appendix E of the Supplementary Appendix show that $\mathcal{R}_{s,l}=\hat{\boldsymbol{F}}_s^l-\boldsymbol{F}_s^l$ is such that

$$\sum_{l=0}^{\infty}\|\mathcal{R}_{s,l}\|=\|\hat{\boldsymbol{F}}_s-\boldsymbol{F}_s\|\,O_p(1)=o_p(1), \tag{B.51}$$

an argument which will be used repeatedly in the proofs.

- Now consider the case where $I_i=\tilde{I}_s$ first, and analyze $\mathcal{A}_2^b$. From (B.48),

$$\mathcal{A}_2^b=T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\xi}_{t-1}^b=T^{-1}\boldsymbol{\xi}_{[\tau_{s-1}T]-1}^b-T^{-1}\boldsymbol{\xi}_{[\tau_sT]}^b+T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\xi}_t^b=T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\xi}_t^b+o_p^b(1)$$
$$=\mathcal{A}_1^b+o_p^b(1). \tag{B.52}$$

Therefore, we now derive the limit of $\mathcal{A}_1^b$. Note that

$$\boldsymbol{\xi}_t^b=\hat{\boldsymbol{\mu}}_s+\hat{\boldsymbol{F}}_s\boldsymbol{\xi}_{t-1}^b+\boldsymbol{\eta}_t^b=\hat{\boldsymbol{F}}_s^{t-[\tau_{s-1}T]}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b+\tilde{\boldsymbol{\xi}}_t^b+\left(\sum_{l=0}^{t-[\tau_{s-1}T]-1}\hat{\boldsymbol{F}}_s^l\right)\hat{\boldsymbol{\mu}}_s,$$

where $\tilde{\boldsymbol{\xi}}_t^b=\sum_{l=0}^{t-[\tau_{s-1}T]-1}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l}^b$. Therefore, $\mathcal{A}_1^b=\sum_{i=1}^4\mathcal{A}_{1,i}^b$, where $\Delta\tau_sT=[\tau_sT]-[\tau_{s-1}T]$, and

$$\mathcal{A}_{1,1}^b=T^{-1}\sum_{t=[\tau_{s-1}T]+1}^{[\tau_sT]}\tilde{\boldsymbol{\xi}}_t^b,\quad \mathcal{A}_{1,2}^b=T^{-1}\Delta\tau_sT\sum_{l=0}^{\Delta\tau_sT-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{\mu}}_s,$$
$$\mathcal{A}_{1,3}^b=T^{-1}\sum_{l=1}^{\Delta\tau_sT}\hat{\boldsymbol{F}}_s^l\boldsymbol{\xi}_{[\tau_{s-1}T]}^b,\quad \mathcal{A}_{1,4}^b=-T^{-1}\left(\sum_{l=1}^{\Delta\tau_sT-1}l\hat{\boldsymbol{F}}_s^l\right)\hat{\boldsymbol{\mu}}_s.$$

We show that $\mathcal{A}_{1,1}^b=o_p^b(1)$. First, we show $\mathrm{E}^b(\mathcal{A}_{1,1}^b)=o_p(1)$. Second, we show $\mathrm{Var}^b(\mathcal{A}_{1,1}^b)=o_p(1)$ which by Markov's inequality implies that $\mathcal{A}_{1,1}^b=o_p^b(1)$. Let $\tilde{t}=t-[\tau_{s-1}T]$ and consider $\mathrm{E}^b(\mathcal{A}_{1,1}^b)$ with $\mathrm{E}^b(\tilde{\boldsymbol{\xi}}_t^b)=\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\,\mathrm{E}^b(\boldsymbol{\eta}_{t-l}^b)=\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1}(\hat{\boldsymbol{g}}_{t-l}\odot\mathcal{I})$.

We have $\boldsymbol{\xi}_t=\boldsymbol{\mu}_s+\boldsymbol{F}_s\boldsymbol{\xi}_{t-1}+\boldsymbol{\eta}_t=\hat{\boldsymbol{\mu}}_s+\hat{\boldsymbol{F}}_s\boldsymbol{\xi}_{t-1}+\hat{\boldsymbol{\eta}}_t$. Then,

$$\hat{\boldsymbol{\eta}}_t=\boldsymbol{\eta}_t+(\boldsymbol{\mu}_s-\hat{\boldsymbol{\mu}}_s)+(\boldsymbol{F}_s-\hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1}, \tag{B.53}$$
$$\hat{\boldsymbol{g}}_t=\hat{\boldsymbol{A}}_{s,\#}\hat{\boldsymbol{\eta}}_t=\hat{\boldsymbol{A}}_{s,\#}\boldsymbol{\eta}_t+\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{\mu}_s-\hat{\boldsymbol{\mu}}_s)+\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s-\hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1}. \tag{B.54}$$

Also, we have $\boldsymbol{\eta}_t^b=(\boldsymbol{\eta}_t+(\boldsymbol{\mu}_s-\hat{\boldsymbol{\mu}}_s)+(\boldsymbol{F}_s-\hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1})\odot\boldsymbol{\nu}_t$.

Note that $\hat{\boldsymbol{\mu}}_s-\boldsymbol{\mu}_s=(\hat{\boldsymbol{c}}_{\bar{z},s}-\boldsymbol{c}_{\bar{z},s})_\#=\mathbf{vect}(\hat{d}_s,\hat{\boldsymbol{d}}_s,\mathbf{0}_{p_2})_\#$, where $\hat{d}_s,\hat{\boldsymbol{d}}_s$ are of dimension 1 and $p_1\times 1$, respectively, and this holds because the rows $p_2+1:n$ are not estimated since the equation for $\boldsymbol{r}_t$ is not estimated. Let $\hat{\boldsymbol{a}}_{1,\bullet},\hat{\boldsymbol{A}}_{p_1,\bullet},\hat{\boldsymbol{A}}_{p_2,\bullet}$ be rows 1, $2:p_1+1$ and $p_1+2:n$ of the matrix $\hat{\boldsymbol{A}}_s^{-1}$ respectively. Note that like $\boldsymbol{A}_s^{-1}$, $\hat{\boldsymbol{A}}_s^{-1}$ is upper triangular with $\hat{\boldsymbol{A}}_{p_2,\bullet}=[\mathbf{0}_{p_2\times(p_1+1)},\boldsymbol{I}_{p_2}]$ because the equation for $\boldsymbol{r}_t$ is not estimated, and $\boldsymbol{r}_t$ is assumed contemporaneously exogenous.

Therefore,

$$\hat{\boldsymbol{A}}_{s,\#}^{-1}(\boldsymbol{\mu}_s-\hat{\boldsymbol{\mu}}_s)=\begin{bmatrix}\hat{\boldsymbol{A}}_s^{-1}(\hat{\boldsymbol{c}}_{\tilde{z},s}-\boldsymbol{c}_{\tilde{z},s})\\ \boldsymbol{0}_{n(p-1)}\end{bmatrix}=\begin{bmatrix}\hat{\boldsymbol{a}}_{1,\bullet}(\hat{\boldsymbol{c}}_{\tilde{z},s}-\boldsymbol{c}_{\tilde{z},s})\\ \hat{\boldsymbol{A}}_{p_1,\bullet}(\hat{\boldsymbol{c}}_{\tilde{z},s}-\boldsymbol{c}_{\tilde{z},s})\\ \hat{\boldsymbol{A}}_{p_2,\bullet}(\hat{\boldsymbol{c}}_{\tilde{z},s}-\boldsymbol{c}_{\tilde{z},s})\\ \boldsymbol{0}_{n(p-1)}\end{bmatrix}=\begin{bmatrix}\hat{\boldsymbol{a}}_{1,\bullet}(\hat{\boldsymbol{c}}_{\tilde{z},s}-\boldsymbol{c}_{\tilde{z},s})\\ \hat{\boldsymbol{A}}_{p_1,\bullet}(\hat{\boldsymbol{c}}_{\tilde{z},s}-\boldsymbol{c}_{\tilde{z},s})\\ \boldsymbol{0}_{p_2}\\ \boldsymbol{0}_{n(p-1)}\end{bmatrix}, \tag{B.55}$$

so

$$(\hat{\boldsymbol{A}}_{s,\#}(\hat{\boldsymbol{\mu}}_s-\boldsymbol{\mu}_s))\odot\mathcal{I}=\boldsymbol{0}_{np}. \tag{B.56}$$

By similar arguments, because the $p_1+2:n$ rows of $\hat{\boldsymbol{F}}_s$ are equal to the corresponding rows of $\boldsymbol{F}_s$, $\hat{\boldsymbol{A}}_{s,\#}(\hat{\boldsymbol{F}}_s-\boldsymbol{F}_s)$, the rows $p_1+2:n$ of $\hat{\boldsymbol{A}}_{s,\#}(\hat{\boldsymbol{F}}_s-\boldsymbol{F}_s)$ are equal to zero, therefore

$$(\hat{\boldsymbol{A}}_{s,\#}(\hat{\boldsymbol{F}}_s-\boldsymbol{F}_s)\boldsymbol{\xi}_{t-1})\odot\mathcal{I}=\boldsymbol{0}_{np}. \tag{B.57}$$

Using (B.56)–(B.57), and recalling that $\tilde{t}=t-[\tau_{s-1}T]$, we have:

$$\begin{aligned}
\mathrm{E}^b(\mathcal{A}_{1,1}^b)&=\sum_{i=1}^{3}\mathcal{H}_i,\ \text{ where:}\\
\mathcal{H}_1&=T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1}\left((\hat{\boldsymbol{A}}_{s,\#}\boldsymbol{\eta}_{t-l})\odot\mathcal{I}\right)=T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1}\left((\hat{\boldsymbol{A}}_{s,\#}\boldsymbol{A}_{s,\#}^{-1}\boldsymbol{g}_{t-l})\odot\mathcal{I}\right),\\
\mathcal{H}_2&=T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1}\left((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{\mu}_s-\hat{\boldsymbol{\mu}}_s))\odot\mathcal{I}\right)=\boldsymbol{0}_{np},\\
\mathcal{H}_3&=T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1}\left((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s-\hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1})\odot\mathcal{I}\right)=\boldsymbol{0}_{np}.
\end{aligned}$$

Since $\hat{\boldsymbol{A}}_{s,\#}\boldsymbol{A}_{s,\#}^{-1}=\boldsymbol{I}_{n,\#}+o_p(1)$, it follows that:

$$\begin{aligned}
\mathcal{H}_1&=T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}_s^l\boldsymbol{A}_{s,\#}^{-1}\left(\boldsymbol{g}_{t-l}\odot\mathcal{I}\right)+T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\mathcal{R}_{s,l}\boldsymbol{A}_{s,\#}^{-1}\left(\boldsymbol{g}_{t-l}\odot\mathcal{I}\right)+o_p(1)\\
&=\mathcal{H}_1^{(1)}+\mathcal{H}_1^{(2)}+o_p(1).
\end{aligned}$$

From Assumptions 7 and 9, and using $\sum_{l=0}^{\infty}\|\boldsymbol{F}_s^l\|<\infty$ and $\sum_{l=0}^{\infty}\|\mathcal{R}_{s,l}\|<\infty$, both proven in the Supplementary Appendix (Appendices D–E), it can be shown that $\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}_s^l\boldsymbol{A}_{s,\#}^{-1}\left(\boldsymbol{g}_{t-l}\odot\mathcal{I}\right)$ and that $T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\mathcal{R}_{s,l}\boldsymbol{A}_{s,\#}^{-1}\left(\boldsymbol{g}_{t-l}\odot\mathcal{I}\right)$ are L^1-mixingales satisfying the conditions of Lemma 1, therefore $\mathcal{H}_1^{(1)}=o_p(1)$ and $\mathcal{H}_1^{(2)}=o_p(1)$. Hence $\mathcal{H}_1=o_p(1)$, so $\mathrm{E}^b(\mathcal{A}_{1,1}^b)=o_p(1)$.

Second, we show that $\mathrm{Var}^b(\mathcal{A}_{1,1}^b)=o_p(1)$. To that end, note that

$$\mathrm{E}^b(\boldsymbol{\eta}_{t-l}^b\boldsymbol{\eta}_{t-\kappa}^{b\prime})=\hat{\boldsymbol{A}}_{s,\#}^{-1}\mathrm{E}^b((\hat{\boldsymbol{g}}_{t-l}\odot\boldsymbol{\nu}_{t-l})(\hat{\boldsymbol{g}}_{t-\kappa}\odot\boldsymbol{\nu}_{t-\kappa})')(\hat{\boldsymbol{A}}_{s,\#}^{-1})'=\hat{\boldsymbol{A}}_{s,\#}^{-1}((\hat{\boldsymbol{g}}_{t-l}\hat{\boldsymbol{g}}_{t-\kappa}')\odot\mathrm{E}^b(\boldsymbol{\nu}_{t-l}\boldsymbol{\nu}_{t-\kappa}'))(\hat{\boldsymbol{A}}_{s,\#}^{-1})',$$

For $l\neq\kappa$, $\mathrm{E}^b(\boldsymbol{\nu}_{t-l}\boldsymbol{\nu}_{t-\kappa}')=\mathcal{I}\mathcal{I}'=[\mathbf{diag}(\boldsymbol{0}_{(p_1+1)\times(p_1+1)},\boldsymbol{J}_{p_2})]_{\#}=\mathcal{J}_2$. Therefore, exploiting the upper triangular structure of $\hat{\boldsymbol{A}}_s^{-1}$ with $p_2\times p_2$ lower right block equal to $\boldsymbol{I}_{p_2}$, for $l\neq\kappa$,

$$\mathrm{E}^b(\boldsymbol{\eta}_{t-l}^b\boldsymbol{\eta}_{t-\kappa}^{b\prime})=\hat{\boldsymbol{A}}_{s,\#}^{-1}((\hat{\boldsymbol{g}}_{t-l}\hat{\boldsymbol{g}}_{t-\kappa}')\odot\mathcal{J}_2)(\hat{\boldsymbol{A}}_{s,\#}^{-1})'=((\hat{\boldsymbol{A}}_{s,\#}^{-1}\hat{\boldsymbol{g}}_{t-l})\odot\mathcal{I})((\hat{\boldsymbol{A}}_{s,\#}^{-1}\hat{\boldsymbol{g}}_{t-\kappa})\odot\mathcal{I})'=\boldsymbol{\eta}_{t-l,2}\boldsymbol{\eta}_{t-\kappa,2}',$$

where $\boldsymbol{\eta}_{t,2}=\boldsymbol{g}_{t,2}=[\mathbf{vect}(\boldsymbol{0}_{p_1+1},\boldsymbol{\zeta}_t)]_{\#}$. For $l=\kappa$, $\mathrm{E}^b(\boldsymbol{\eta}_{t-l}^b\boldsymbol{\eta}_{t-l}^{b\prime})=\hat{\boldsymbol{A}}_{s,\#}^{-1}((\hat{\boldsymbol{g}}_{t-l}\hat{\boldsymbol{g}}_{t-l}')\odot\mathrm{E}^b(\boldsymbol{\nu}_{t-l}\boldsymbol{\nu}_{t-l}'))(\hat{\boldsymbol{A}}_{s,\#}^{-1})'$, where $\mathrm{E}^b(\boldsymbol{\nu}_{t-l}\boldsymbol{\nu}_{t-\kappa}')=\mathcal{J}$, so $T^{-1}\sum_{t\in\tilde{I}_s}(\hat{\boldsymbol{g}}_{t-l}\hat{\boldsymbol{g}}_{t-l}')\odot\mathcal{J}=T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{g}_{t-l}\boldsymbol{g}_{t-l}'+o_p(1)$ by Assumption 9, Lemma 8 followed by standard 2SLS theory. So,

$$T^{-1}\sum_{t\in\tilde{I}_s}\mathrm{E}^b(\boldsymbol{\eta}_{t-l}^b\boldsymbol{\eta}_{t-l}^{b\prime})=\hat{\boldsymbol{A}}_{s,\#}^{-1}\left(T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{g}_{t-l}\boldsymbol{g}_{t-l}'\right)(\hat{\boldsymbol{A}}_{s,\#}^{-1})'+o_p(1)=T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\eta}_{t-l}\boldsymbol{\eta}_{t-l}'+o_p(1).$$

Hence,

$$\mathrm{Var}^b(\mathcal{A}_{1,1}^b)=T^{-2}\sum_{t\in\tilde{I}_s}\mathrm{E}^b(\tilde{\boldsymbol{\xi}}_t^b\tilde{\boldsymbol{\xi}}_t^{b\prime})=\mathcal{V}_1+\mathcal{V}_2+o_p(1) \tag{B.58}$$

$$\mathcal{V}_1 = T^{-2}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l}\boldsymbol{\eta}'_{t-l}(\hat{\boldsymbol{F}}_s^l)' \tag{B.59}$$

$$\mathcal{V}_2 = T^{-2}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0,l\neq\kappa}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l,2}\boldsymbol{\eta}'_{t-\kappa,2}(\hat{\boldsymbol{F}}_s^\kappa)' \tag{B.60}$$

Consider $\mathcal{V}_1$. We analyze first $\mathcal{V}_1^*$, where

$$\begin{aligned}\mathcal{V}_1^* &= T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}_s^l\left(\boldsymbol{A}_{s,\#}^{-1}\boldsymbol{g}_{t-l}\boldsymbol{g}'_{t-l}(\boldsymbol{A}_{s,\#}^{-1})'\right)(\boldsymbol{F}_s^l)' = T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l}\boldsymbol{\eta}'_{t-l}\boldsymbol{F}_s^{l'}\\ &= \mathcal{B}_{1,1}^{(1)}+\mathcal{B}_{1,1}^{(2)}+o_p(1) = \mathbb{B}_1(\tau_{s-1},\tau_s)+o_p(1),\end{aligned}$$

where the last three quantities above were already defined and analyzed in the proof of Lemma 2 in the Supplementary Appendix, Appendix C, where it was shown that

$$\mathcal{B}_{1,1}^{(1)}\xrightarrow{p}\mathbb{B}_1(\tau_{s-1},\tau_s) = \sum_{l=0}^{\infty}\boldsymbol{F}_s^l\left(\boldsymbol{A}_s^{-1}\int_{\tau_{s-1}}^{\tau_s}\overline{\boldsymbol{\Sigma}}(\tau)\mathrm{d}\tau\boldsymbol{A}_s^{-1'}\right)_{\#}\boldsymbol{F}_s^{l\prime}$$

and that $\mathcal{B}_{1,1}^{(2)} = o_p(1)$. By similar arguments that were employed to analyze those terms,

$$\mathcal{V}_1^{**} = T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l}\boldsymbol{\eta}'_{t-l}(\boldsymbol{F}_s^l)' = \mathcal{B}_{1,1}^{(1)}+o_p(1) = \mathbb{B}_1(\tau_{s-1},\tau_s)+o_p(1).$$

Now consider $\mathcal{V}_1$, where

$$\begin{aligned}\mathcal{V}_1 &= T^{-1}(\mathcal{V}_1^{(1)}+\mathcal{V}_1^{(2)}+(\mathcal{V}_1^{(2)})'+\mathcal{V}_1^{(3)})+o_p(1),\\ \mathcal{V}_1^{(1)} &= T^{-1}\sum_{t\in I_i}\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l}\boldsymbol{\eta}'_{t-l}(\boldsymbol{F}_s^l)' = \mathcal{V}_1^{**} = O_p(1),\\ \mathcal{V}_1^{(2)} &= T^{-1}\sum_{t\in I_i}\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l}\boldsymbol{\eta}'_{t-l}(\mathcal{R}_{s,l})',\\ \mathcal{V}_1^{(3)} &= T^{-1}\sum_{t\in I_i}\sum_{l=0}^{\tilde{t}-1}\mathcal{R}_{s,l}\boldsymbol{\eta}_{t-l}\boldsymbol{\eta}'_{t-l}(\mathcal{R}_{s,l})'.\end{aligned}$$

Similarly to $\mathcal{V}_1^{**}$, because $\hat{\boldsymbol{F}}_s^l - \boldsymbol{F}_s^l = \mathcal{R}_{s,l} = o_p(1)$, we can show that $\mathcal{V}_1^{(2)} = o_p(1)$. From (B.51), $\mathcal{R}_{s,l}$ is such that $\sum_{l=0}^{\infty}\mathcal{R}_{s,l} = \|\hat{\boldsymbol{F}}_s-\boldsymbol{F}_s\|\,O_p(1) = o_p(1)$. Therefore, by the same arguments as for $\mathcal{V}_1^{(2)} = o_p(1)$, one can show that $\mathcal{V}_1^{(3)} = o_p(1)$, therefore $T\mathcal{V}_1 = \mathbb{B}_1(\tau_{s-1},\tau_s)+o_p(1)$, and $\mathcal{V}_1 = o_p(1)$.

By similar arguments to the analysis of the term $\mathcal{B}_{1,1}$ in the Supplementary Appendix, Appendix C, proof of Lemma 2, $T\mathcal{V}_2 = o_p(1)$. Substituting $\mathcal{V}_2 = o_p(1)$ and $\mathcal{V}_1 = o_p(1)$ into (B.58), it follows that $\mathrm{Var}^b(\mathcal{A}_{1,1}^b) = \mathcal{V}_1+\mathcal{V}_2 = o_p(1)$, and, by Markov's inequality, that $\mathcal{A}_{1,1}^b = o_p^b(1)$. It also follows that:

$$T\mathrm{Var}^b(\mathcal{A}_{1,1}^b) = \mathbb{B}_1(\tau_{s-1},\tau_s)+o_p(1), \tag{B.61}$$

a stronger result that we need later in this proof.

Consider now $\mathcal{A}_{1,2}^b, \mathcal{A}_{1,3}^b, \mathcal{A}_{1,4}^b$. We have $\mathcal{A}_{1,2}^b = T^{-1}\Delta\tau_sT\sum_{l=0}^{\Delta\tau_sT-1}\boldsymbol{F}_s^l\boldsymbol{\mu}_s+o_p(1)$ and by Assumption 7, it follows that $\mathcal{A}_{1,2}^b\xrightarrow{p}\Delta\tau_s(\boldsymbol{I}_{np}-\boldsymbol{F}_s)^{-1}\boldsymbol{\mu}_s = \int_{\tau_{s-1}}^{\tau_s}\mathbb{Q}_1(\tau)\mathrm{d}\tau$. Now consider $\mathcal{A}_{1,3}^b$. Because $\|T^{-1}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b\| = o_p^b(1)$ and $\|\sum_{l=1}^{\infty}\hat{\boldsymbol{F}}_s^l\| = O_p(1)$,

$$\|\mathcal{A}_{1,3}^b\| = \|T^{-1}\sum_{l=1}^{\Delta\tau_sT}\hat{\boldsymbol{F}}_s^l\boldsymbol{\xi}_{[\tau_{s-1}T]}^b\| \le \|\sum_{l=1}^{\infty}\hat{\boldsymbol{F}}_s^l\|\;\|T^{-1}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b\| = O_p(1)o_p^b(1) = o_p^b(1).$$

Since $\sum_{l=1}^{\Delta\tau_sT-1}l\hat{\boldsymbol{F}}_s^l = O_p(1)$ and $\hat{\boldsymbol{\mu}}_s-\boldsymbol{\mu}_s = o_p(1)$, $\mathcal{A}_{1,4}^b = o_p(1)$. Combining these results, we obtain:

$$\mathcal{A}_1^b = \Delta\tau_s(\boldsymbol{I}_{np}-\boldsymbol{F}_s)^{-1}\boldsymbol{\mu}_s + o_p^b(1) = \int_{\tau_{s-1}}^{\tau_s}\mathbb{Q}_1(\tau)\mathrm{d}\tau + o_p^b(1) = \mathcal{A}_1+o_p^b(1);$$

where $\mathcal{A}_1 = \int_{\tau_{s-1}}^{\tau_s} \mathbb{Q}_1(\tau)\mathrm{d}\tau + o_p(1)$ from (C.1) in Supplementary Appendix, Appendix C, and $\mathcal{A}_i$, $i = 1, 2$, are the sample equivalents of $\mathcal{A}_i^b$. From (B.52), it follows that $\mathcal{A}_2^b = \mathcal{A}_1 + o_p^b(1) = \mathcal{A}_2 + o_p^b(1) \xrightarrow{p^b} \int_{\tau_{s-1}}^{\tau_s} \mathbb{Q}_1(\tau)\mathrm{d}\tau$.

- Next, analyze $\mathcal{B}_3^b$. First, note that because $T^{-1}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b'} = o_p^b(1)$ as shown in the preliminaries of this proof,

$$\mathcal{B}_1^b = T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b'} = T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\xi}_{t-1}^b\boldsymbol{\xi}_{t-1}^{b'} + o_p^b(1) = \mathcal{B}_3^b + o_p^b(1), \tag{B.62}$$

so we analyze instead $\mathcal{B}_1^b$. Note that $\mathcal{B}_1^b = \sum_{i=1}^3 \mathcal{B}_{1,i}^b + \sum_{j=1}^3 \left\{ \mathcal{B}_{1,3+j}^b + \mathcal{B}_{1,3+j}^{b\prime} \right\}$, where

$$\begin{aligned}
\mathcal{B}_{1,1}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\tilde{\boldsymbol{\xi}}_t^b\tilde{\boldsymbol{\xi}}_t^{b\prime} = T^{-1}\sum_{t\in\tilde{I}_s}\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l}^b\right)\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l}^b\right)'\\
\mathcal{B}_{1,2}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{\mu}}_s\right)\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{\mu}}_s\right)'\\
\mathcal{B}_{1,3}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\hat{\boldsymbol{F}}_s^{\tilde{t}}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b\boldsymbol{\xi}_{[\tau_{s-1}T]}^{b\prime}(\hat{\boldsymbol{F}}_s^{\tilde{t}})'\\
\mathcal{B}_{1,4}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\tilde{\boldsymbol{\xi}}_t^b\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{\mu}}_s\right)'\\
\mathcal{B}_{1,5}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\tilde{\boldsymbol{\xi}}_t^b\boldsymbol{\xi}_{[\tau_{s-1}T]}^{b\prime}(\hat{\boldsymbol{F}}_s^{\tilde{t}})'\\
\mathcal{B}_{1,6}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{\mu}}_s\right)\boldsymbol{\xi}_{[\tau_{s-1}T]}^{b\prime}(\hat{\boldsymbol{F}}_s^{\tilde{t}})'.
\end{aligned}$$

Recall that

$$\mathbb{B}_1(\tau_{s-1},\tau_s) = \sum_{l=0}^{\infty}\boldsymbol{F}_s^l\left(\boldsymbol{A}_s^{-1}\int_{\tau_{s-1}}^{\tau_s}\overline{\boldsymbol{\Sigma}}(\tau)\mathrm{d}\tau\boldsymbol{A}_s^{-1'}\right)_{\#}\boldsymbol{F}_s^{l\prime}.$$

We show $\mathcal{B}_{1,1}^b - \mathbb{B}_1(\tau_{s-1},\tau_s) = o_p^b(1)$ by showing that $\mathrm{E}^b(\mathcal{B}_{1,1}^b - \mathbb{B}_1(\tau_{s-1},\tau_s)) = o_p(1)$ and $\mathrm{Var}^b(\mathbf{vect}\mathcal{B}_{1,1}^b) = o_p(1)$.

$$\begin{aligned}
\mathcal{B}_{1,1}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l}^b\boldsymbol{\eta}_{t-\kappa}^{b\prime}(\hat{\boldsymbol{F}}_s^\kappa)' = T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1}\boldsymbol{g}_{t-l}^b\boldsymbol{g}_{t-\kappa}^{b\prime}(\hat{\boldsymbol{A}}_{s,\#}^{-1})'(\hat{\boldsymbol{F}}_s^\kappa)'\\
\mathrm{E}^b(\mathcal{B}_{1,1}^b) &= T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}_s^l\,\mathrm{E}^b(\boldsymbol{\eta}_{t-l}^b\boldsymbol{\eta}_{t-\kappa}^{b\prime})(\hat{\boldsymbol{F}}_s^\kappa)' = T\mathrm{Var}^b(\mathcal{A}_{1,1}^b) = \mathbb{B}_1(\tau_{s-1},\tau_s) + o_p(1),
\end{aligned}$$

where the last equality above follows from (B.61). We have:

$$\begin{aligned}
\mathbf{vect}\mathcal{B}_{1,1}^b &= T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\left((\hat{\boldsymbol{F}}_s^\kappa\hat{\boldsymbol{A}}_{s,\#}^{-1})\otimes(\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1})\right)\left(\boldsymbol{g}_{t-\kappa}^b\otimes\boldsymbol{g}_{t-l}^b\right)\\
&= T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\left((\hat{\boldsymbol{F}}_s^\kappa\hat{\boldsymbol{A}}_{s,\#}^{-1})\otimes(\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1})\right)\left((\hat{\boldsymbol{g}}_{t-\kappa}\odot\boldsymbol{v}_{t-\kappa})\otimes(\hat{\boldsymbol{g}}_{t-l}\odot\boldsymbol{v}_{t-l})\right)\\
&= T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\left((\hat{\boldsymbol{F}}_s^\kappa\hat{\boldsymbol{A}}_{s,\#}^{-1})\otimes(\hat{\boldsymbol{F}}_s^l\hat{\boldsymbol{A}}_{s,\#}^{-1})\right)\left((\hat{\boldsymbol{g}}_{t-\kappa}\otimes\hat{\boldsymbol{g}}_{t-l})\odot(\boldsymbol{v}_{t-\kappa}\otimes\boldsymbol{v}_{t-l})\right).
\end{aligned}$$

$$\begin{aligned}
\mathrm{Var}^b(\mathbf{vect}\mathcal{B}_{1,1}^b) &= \mathrm{E}^b(\mathbf{vect}\mathcal{B}_{1,1}^b(\mathbf{vect}\mathcal{B}_{1,1}^b)') - \mathrm{E}^b(\mathbf{vect}\mathcal{B}_{1,1}^b)\,\mathrm{E}^b(\mathbf{vect}\mathcal{B}_{1,1}^b)'\\
&= \mathrm{E}^b(\mathbf{vect}\mathcal{B}_{1,1}^b(\mathbf{vect}\mathcal{B}_{1,1}^b)') - \mathbf{vect}\mathbb{B}_1(\tau_{s-1},\tau_s)\,(\mathbf{vect}\mathbb{B}_1(\tau_{s-1},\tau_s))' + o_p(1).
\end{aligned}$$

We need to show that $\mathrm{E}^b(\mathbf{vect}\,\mathcal{B}^b_{1,1}(\mathbf{vect}\,\mathcal{B}^b_{1,1})') \xrightarrow{p} \mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s)\,(\mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s))'$. Letting $\tilde{t}^* = t^* - [\tau_{s-1}T]$,

$$\begin{aligned}\mathrm{E}^b(\mathbf{vect}\,\mathcal{B}^b_{1,1}(\mathbf{vect}\,\mathcal{B}^b_{1,1})') &= \left[T^{-1}\sum_{t\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\left((\hat{\boldsymbol{F}}^{\kappa}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\otimes(\hat{\boldsymbol{F}}^{l}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\right)(\boldsymbol{g}^b_{t-\kappa}\otimes\boldsymbol{g}^b_{t-l})\right]\\ &\quad\times\left[T^{-1}\sum_{t^*\in\tilde{I}_s}\sum_{l^*,\kappa^*=0}^{\tilde{t}^*-1}\left((\hat{\boldsymbol{F}}^{\kappa^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\otimes(\hat{\boldsymbol{F}}^{l^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\right)(\boldsymbol{g}^b_{t^*-\kappa^*}\otimes\boldsymbol{g}^b_{t^*-l^*})\right]\\ &= T^{-2}\sum_{t,t^*\in\tilde{I}_s}\sum_{l,\kappa=0}^{\tilde{t}-1}\sum_{l^*,\kappa^*=0}^{\tilde{t}^*-1}((\hat{\boldsymbol{F}}^{\kappa}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\otimes(\hat{\boldsymbol{F}}^{l}_s\hat{\boldsymbol{A}}^{-1}_{s,\#}))\mathcal{G}((\hat{\boldsymbol{F}}^{\kappa^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})'\otimes(\hat{\boldsymbol{F}}^{l^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})')\\ &= \sum_{i=1}^{9}\mathcal{O}_i,\\ \mathcal{G} &= \mathrm{E}^b\left((\boldsymbol{g}^b_{t-\kappa}\otimes\boldsymbol{g}^b_{t-l})(\boldsymbol{g}^b_{t^*-\kappa^*}\otimes\boldsymbol{g}^b_{t^*-l^*})'\right),\end{aligned}$$

where $\mathcal{O}_i$ are the terms corresponding to nine cases when $\mathcal{G}\neq\boldsymbol{O}_{(n^2p^2)\times(n^2p^2)}$. Case (1) is when $t-\kappa = t-l$, $t^*-\kappa^* = t^*-l^*$, $t-\kappa\neq t^*-\kappa^*$; we show below that $\mathcal{O}_1 = \mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s)(\mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s))' + o_p(1)$. For brevity, the rest of the cases are defined and analyzed in Supplementary Appendix, Appendix F, where we show that

$$\mathcal{O}_i = o_p(1) \text{ for } i = 2,\ldots,9. \tag{B.63}$$

By Assumption 10, $\mathrm{E}^b[(\boldsymbol{\nu}_t\boldsymbol{\nu}_t')\otimes(\boldsymbol{\nu}_{t-l}\boldsymbol{\nu}_{t-l}')] = [\mathrm{E}^b(\boldsymbol{\nu}_t\boldsymbol{\nu}_t')]\otimes[\mathrm{E}^b(\boldsymbol{\nu}_{t-l}\boldsymbol{\nu}_{t-l})']$, because we know $\mathrm{E}^b(\nu_t^2\nu_{t-l}^2) = \mathrm{E}^b(\nu_t^2)\,\mathrm{E}^b(\nu_{t-l}^2) = 1$, $\mathrm{E}^b(\nu_t\nu_{t-l}) = 0$ and $\mathrm{E}^b(\nu_t^2\nu_{t-l}) = 0$ (these are elements of $\mathrm{E}^b((\boldsymbol{\nu}_t\boldsymbol{\nu}_t')\otimes(\boldsymbol{\nu}_{t-l}\boldsymbol{\nu}_{t-l}'))$. Hence, conditional on the data, we have, by Assumption 10,

$$\begin{aligned}\mathcal{G} &= \mathrm{E}^b((\boldsymbol{g}^b_{t-\kappa}\otimes\boldsymbol{g}^b_{t-\kappa})(\boldsymbol{g}^b_{t^*-\kappa^*}\otimes\boldsymbol{g}^b_{t^*-\kappa^*})')\\ &= \mathrm{E}^b[[(\hat{\boldsymbol{g}}_{t-\kappa}\otimes\hat{\boldsymbol{g}}_{t-\kappa})\odot(\boldsymbol{\nu}_{t-\kappa}\otimes\boldsymbol{\nu}_{t-\kappa})][(\hat{\boldsymbol{g}}_{t^*-\kappa^*}\otimes\hat{\boldsymbol{g}}_{t^*-\kappa^*})\odot(\boldsymbol{\nu}_{t^*-\kappa^*}\otimes\boldsymbol{\nu}_{t^*-\kappa^*})]]\\ &= [(\hat{\boldsymbol{g}}_{t-\kappa}\otimes\hat{\boldsymbol{g}}_{t-\kappa})\odot\mathrm{E}^b(\boldsymbol{\nu}_{t-\kappa}\otimes\boldsymbol{\nu}_{t-\kappa})][(\hat{\boldsymbol{g}}_{t^*-\kappa^*}\otimes\hat{\boldsymbol{g}}_{t^*-\kappa^*})\odot\mathrm{E}^b(\boldsymbol{\nu}_{t^*-\kappa^*}\otimes\boldsymbol{\nu}_{t^*-\kappa^*})]\\ &= (\hat{\boldsymbol{g}}_{t-\kappa}\otimes\hat{\boldsymbol{g}}_{t-\kappa})(\hat{\boldsymbol{g}}_{t^*-\kappa^*}\otimes\hat{\boldsymbol{g}}_{t^*-\kappa^*}),\end{aligned}$$

hence

$$\begin{aligned}\mathcal{O}_1 &= \left[T^{-1}\sum_{t\in\tilde{I}_s}\sum_{\kappa=0}^{\tilde{t}-1}\left((\hat{\boldsymbol{F}}^{\kappa}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\otimes(\hat{\boldsymbol{F}}^{\kappa}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\right)(\boldsymbol{g}^b_{t-\kappa}\otimes\boldsymbol{g}^b_{t-\kappa})\right]\\ &\quad\times\left[T^{-1}\sum_{t^*\in\tilde{I}_s}\sum_{\kappa^*=0}^{\tilde{t}-1}\left((\hat{\boldsymbol{F}}^{\kappa^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\otimes(\hat{\boldsymbol{F}}^{\kappa^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\right)(\boldsymbol{g}^b_{t^*-\kappa^*}\otimes\boldsymbol{g}^b_{t^*-\kappa^*})\right]'\\ &= \sum_{\kappa,\kappa^*=0}^{\Delta\tau_sT-1}\left((\hat{\boldsymbol{F}}^{\kappa}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\otimes(\hat{\boldsymbol{F}}^{\kappa}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\right)\left(T^{-1}\sum_{t\in\tilde{I}_s}(\hat{\boldsymbol{g}}_{t-\kappa}\otimes\hat{\boldsymbol{g}}_{t-\kappa})\right)\\ &\quad\times\left(T^{-1}\sum_{t\in\tilde{I}_s}(\hat{\boldsymbol{g}}_{t^*-\kappa^*}\otimes\hat{\boldsymbol{g}}_{t^*-\kappa^*})\right)\left((\hat{\boldsymbol{F}}^{\kappa^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\otimes(\hat{\boldsymbol{F}}^{\kappa^*}_s\hat{\boldsymbol{A}}^{-1}_{s,\#})\right)' + o_p(1)\\ &= (\mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s)+o_p(1))(\mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s)+o_p(1))' = \mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s)(\mathbf{vect}\,\mathbb{B}_1(\tau_{s-1},\tau_s))' + o_p(1),\end{aligned}$$

where the last two lines follow because $\boldsymbol{g}_{t-\kappa}\otimes\boldsymbol{g}_{t-\kappa} = \mathbf{vect}(\boldsymbol{g}_{t-\kappa}\boldsymbol{g}_{t-\kappa}')$, and

$$T^{-1}\sum_{t\in\tilde{I}_s}(\hat{\boldsymbol{g}}_{t-\kappa}\otimes\hat{\boldsymbol{g}}_{t-\kappa}) = T^{-1}\sum_{t\in\tilde{I}_s}\mathbf{vect}\,(\hat{\boldsymbol{g}}_{t-\kappa}\hat{\boldsymbol{g}}_{t-\kappa}') = \operatorname*{plim}_{T\to\infty}\mathbf{vect}\,T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{g}_{t-\kappa}\boldsymbol{g}_{t-\kappa}' + o_p(1),$$

which follows by standard 2SLS theory and Lemma 8. So, $\mathcal{O}_1 = \mathbf{vect}\mathbb{B}_1(\tau_{s-1},\tau_s)(\mathbf{vect}\mathbb{B}_1(\tau_{s-1},\tau_s))' + o_p(1)$.

Therefore, $\mathrm{Var}^b(\mathbf{vect}\mathcal{B}^b_{1,1}) = o_p(1)$, so by Markov's inequality,

$$\mathcal{B}^b_{1,1} = \mathbb{B}_1(\tau_{s-1},\tau_s) + o^b_p(1).$$

Next, because $\hat{\boldsymbol{\mu}}_s = \boldsymbol{\mu}_s + o_p(1)$, and $\hat{\boldsymbol{F}}_s = \boldsymbol{F}_s + o_p(1)$, and $\sum_{l=0}^{\infty}\|\hat{\boldsymbol{F}}^l_s - \boldsymbol{F}^l_s\| = o_p(1)$ as shown in Supplementary Appendix, Appendix E, $\mathcal{B}^b_{1,2} = \mathcal{B}_{1,2} + o_p(1) = \mathbb{B}_2(\tau_{s-1},\tau_s) + o_p(1)$, where $\mathbb{B}_2(\tau_{s-1},\tau_s) = \int_{\tau_{s-1}}^{\tau_s}\mathbb{Q}_1(\tau)\mathbb{Q}_1'(\tau)\mathrm{d}\tau$, and $\mathcal{B}_{1,2}$ is the sample equivalent of $\mathcal{B}^b_{1,2}$ (and in general, $\mathcal{B}_{1,i}$, $i = 1, \ldots, 6$ are the sample equivalents of $\mathcal{B}^b_{1,i}$, defined in the proof of Lemma 2 in Supplementary Appendix, Appendix C). Also, we have $\mathcal{B}^b_{1,3} = \mathcal{B}_{1,3} + o^b_p(1) = o^b_p(1)$, because, as shown in the preliminaries, $T^{-\alpha}\boldsymbol{\xi}^b_t\boldsymbol{\xi}^{b'}_t = o^b_p(1)$, and $\hat{\boldsymbol{F}}^l_s$ is exponentially decaying with l.

Consider $\mathcal{B}^b_{1,4}$, which is equal to

$$T^{-1}\sum_{t\in\tilde{I}_s}\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}^l_s\boldsymbol{\eta}^b_{t-l}\right)\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}^l_s\hat{\boldsymbol{\mu}}_s\right)' = T^{-1}\sum_{t\in\tilde{I}_s}\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}^l_s((\hat{\boldsymbol{A}}^{-1}_{s,\#}\hat{\boldsymbol{g}}_{t-l})\odot\boldsymbol{\nu}_{t-l})\right)\left(\sum_{l=0}^{\tilde{t}-1}\hat{\boldsymbol{F}}^l_s\hat{\boldsymbol{\mu}}_s\right)'.$$

We show $\mathcal{B}^b_{1,4} = o^b_p(1)$. To that end, note that

$$\mathrm{E}^b(\mathcal{B}^b_{1,4}) = T^{-1}\sum_{t\in\tilde{I}_s}\left(\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}^l_s\boldsymbol{\eta}_{t-l,2}\right)\left(\sum_{l=0}^{\tilde{t}-1}\boldsymbol{F}^l_s\boldsymbol{\mu}_s\right)' + o_p(1) = o_p(1),$$

by similar arguments as for its sample equivalent $\mathcal{B}_{1,4}$ defined in the proof of Lemma 2. So, $\mathrm{E}^b(\mathcal{B}^b_{1,4}) = o_p(1)$. Moreover, by similar arguments as before, it can be shown that $\|\mathrm{Var}^b(\mathbf{vect}\,\mathcal{B}^b_{1,4})\| = o_p(1)$. Hence, by Markov's inequality, $\mathcal{B}^b_{1,4} = o^b_p(1)$. Similarly, because $T^{-\alpha}\boldsymbol{\xi}^b_{[\tau_{s-1}T]} = o^b_p(1)$ for any $\alpha > 0$, it can be shown that $\mathcal{B}^b_{1,5} = o^b_p(1)$, and $\mathcal{B}^b_{1,6} = o^b_p(1)$. Putting all the results for $\mathcal{B}^b_{1,i}$ together, $i = 1, \ldots, 6$ we conclude $\mathcal{B}^b_1 = \mathbb{B}_1(\tau_{s-1}, \tau_s) + \mathbb{B}_2(\tau_{s-1}, \tau_s) + o^b_p(1) = \mathbb{B}(\tau_{s-1}, \tau_s) + o^b_p(1) = \mathcal{B}_1 + o^b_p(1)$, where $\mathbb{B}(\tau_{s-1}, \tau_s) = \mathbb{B}_1(\tau_{s-1}, \tau_s) + \mathbb{B}_2(\tau_{s-1}, \tau_s)$, and $\mathcal{B}_1$ is the sample equivalents of $\mathcal{B}^b_1$ defined in the proof of Lemma 2.

Because $\mathcal{B}^b_3 = \mathcal{B}^b_1 + o^b_p(1)$, it follows that $\mathcal{B}^b_3 = \mathbb{B}(\tau_{s-1}, \tau_s) + o^b_p(1)$, where the same result was shown to hold for $\mathcal{B}_3$ defined in the proof of Lemma 2. Now consider $\mathcal{B}^b_2$. Using $\boldsymbol{\xi}^b_t = \hat{\boldsymbol{\mu}}_s + \hat{\boldsymbol{F}}_s\boldsymbol{\xi}^b_{t-1} + \boldsymbol{\eta}^b_t$, it follows that:

$$\mathcal{B}^b_2 = \hat{\boldsymbol{\mu}}_s\mathcal{A}^{b'}_2 + \hat{\boldsymbol{F}}_s\mathcal{B}^b_3 + T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\xi}^b_{t-1}\boldsymbol{\eta}^{b'}_t + o^b_p(1).$$

By similar arguments as for some elements of $\mathcal{B}_1$, it can be shown that $T^{-1}\sum_{t\in\tilde{I}_s}\boldsymbol{\xi}^b_{t-1}\boldsymbol{\eta}^{b'}_t = o^b_p(1)$. Therefore, $\mathcal{B}^b_2 = \int_{\tau_{s-1}}^{\tau_s}\boldsymbol{\mu}(\tau)\mathbb{Q}_1(\tau)\mathrm{d}\tau + \int_{\tau_{s-1}}^{\tau_s}\boldsymbol{F}(\tau)\mathbb{Q}_2(\tau)\mathrm{d}\tau + o^b_p(1)$. So, for $I_i = \tilde{I}_s$,

$$\hat{\mathbf{Q}}^b_{(i)} = \int_{\tau_{s-1}}^{\tau_s}\boldsymbol{\Upsilon}'(\tau)\mathbb{Q}_{\boldsymbol{z}}(\tau)\boldsymbol{\Upsilon}(\tau)\mathrm{d}\tau + o^b_p(1) = \mathbb{Q}_{(i)} + o^b_p(1).$$

For other regimes, by similar arguments as in the end of the proof of Lemma 2,

$$\hat{\mathbf{Q}}^b_{(i)} = \int_{\lambda_{i-1}}^{\lambda_i}\boldsymbol{\Upsilon}'(\tau)\mathbb{Q}_{\boldsymbol{z}}(\tau)\boldsymbol{\Upsilon}(\tau)\mathrm{d}\tau + o^b_p(1) = \mathbb{Q}_{(i)} + o^b_p(1),$$

concluding the proof. □

Lemma 10. *If Assumptions 1–10 and 9′ hold for the WR bootstrap,* $T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}^b_t\boldsymbol{g}^{b'}_t\mathcal{S}^b_\dagger \overset{d^b_p}{\Rightarrow} \widetilde{\mathbb{M}}_i$ *in probability uniformly in* $\boldsymbol{\lambda}_k$*, where* $\widetilde{\mathbb{M}}_i$ *is defined as in Lemma 6, and* $\mathcal{S}^b_\dagger = \mathcal{S}_u$ *or* $\mathcal{S}^b_\dagger = (\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)})_\#$*. If* $m = 0$*, then* $\mathcal{S}^b_\dagger = \mathcal{S}_u$ *or* $\mathcal{S}^b_\dagger = \hat{\boldsymbol{\beta}}_{\boldsymbol{x},\#}$*.*

Lemma 11. *Let Assumptions 1–10 hold for the WF bootstrap. Then,* $T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t\boldsymbol{g}^{b'}_t\mathcal{S}^b_\dagger \overset{d^b_p}{\Rightarrow} \widetilde{\mathbb{M}}_i$ *in probability uniformly in* $\boldsymbol{\lambda}_k$*, where* $\mathcal{S}^b_\dagger$ *is as defined in Lemma 10.*

For the proofs of Lemmas 10–11, it suffices to consider $\mathcal{S}^b_\dagger = \mathcal{S}_u$ or $\mathcal{S}^b_\dagger = \hat{\boldsymbol{\beta}}_{\boldsymbol{x},\#}$, therefore considering $m = 0$. If $\mathcal{S}^b_\dagger = (\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)})_\#$, by Lemma 7 followed by standard 2SLS theory, $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)} = \boldsymbol{\beta}^0_{\boldsymbol{x},(i)} + O_p(T^{-1/2})$ so $\mathcal{S}^b_\dagger = \mathcal{S}_\dagger + O_p(T^{-1/2})$, and the results follow in a similar fashion.

Additionally to the notation already defined at the beginning of the proof of Lemma 9, we use the following results and notation, some relevant for both Lemmas 10 and 11. Consider the partition $\tilde{I}_s$, then for the WR bootstrap we have $\tilde{\boldsymbol{z}}^b_t = \hat{\boldsymbol{c}}_{\tilde{\boldsymbol{z}},s} + \sum_{i=1}^p\hat{\boldsymbol{C}}_{i,s}\tilde{\boldsymbol{z}}^b_{t-i} + \boldsymbol{e}^b_t$, and for both WR and WF bootstraps, we have $\boldsymbol{e}^b_t = \hat{\boldsymbol{A}}^{-1}_s\boldsymbol{\epsilon}^b_t$. We have for the WR bootstrap:

$$\boldsymbol{\xi}^b_t = \hat{\boldsymbol{\mu}}_s + \hat{\boldsymbol{F}}_s\boldsymbol{\xi}^b_{t-1} + \boldsymbol{\eta}^b_t = \hat{\boldsymbol{F}}_s^{t-[\tau_{s-1}T]}\boldsymbol{\xi}^b_{[\tau_{s-1}T]} + \sum_{l=0}^{t-[\tau_{s-1}T]-1}\hat{\boldsymbol{F}}^l_s\boldsymbol{\eta}^b_{t-l} + \left(\sum_{l=0}^{t-[\tau_{s-1}T]-1}\hat{\boldsymbol{F}}^l_s\right)\hat{\boldsymbol{\mu}}_s \tag{B.64}$$

except that in (B.64) when $s = 1$ and we are in the first regime $\tilde{I}_1 = [1, \ldots, [\tau_1 T]]$, we have that $\boldsymbol{\xi}^b_0 = \boldsymbol{\xi}_0$, where $\boldsymbol{\xi}^b_t = \mathbf{vect}_{j=0:(p-1)}(\tilde{\boldsymbol{z}}^b_{t-j})$. Let $\mathcal{F}^b_t = \{\nu_t, \nu_{t-1}, \ldots, \nu_1\}$. Recall that, by Assumption 10,

$$\mathrm{E}^b(\boldsymbol{\nu}_t) = \mathrm{E}^b(\boldsymbol{\nu}_t|\mathcal{F}^b_{t-1}) = \mathbf{vect}(\mathbf{0}_{p_1+1}, \boldsymbol{\iota}_{p_2}, \mathbf{0}_{n(p-1)\times 1}) = \mathcal{I} \tag{B.65}$$

$$\mathrm{E}^b(\boldsymbol{v}_t\boldsymbol{v}_t') = \mathrm{E}^b(\boldsymbol{v}_t\boldsymbol{v}_t'|\mathcal{F}_{t-1}^b) = \left(\mathbf{diag}(\boldsymbol{J}_{p_1+1}, \boldsymbol{J}_{p_2})\right)_{\#} = \mathcal{J} \tag{B.66}$$

$$\mathrm{E}^b(\boldsymbol{v}_t\boldsymbol{v}_{t-j}') = \mathrm{E}^b(\boldsymbol{v}_t\boldsymbol{v}_{t-j}'|\mathcal{F}_{t-1}^b) = (\mathbf{diag}(\mathbf{0}_{p_1+1}, \boldsymbol{J}_{p_2}))_{\#} = \mathcal{J}_2. \tag{B.67}$$

Furthermore, recall that $\boldsymbol{\xi}_t = \boldsymbol{\mu}_s + \boldsymbol{F}_s\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t = \hat{\boldsymbol{\mu}}_s + \hat{\boldsymbol{F}}_s\boldsymbol{\xi}_{t-1} + \hat{\boldsymbol{\eta}}_t$, and therefore $\hat{\boldsymbol{\eta}}_t = \boldsymbol{\eta}_t + (\boldsymbol{\mu}_s - \hat{\boldsymbol{\mu}}_s) + (\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1}$. By backward substitution of $\boldsymbol{\xi}_t = \boldsymbol{\mu}_s + \boldsymbol{F}_s\boldsymbol{\xi}_{t-1} + \boldsymbol{\eta}_t$, we have that: $\boldsymbol{\xi}_{t-1} = \boldsymbol{F}_s^{\tilde{t}-1}\boldsymbol{\xi}_{[\tau_{s-1}T]} + \sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l-1} + \left(\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\right)\boldsymbol{\mu}_s$, where $\tilde{t} = t - [\tau_{s-1}T]$. We also have $\boldsymbol{\eta}_t^b = \hat{\boldsymbol{A}}_{s,\#}^{-1}\,\boldsymbol{g}_t^b = \hat{\boldsymbol{A}}_{s,\#}^{-1}\,(\hat{\boldsymbol{g}}_t \odot \boldsymbol{v}_t)$, where recall that $\odot$ is the element-wise multiplication. Hence:

$$\begin{aligned}
\hat{\boldsymbol{g}}_t &= \hat{\boldsymbol{A}}_{s,\#}\,\hat{\boldsymbol{\eta}}_t = \hat{\boldsymbol{A}}_{s,\#}\boldsymbol{\eta}_t + \hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{\mu}_s - \hat{\boldsymbol{\mu}}_s) + \hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1}\\
&= \hat{\boldsymbol{A}}_{s,\#}\boldsymbol{\eta}_t + \hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{\mu}_s - \hat{\boldsymbol{\mu}}_s) + \hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\boldsymbol{F}_s^{\tilde{t}-1}\boldsymbol{\xi}_{[\tau_{s-1}T]}\\
&+ \hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\left(\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l-1}\right) + \hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\left(\left(\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\right)\boldsymbol{\mu}_s\right)
\end{aligned} \tag{B.68}$$

$$\begin{aligned}
\boldsymbol{g}_t^b &= ((\hat{\boldsymbol{A}}_{s,\#}\boldsymbol{\eta}_t) \odot \boldsymbol{v}_t) + ((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{\mu}_s - \hat{\boldsymbol{\mu}}_s)) \odot \boldsymbol{v}_t) + ((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1}) \odot \boldsymbol{v}_t)\\
&= \boldsymbol{g}_{t,A}^b + \boldsymbol{g}_{t,B}^b + \boldsymbol{g}_{t,C}^b,
\end{aligned} \tag{B.69}$$

$$\begin{aligned}
\boldsymbol{\eta}_t^b &= \hat{\boldsymbol{A}}_{s,\#}^{-1}\,((\hat{\boldsymbol{A}}_{s,\#}\boldsymbol{\eta}_t) \odot \boldsymbol{v}_t) + \hat{\boldsymbol{A}}_{s,\#}^{-1}\,((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{\mu}_s - \hat{\boldsymbol{\mu}}_s)) \odot \boldsymbol{v}_t) + \hat{\boldsymbol{A}}_{s,\#}^{-1}\,((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1}) \odot \boldsymbol{v}_t)\\
&= \boldsymbol{\eta}_{t,A}^b + \boldsymbol{\eta}_{t,B}^b + \boldsymbol{\eta}_{t,C}^b.
\end{aligned} \tag{B.70}$$

Finally, for a vector $\boldsymbol{o}$, we denote $\boldsymbol{o}^{(j_1:j_2)}$ its sub-vector with elements j_1 to j_2 selected in order, and for a matrix $\boldsymbol{O}$, we denote by $\boldsymbol{O}^{(j_1:j_2,j_1^*:j_2^*)}$ its sub-matrix consisting of rows j_1 to j_2 and columns j_1^* to j_2^*.

Proof of Lemma 10.

As for the proof of Lemma 9, consider the interval $I_i = \tilde{I}_s$. Let $\mathcal{S}_\dagger^b = \mathcal{S}_u$ or $\mathcal{S}_\dagger^b = \hat{\boldsymbol{\beta}}_{\boldsymbol{x},\#}$. We derive the asymptotic distribution of $T^{-1/2}\sum_{t\in\tilde{I}_s}\boldsymbol{z}_t^b\boldsymbol{g}_t^{b\prime}\mathcal{S}_\dagger^b$,

$$T^{-1/2}\sum_{t\in\tilde{I}_s}\boldsymbol{z}_t^b\boldsymbol{g}_t^{b\prime}\,\mathcal{S}_\dagger^b = \begin{bmatrix} T^{-1/2}\sum_{t\in\tilde{I}_s}\boldsymbol{g}_t^{b\prime}\,\mathcal{S}_\dagger^b\\ T^{-1/2}\sum_{t\in\tilde{I}_s}\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t\boldsymbol{g}_t^{b\prime}\,\mathcal{S}_\dagger^b\\ T^{-1/2}\sum_{t\in\tilde{I}_s}\mathcal{S}\boldsymbol{\xi}_{t-1}^b\boldsymbol{g}_t^{b\prime}\,\mathcal{S}_\dagger^b \end{bmatrix} \equiv \begin{bmatrix}\mathcal{E}_1^b\\ \mathcal{E}_2^b\\ \mathcal{E}_3^b\end{bmatrix}. \tag{B.71}$$

- Consider first $\mathcal{E}_1^b$. By (B.69),

$$\begin{aligned}
\mathcal{E}_1^b &= T^{-1/2}\sum_{t\in\tilde{I}_s}\boldsymbol{g}_t^{b\prime}\mathcal{S}_\dagger^b = T^{-1/2}\sum_{t\in I_i}\boldsymbol{g}_{t,A}^{b\prime}\,\mathcal{S}_\dagger^b + T^{-1/2}\sum_{t\in\tilde{I}_s}\boldsymbol{g}_{t,B}^{b\prime}\,\mathcal{S}_\dagger^b + T^{-1/2}\sum_{t\in\tilde{I}_s}\boldsymbol{g}_{t,C}^{b\prime}\,\mathcal{S}_\dagger^b\\
&= T^{-1/2}\sum_{t\in\tilde{I}_s}\mathcal{S}_\dagger^{b\prime}((\hat{\boldsymbol{A}}_{s,\#}\boldsymbol{\eta}_t) \odot \boldsymbol{v}_t) + T^{-1/2}\sum_{t\in\tilde{I}_s}\mathcal{S}_\dagger^{b\prime}((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{\mu}_s - \hat{\boldsymbol{\mu}}_s)) \odot \boldsymbol{v}_t)\\
&+ T^{-1/2}\sum_{t\in\tilde{I}_s}\mathcal{S}_\dagger^{b\prime}((\hat{\boldsymbol{A}}_{s,\#}(\boldsymbol{F}_s - \hat{\boldsymbol{F}}_s)\boldsymbol{\xi}_{t-1}) \odot \boldsymbol{v}_t)\\
&= \mathcal{E}_{1,1}^b + \mathcal{E}_{1,2}^b + \mathcal{E}_{1,3}^b.
\end{aligned}$$

For $\mathcal{S}_\dagger = \mathcal{S}_u$ we have $\mathcal{S}_\dagger^b = \mathcal{S}_u$. For $\mathcal{S}_\dagger = \hat{\boldsymbol{\beta}}_{\boldsymbol{x},\#}$ we have $\|\mathcal{S}_\dagger^b - \mathcal{S}_\dagger\| = o_p(1)$, $\|\hat{\boldsymbol{\mu}}_s - \boldsymbol{\mu}_s\| = o_p(1)$, $\|\hat{\boldsymbol{A}}_s - \boldsymbol{A}_s\| = o_p(1)$, $\|\hat{\boldsymbol{A}}_s^{-1} - \boldsymbol{A}_s^{-1}\| = o_p(1)$ and $\sum_{l=0}^{\infty}\|\hat{\boldsymbol{F}}_s^l - \hat{\boldsymbol{F}}_s^l\| = o_p(1)$, whenever a $O_p^b(1)$ term is written with the estimated quantities instead of the true one, the difference is $o_p^b(1)$, so asymptotically negligible. Therefore, we proceed in the rest of the proof by replacing the estimated parameters mentioned above with their true values, and denote the remainder by $o_p^b(1)$.

Using these replacements, one can show that $\mathcal{E}_{1,2}^b = o_p^b(1)$ and $\mathcal{E}_{1,3}^b = o_p^b(1)$. So, we have $\mathcal{E}_1^b = \mathcal{E}_{1,1}^b = \mathcal{S}_\dagger'\,T^{-1/2}\sum_{t\in\tilde{I}_s}((\boldsymbol{A}_{s,\#}\boldsymbol{\eta}_t) \odot \boldsymbol{v}_t) + o_p^b(1)$. Since $\boldsymbol{A}_{s,\#}\boldsymbol{\eta}_t = \boldsymbol{g}_t = \boldsymbol{\epsilon}_{t,\#}$, it follows that

$$\mathcal{E}_1^b = \mathcal{S}_\dagger'\,T^{-1/2}\sum_{t\in\tilde{I}_s}(\boldsymbol{g}_t \odot \boldsymbol{v}_t) + o_p^b(1). \tag{B.72}$$

First, let $\mathcal{S}_\dagger = \mathcal{S}_u$. Then $\mathcal{E}_1^b = T^{-1/2}\sum_{t\in\tilde{I}_s}u_t\nu_t + o_p^b(1) = T^{-1/2}\sum_{t\in\tilde{I}_s}d_{u,t}l_{u,t}\nu_t + o_p^b(1) = \mathcal{E}_{1,1}^b + o_p^b(1)$, where recall that $d_{u,t} = d_{1,t}$ and $l_{u,t}$ is the first element of $\boldsymbol{l}_t$.

We now derive the limiting distribution of $\mathcal{E}_{1,1}^b$, in two parts: in part (i), we show that Lemma 3 holds for $\tilde{\mathcal{E}}_{1,1}^b = T^{-1/2}\sum_{t=1}^{[Tr]}l_{u,t}\nu_t$, i.e. $\tilde{\mathcal{E}}_{1,1}^b \overset{d_p^b}{\Rightarrow} \boldsymbol{B}_0^{(1)}(r)$ in probability, where $\boldsymbol{B}_0^{(1)}(r)$ is the first element of $\boldsymbol{B}_0(r)$ defined just before Lemma 6; in part (ii), we show that the condition of Theorem 2.1 of Hansen (1992) holds, that is, the bootstrap unconditional variance of $\mathcal{E}_{1,1}^b$ converges in probability to the unconditional variance of $T^{-1/2}\sum_{t\in\tilde{I}_s}d_{u,t}l_{u,t} = \mathcal{E}_1$. Note that here

$\mathcal{E}_1 = \mathcal{S}_\dagger' \left(T^{-1/2} \sum_{t\in \tilde{I}_s} \boldsymbol{g}_t\right)$ is the sample equivalent of $\mathcal{E}_1^b$, defined in the proof of Lemma 6 in the Supplementary Appendix, Appendix C.

Part (i). First, ν_t is i.i.d, so conditional on the data, $\mathrm{E}^b(l_{u,t}\nu_t|\mathcal{F}_{t-1}^b) = l_{u,t}\,\mathrm{E}^b(\nu_t|\mathcal{F}_{t-1}^b) = 0$, so $l_{u,t}\nu_t$ is a m.d.s. Second, for some $C > 0$, $\sup_t \mathrm{E}(\mathrm{E}^b|l_{u,t}\nu_t|^{2+\delta^*}) \le \sup_t \mathrm{E}|l_{u,t}|^{2+\delta^*} \sup_t \mathrm{E}^b|\nu_t|^{2+\delta^*} < C$ by Assumption 9(iii) and Assumption 10(ii), so $\sup_t \mathrm{E}^b|l_{u,t}\nu_t|^{2+\delta^*} < o_p(1) + C$. Third, by Lemma 8(iv) and Assumption 10(ii),

$$\mathrm{E}^b(\mathcal{E}_{1,1}^b)^2 = T^{-1}\sum_{t=1}^{[Tr]} \mathrm{E}^b(l_{u,t}^2\nu_t^2) = T^{-1}\sum_{t=1}^{[Tr]} l_{u,t}^2 \xrightarrow{p} r \text{ (uniformly in } r).$$

Fourth, because $\mathrm{E}^b(\nu_t^2|\mathcal{F}_{t-1}^b) = \mathrm{E}^b(\nu_t^2) = 1$, the conditional and unconditional bootstrap second moments are the same, so $\mathrm{E}^b[(\mathcal{E}_{1,1}^b)^2|\mathcal{F}_{t-1}^b] - \mathrm{E}^b(\mathcal{E}_{1,1}^b)^2 = 0$. This shows that $\tilde{\mathcal{E}}_{1,1}^b = T^{-1/2}\sum_{t=1}^{[Tr]} l_{u,t}\nu_t \overset{d_p^b}{\Rightarrow} \boldsymbol{B}_0^{(1)}(r)$ in probability (uniformly in r).

Part (ii). By Assumption 9(ii), $\mathrm{E}(d_{u,t}^2 l_{u,t}^2) = d_{u,t}^2$. Therefore, by Lemma 8(iv), uniformly in r,

$$\mathrm{E}^b(\mathcal{E}_{1,1}^b)^2 - \mathrm{E}(\mathcal{E}_1^2) = T^{-1}\sum_{t=1}^{[Tr]} [d_{u,t}^2 l_{u,t}^2 - \mathrm{E}(d_{u,t}^2 l_{u,t}^2)] \xrightarrow{p} 0.$$

Therefore, by Theorem 2.1 in Hansen (1992), $T^{-1/2}\sum_{t=1}^{[Tr]} d_{u,t} l_{u,t}\nu_t \overset{d_p^b}{\Rightarrow} \int_0^r d_u(\tau) d\boldsymbol{B}_0^{(1)}(\tau) = \mathbb{M}_1(\tau_{s-1},\tau_s)$ in probability, where $\mathbb{M}_1(\tau_{s-1},\tau_s)$ is defined just before Lemma 6. So for $\mathcal{S}_\dagger = S_u$, $\mathcal{E}_1^b \overset{d_p^b}{\Rightarrow} \mathbb{M}_1(\tau_{s-1},\tau_s)$ in probability.

Now let $\mathcal{S}_\dagger = \boldsymbol{\beta}_{x,\#}$ and note that $\mathcal{E}_1^b = \boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime} T^{-1/2}\sum_{t\in\tilde{I}_s} \boldsymbol{v}_t\nu_t$. Recall that by the decomposition of $\boldsymbol{S}$ and a decomposition of $\boldsymbol{D}_t$ exactly as $\boldsymbol{D}(\tau)$ in (B.40), we have:

$$\boldsymbol{g}_t \odot \boldsymbol{\nu}_t = \boldsymbol{\epsilon}_{t,\#} \odot \boldsymbol{\nu}_t = (\boldsymbol{S}\boldsymbol{D}_t\boldsymbol{l}_t)_\# \odot \boldsymbol{\nu}_t = \left[\mathbf{vect}(d_{u,t}l_{u,t}\nu_t\,,\ \boldsymbol{s}_{p_1} d_{u,t}l_{u,t}\nu_t + \boldsymbol{S}_{p_1}\boldsymbol{D}_{v,t}\boldsymbol{l}_{v,t}\nu_t\,,\ \boldsymbol{S}_{p_2}\boldsymbol{l}_{\zeta,t})\right]_\#, \tag{B.73}$$

so $\mathcal{E}_1^b = \boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime} T^{-1/2}\sum_{t\in\tilde{I}_s} \boldsymbol{v}_t\nu_t = (\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{s}_{p_1})\left(T^{-1/2}\sum_{t\in\tilde{I}_s} d_{u,t}l_{u,t}\nu_t\right) + (\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{S}_{p_1})\left(T^{-1/2}\sum_{t\in\tilde{I}_s} \boldsymbol{D}_{v,t}\boldsymbol{l}_{v,t}\nu_t\right)$.

Because $\mathrm{E}(\boldsymbol{l}_{v,t}\boldsymbol{l}_{v,t}') = \boldsymbol{I}_{p_1}$, by similar arguments as for $T^{-1/2}\sum_{t=1}^{[Tr]} d_{u,t}l_{u,t}\nu_t \overset{d_p^b}{\Rightarrow} \int_0^r d_u(\tau)\mathrm{d}\boldsymbol{B}_0^{(1)}(\tau)$ in probability, it can be shown that $(\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{S}_{p_1})\left(T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{D}_{v,t}\boldsymbol{l}_{v,t}\nu_t\right) \overset{d_p^b}{\Rightarrow} (\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{S}_{p_1})\int_0^r \boldsymbol{D}_v(\tau)\mathrm{d}\boldsymbol{B}_0^{(2:p_1+1)}(\tau)$ in probability, where $\boldsymbol{B}_0^{(2:p_1+1)}(\cdot)$ refers to selecting elements $2 : (p_1+1)$ in order from $\boldsymbol{B}_0(\cdot)$. Moreover, because $u_t\nu_t$, $\boldsymbol{v}_t\nu_t$ share the same ν_t which is i.i.d and for which $\mathrm{E}^b(\nu_t^2) = 1$, $(\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{s}_{p_1})\left(T^{-1/2}\sum_{t=1}^{[Tr]} d_{u,t}l_{u,t}\nu_t\right)$ and $(\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{S}_{p_1})\left(T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{D}_{v,t}\boldsymbol{l}_{v,t}\nu_t\right)$ also jointly converge, and their unconditional bootstrap covariance converges to the unconditional covariance of their respective limits. Therefore, also for $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$,

$$\begin{aligned}\mathcal{E}_1^b &\overset{d_p^b}{\Rightarrow} (\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{s}_{p_1})\int_{\tau_{s-1}}^{\tau_s} d_u^2(\tau)\mathrm{d}\boldsymbol{B}_0^{(1)}(\tau) + (\boldsymbol{\beta}_{\boldsymbol{x}}^{0\prime}\boldsymbol{S}_{p_1})\int_{\tau_{s-1}}^{\tau_s} \boldsymbol{D}_v(\tau)\mathrm{d}\boldsymbol{B}_0^{(2:p_1+1)}(\tau)\\ &= (\mathcal{S}_\dagger'\boldsymbol{S}_\#)\int_{\tau_{s-1}}^{\tau_s} \boldsymbol{D}(\tau)\mathrm{d}\,\boldsymbol{B}_{0,\#}(\tau) = \mathbb{M}_1(\tau_{s-1},\tau_s)\end{aligned} \tag{B.74}$$

in probability with variance matrix $\boldsymbol{V}_{\mathbb{M}_1(\tau_{s-1},\tau_s)}$ given in the Supplementary Appendix, Appendix C, proof of Lemma 6.

- Next, consider $\mathcal{E}_3^b$. From (B.64) we have that: $\boldsymbol{\xi}_{t-1}^b = \hat{\boldsymbol{F}}_s^{\tilde{t}-1}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b + \sum_{l=0}^{\tilde{t}-2}\hat{\boldsymbol{F}}_s^l\boldsymbol{\eta}_{t-l-1}^b + \left(\sum_{l=0}^{\tilde{t}-2}\hat{\boldsymbol{F}}_s^l\right)\hat{\boldsymbol{\mu}}_s$. Define $\tilde{I}_s^- = [[\tau_{s-1}T]+2, [\tau_s T]]$. Then, replacing estimated parameters with the true ones and denoting the remainder by $o_p^b(1)$ for reasons discussed earlier,

$$\begin{aligned}\mathcal{E}_3^b &= T^{-1/2}\sum_{t\in\tilde{I}_s}\mathcal{S}\boldsymbol{\xi}_{t-1}^b\boldsymbol{g}_t^{b\prime}\,\mathcal{S}_\dagger\\ &= T^{-1/2}(\mathcal{S}_\dagger'\boldsymbol{g}_{[\tau_{s-1}T]+1}^b)(\mathcal{S}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b) + T^{-1/2}\sum_{t\in\tilde{I}_s^-}(\mathcal{S}_\dagger'\boldsymbol{g}_t^b)\left[\mathcal{S}\boldsymbol{F}_s^{\tilde{t}-1}\boldsymbol{\xi}_{[\tau_{s-1}T]}^b\right]\\ &+ T^{-1/2}\sum_{t\in\tilde{I}_s^-}(\mathcal{S}_\dagger'\boldsymbol{g}_t^b)\left[\mathcal{S}\left(\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\right)\boldsymbol{\mu}_s\right]\\ &+ T^{-1/2}\sum_{t\in\tilde{I}_s^-}(\mathcal{S}_\dagger'\boldsymbol{g}_t^b)\left[\mathcal{S}\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l-1}^b\right] + o_p^b(1) = \mathcal{E}_{3,1}^b + \mathcal{E}_{3,2}^b + \mathcal{E}_{3,3}^b + \mathcal{E}_{3,4}^b + o_p^b(1).\end{aligned} \tag{B.75}$$

First, note that by (B.48), $\|\boldsymbol{\xi}^b_{[\tau_{s-1}T]}\| = O^b_p(T^{-\alpha})$, and also that $\|\boldsymbol{g}^b_{[\tau_{s-1}T]+1}\| = O^b_p(T^{-\alpha})$, for any $\alpha > 0$. Therefore, $\mathcal{E}^b_{3,1} = o^b_p(1)$. For the same reason and by the fact that $\|\boldsymbol{F}^l_s\|$ is exponentially decaying with l,

$$\|\mathcal{E}^b_{3,2}\| \le \|T^{-1/2}\sum_{t\in \tilde{I}^-_s}\mathcal{S}'_\dagger \boldsymbol{g}^b_t\| \left(\|\mathcal{S}\| \sup_l \|\boldsymbol{F}^l_s\| \;\; \|\boldsymbol{\xi}^b_{[\tau_{s-1}T]}\|\right) = \|\mathcal{E}^b_1\|\left(\|\mathcal{S}\| \sup_l \|\boldsymbol{F}^l_s\|\right) o^b_p(1) = o^b_p(1).$$

Next, note that by similar derivations as for (C.13) in Supplementary Appendix, Appendix C, and artificially setting $\boldsymbol{g}_{t-l} = 0$, $\boldsymbol{v}_{t-l} = 0$ for all $t < l$ (as in Boswijk et al. (2016)) we have, for $\tilde{n} = [\tau_{s-1}T] + 2$,

$$\begin{aligned}\mathcal{E}^b_{3,4} &= \sum_{l=1}^{\Delta\tau_s T-2} \mathcal{S}\boldsymbol{F}^l_s \left(T^{-1/2}\sum_{t\in\tilde{I}^-_s}(\mathcal{S}'_\dagger \boldsymbol{g}^b_t)\boldsymbol{\eta}^b_{t-l-1}\right)\\ &\quad - T^{-1/2}\sum_{l=1}^{\Delta\tau_s T-2}\mathcal{S}\boldsymbol{F}^l_s \sum_{j=0}^{l-1}(\mathcal{S}'_\dagger \boldsymbol{g}^b_{\tilde{n}+j})\boldsymbol{\eta}^b_{\tilde{n}+j-(l+1)} \equiv \tilde{\mathcal{E}}^b_{3,4}(\Delta\tau_s T-2) - \mathcal{L}.\end{aligned} \tag{B.76}$$

We now show that $\mathcal{L} = o^b_p(1)$. Let $\mathcal{S}'_\dagger \boldsymbol{g}_t = u_t$. Then,

$$(\mathcal{S}'_\dagger \boldsymbol{g}_t \nu_t)(\boldsymbol{g}_{t-l}\odot \boldsymbol{v}_{t-l})(\boldsymbol{g}_{t-l}\odot \boldsymbol{v}_{t-l})'(\mathcal{S}'_\dagger \boldsymbol{g}_t \nu_t) = \nu_t^2 u_t^2 \begin{bmatrix} u^2_{t-l}\nu^2_{t-l} & u_{t-l}\boldsymbol{v}'_{t-l}\nu^2_{t-l} & u_{t-l}\boldsymbol{\zeta}'_{t-l}\nu_{t-l}\\ (u_{t-l}\boldsymbol{v}'_{t-l}\nu^2_{t-l})' & \boldsymbol{v}_{t-l}\boldsymbol{v}'_{t-l}\nu^2_{t-l} & \boldsymbol{v}_{t-l}\boldsymbol{\zeta}'_{t-l}\nu_{t-l}\\ (u_{t-l}\boldsymbol{\zeta}'_{t-l}\nu_{t-l})' & (\boldsymbol{v}_{t-l}\boldsymbol{\zeta}'_{t-l}\nu_{t-l})' & \boldsymbol{\zeta}_{t-l}\boldsymbol{\zeta}'_{t-l}\end{bmatrix}_\#$$

$$= \begin{bmatrix} u^2_t u^2_{t-l} & u^2_t u_{t-l}\boldsymbol{v}'_{t-l} & u^2_t u_{t-l}\boldsymbol{\zeta}'_{t-l}\\ u^2_t(u_{t-l}\boldsymbol{v}'_{t-l})' & u^2_t \boldsymbol{v}_{t-l}\boldsymbol{v}'_{t-l} & u^2_t\boldsymbol{v}_{t-l}\boldsymbol{\zeta}'_{t-l}\\ u^2_t(u_{t-l}\boldsymbol{\zeta}'_{t-l})' & (u^2_t\boldsymbol{v}_{t-l}\boldsymbol{\zeta}'_{t-l})' & u^2_t\boldsymbol{\zeta}_{t-l}\boldsymbol{\zeta}'_{t-l}\end{bmatrix}_\# \odot \begin{bmatrix} \nu^2_t\nu^2_{t-l} & (\nu^2_t\nu^2_{t-l})\boldsymbol{\iota}'_{p_1} & (\nu^2_t\nu_{t-l})\boldsymbol{\iota}'_{p_2}\\ (\nu^2_t\nu^2_{t-l})\boldsymbol{\iota}_{p_1} & (\nu^2_t\nu^2_{t-l})\boldsymbol{J}_{p_1} & (\nu^2_t\nu_{t-l})\boldsymbol{\iota}_{p_1}\boldsymbol{\iota}'_{p_2}\\ (\nu^2_t\nu_{t-l})\boldsymbol{\iota}_{p_2} & (\nu^2_t\nu_{t-l})\boldsymbol{\iota}_{p_2}\boldsymbol{\iota}'_{p_1} & \nu^2_t\boldsymbol{J}_{p_2}\end{bmatrix}_\#.$$

Therefore, for $l \ge 1$,

$$\mathrm{E}(\mathrm{E}^b[(\mathcal{S}'_\dagger \boldsymbol{g}_t\nu_t)(\boldsymbol{g}_{t-l}\odot\boldsymbol{v}_{t-l})(\boldsymbol{g}_{t-l}\odot\boldsymbol{v}_{t-l})'(\mathcal{S}'_\dagger\boldsymbol{g}_t\nu_t)]) = \mathrm{E}(u^2_t((\boldsymbol{g}_{t-l}\boldsymbol{g}'_{t-l})\odot\mathcal{J})). \tag{B.77}$$

By Assumption 9, the non-zero elements of $\mathrm{E}(u^2_t\otimes(\boldsymbol{g}_{t-l}\boldsymbol{g}'_{t-l}))\odot\mathcal{J}$, do not depend on t, and are elements of linear functions $\boldsymbol{\rho}_{0,l}$, so they are uniformly bounded in l. Therefore, for element $\mathcal{L}^{(a,b)}$ of the matrix $\mathcal{L}$, and constants $c > 0$, $c_1 > 0$,

$$\begin{aligned}&\sup_{\tilde{n}+j}\mathrm{E}(\mathrm{E}^b|\mathcal{L}^{(a,b)}|)\\ &\le T^{-1/2}\sum_{l=1}^{\Delta\tau_s T-2}|(\mathcal{S}\boldsymbol{F}^l_s\boldsymbol{A}^{-1}_s)^{(a,b)}|\sum_{j=0}^{l-1}\sup_{\tilde{n}+j}\mathrm{E}\,\mathrm{E}^b|\{(\mathcal{S}^{b\prime}_\dagger\boldsymbol{g}_{\tilde{n}+j})\nu_{\tilde{n}+j}\,[\boldsymbol{g}_{\tilde{n}+j-(l+1)}\odot\boldsymbol{v}_{\tilde{n}+j-(l+1)}]\}^{(a,b)}|\\ &\le T^{-1/2}\sum_{l=1}^{\Delta\tau_s T-2}\|\boldsymbol{A}^{-1}_s\|\;\|\mathcal{S}\|\;\|\boldsymbol{F}^l_s\|\sum_{j=0}^{l-1}c \le c_1 T^{-1/2}\sum_{l=0}^{\infty} l\|\boldsymbol{F}^l_s\| \to 0.\end{aligned}$$

Therefore, $\mathcal{L} = o^b_p(1)$ for $\mathcal{S}_\dagger = \mathcal{S}_u$, and by similar arguments, $\mathcal{L} = o^b_p(1)$ for $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$.

Next, we analyze $\tilde{\mathcal{E}}^b_{3,4}(\Delta_{\tau_s}T-2)$. To that end, let for now $\mathcal{S}_\dagger = \mathcal{S}_u$, and note that a crucial term in $\tilde{\mathcal{E}}^b_{3,4}(\Delta_{\tau_s}T-2)$ is $\mathcal{L}^b_1(l) = T^{-1/2}\sum_{t=1}^{[Tr]}u_t\nu_t(\boldsymbol{g}_{t-l}\odot\boldsymbol{v}_{t-l})$ for $l \ge 1$. We can write $\boldsymbol{S}$ and $\boldsymbol{D}(\tau)$ in (B.40) as $\boldsymbol{S} = \mathbf{diag}(\boldsymbol{S}_{p_1+1}, \boldsymbol{S}_{p_2})$, $\boldsymbol{D}(\tau) = \mathbf{diag}(\boldsymbol{D}_{p_1+1}(\tau), \boldsymbol{D}_{\zeta}(\tau))$, where $\boldsymbol{S}_{p_1+1} = \begin{bmatrix} 1 & \boldsymbol{0}_{1\times p_1}\\ \boldsymbol{s}_{p_1} & \boldsymbol{S}_{p_1}\end{bmatrix}$, $\boldsymbol{D}_{p_1+1}(\tau) = \mathbf{diag}(d_u(\tau), \boldsymbol{D}_v(\tau))$. Then, we have:

$$\boldsymbol{g}_{t-l}\odot\boldsymbol{v}_{t-l} = \begin{bmatrix} d_{u,t-l}l_{u,t-l}\nu_{t-l}\\ \boldsymbol{s}_{p_1}d_{u,t}l_{u,t-l}\nu_{t-l} + \boldsymbol{S}_{p_1}\boldsymbol{D}_{v,t}\boldsymbol{l}_{v,t-l}\nu_{t-l}\\ \boldsymbol{S}_{p_2}\boldsymbol{D}_{\zeta,t-l}\boldsymbol{l}_{\zeta,t-l}\end{bmatrix}_\# = \begin{bmatrix}\boldsymbol{S}_{p_1+1}\boldsymbol{D}_{p_1+1,t-l}\boldsymbol{n}_{t-l}\nu_{t-l}\\ \boldsymbol{S}_{p_2}\boldsymbol{D}_{\zeta,t-l}\boldsymbol{l}_{\zeta,t-l}\end{bmatrix}_\#,$$

where $\boldsymbol{n}_t$ is defined in Assumption 9′. Letting $\mathcal{E}^{(1)}_{t,l} = l_{u,t}\boldsymbol{n}_{t-l}$, $\mathcal{E}^{(2)}_{t,l} = l_{u,t}\boldsymbol{l}_{\zeta,t-l}$ and $\mathcal{E}^{(1),b}_{t,l} = l_{u,t}\boldsymbol{n}_{t-l}\nu_t\nu_{t-l}$, $\mathcal{E}^{(2),b}_{t,l} = l_{u,t}\boldsymbol{l}_{\zeta,t-l}\nu_t$, we have:

$$\mathcal{L}^b_1(l) = T^{-1/2}\sum_{t=1}^{[Tr]}u_t\nu_t(\boldsymbol{g}_{t-l}\odot\boldsymbol{v}_{t-l}) = \begin{bmatrix} d_{u,t}\boldsymbol{S}_{p_1+1}\boldsymbol{D}_{p_1+1,t-l}\mathcal{E}^{(1),b}_{t,l}\\ d_{u,t}\boldsymbol{S}_{p_2}\boldsymbol{D}_{\zeta,t-l}\mathcal{E}^{(2),b}_{t,l}\end{bmatrix}_\#. \tag{B.78}$$

We now proceed as for $\mathcal{E}^b_1$, in two parts: in part (i), we derive the limiting distribution of $\boldsymbol{B}^{(i),b}_{l,T,A}(r) = T^{-1/2}\sum_{t=1}^{[Tr]}\mathcal{E}^{(i),b}_{t,l}$ $(i = 1, 2)$ and its equivalent for $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$ for each l, by verifying Lemma 3 (we verify this for both definitions of $\mathcal{S}_\dagger$ and therefore replace $\mathcal{E}^{(i),b}_{t,l}$ with the appropriate quantities when $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$); in part (ii), we derive the limiting distribution of $\tilde{\mathcal{E}}^b_{3,4}(n^*)$ using Theorem 2.1 in Hansen (1992) for fixed n^*. Then we take the limit as $n^* \to \infty$.

Part (i). Let $\mathcal{S}_{\dagger} = \mathcal{S}_u$. First, we apply Lemma 3 to $\boldsymbol{B}^{(i),b}_{l,T,A}(r) = T^{-1/2}\sum_{t=1}^{[Tr]}\mathcal{E}^{(i),b}_{t,l}$, for $l \geq 1$, $i = 1,2$ where note that even though $\boldsymbol{V}_{\boldsymbol{B}^{(i,),b}_{l,T,A}(r)} = \text{plim}_{T\to\infty}\text{Var}^b(\boldsymbol{B}^{(i),b}_{l,T,A}(r))$ does not converge to $r\boldsymbol{I}_j$, $j = 1+p_1, p_2$ for $i = 1,2$ respectively, as one condition in Lemma 3 requires, it is symmetric and positive semi-definite, so by a decomposition of $\boldsymbol{V}_{\boldsymbol{B}^{(i),b}_{l,T,A}} = \boldsymbol{E}_i^{1/2}\boldsymbol{E}_i^{1/2'}$, $\boldsymbol{E}_i^{-1/2}\boldsymbol{B}^{(i),b}_{l,T,A}(r)$ converges to a process whose limiting variance is r times the identity matrix, where $\boldsymbol{E}_i^{-1}$ is the generalized inverse. Therefore, in the rest of the analysis, we no longer need to verify this condition, except for deriving the limit of the unconditional bootstrap variance, and proceed to verify the rest of the conditions. First, $\text{E}^b(\mathcal{E}^{(1),b}_{t,l}) = \mathcal{E}^{(1)}_{t,l}\,\text{E}^b(\nu_t\nu_{t-l}|\mathcal{F}^b_{t-1}) = \boldsymbol{0}_{p_1+1}$, $\text{E}^b(\mathcal{E}^{(2),b}_{t,l}) = \mathcal{E}^{(2)}_{t,l}\,\text{E}^b(\nu_t|\mathcal{F}^b_{t-1}) = \boldsymbol{0}_{p_2}$, so $\mathcal{E}^{(i),b}_{t,l}$ is a m.d.s. Second, for $\phi^{(i),b}_t$ denoting a typical element of $\mathcal{E}^{(i),b}_{t,l}$, and $\phi^{(i)}_t$ denoting the corresponding element of $\mathcal{E}^{(i)}_{t,l}$, we have, for some $\delta^* > 0$, that $\sup_t \text{E}(\text{E}^b|\phi^{(1),b}_t|^{2+\delta^*}) = \sup_t \text{E}(|\phi^{(1)}_t|^{2+\delta^*}\sup_t \text{E}^b|\nu_t\nu_{t-l}|^{2+\delta^*}) < \infty$ by Markov's inequality, Assumption 9(iii) and Assumption 10(ii) for $\mathcal{E}^{(1),b}_{t,l}$, or we have that $\sup_t \text{E}(\text{E}^b|\phi^{(2),b}_t|^{2+\delta^*}) = \sup_t \text{E}|\phi^{(2)}_t|^{2+\delta^*}\sup_t \text{E}^b|\nu_t|^{2+\delta^*} < \infty$ by the same assumptions, for $\mathcal{E}^{(2),b}_{t,l}$.

Third, to facilitate showing that $\text{Var}^b(\boldsymbol{B}^{(i),b}_{l,T,A}(r)|\mathcal{F}^b_{t-1}) - \text{Var}^b(\boldsymbol{B}^{(i),b}_{l,T,A}(r)) \xrightarrow{p} 0$, note that, from (B.77) we have, $\text{Var}^b(\mathcal{E}^{(1),b}_{t,l}) = (\mathcal{E}^{(1)}_{t,l}\mathcal{E}^{(1)'}_{t,l})$, $\text{Var}^b(\mathcal{E}^{(1),b}_{t,l}|\mathcal{F}^b_{t-1}) = \mathcal{E}^{(1)}_{t,l}\nu^2_{t-l}$, $\text{Var}^b(\mathcal{E}^{(2),b}_{t,l}) = \text{Var}^b(\mathcal{E}^{(2),b}_{t,l}|\mathcal{F}^b_{t-1}) = (\mathcal{E}^{(2)}_{t,l}\mathcal{E}^{(2)'}_{t,l})$. Therefore, we have by Lemma 8,

$$\text{Var}^b(\boldsymbol{B}^{(1),b}_{l,T,A}(r)) = T^{-1}\sum_{t=1}^{[Tr]}(l^2_{u,t}\boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l}) \xrightarrow{p} \text{Var}(\boldsymbol{B}_l(r)^{1:p_1+1}) = \boldsymbol{\rho}^{(1:p_1+1,1:p_1+1)}_{l,l},$$
$$\text{Var}^b(\boldsymbol{B}^{(2),b}_{l,T,A}(r)) = T^{-1}\sum_{t=1}^{[Tr]}(l^2_{u,t}\boldsymbol{l}_{\zeta,t-l}\boldsymbol{l}'_{\zeta,t-l}) \xrightarrow{p} \text{Var}(\boldsymbol{B}_l(r)^{p_1+2:n}) = \boldsymbol{\rho}^{(p_1+2:n,p_1+2:n)}_{l,l}, \tag{B.79}$$

where $\boldsymbol{B}_l(r)$ was defined just before Lemma 5, $\boldsymbol{B}_l(r)^{(1:p_1+1)}$, $\boldsymbol{B}_l(r)^{(p_1+2:n)}$ are the vectors stacking elements $1:p_1+1$, and $p_1+2:n$ respectively of $\boldsymbol{B}_l(r)$ in order, and $\boldsymbol{\rho}^{(1:p_1+1,1:p_1+1)}_{l,l}$, $\boldsymbol{\rho}^{(p_1+2:n,p_1+2:n)}_{l,l}$ are the left upper $p_1+1\times p_1+1$, $p_2\times p_2$ respectively, blocks of $\boldsymbol{\rho}_{l,l}$.

Regarding the last condition in Lemma 3, notice that this is satisfied when $i = 2$ (the conditional and unconditional bootstrap moments are the same). To verify the last condition in Lemma 3 for $i = 1$, because ν^2_{t-l} is i.i.d. and $\sup_t \text{E}^b|\nu_t|^{4+\delta^*} < \infty$, by Lemma 1 and Lemma 8(iv), we have:

$$\text{Var}^b(\boldsymbol{B}^{(1)b}_{l,T,A}(r)|\mathcal{F}^b_{t-1}) - \text{Var}^b(\boldsymbol{B}^{(1),b}_{l,T,A}(r)) = T^{-1}\sum_{t=1}^{[Tr]}(l^2_{u,t}\boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l})(\nu^2_{t-l}-1) = o^b_p(1)$$

by Chebyshev inequality since for any $C > 0$, we have $P^b(\|T^{-1}\sum_{t=1}^{[Tr]}(l^2_{u,t}\boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l})(\nu^2_{t-l}-1)\| \geq C) \leq C^{-2}\,T^{-1}\|T^{-1}\sum_{t,t^*=1}^{[Tr]} l^2_{u,t}l^2_{u,t^*}\boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l}\boldsymbol{n}_{t^*-l}\boldsymbol{n}'_{t^*-l}\|\,\text{E}^b|(\nu^2_{t-l}-1)(\nu^2_{t^*-l}-1)| \xrightarrow{p} 0$, where $\text{E}^b|(\nu^2_{t-l}-1)(\nu^2_{t^*-l}-1)| < \infty$ by Assumption 10(ii) and $\|T^{-1}\sum_{t,t^*=1}^{[Tr]} l^2_{u,t}l^2_{u,t^*}\boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l}\boldsymbol{n}_{t^*-l}\boldsymbol{n}'_{t^*-l}\| = O_p(1)$ by Assumption 9′(iv) which requires the existence of the moments of 8th order for $\boldsymbol{l}_t$. Notice that Assumption 9′(iv) is only needed for the WR bootstrap, but not for the WF bootstrap for which Assumption 9(iii) (which requires moments of 4th order only) is enough (as in Boswijk et al. (2016)) since $\boldsymbol{z}_t$ is held fixed when y^b_t and $\boldsymbol{x}^b_t$ are generated (see Section 2.3). Moreover, notice that for the WR bootstrap, if ν_t is *i.i.d.* from the Rademacher distribution then $\text{Var}^b(\boldsymbol{B}^{(1)b}_{l,T,A}(r)|\mathcal{F}^b_{t-1}) - \text{Var}^b(\boldsymbol{B}^{(1),b}_{l,T,A}(r)) = 0$ (because $\nu_t = \pm 1$) and therefore Assumption 9′(iv) is not needed.

Therefore,

$$\boldsymbol{B}^{(1),b}_{l,T,A}(r) = T^{-1/2}\sum_{t=1}^{[Tr]}\mathcal{E}^{(1),b}_{t,l} \overset{d^b_p}{\Rightarrow} \boldsymbol{B}^{(1:p_1+1)}_l(r), \tag{B.80}$$
$$\boldsymbol{B}^{(2),b}_{l,T,A}(r) = T^{-1/2}\sum_{t=1}^{[Tr]}\mathcal{E}^{(2),b}_{t,l} \overset{d^b_p}{\Rightarrow} \boldsymbol{B}^{(p_1+2:n)}_l(r) \tag{B.81}$$

in probability. Moreover, because $\boldsymbol{B}^{(1),b}_{l,T,A}(r)$ and $\boldsymbol{B}^{(2),b}_{l,T,A}(r)$ share the same ν_t (which is i.i.d.), they also jointly converge weakly in probability: $\boldsymbol{B}^b_{l,T,A}(r) = \textbf{vect}(\boldsymbol{B}^{(1),b}_{l,T,A}(r), \boldsymbol{B}^{(2),b}_{l,T,A}(r)) = T^{-1/2}\sum_{t=1}^{[Tr]} l_{u,t}\nu_t(\boldsymbol{l}_{t-l}\odot\tilde{\boldsymbol{\nu}}_{t-l}) \overset{d^b_p}{\Rightarrow} \boldsymbol{B}^{(1:n)}_l(r)$ in probability. Notice that so far, all the proofs went through using Assumptions 1–10. The joint convergence above requires Assumption 9′(iii) which imposes the block diagonal structure for $\rho^{(1:n,1:n)}_{l,l}$ (equivalently, it imposes $\text{E}[l^2_{u,t}\boldsymbol{n}_{t-l}\boldsymbol{l}'_{\zeta,t-l}] = \boldsymbol{0}_{p_2\times p_2}$ for $l \geq 1$ and $\boldsymbol{n}_t = \textbf{vect}(l_{u,t}, \boldsymbol{l}_{\upsilon,t})$.

Now let $\mathcal{S}_{\dagger} = \boldsymbol{\beta}_{\boldsymbol{x},\#}$. Then from (B.40):

$$\mathcal{L}^b_2(l) \equiv T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{\beta}'_{\boldsymbol{x}}\nu_t\nu_t(\boldsymbol{g}_{t-l}\odot\boldsymbol{\nu}_{t-l}) = T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{\beta}'_{\boldsymbol{x}}\boldsymbol{v}_t\nu_t\begin{bmatrix}\boldsymbol{S}_{p_1+1}\boldsymbol{D}_{p_1+1,t-l}\boldsymbol{n}_{t-l}\nu_{t-l}\\ \boldsymbol{S}_{p_2}\boldsymbol{D}_{\zeta,t-l}\boldsymbol{l}_{\zeta,t-l}\end{bmatrix}_{\#}$$

$$= (\beta_x' s_{p_1}) T^{-1/2} \sum_{t=1}^{[Tr]} d_{u,t}\, l_{u,t}\, \nu_t \begin{bmatrix} S_{p_1+1} D_{p_1+1,t-l} n_{t-l} \nu_{t-l} \\ S_{p_2} D_{\zeta,t-l} l_{\zeta,t-l} \end{bmatrix}_{\#}$$

$$+ T^{-1/2} \sum_{t=1}^{[Tr]} \beta_x' S_{p_1} D_{v,t} l_{v,t} \nu_t \begin{bmatrix} S_{p_1+1} D_{p_1+1,t-l} n_{t-l} \nu_{t-l} \\ S_{p_2} D_{\zeta,t-l} l_{\zeta,t-l} \end{bmatrix}_{\#} \quad (B.82)$$

$$= (\beta_x' s_{p_1}) \mathcal{L}_1^b(l) + T^{-1/2} \sum_{t=1}^{[Tr]} \begin{bmatrix} \beta_x' S_{p_1} D_{v,t} l_{v,t} S_{p_1+1} D_{p_1+1,t-l} n_{t-l} \nu_{t-l} \nu_t \\ \beta_x' S_{p_1} D_{v,t} l_{v,t} S_{p_2} D_{\zeta,t-l} l_{\zeta,t-l} \nu_t \end{bmatrix}_{\#}$$

$$= (\beta_x' s_{p_1}) \mathcal{L}_1^b(l) + \begin{bmatrix} (\beta_x' \otimes I_{p_1+1})(S_{p_1} \otimes S_{p_1+1}) T^{-1/2} \sum_{t=1}^{[Tr]} (D_{v,t} \otimes D_{p_1+1,t-l})(l_{v,t} \otimes n_{t-l}) \nu_t \nu_{t-l} \\ (\beta_x' \otimes I_{p_2})(S_{p_1} \otimes S_{p_2}) T^{-1/2} \sum_{t=1}^{[Tr]} (D_{v,t} \otimes D_{\zeta,t-l})(l_{v,t} \otimes l_{\zeta,t-l}) \nu_t \end{bmatrix}_{\#}$$

$$= (\beta_x' s_{p_1}) \mathcal{L}_1^b(l) + \begin{bmatrix} \mathcal{L}_{2,A}^b \\ \mathcal{L}_{2,B}^b \end{bmatrix} = \mathcal{L}_{2,1}^b + \mathcal{L}_{2,2}^b. \quad (B.83)$$

The distribution of $\mathcal{L}_{2,1}^b = (\beta_x' s_{p_1}) \mathcal{L}_1^b(l)$ follows from the joint convergence from (B.80)–(B.81) and part (ii) below. Following similar steps as for $\mathcal{L}_1^b(l)$ above (where $\mathcal{S}_\dagger = \mathcal{S}_u$), it can be shown that under Assumptions 1–10:

$$B_{l,T,B}^{(1),b}(r) \equiv T^{-1/2} \sum_{t=1}^{[Tr]} (l_{v,t} \otimes n_{t-l}) \nu_t \nu_{t-l} \overset{d_p^b}{\Rightarrow} \mathbf{vect}(B_l^{(n+1:n+p_1+1)}(r), \ldots, B_l^{(np_1+1:np_1+p_1+1)}(r)), \quad (B.84)$$

$$B_{l,T,B}^{(2),b}(r) \equiv T^{-1/2} \sum_{t=1}^{[Tr]} (l_{v,t} \otimes l_{\zeta,t-l}) \nu_t \overset{d_p^b}{\Rightarrow} \mathbf{vect}(B_l^{(n+p_1+2:2n)}(r), \ldots, B_l^{(np_1+p_1+2:n(p_1+1))}(r)), \quad (B.85)$$

in probability. And by Assumption 10(i) and Assumption 9′ (iii) which imposes $\mathrm{E}[b_t n_{t-l} l_{\zeta,t-l}'] = \mathbf{0}_{p_1 \times p_2}$ for b_t being any element of $l_{v,t} l_{v,t}'$, we have:

$$B_{l,T,B}^b(r) = \mathbf{vect}(B_{l,T,B}^{(1),b}(r), B_{l,T,B}^{(2),b}(r)) \overset{d_p^b}{\Rightarrow} B_l^{(n+1:n(p_1+1))}(r) \text{ in probability.} \quad (B.86)$$

Next we have to verify the second condition of Lemma 3 for $B_{l,T}^* \equiv \mathbf{vect}(B_{l,T,A}^b(r), B_{l,T,B}^b(r))$. To that end, define for a matrix O whose rows and columns are multiples of n, the operation

$$\mathbf{block}_{\kappa,\kappa^*}(O) = O^{(n(\kappa-1)+1:n\kappa,\, n(\kappa^*-1)+1:n\kappa^*)}, \quad (B.87)$$

that is, the operation that selects the (κ, κ^*) $n \times n$ sub-matrix of the matrix O. Also, for a $\bar{n} \times \bar{n}$ square matrix O_1, define the operation that makes O_1 block diagonal at row j as follows:

$$\mathbf{blockdiag}_j(O_1) = \mathbf{diag}(O_1^{(1:j,1:j)}, O_1^{(j+1:\bar{n},j+1:\bar{n})}). \quad (B.88)$$

Then, by similar arguments to $\mathcal{S}_\dagger = \mathcal{S}_u$,

$$\mathbf{vect}(B_{l,T,A}^b(r), B_{l,T,B}^b(r)) \overset{d_p^b}{\Rightarrow} B_l^{(1:n(p_1+1))}(r), \quad (B.89)$$

in probability, where the relevant bootstrap condition we have to show, by analogy to $\mathcal{S}_\dagger = \mathcal{S}_u$, is that $\mathrm{E}^b(B_{l,T,A}^b(r_1) B_{l,T,B}^b(r_2)') - \min(r_1, r_2)\,\mathrm{E}(B_l^{(1:n)}(r_1)(B_l^{(n+1:n(p_1+1))}(r_2))') = o_p(1)$, a condition proven below for $r_1 = r_2 = r$, because when $r_1 \neq r_2$, the proof follows in a similar fashion. Recalling that $B_{l,T}^* \equiv \mathbf{vect}(B_{l,T,A}^b(r), B_{l,T,B}^b(r))$, note that

$$\mathrm{E}^b(B_{l,T,A}^b(r) B_{l,T,B}^{b\prime}(r)) = [\mathbf{block}_{1,2}(\mathrm{Var}^b(B_{l,T}^*)), \mathbf{block}_{1,3}(\mathrm{Var}^b(B_{l,T}^*)), \ldots, \mathbf{block}_{1,p_1+1}(\mathrm{Var}^b(B_{l,T}^*))],$$

so we proceed with each block $2, \ldots, p_1 + 1$, and let $b_t = l_{u,t} l_{v_\kappa,t}$, where l_{v_κ} is the element κ of $l_{v,t}$, for $\kappa = 1, \ldots, p_1$. Then for $l \geq 1$, we have:

$$\mathbf{block}_{1,\kappa+1} \mathrm{Var}^b(B_{l,T}^*) = T^{-1} \sum_{t=1}^{[Tr]} \mathrm{E}^b \begin{bmatrix} b_t n_{t-l} n_{t-l}' \nu_t^2 \nu_{t-l}^2 & b_t n_{t-l} l_{\zeta,t-l}' \nu_t^2 \nu_{t-l} \\ b_t l_{\zeta,t-l} n_{t-l}' \nu_t^2 \nu_{t-l} & b_t l_{\zeta,t-l} l_{\zeta,t-l}' \nu_t^2 \end{bmatrix}$$

$$= T^{-1} \sum_{t=1}^{[Tr]} \mathbf{blockdiag}_{p_1+1}(b_t l_{t-l} l_{t-l}') \overset{p^b}{\to} \mathbf{blockdiag}_{p_1+1}(\mathbf{block}_{1,\kappa+1}(\rho_{l,l})) = \mathbf{block}_{1,\kappa+1}(\rho_{l,l}),$$

where the last equality follows by Lemma 8(iv) and by Assumption 9′(iii) which imposes a diagonal structure on $\rho_{l,l}^{(1:n,n(\kappa-1)+1:n\kappa)}$. Finally, we have:

$$\mathbf{block}_{1,\kappa+1}\mathrm{Var}^b(B^*_{l,T}|\mathcal{F}^b_{t-1}) - \mathbf{block}_{1,\kappa+1}\mathrm{Var}^b(B^*_{l,T}) = T^{-1}\sum_{t=1}^{[Tr]}\begin{bmatrix} b_t\boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l}(\nu^2_{t-l}-1) & b_t\boldsymbol{n}_{t-l}\boldsymbol{l}'_{\zeta,t-l}\nu_{t-l} \\ b_t\boldsymbol{l}_{\zeta,t-l}\boldsymbol{n}'_{t-l}\nu_{t-l} & \mathbf{0}_{p_2\times p_2}\end{bmatrix} = o^b_p(1),$$

by Chebyshev's inequality and Assumption 9′(iv) and Assumption 10(ii). The same comments mentioned before (B.80) apply here as well.

Part (ii). First, let $\mathcal{S}_\dagger = \mathcal{S}_u$, and recall that, from (B.78), we need the distribution of $\mathcal{L}^b_1(l)$. By Hansen (1992), Theorem 2.1, because $\|d_{u,t}\boldsymbol{S}_\#\boldsymbol{D}_{t-l,\#}\|$ is bounded by Assumption 9(ii), and $\boldsymbol{D}(\tau - \frac{l}{T}) = \boldsymbol{D}_{t-l}$ when $\tau \in [\frac{t}{T}, \frac{t+1}{T})$, we have:

$$\mathcal{L}^b_1(l) = \int_0^r d_u(\tau)\boldsymbol{S}_\#\boldsymbol{D}_\#(\tau - \frac{l}{T})\mathrm{d}\boldsymbol{B}^b_{l,T,A,\#}(\tau) \overset{d^b_p}{\Rightarrow} \int_0^r d_u(\tau)\boldsymbol{S}_\#\boldsymbol{D}_\#(\tau)\mathrm{d}\boldsymbol{B}^{(1:n)}_{l,\#}(\tau)$$
$$= ((\mathcal{S}'_\dagger\boldsymbol{S}_\#)\otimes\boldsymbol{S}_\#)\int_0^r (\boldsymbol{D}_\#(\tau)\otimes\boldsymbol{D}_\#(\tau))\,\mathrm{d}\boldsymbol{B}_{l,\#}(\tau), \tag{B.90}$$

in probability, where the convergence follows because $\mathrm{Var}^b(\mathcal{L}^b_1(l)) - \mathrm{Var}(\mathcal{L}_1(l)) \overset{p^b}{\to} 0$, which can be shown by similar arguments to (B.79), and using Lemma 8(iii) instead of Lemma 8(iv), where $\mathcal{L}_1(l)$ is the sample counterpart of $\mathcal{L}^b_1(l)$. Similarly, for $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$, we have from (B.83):

$$\mathcal{L}^b_2(l) \overset{d^b_p}{\Rightarrow} (\boldsymbol{\beta}'_{\boldsymbol{x}}\boldsymbol{s}_{p_1})\int_0^r d_u(\tau)\boldsymbol{S}_\#\boldsymbol{D}_{\tau,\#}\mathrm{d}\boldsymbol{B}^{(1:n)}_{l,\#}(\tau) + (\boldsymbol{\beta}'_{\boldsymbol{x}}\boldsymbol{S}_{p_1}\otimes\boldsymbol{S}_\#)\int_0^r [\boldsymbol{D}_{\boldsymbol{v}}(\tau)\otimes\boldsymbol{D}_\#(\tau)]\mathrm{d}\boldsymbol{B}^{(n+1:n(p_1+1))}_{l,\#}(\tau)$$
$$= [(\mathcal{S}'_\dagger\boldsymbol{S}_\#)\otimes\boldsymbol{S}_\#]\int_0^r (\boldsymbol{D}_\#(\tau)\otimes\boldsymbol{D}_\#(\tau))\,\mathrm{d}\boldsymbol{B}_{l,\#}(\tau) \text{ in probability.} \tag{B.91}$$

Next, we derive the distribution of $\mathcal{E}_{3,4}(n^*)$. This follows by similar arguments as above if we can verify that the off-diagonal elements of the bootstrap covariance $\mathrm{Cov}^b(\boldsymbol{B}^*_{l,T}(r), \boldsymbol{B}^*_{l^*,T}(r))$ converge in probability to the counterpart elements of the covariance $\mathrm{Cov}(\boldsymbol{B}^{(1:n(p_1+1))}_l(r), \boldsymbol{B}^{(1:n(p_1+1))}_{l^*}(r))$ for $l \neq l^*$. We only do so for $\mathbf{block}_{1,1}(\mathrm{Cov}^b(\boldsymbol{B}^*_{l,T}(r), \boldsymbol{B}^*_{l^*,T}(r))$; the rest follows by similar reasoning.

$$\mathbf{block}_{1,1}(\mathrm{Cov}^b(\boldsymbol{B}^*_{l,T}(r), \boldsymbol{B}^*_{l^*,T}(r))) = T^{-1}\sum_{t=1}^{[Tr]} l^2_{u,t}\begin{bmatrix} \boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l^*}\,\mathrm{E}^b(\nu^2_t\nu_{t-l}\nu_{t-l^*}) & \boldsymbol{n}_{t-l}\boldsymbol{l}_{\zeta,t-l^*}\,\mathrm{E}^b(\nu^2_t\nu_{t-l}) \\ \boldsymbol{l}_{\zeta,t-l}\boldsymbol{n}'_{t-l^*}\,\mathrm{E}^b(\nu^2_t\nu_{t-l^*}) & \boldsymbol{l}_{\zeta,t-l}\boldsymbol{l}_{\zeta,t-l^*}\,\mathrm{E}^b(\nu^2_t)\end{bmatrix}$$
$$+ T^{-1}\sum_{t,t^*=1,t\neq t^*}^{[Tr]} l_{u,t}l_{u,t^*}\begin{bmatrix} \boldsymbol{n}_{t-l}\boldsymbol{n}'_{t^*-l^*}\,\mathrm{E}^b(\nu_t\nu_{t^*}\nu_{t-l}\nu_{t^*-l^*}) & \boldsymbol{n}_{t-l}\boldsymbol{l}_{\zeta,t^*-l^*}\,\mathrm{E}^b(\nu_t\nu_{t^*}\nu_{t-l}) \\ \boldsymbol{l}_{\zeta,t-l}\boldsymbol{n}'_{t^*-l^*}\,\mathrm{E}^b(\nu_t\nu_{t^*}\nu_{t^*-l^*}) & \boldsymbol{l}_{\zeta,t-l}\boldsymbol{l}_{\zeta,t^*-l^*}\,\mathrm{E}^b(\nu_t\nu_{t^*})\end{bmatrix}$$
$$= T^{-1}\sum_{t=1}^{[Tr]} l^2_{u,t}(\boldsymbol{l}_{t-l}\boldsymbol{l}'_{t-l^*})\odot\mathbf{diag}(\mathbf{0}_{p_1+1,p_1+1},\boldsymbol{J}_2) \overset{p}{\to} \mathbf{block}_{1,1}(\boldsymbol{\rho}_{l,l^*}) = \mathbf{block}_{1,1}(\mathrm{Cov}(\boldsymbol{B}_l(r),\boldsymbol{B}_{l^*}(r))),$$

because of Assumption 9′(ii) which imposes that $\mathrm{E}[l^2_{u,t}\boldsymbol{n}_{t-l}\boldsymbol{n}_{t-l^*}] = \mathbf{0}_{(p_1+1)\times(p_1+1)}$ for $l, l^* \geq 1, l \neq l^*$, and Assumption 9′(iii) which imposes that $\mathrm{E}[l^2_{u,t}\boldsymbol{n}_{t-l}\boldsymbol{l}'_{\zeta,t-l^*}] = \mathbf{0}_{(p_1+1)\times p_2}$ for $l, l^* \geq 1, l \neq l^*$. In the general setting, for $\mathcal{S}_\dagger = \mathcal{S}_u$ or $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$, by analogy we need $\mathrm{E}[(\boldsymbol{n}_t\boldsymbol{n}'_t)\otimes(\boldsymbol{n}_{t-l}\boldsymbol{n}'_{t-l^*})] = \mathbf{0}_{(p_1+1)^2\times(p_1+1)^2}$ for $l, l^* \geq 1, l \neq l^*$, and $\mathrm{E}[(\boldsymbol{n}_t\boldsymbol{n}'_t)\otimes(\boldsymbol{n}_{t-l}\boldsymbol{l}'_{\zeta,t-l^*})] = \mathbf{0}_{(p_1+1)^2\times((p_1+1)p_2)}$ for $l, l^* \geq 1, l \neq l^*$, which are also satisfied by Assumption 9′(ii)–(iii).

Using (B.90)–(B.91) in the expression $\mathcal{E}^b_{3,4}(n^*) = \sum_{l=0}^{n^*}\mathcal{S}\boldsymbol{F}^l_s\boldsymbol{A}^{-1}_s\left(T^{-1/2}\sum_{t\in\tilde{I}_s}(\mathcal{S}'_\dagger\boldsymbol{g}_t)\nu_t(\boldsymbol{g}_{t-l-1}\odot\boldsymbol{v}_{t-l-1})\right)$, it follows that, for a fixed n^*,

$$\tilde{\mathcal{E}}^b_{3,4}(n^*) \overset{d^b_p}{\Rightarrow} \sum_{l=0}^{n^*}\mathcal{S}\boldsymbol{F}^l_s\boldsymbol{A}^{-1}_s((\mathcal{S}'_\dagger\boldsymbol{S}_\#)\otimes\boldsymbol{S}_\#)\int_{\tau_{s-1}}^{\tau_s}(\boldsymbol{D}_\#(\tau)\otimes\boldsymbol{D}_\#(\tau))\,\mathrm{d}\boldsymbol{B}_{l+1,\#}(\tau) \text{ in probability.}$$

Now as in the proof of Lemma 6, setting $n^* = T^\alpha$ for some $\alpha \in (0,1)$, and noting that the remainder $\tilde{\mathcal{E}}^b_{3,4}(\Delta\tau_sT - 2) - \tilde{\mathcal{E}}^b_{3,4}(n^*) = o^b_p(1)$, it can be shown that:

$$\tilde{\mathcal{E}}^b_{3,4}(\Delta\tau_sT - 2) \overset{d^b_p}{\Rightarrow} \sum_{l=0}^{\infty}((\mathcal{S}'_\dagger\boldsymbol{S}_\#)\otimes(\mathcal{S}\boldsymbol{F}^l_s\boldsymbol{A}^{-1}_s\boldsymbol{S}_\#))\int_{\tau_{s-1}}^{\tau_s}(\boldsymbol{D}_\#(\tau)\otimes\boldsymbol{D}_\#(\tau))\,\mathrm{d}\boldsymbol{B}_{l+1,\#}(\tau) = \mathbb{M}_{3,2}(\tau_{s-1},\tau_s),$$

in probability, where $\mathbb{M}_{3,2}(\tau_{s-1},\tau_s)$ as defined above is also the asymptotic distribution of the sample counterpart of $\tilde{\mathcal{E}}^b_{3,4}(\Delta\tau_sT - 2)$, that is $\tilde{\mathcal{E}}_{3,4}$, featuring in Supplementary Appendix, Appendix C on page 7. The variance of $\mathbb{M}_{3,2}(\tau_{s-1},\tau_s)$ exists and is derived in the Supplementary Appendix right after (C.16). Therefore,

$$\mathcal{E}^b_{3,4} \overset{d^b_p}{\Rightarrow} \mathbb{M}_{3,2}(\tau_{s-1},\tau_s) \text{ in probability.} \tag{B.92}$$

Now consider $\mathcal{E}^b_{3,3} = T^{-1/2}\sum_{t\in\tilde{I}_s^-}(\mathcal{S}^{b\prime}_{\dagger}\boldsymbol{g}^b_t)\left[\mathcal{S}\left(\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}^l_s\right)\boldsymbol{\mu}_s\right]$ in (B.75). By similar analysis as for $\mathcal{E}^b_{3,4}$, it can be shown that:

$$\mathcal{E}^b_{3,3} \overset{d^b_p}{\Rightarrow} \sum_{l=0}^{\infty}\left((\mathcal{S}'_{\dagger}\boldsymbol{S}_{\#})\otimes(\mathcal{S}\boldsymbol{F}^l_s)\right)\left(\left[\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}_{\#}(\tau)\mathrm{d}\boldsymbol{B}_{0,\#}(\tau)\right]\otimes\boldsymbol{\mu}_s\right) = \mathbb{M}_{3,1}(\tau_{s-1},\tau_s) \text{ in probability.} \tag{B.93}$$

Also, $\mathcal{E}^b_{3,3}+\mathcal{E}^b_{3,4}$ can be shown to jointly converge to $\mathbb{M}_3(\tau_{s-1},\tau_s) = \mathbb{M}_{3,1}(\tau_{s-1},\tau_s)+\mathbb{M}_{3,2}(\tau_{s-1},\tau_s)$, provided that $\mathrm{Cov}^b(\boldsymbol{B}^b_{0,T}(r),\boldsymbol{B}^*_{l,T}(r)) - \mathrm{Cov}(\boldsymbol{B}_0(r),\boldsymbol{B}^{(1:n(p_1+1))}_l(r)) \overset{p}{\to} 0$ uniformly in r for all $l\geq 1$, where $\boldsymbol{B}^b_{0,T} = T^{-1/2}\sum_{t=1}^{[Tr]}\boldsymbol{l}_t\odot\tilde{\boldsymbol{\nu}}_t$, and recall that $\boldsymbol{B}^*_{l,T}(r) = \mathbf{vect}(\boldsymbol{B}^b_{l,T,A}(r),\boldsymbol{B}^b_{l,T,B}(r))$. Now consider $\mathrm{Cov}^b(\boldsymbol{B}^{b,(1)}_{0,T}(r),\boldsymbol{B}^b_{l,T,A}(r)) = \mathrm{Cov}^b(\boldsymbol{B}_{0,T}(r),\boldsymbol{B}^*_{l,T}(r))$; the proof for the rest of the elements is similar.

$$\begin{aligned}\mathrm{Cov}^b(\boldsymbol{B}^{b,(1)}_{0,T}(r),\boldsymbol{B}^b_{l,T,A}(r)) &= T^{-1/2}\sum_{t=1}^{[Tr]}(l_{u,t}\nu_t)(T^{-1/2}\sum_{t=1}^{[Tr]}l_{u,t}\nu_t(\boldsymbol{l}_{t-l}\odot\tilde{\boldsymbol{\nu}}_{t-l}))\\ &= T^{-1}\sum_{t=1}^{[Tr]}l^2_{u,t}\nu^2_t(\boldsymbol{l}_{t-l}\odot\tilde{\boldsymbol{\nu}}_{t-l}) + T^{-1}\sum_{t,t^*=1,t\neq t^*}^{[Tr]}l_{u,t}l_{u,t^*}\nu_t\nu_{t^*}(\boldsymbol{l}_{t^*-l}\odot\tilde{\boldsymbol{\nu}}_{t^*-l})\\ &= \mathcal{L}^b_{3,1}+\mathcal{L}^b_{3,2}.\end{aligned}$$

Now note that by Lemma 8(iv),

$$\begin{aligned}\mathrm{E}^b(\mathcal{L}^b_{3,1}) &= T^{-1}\sum_{t=1}^{[Tr]}l^2_{u,t}\nu^2_t(\boldsymbol{l}_{t-l}\odot\tilde{\boldsymbol{\nu}}_{t-l}) = T^{-1}\sum_{t=1}^{[Tr]}\begin{bmatrix}l^2_{u,t}l_{u,t-l}\\ l^2_{u,t}\boldsymbol{l}_{\boldsymbol{\nu},t-l}\\ l^2_{u,t}\boldsymbol{l}_{\boldsymbol{\zeta},t-l}\end{bmatrix}\odot\mathrm{E}^b\begin{bmatrix}\nu^2_t\nu_{t-l}\\ \nu^2_t\nu_{t-l}\boldsymbol{\iota}_{p_1}\\ \nu^2_t\boldsymbol{\iota}_{p_2}\end{bmatrix}\\ &= T^{-1}\sum_{t=1}^{[Tr]}l^2_{u,t}\mathbf{vect}(\mathbf{0}_{1+p_1},\boldsymbol{l}_{\zeta,t-l}) \overset{p}{\to} r[\boldsymbol{\rho}^{(1:n,1:1)}_l\odot\mathcal{I}],\\ \mathrm{E}^b(\mathcal{L}^b_{3,2}) &= T^{-1}\sum_{t,t^*=1,t\neq t^*}^{[Tr]}\begin{bmatrix}l_{u,t}l_{u,t^*}l_{u,t^*-l}\\ l_{u,t}l_{u,t^*}\boldsymbol{l}_{\boldsymbol{\nu},t^*-l}\\ l_{u,t}l_{u,t^*}\boldsymbol{l}_{\boldsymbol{\zeta},t^*-l}\end{bmatrix}\odot\mathrm{E}^b\begin{bmatrix}\nu_t\nu_{t^*}\nu_{t^*-l}\\ \nu_t\nu_{t^*}\nu_{t^*-l}\boldsymbol{\iota}_{p_1}\\ \nu_t\nu_{t^*}\boldsymbol{\iota}_{p_2}\end{bmatrix} = \mathbf{0}_n.\end{aligned}$$

Therefore, $\mathrm{Cov}^b(\boldsymbol{B}^{b,(1)}_{0,T}(r),\boldsymbol{B}^b_{l,T,A}(r)) - \mathrm{Cov}(\boldsymbol{B}^{(1)}_0(r),\boldsymbol{B}^{(1:n)}_l(r)) = o_p(1)$, by the restriction in Assumption 9′(i), which ensures that $\boldsymbol{\rho}^{(1:n,1:1)}_l = \boldsymbol{\rho}^{(1:n,1:1)}_l\odot\mathcal{I}$; also note that for the rest of the terms of the covariance above, by analogy, we need $\mathrm{E}^b[(\boldsymbol{n}_t\boldsymbol{n}'_t)\otimes\boldsymbol{n}_{t-l}] = \mathbf{0}_{(p_1+1)^2\times(p_1+1)}$ for $l\geq 1$, imposed in Assumption 9′(i). So,

$$\mathcal{E}^b_{3,3}+\mathcal{E}^b_{3,4} \overset{d^b_p}{\Rightarrow} \mathbb{M}_3(\tau_{s-1},\tau_s) = \mathbb{M}_{3,1}(\tau_{s-1},\tau_s)+\mathbb{M}_{3,2}(\tau_{s-1},\tau_s)$$

in probability. Because we showed that $\mathcal{E}^b_{3,1} = o^b_p(1)$ and $\mathcal{E}^b_{3,2} = o_p(1)$, it follows that: $\mathcal{E}^b_3 \overset{d^b_p}{\Rightarrow} \mathbb{M}_3(\tau_{s-1},\tau_s)$, in probability and that $\mathbf{vect}(\mathcal{E}^b_1,\mathcal{E}^b_3) \overset{d^b_p}{\Rightarrow} \mathbf{vect}(\mathbb{M}_1(\tau_{s-1},\tau_s),\mathbb{M}_3(\tau_{s-1},\tau_s))$ in probability.

• Now consider $\mathcal{E}^b_2$. Note that $\boldsymbol{r}_t = \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t$ is not bootstrapped, and recall that $\boldsymbol{\xi}_t = \boldsymbol{\mu}_s+\boldsymbol{\eta}_t+\boldsymbol{F}_s\boldsymbol{\xi}_{t-1}$. Therefore, replacing again estimated parameters by the true values, because the rest of the terms are $o^b_p(1)$ (therefore also replacing, as before, $\boldsymbol{g}^b_t$ with $\boldsymbol{g}_t\odot\boldsymbol{\nu}_t$),

$$\begin{aligned}\mathcal{E}^b_2 &= \left\{T^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}'_{\dagger}(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\right\}\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\mu}_s + T^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}'_{\dagger}(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_s\boldsymbol{\xi}_{t-1} + T^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}'_{\dagger}(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\eta}_t\\ &= \mathcal{E}^b_{2,1}+\mathcal{E}^b_{2,2}+\mathcal{E}^b_{2,3}+o^b_p(1).\end{aligned}$$

Now consider $\mathcal{E}^b_{2,1}$. From (B.72) and (B.74), without any restrictions on $\boldsymbol{\rho}_i$, $\boldsymbol{\rho}_{ij}$ except those in Assumption 9,

$$\begin{aligned}\mathcal{E}^b_{2,1} &= [\mathcal{E}^b_1+o^b_p(1)]\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\mu}_s \overset{d^b_p}{\Rightarrow} \left\{(\mathcal{S}'_{\dagger}\boldsymbol{S}_{\#})\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}(\tau)\mathrm{d}\,\boldsymbol{B}_{0,\#}(\tau)\right\}\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\mu}_s\\ &= ((\mathcal{S}'_{\dagger}\boldsymbol{S}_{\#})\otimes(\mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}^0_s))\left(\left[\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}_{\#}(\tau)\mathrm{d}\boldsymbol{B}_{0,\#}(\tau)\right]\otimes\boldsymbol{\mu}_s\right),\end{aligned} \tag{B.94}$$

in probability, where the latter is the first term in $\mathbb{M}_{2,1}(\tau_{s-1},\tau_s)$ defined before Lemma 6 (the rest of the terms appear from the distribution of $\mathcal{E}^b_{2,2}$ as seen below).

Now consider $\mathcal{E}^b_{2,3} = T^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}'_{\dagger}(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\eta}_t$. Recall from the arguments above (B.55) that $\boldsymbol{A}^{-1}_s$ is also upper triangular with rows $p_1+2:n$ equal to $[\mathbf{0}_{p_2}\quad\mathbf{0}_{p_2\times p_1}\quad\boldsymbol{I}_{p_2}]$. Therefore, as shown on page 10 of the Supplementary

Appendix, $\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\eta}_t = \mathcal{S}_{\boldsymbol{r}}\boldsymbol{A}_{s,\#}^{-1}\boldsymbol{g}_t = \boldsymbol{\zeta}_t$, so, for $\mathcal{S}_\dagger = \mathcal{S}_u$,

$$\begin{aligned}\mathcal{E}_{2,3}^b &= T^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}_\dagger'(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\mathcal{S}_{\boldsymbol{r}}\boldsymbol{\eta}_t = T^{-1/2}\sum_{t\in\tilde{I}_s}u_t\nu_t\boldsymbol{\zeta}_t\\ &= T^{-1/2}\sum_{t\in\tilde{I}_s}(\boldsymbol{S}_{p_2}\boldsymbol{D}_{\zeta,t}d_{u,t})l_{u,t}\nu_t\boldsymbol{l}_{\zeta,t},\end{aligned}$$

and consider first $\boldsymbol{B}_{u\zeta,T}^b(r) = T^{-1/2}\sum_{t=1}^{[Tr]}l_{u,t}\boldsymbol{l}_{\zeta,t}\nu_t$.

Part (i). Since ν_t is i.i.d, $\mathrm{E}^b(l_{u,t}l_{\zeta_\kappa,t}\nu_t|\mathcal{F}_{t-1}^b) = 0$, for any element $l_{\zeta_\kappa,t}$ of $\boldsymbol{l}_{\zeta,t}$, $\kappa = 1,\ldots,p_2$. Also, for some $c > 0$, $\sup_t \mathrm{E}\,\mathrm{E}^b|l_{u,t}l_{\zeta_\kappa,t}\nu_t|^{2+\delta^*} \le \sup_t \mathrm{E}|l_{u,t}l_{\zeta_\kappa,t}|^{2+\delta^*}\sup_t \mathrm{E}^b|\nu_t|^{2+\delta^*} < C$. Because $\mathrm{E}^b(\nu_t^2|\mathcal{F}_{t-1}^b) = \mathrm{E}^b(\nu_t^2) = 1$ by Assumption 10, we have $\mathrm{E}^b(l_{u,t}^2 l_{\zeta_\kappa,t}^2\nu_t^2|\mathcal{F}_{t-1}^b) = \mathrm{E}^b(l_{u,t}^2 l_{\zeta_\kappa,t}^2\nu_t^2)$, therefore, the conditional and unconditional bootstrap second moments are the same, and it remains to verify that $\mathrm{Var}^b(\boldsymbol{B}_{u\zeta,T}^b(r)) - \mathrm{Var}(\boldsymbol{B}_{u\zeta}(r)) = o_p(1)$, where $\boldsymbol{B}_{u\zeta}(r)$ was defined just before Lemma 6.

$$\mathrm{Var}^b(\boldsymbol{B}_{u\zeta,T}^b(r)) = T^{-1}\sum_{t=1}^{[Tr]}l_{u,t}^2\boldsymbol{l}_{\zeta,t}\boldsymbol{l}_{\zeta,t}'\,\mathrm{E}^b(\nu_t^2) = T^{-1}\sum_{t=1}^{[Tr]}l_{u,t}^2\boldsymbol{l}_{\zeta,t}\boldsymbol{l}_{\zeta,t}' \xrightarrow{p} r\boldsymbol{\rho}_{u,\zeta,0,0} = \mathrm{Var}(\boldsymbol{B}_{u\zeta}(r)),$$

where $\boldsymbol{\rho}_{u,\zeta,0,0}$ was defined before Lemma 5, and the convergence follows by Lemma 8(iv).

Part (ii). Because $\mathrm{Var}^b(\boldsymbol{B}_{u\zeta,T}^b(r)) \xrightarrow{p} \mathrm{Var}(\boldsymbol{B}_{u\zeta}(r))$, using Lemma 8(iii), it follows by Hansen (1992), Theorem 2.1, that:

$$\mathcal{E}_{2,3}^b = T^{-1/2}\sum_{t\in\tilde{I}_s}(\boldsymbol{S}_{p_2}\boldsymbol{D}_{\zeta,t}d_{u,t})l_{u,t}\nu_t\boldsymbol{l}_{\zeta,t} \overset{d_p^b}{\Rightarrow} \boldsymbol{S}_{p_2}\int_{\tau_{s-1}}^{\tau_s}d_u(\tau)\boldsymbol{D}_\zeta(\tau)\mathrm{d}\boldsymbol{B}_{u\zeta}(\tau) = \mathbb{M}_{2,3}^{(1)}(\tau_{s-1},\tau_s), \tag{B.95}$$

in probability, where $\mathbb{M}_{2,3}^{(1)}(\tau_{s-1},\tau_s)$ was defined right before Lemma 6. Similarly, it can be shown that for $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$, without restrictions on $\boldsymbol{\rho}_{0,0}$ besides those imposed in Assumption 9,

$$\mathcal{E}_{2,3}^b \overset{d_p^b}{\Rightarrow} \mathbb{M}_{2,3}^{(2)}(\tau_{s-1},\tau_s) \text{ in probability.} \tag{B.96}$$

Next, consider $\mathcal{E}_{2,2}^b$. By backward substituting ξ_{t-1},

$$\begin{aligned}\mathcal{E}_{2,2}^b &= \mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_sT^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}_\dagger'(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\boldsymbol{\xi}_{t-1}\\ &= \mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_sT^{-1/2}[\mathcal{S}_\dagger'(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\boldsymbol{\xi}_{[\tau_{s-1}T]} + S_{\boldsymbol{r}}\boldsymbol{F}_sT^{-1/2}\sum_{t\in\tilde{I}_s^-}\boldsymbol{F}_s^{\tilde{t}-1}[\mathcal{S}_\dagger'(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\left[\boldsymbol{\xi}_{[\tau_{s-1}T]}\right]\\ &\quad + S_{\boldsymbol{r}}\,T^{-1/2}\sum_{t\in\tilde{I}_s^-}[\mathcal{S}_\dagger'(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\left[\left(\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\right)\boldsymbol{\mu}_s\right] + S_{\boldsymbol{r}}\,T^{-1/2}\sum_{t\in\tilde{I}_s^-}[\mathcal{S}_\dagger'(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\left[\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\boldsymbol{\eta}_{t-l-1}\right]\\ &= \sum_{i=1}^4\mathcal{E}_{2,2,i}^b.\end{aligned} \tag{B.97}$$

First, note that $\mathcal{E}_{2,2,1}^b = o_p^b(1)$ because $T^{-1/2}\boldsymbol{g}_t\odot\boldsymbol{\nu}_t = o_p^b(1)$ (as shown before in the proof of $\mathcal{E}_1^b$) and $\boldsymbol{\xi}_{[\tau_{s-1}T]} = O_p(1)$ (see the proof of Lemma 2 in the Supplementary Appendix, Appendix C). Next, because $\sum_{l=0}^\infty\|\boldsymbol{F}_s^l\|$ is bounded, $\mathrm{Var}^b(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t) = O_p(1)$ and $\boldsymbol{\xi}_{[\tau_{s-1}T]} = O_p(1)$, we have $\mathcal{E}_{2,2,2}^b = o_p^b(1)$.

Next, by similar arguments as for $\mathcal{E}_{3,3}^b$, and noting that no restrictions are needed on $\boldsymbol{\rho}_i, \boldsymbol{\rho}_{ij}$ besides those in Assumption 9 (because $\mathcal{E}_{2,3}^b$ has at the basis the same random process as $\mathcal{E}_1^b$),

$$\begin{aligned}\mathcal{E}_{2,2,3}^b &= \mathcal{S}_r\boldsymbol{F}_sT^{-1/2}\sum_{t\in\tilde{I}_s^-}[\mathcal{S}_\dagger'(\boldsymbol{g}_t\odot\boldsymbol{\nu}_t)]\left[\left(\sum_{l=0}^{\tilde{t}-2}\boldsymbol{F}_s^l\right)\boldsymbol{\mu}_s\right]\\ &\overset{d_p^b}{\Rightarrow}\sum_{l=1}^\infty((\mathcal{S}_\dagger'\boldsymbol{S}_\#)\otimes(\mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_s^l))\left(\left[\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}_\#(\tau)\mathrm{d}\boldsymbol{B}_{0,\#}(\tau)\right]\otimes\boldsymbol{\mu}_s\right)\\ &= \mathbb{M}_{2,1}(\tau_{s-1},\tau_s) - ((\mathcal{S}_\dagger'\boldsymbol{S}_\#)\otimes(\mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_s^0))\left(\left[\int_{\tau_{s-1}}^{\tau_s}\boldsymbol{D}_\#(\tau)\mathrm{d}\boldsymbol{B}_{0,\#}(\tau)\right]\otimes\boldsymbol{\mu}_s\right)\end{aligned} \tag{B.98}$$

in probability. Now consider $\mathcal{E}^b_{2,2,4}$,

$$\mathcal{E}^b_{2,2,4} = S_{\mathbf{r}}\, T^{-1/2} \sum_{t\in\tilde{I}_s^-} [\mathcal{S}'_\dagger(\boldsymbol{g}_t \odot \boldsymbol{\nu}_t)] \left[\sum_{l=0}^{\tilde{t}-2} \boldsymbol{F}_s^l \boldsymbol{\eta}_{t-l-1}\right].$$

By similar arguments as for $\mathcal{E}^b_{3,4}$ in (B.76) and just below it,

$$\mathcal{E}^b_{2,2,4} = S_{\mathbf{r}} \sum_{l=0}^{\Delta\tau_s T-2} \boldsymbol{F}_s^l \left(T^{-1/2} \sum_{t\in\tilde{I}_s^-} \mathcal{S}'_\dagger(\boldsymbol{g}_t \odot \boldsymbol{\nu}_t)\boldsymbol{\eta}_{t-l-1}\right) + o^b_p(1) = \mathcal{E}^b_5(\Delta\tau_s T-2) + o^b_p(1).$$

Next, we analyze $\mathcal{E}^b_5(n^*)$, first for a fixed n^*. Let $\mathcal{S}_\dagger = \mathcal{S}_u$. Note that a crucial term in $\mathcal{E}^b_5(n^*)$ is $\mathcal{L}^b_5(l) = T^{-1/2}\sum_{t=1}^{[Tr]} u_t \nu_t \boldsymbol{g}_{t-l}$ for $l \geq 1$, because $\boldsymbol{\eta}_{t-l} = \boldsymbol{A}_s^{-1}\boldsymbol{g}_{t-l}$. By the structure of $\boldsymbol{S}$ and $\boldsymbol{D}_t$ in (B.40),

$$\boldsymbol{g}_{t-l} = \begin{bmatrix} d_{u,t-l} l_{u,t-l} \\ \boldsymbol{s}_{p_1} d_{u,t-l} l_{u,t-l} + \boldsymbol{S}_{p_1}\boldsymbol{D}_{\boldsymbol{v},t-l}\boldsymbol{l}_{\boldsymbol{v},t} \\ \boldsymbol{S}_{p_2}\boldsymbol{D}_{\zeta,t-l}\boldsymbol{l}_{\zeta,t-l}\end{bmatrix}_{\#} = \boldsymbol{S}_{\#}\boldsymbol{D}_{t-l,\#}\boldsymbol{l}_{t-l,\#}.$$

Then, letting $\mathcal{E}_{t,l,5} = l_{u,t}\boldsymbol{l}_{t-l}$ and $\mathcal{E}^b_{t,l,5} = l_{u,t}\boldsymbol{l}_{t-l}\nu_t$,

$$\mathcal{L}^b_5(l) = T^{-1/2}\sum_{t=1}^{[Tr]} d_{u,t}\boldsymbol{S}_{\#}\boldsymbol{D}_{t-l,\#} l_{u,t}\boldsymbol{l}_{t-l,\#}\nu_t = T^{-1/2}\sum_{t=1}^{[Tr]} (d_{ut}\boldsymbol{S}_{\#}\boldsymbol{D}_{t-l,\#})(l_{u,t}\boldsymbol{l}_{t-l}\nu_t)_{\#}. \tag{B.99}$$

Part (i). First, consider $\boldsymbol{B}^b_{l,T,C}(r) = T^{-1/2}\sum_{t=1}^{[Tr]} l_{u,t}\boldsymbol{l}_{t-l}\nu_t$, for $l \geq 1$. Because ν_t is i.i.d, it is m.d.s under the bootstrap measure conditional on the data, so by arguments similar to before, $\boldsymbol{B}^b_{l,T,C}(r) \overset{d^b_p}{\Rightarrow} \boldsymbol{B}^{(1:n)}_l(r)$ in probability, provided that $\mathrm{Var}^b(\boldsymbol{B}^b_{l,T,C}(r)) \overset{p}{\to} \mathrm{Var}(\boldsymbol{B}^{(1:n)}_l(r))$, which we verify below:

$$\mathrm{Var}^b(\boldsymbol{B}^b_{l,T,C}(r)) = T^{-1}\sum_{t=1}^{[Tr]} l^2_{u,t}\boldsymbol{l}_{t-l}\boldsymbol{l}'_{t-l}\,\mathrm{E}^b(\nu_t^2) = T^{-1}\sum_{t=1}^{[Tr]} l^2_{u,t}\boldsymbol{l}_{t-l}\boldsymbol{l}'_{t-l} \overset{p}{\to} \mathrm{Var}(\boldsymbol{B}^{(1:n)}_l(r)) = \mathbf{block}_{1,1}(\boldsymbol{\rho}_{l,l}).$$

The previous to last statement above follows by Lemma 8(iv) without restrictions on the form of $\boldsymbol{\rho}_{l,l}$ besides the ones in Assumption 9. Therefore, $\boldsymbol{B}^b_{l,T,C}(r) \overset{d^b_p}{\Rightarrow} \boldsymbol{B}^{(1:n)}_l(r))$ in probability.

Part (ii). By Hansen (1992), Theorem 2.1, and Lemma 8(iii), $\mathcal{L}^b_5(l)$ defined in (B.99) is such that:

$$\mathcal{L}^b_5(l) \overset{d^b_p}{\Rightarrow} \int_0^r d_u(\tau)\boldsymbol{S}_{\#}\boldsymbol{D}_{\#}(\tau)\mathrm{d}\boldsymbol{B}^{(1:n)}_{l,\#}(\tau) = ((\mathcal{S}'_\dagger\boldsymbol{S}_{\#})\otimes\boldsymbol{S}_{\#})\int_0^r (\boldsymbol{D}_{\#}(\tau)\otimes\boldsymbol{D}_{\#}(\tau))\,\mathrm{d}\boldsymbol{B}_{l,\#}(\tau)$$

in probability. For $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$, the same result can be shown by similar arguments, and with no restrictions on $\boldsymbol{\rho}_{l,l}$ besides being finite.

Now let $\mathcal{S}_\dagger = \mathcal{S}_u$ again. To derive the limiting distribution of $\mathcal{E}^b_5(n^*)$, we need not only that $\boldsymbol{B}^b_{l,T,C}(r) \overset{d^b_p}{\Rightarrow} \boldsymbol{B}^{(1:n)}_l(r))$ in probability, but also that $\mathbf{vect}(\boldsymbol{B}^b_{l,T,C}(r), \boldsymbol{B}^b_{l^*,T,C}(r)) \overset{d^b_p}{\Rightarrow} \mathbf{vect}(\boldsymbol{B}^{(1:n)}_l(r), \boldsymbol{B}^{(1:n)}_{l^*}(r))$ in probability, which can be shown using Lemmas 3 and 8(iv), because
$\mathrm{Cov}^b(\boldsymbol{B}^b_{l,T,C}(r), \boldsymbol{B}^b_{l^*,T,C}(r)) \overset{p}{\to} \mathrm{Cov}(\boldsymbol{B}^{(1:n)}_l(r), \boldsymbol{B}^{(1:n)}_{l^*}(r)) = \mathbf{block}_{1,1}(\boldsymbol{\rho}_{l,l^*})$. The latter condition holds because:

$$\begin{aligned}\mathrm{E}^b(\boldsymbol{B}^b_{l,T,C}(r)(\boldsymbol{B}^b_{l^*,T,C}(r))') &= T^{-1}\sum_{t=1}^{[Tr]} l^2_{u,t}\boldsymbol{l}_{t-l}\boldsymbol{l}'_{t-l^*}\,\mathrm{E}^b(\nu_t^2) + T^{-1}\sum_{t,t^*=1,t\neq t^*}^{[Tr]} l_{u,t}l_{u,t^*}\boldsymbol{l}_{t-l}\boldsymbol{l}'_{t^*-l^*}\,\mathrm{E}^b(\nu_t\nu_{t^*})\\ &= T^{-1}\sum_{t=1}^{[Tr]} l^2_{u,t}\boldsymbol{l}_{t-l}\boldsymbol{l}'_{t-l^*} \overset{p}{\to} \mathbf{block}_{1,1}(\boldsymbol{\rho}_{l,l^*}),\end{aligned}$$

where the last statement follows by Lemma 8(iv), and under Assumptions 1–10. By analogy, no other restrictions besides Assumptions 1–10 are needed also when $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$.

Therefore, by Hansen (1992) and Lemma 8(iii), for a fixed n^*,

$$\begin{aligned}\mathcal{E}^b_5(n^*) &= S_{\mathbf{r}}\sum_{l=0}^{n^*}\boldsymbol{F}_s^l\boldsymbol{A}_s^{-1}\left(T^{-1/2}\sum_{t\in\tilde{I}_s^-}(\mathcal{S}'_\dagger\boldsymbol{g}^b_t)\boldsymbol{g}_{t-l-1}\right)\\ &\overset{d^b_p}{\Rightarrow} \sum_{l=0}^{n^*}((\mathcal{S}'_\dagger\boldsymbol{S}_{\#})\otimes(\mathcal{S}_r\boldsymbol{F}_s^{l+1}\boldsymbol{A}_s^{-1}\boldsymbol{S}_{\#}))\int_{\tau_{s-1}}^{\tau_s}(\boldsymbol{D}_{\#}(\tau)\otimes\boldsymbol{D}_{\#}(\tau))\,\mathrm{d}\boldsymbol{B}_{l+1,\#}(\tau)\end{aligned}$$

in probability. Letting as before $n^* = T^\alpha$, it can be shown that $\mathcal{E}_5^b(\Delta\tau_s T - 2) = \mathcal{E}_5^b(n^*) + o_p^b(1)$, and therefore

$$\begin{aligned}\mathcal{E}_{2,2,4}^b &= \mathcal{E}_5^b(\Delta\tau_s T - 2) + o_p^b(1) \overset{d_p^b}{\Rightarrow} \sum_{l=0}^{\infty}((\mathcal{S}_\dagger' \boldsymbol{S}_\#) \otimes (\mathcal{S}_r \boldsymbol{F}_s^{l+1} \boldsymbol{A}_s^{-1} \boldsymbol{S}_\#)) \int_{\tau_{s-1}}^{\tau_s} (\boldsymbol{D}_\#(\tau) \otimes \boldsymbol{D}_\#(\tau))\, d\boldsymbol{B}_{l+1,\#}(\tau) \\ &= \mathbb{M}_{2,2}(\tau_{s-1}, \tau_s), \text{ in probability,}\end{aligned} \tag{B.100}$$

where $\mathbb{M}_{2,2}(\tau_{s-1}, \tau_s)$ is defined just before Lemma 6. Substituting $\mathcal{E}_{2,2,1}^b = o_p^b(1)$, $\mathcal{E}_{2,2,2}^b = o_p^b(1)$, and (B.98) and (B.100) into (B.97), and then using (B.94), it follows that:

$$\begin{aligned}&\mathcal{E}_{2,2}^b \overset{d_p^b}{\Rightarrow} \mathbb{M}_{2,1}(\tau_{s-1}, \tau_s) - ((\mathcal{S}_\dagger' \boldsymbol{S}_\#) \left(\left[\int_{\tau_{s-1}}^{\tau_s} \boldsymbol{D}_\#(\tau) d\boldsymbol{B}_{0,\#}(\tau) \right] \otimes \boldsymbol{\mu}_s \right) \otimes (\mathcal{S}_{\boldsymbol{r}} \boldsymbol{F}_s^0)) + \mathbb{M}_{2,2}(\tau_{s-1}, \tau_s) \\ &\mathcal{E}_{2,1}^b + \mathcal{E}_{2,2}^b \overset{d_p^b}{\Rightarrow} \mathbb{M}_{2,1}(\tau_{s-1}, \tau_s) + \mathbb{M}_{2,2}(\tau_{s-1}, \tau_s)\end{aligned} \tag{B.101}$$

in probability, because the joint convergence of $\mathcal{E}_{2,2,3}^b, \mathcal{E}_{2,2,4}^b$, and of $\mathcal{E}_{2,1}^b, \mathcal{E}_{2,2}^b$ can be shown as above under Assumptions 1–10. Because all these terms share the same ν_t, it can be shown that they also jointly converge with $\mathcal{E}_{2,3}^b$ and their bootstrap covariance to the covariances of the relevant limits, under Assumptions 1–10.

Therefore, for $\mathcal{S}_\dagger = \mathcal{S}_u$,

$$\mathcal{E}_2^b = \sum_{i=1}^{3} \mathcal{E}_{2,i}^b \overset{d_p^b}{\Rightarrow} \mathbb{M}_{2,1}(\tau_{s-1}, \tau_s) + \mathbb{M}_{2,2}(\tau_{s-1}, \tau_s) + \mathbb{M}_{2,3}^{(1)}(\tau_{s-1}, \tau_s) = \mathbb{M}_2(\tau_{s-1}, \tau_s)$$

in probability. Similarly, for $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$, $\mathcal{E}_2^b \overset{d_p^b}{\Rightarrow} \mathbb{M}_2(\tau_{s-1}, \tau_s)$ in probability, completing the proof for the distribution of $\mathcal{E}_2^b$, which we note was proven only under Assumptions 1–10. Note that Assumption 9′(iv) is not needed here for the WR bootstrap (because $\boldsymbol{r}_t = \mathcal{S}_{\boldsymbol{r}} \boldsymbol{\xi}_t$ is not bootstrapped in $\mathcal{E}_2^b$).

Now note that because $\mathcal{E}_1^b$ featured as part of $\mathcal{E}_2^b$, their joint convergence was already shown, and recall that it also followed under Assumptions 1–10. It remains to verify the condition:

$$\mathrm{Cov}^b(\mathbf{vect}(\mathcal{E}_2^b, \mathcal{E}_3^b)) - \mathrm{Cov}(\mathbf{vect}(\mathbb{M}_2(\tau_{s-1}, \tau_s), \mathbb{M}_3(\tau_{s-1}, \tau_s))) \overset{p}{\to} 0,$$

because then $\mathbf{vect}_{i=1:3}(\mathcal{E}_i^b) \overset{d_p^b}{\Rightarrow} \mathbf{vect}_{i=1:3}(\mathbb{M}_i(\tau_{s-1}, \tau_s)) = \mathbb{M}(\tau_{s-1}, \tau_s)$ in probability. This condition follows by similar arguments as before, if we show that (C 1) $\mathrm{Cov}^b(\mathcal{E}_{2,3}^b, \mathcal{E}_{3,4}^b)$ converges to the joint covariance of their respective limits, and that (C 2) $\mathrm{Cov}^b(\mathcal{E}_{2,2,4}^b, \mathcal{E}_{3,4}^b)$ converges to the joint covariance of their respective limits. For (C 1), by arguments as before, it suffices to show $\mathrm{Cov}^b(\boldsymbol{B}_{l,T,A}^b(r), \boldsymbol{B}_{u\zeta,T}^b(r)) - \mathrm{Cov}(\boldsymbol{B}_l^{(1:n)}(r), \boldsymbol{B}_{u\zeta}(r)) \overset{p}{\to} 0$ (here, we set $\mathcal{S}_\dagger = \mathcal{S}_u$ for all terms and that is why we consider the first $n \times 1$ elements of $\boldsymbol{B}_l(r)$; the proofs forthe case $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$ are similar and are briefly discussed below). Note:

$$\begin{aligned}\mathrm{Cov}^b(\boldsymbol{B}_{l,T,A}^b(r), \boldsymbol{B}_{u\zeta,T}^b(r)) &= \mathrm{E}^b \left((T^{-1} \sum_{t=1}^{[Tr]} (l_{u,t} \boldsymbol{l}_{t-l}) \odot \mathbf{vect}(\nu_t \nu_{t-l} \boldsymbol{\iota}_{p_1+1}, \nu_t \boldsymbol{\iota}_{p_2}))(T^{-1} \sum_{t^*=1}^{[Tr]} l_{u,t^*} \boldsymbol{l}_{\zeta,t^*}' \nu_{t^*}) \right) \\ &= T^{-1} \sum_{t=1}^{[Tr]} (l_{u,t}^2 \boldsymbol{l}_{t-l} \boldsymbol{l}_{\zeta,t}') \odot \mathrm{E}^b \begin{bmatrix} \nu_t^2 \nu_{t-l} \boldsymbol{\iota}_{p_1+1} \boldsymbol{\iota}_{p_2}' \\ \nu_t^2 \boldsymbol{\iota}_{p_2} \boldsymbol{\iota}_{p_2}' \end{bmatrix} + T^{-1} \sum_{t,t^*=1, t\neq t^*}^{[Tr]} (l_{u,t} l_{u,t^*} \boldsymbol{l}_{t-l} \boldsymbol{l}_{\zeta,t^*}') \mathrm{E}^b \begin{bmatrix} \nu_t \nu_{t^*} \nu_{t-l} \boldsymbol{\iota}_{p_1+1} \boldsymbol{\iota}_{p_2}' \\ \nu_t \nu_{t^*} \boldsymbol{\iota}_{p_2} \boldsymbol{\iota}_{p_2}' \end{bmatrix} \\ &= T^{-1} \sum_{t=1}^{[Tr]} (l_{u,t}^2 \boldsymbol{l}_{t-l} \boldsymbol{l}_{\zeta,t}') \odot \mathbf{vect}(\mathbf{0}_{(p_1+1)\times p_2}, \boldsymbol{J}_2) \overset{p}{\to} T^{-1} \sum_{t=1}^{[Tr]} \mathrm{E}(l_{u,t}^2 \boldsymbol{l}_{t-l} \boldsymbol{l}_{\zeta,t}') \odot \mathbf{vect}(\mathbf{0}_{(p_1+1)\times p_2}, \boldsymbol{J}_2) + o_p(1),\end{aligned}$$

which shows why we need $\mathrm{E}(l_{u,t}^2 \boldsymbol{n}_{t-l} \boldsymbol{l}_{\zeta,t}') = \mathbf{0}_{(p_1+1)\times p_2}$, imposed in Assumption 9′(iii). In the general case of $\mathcal{S}_\dagger = \mathcal{S}_u$ or $\mathcal{S}_\dagger = \boldsymbol{\beta}_{\boldsymbol{x},\#}$, by analogy, the condition needed and imposed in Assumption 9′(iii) is that for $l \geq 1$, $\mathrm{E}((\boldsymbol{n}_t \boldsymbol{n}_t') \otimes (\boldsymbol{n}_{t-l} \boldsymbol{l}_{\zeta,t}')) = \mathbf{0}_{(p_1+1)^2 \times (p_1+1)p_2}$.

For (C 2), notice that from (B.100),

$$\mathcal{E}_{2,2,4}^b \overset{d_p^b}{\Rightarrow} \mathbb{M}_{3,2}(\tau_{s-1}, \tau_s) = \sum_{l=0}^{\infty}((\mathcal{S}_\dagger' \boldsymbol{S}_\#) \otimes (\mathcal{S}_r \boldsymbol{F}_s^{l+1} \boldsymbol{A}_s^{-1} \boldsymbol{S}_\#)) \int_{\tau_{s-1}}^{\tau_s} (\boldsymbol{D}_\#(\tau) \otimes \boldsymbol{D}_\#(\tau))\, d\boldsymbol{B}_{l+1,\#}(\tau) \tag{B.102}$$

$$\begin{aligned}&= [\boldsymbol{\iota}_{np}' \otimes (\mathcal{S}_r \boldsymbol{F}_s)] \sum_{l=0}^{\infty}((\mathcal{S}_\dagger' \boldsymbol{S}_\#) \otimes (\boldsymbol{F}_s^l \boldsymbol{A}_s^{-1} \boldsymbol{S}_\#)) \int_{\tau_{s-1}}^{\tau_s} (\boldsymbol{D}_\#(\tau) \otimes \boldsymbol{D}_\#(\tau))\, d\boldsymbol{B}_{l+1,\#}(\tau) \\ &= [\boldsymbol{\iota}_{np}' \otimes (\mathcal{S}_r \boldsymbol{F}_s)]\, \mathbb{P}(\tau_{s-1}, \tau_s), \text{ say, (in probability)}\end{aligned} \tag{B.103}$$

Please cite this article as: O. Boldea, A. Cornea-Madeira and A.R. Hall, Bootstrapping structural change tests. Journal of Econometrics (2019), https://doi.org/10.1016/j.jeconom.2019.05.019.

while from (B.92),

$$\mathcal{E}^b_{3,4} \overset{d^b_p}{\Rightarrow} \sum_{l=0}^{\infty}((\mathcal{S}'_\dagger \boldsymbol{S}_\#) \otimes (\mathcal{S}\boldsymbol{F}^l_s \boldsymbol{A}_s^{-1}\boldsymbol{S}_\#)) \int_{\tau_{s-1}}^{\tau_s} (\boldsymbol{D}_\#(\tau) \otimes \boldsymbol{D}_\#(\tau))\, \mathrm{d}\boldsymbol{B}_{l+1,\#}(\tau) = [\boldsymbol{\iota}'_{np} \otimes \mathcal{S}]\mathbb{P}(\tau_{s-1}, \tau_s)$$

in probability. Therefore, they jointly converge. It follows that for $I_i = \tilde{I}_s$,

$$T^{-1/2} \sum_{t\in I_i} \boldsymbol{z}^b_t \boldsymbol{g}^{b\prime}_t \mathcal{S}^b_\dagger = \mathbf{vect}_{i=1:3}(\mathcal{E}^b_i) \overset{d^b_p}{\Rightarrow} \mathbb{M}(\tau_{s-1}, \tau_s) \text{ in probability.}$$

Using exactly the same arguments as in the end of the proof of Lemma 6, $T^{-1/2}\sum_{t\in I_i} \boldsymbol{z}^b_t \boldsymbol{g}^{b\prime}_t \mathcal{S}^b_\dagger \overset{d^b_p}{\Rightarrow} \tilde{\mathbb{M}}_i$ in probability for $I_i \neq \tilde{I}_s$, completing the proof. □

Proof of Lemma 11. As for the proof of Lemma 9, consider the interval $I_i = \tilde{I}_s$. Let $\mathcal{S}^b_\dagger = \mathcal{S}_u$ or $\mathcal{S}^b_\dagger = \hat{\boldsymbol{\beta}}_{\boldsymbol{x},\#}$. We need the asymptotic distribution of $\mathcal{Z}^b_T = T^{-1/2}\sum_{t\in\tilde{I}_s} \boldsymbol{z}_t \boldsymbol{g}^{b\prime}_t \mathcal{S}^b_\dagger$.

$$\mathcal{Z}^b_T = T^{-1/2}\sum_{t\in\tilde{I}_s} \boldsymbol{z}_t \boldsymbol{g}^{b\prime}_t\, \mathcal{S}^b_\dagger = \begin{bmatrix} T^{-1/2}\sum_{t\in\tilde{I}_s} \boldsymbol{g}^{b\prime}_t\, \mathcal{S}^b_\dagger \\ T^{-1/2}\sum_{t\in\tilde{I}_s} \mathcal{S}_{\boldsymbol{r}} \boldsymbol{\xi}_t \boldsymbol{g}^{b\prime}_t\, \mathcal{S}^b_\dagger \\ T^{-1/2}\sum_{t\in\tilde{I}_s} \mathcal{S}\boldsymbol{\xi}_{t-1} \boldsymbol{g}^{b\prime}_t\, \mathcal{S}^b_\dagger \end{bmatrix} \equiv \begin{bmatrix} \mathcal{F}^b_1 \\ \mathcal{F}^b_2 \\ \mathcal{F}^b_3 \end{bmatrix}.$$

Note that $\mathcal{F}^b_1 = \mathcal{E}^b_1$, and $\mathcal{F}^b_2 = \mathcal{E}^b_2$, defined in (B.71) and analyzed in the proof of Lemma 10. Also note that, using (B.103) – as in the proof of Lemma 10 – and replacing as in the proof of Lemma 10, estimated parameters with true values because their difference is asymptotically negligible,

$$\mathcal{F}^b_3 = \mathcal{S}\left\{T^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}'_\dagger(\boldsymbol{g}_t \odot \boldsymbol{v}_t)]\boldsymbol{\xi}_{t-1}\right\} + o^b_p(1)$$

$$\mathcal{E}^b_{2,2} = \mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_s\left\{T^{-1/2}\sum_{t\in\tilde{I}_s}[\mathcal{S}'_\dagger(\boldsymbol{g}_t \odot \boldsymbol{v}_t)]\boldsymbol{\xi}_{t-1}\right\} + o^b_p(1),$$

where $\mathcal{E}^b_{2,2}$ is given in (B.97). Since they involve the same underlying random quantity, just scaled differently ($\mathcal{S}$ versus $\mathcal{S}_{\boldsymbol{r}}\boldsymbol{F}_s$), the desired distribution for $\mathcal{F}^b_3$ follows directly from the analysis of $\mathcal{E}^b_{2,2}$ in Lemma 10. Careful inspection of the proof of Lemma 10 (focusing on the analysis of $\mathcal{E}^b_1$ and $\mathcal{E}^b_2$ only) also shows that $\mathcal{Z}^b_T \overset{d^b_p}{\Rightarrow} \mathbb{M}(\tau_{s-1}, \tau_s)$ in probability, and indicates that this result holds under Assumptions 1–10, without the need for 9′. By a similar argument as for the proof of Lemma 6 in the Supplementary Appendix, Appendix C, when $I_i \neq \tilde{I}_s$, $\mathcal{Z}^b_T \overset{d^b_p}{\Rightarrow} \tilde{\mathbb{M}}_i$ in probability, completing the proof. □

Proof of Theorem 1. We consider only the WR bootstrap; for the WF bootstrap, the results follow in a similar fashion. Let for simplicity $I_i = I_{i,\lambda_k}$. From (8)–(10) and for the Eicker–White estimator $\hat{\boldsymbol{M}}_{(i)}$,

$$Wald_{T\lambda_k} = T\,\hat{\boldsymbol{\beta}}'_{\lambda_k}\boldsymbol{R}'_k\left(\boldsymbol{R}_k\hat{\boldsymbol{V}}_{\lambda_k}\boldsymbol{R}'_k\right)^{-1}\boldsymbol{R}_k\hat{\boldsymbol{\beta}}_{\lambda_k},\ \text{ where } \hat{\boldsymbol{V}}_{\lambda_k} = \mathbf{diag}_{i=1:k+1}(\hat{\boldsymbol{Q}}^{-1}_{(i)}\,\hat{\boldsymbol{M}}_{(i)}\,\hat{\boldsymbol{Q}}^{-1}_{(i)}) \tag{B.104}$$
$$\hat{\boldsymbol{Q}}_{(i)} = T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}'_t\boldsymbol{z}_t\boldsymbol{z}'_t\hat{\boldsymbol{\Upsilon}}_t,\ \text{ and } \hat{\boldsymbol{M}}_{(i)} = \widehat{EW}\left[\,\hat{\boldsymbol{\Upsilon}}'_t\boldsymbol{z}_t(\hat{u}_t + \hat{\boldsymbol{v}}'_t\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)});\ I_i\right].$$

From (19)–(21),

$$Wald^b_{T\lambda_k} = T\,\hat{\boldsymbol{\beta}}^{b\prime}_{\lambda_k}\boldsymbol{R}'_k\left(\boldsymbol{R}_k\hat{\boldsymbol{V}}^b_{\lambda_k}\boldsymbol{R}'_k\right)^{-1}\boldsymbol{R}_k\hat{\boldsymbol{\beta}}^b_{\lambda_k}\ \text{ where } \hat{\boldsymbol{V}}^b_{\lambda_k} = \mathbf{diag}_{i=1:k+1}((\hat{\boldsymbol{Q}}^b_{(i)})^{-1}\hat{\boldsymbol{M}}^b_{(i)}\,(\hat{\boldsymbol{Q}}^b_{(i)})^{-1}) \tag{B.105}$$
$$\hat{\boldsymbol{Q}}^b_{(i)} = T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}^{b\prime}_t\boldsymbol{z}^b_t\boldsymbol{z}^{b\prime}_t\hat{\boldsymbol{\Upsilon}}^b_t,\ \text{ and } \hat{\boldsymbol{M}}^b_{(i)} = \widehat{EW}\left[\,\hat{\boldsymbol{\Upsilon}}^{b\prime}_t\boldsymbol{z}^b_t(\hat{u}^b_t + \hat{\boldsymbol{v}}^{b\prime}_t\hat{\boldsymbol{\beta}}^b_{\boldsymbol{x},(i)});\ I_i\right].$$

From Lemma 2, $\hat{\boldsymbol{Q}}_{(i)} \overset{p}{\to} \mathbb{Q}_i$ and from Lemmas 9 and 11, $\hat{\boldsymbol{Q}}^b_{(i)} \overset{p^b}{\to} \mathbb{Q}_i$.

Now consider $\hat{\boldsymbol{\beta}}_{\lambda_k} \equiv \mathbf{vect}(\hat{\boldsymbol{\beta}}_{i,\lambda_k}) = \mathbf{vect}(\hat{\boldsymbol{\beta}}_{(i)})$ defined on page 4 of this paper. Let $\hat{\boldsymbol{Q}}_{j^*} = T^{-1}\sum_{t\in I^*_j}\boldsymbol{z}_t\boldsymbol{z}'_t$. By Lemma 2, $\hat{\boldsymbol{Q}}_{j^*} \overset{p}{\to} \int_{\pi^0_{j-1}}^{\pi^0_j}\mathbb{Q}_{\boldsymbol{z}}(\tau)\mathrm{d}\tau = \mathbb{Q}_{\boldsymbol{z},j^*}$. Therefore, from the proof of Theorem C1 in the Supplementary Appendix (Appendix C),

$$T^{1/2}(\hat{\boldsymbol{\beta}}_{i,\lambda_k} - \boldsymbol{\beta}^0) = \mathbb{Q}^{-1}_i\boldsymbol{\Upsilon}^{0\prime}_t\left(T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t u_t + T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t\boldsymbol{v}'_t\boldsymbol{\beta}^0_{\boldsymbol{x}}\right.$$

$$- \quad T^{-1}\sum_{t\in I_i} \boldsymbol{z}_t\boldsymbol{z}_t' \left\{ \sum_{j=1}^{h+1} 1_{t\in I_j^*}\mathbb{Q}_{\boldsymbol{z},j^*}^{-1} T^{-1/2}\sum_{t\in I_j^*}\boldsymbol{z}_t\boldsymbol{v}_t'\boldsymbol{\beta}_{\boldsymbol{x}}^0 \right\}\Bigg) + o_p(1). \tag{B.106}$$

From Lemmas 9 and 10, $\hat{\boldsymbol{\Upsilon}}_t^b = \boldsymbol{\Upsilon}_t^0 + o_p^b(1)$. Also, $\hat{\boldsymbol{Q}}_{j^*}^b = T^{-1}\sum_{t\in I_j^*}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime} \xrightarrow{p^b} \mathbb{Q}_{\boldsymbol{z},j^*}$ (in probability) by the proof of Lemma 9, therefore:

$$T^{1/2}(\hat{\boldsymbol{\beta}}_{i,\lambda_k}^b - \boldsymbol{\beta}^0) = \mathbb{Q}_i^{-1}\boldsymbol{\Upsilon}_t^{0\prime}\Bigg(T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t^b(u_t^b + \boldsymbol{v}_t^{b\prime}\boldsymbol{\beta}_{\boldsymbol{x}}^0)$$

$$- \quad T^{-1}\sum_{t\in I_i}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}\left\{\sum_{j=1}^{h+1}1_{t\in I_j^*}\mathbb{Q}_{\boldsymbol{z},j^*}^{-1}T^{-1/2}\sum_{t\in I_j^*}\boldsymbol{z}_t^b\boldsymbol{v}_t^{b\prime}\boldsymbol{\beta}_{\boldsymbol{x}}^0\right\}\Bigg) + o_p^b(1). \tag{B.107}$$

From Lemmas 6 and 10, we have that $T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t^b\boldsymbol{v}_t^{b\prime}\boldsymbol{\beta}_{\boldsymbol{x}}^0 - T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t\boldsymbol{v}_t'\boldsymbol{\beta}_{\boldsymbol{x}}^0 = o_p^b(1)$ and that $T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t^b u_t^b - T^{-1/2}\sum_{t\in I_i}\boldsymbol{z}_t u_t = o_p^b(1)$. Therefore, from (B.106)–(B.107), $\hat{\boldsymbol{\beta}}_{i,\lambda_k}^b - \hat{\boldsymbol{\beta}}_{i,\lambda_k} = o_p^b(1)$ (recall that we denoted $\hat{\boldsymbol{\beta}}_{(i)}^b \equiv \hat{\boldsymbol{\beta}}_{i,\lambda_k}^b$ on page 6 of this paper). Next consider $\hat{\boldsymbol{M}}_{(i)}^b$ for the WR bootstrap under a stable regime $I_i = \tilde{I}_s = [\,[\tau_{s-1}T] + 1, [\tau_s T]\,]$,

$$\begin{aligned}
\hat{\boldsymbol{M}}_{(i)}^b &= T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}^{b\prime}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}\,\hat{\boldsymbol{\Upsilon}}^b(\hat{u}_t^b + \hat{\boldsymbol{v}}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b)^2 = T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}'\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}\,\hat{\boldsymbol{\Upsilon}}(\hat{u}_t^b + \hat{\boldsymbol{v}}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}})^2 + o_p^b(1)\\
&= T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}'\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}\,\hat{\boldsymbol{\Upsilon}}(\hat{u}_t^b)^2 + T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}'\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}\,\hat{\boldsymbol{\Upsilon}}(\hat{\boldsymbol{v}}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}})^2 + 2T^{-1}\sum_{t\in I_i}\hat{\boldsymbol{\Upsilon}}'\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}\,\hat{\boldsymbol{\Upsilon}}\hat{u}_t^b\hat{\boldsymbol{v}}_t^{b\prime}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}} + o_p^b(1)\\
&= \sum_{i=1}^{3}\hat{\boldsymbol{\Upsilon}}'\boldsymbol{S}_i^b\,\hat{\boldsymbol{\Upsilon}} + o_p^b(1),
\end{aligned}$$

where the $o_p^b(1)$ term comes from the fact that the difference $\hat{\boldsymbol{\beta}}_{\boldsymbol{x},(i)}^b - \hat{\boldsymbol{\beta}}_{\boldsymbol{x}}$ is $o_p^b(1)$. Therefore, the terms involving these differences are of lower order than the term after the second equality above.

We have:

$$\begin{aligned}
\boldsymbol{S}_1^b &= T^{-1}\sum_{t\in I_i}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}(u_t^b)^2 + T^{-1}\sum_{t\in I_i}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}(\boldsymbol{w}_t^{b\prime}(\hat{\boldsymbol{\beta}}_{(i)}^b - \hat{\boldsymbol{\beta}}))^2 - 2T^{-1}\sum_{t\in I_i}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}u_t^b(\boldsymbol{w}_t^{b\prime}(\hat{\boldsymbol{\beta}}_{(i)}^b - \hat{\boldsymbol{\beta}}))\\
&= \sum_{i=1}^{3}\boldsymbol{S}_{1,i}^b.
\end{aligned}$$

Letting $\bar{\boldsymbol{v}}_t^b = \textbf{vect}(\boldsymbol{v}_t^b, \boldsymbol{0}_{q_1})$, we have $\boldsymbol{w}_t^b = \hat{\boldsymbol{\Upsilon}}'\boldsymbol{z}_t^b + \bar{\boldsymbol{v}}_t^b$, so

$$\begin{aligned}
\boldsymbol{S}_{1,2}^b &= [\boldsymbol{I}_q \otimes (\hat{\boldsymbol{\beta}}_{(i)}^b - \hat{\boldsymbol{\beta}})]'[T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) \otimes (\boldsymbol{w}_t^b\boldsymbol{w}_t^{b\prime})][\boldsymbol{I}_q \otimes (\hat{\boldsymbol{\beta}}_{(i)}^b - \hat{\boldsymbol{\beta}})]\\
&= [\boldsymbol{I}_q \otimes (\hat{\boldsymbol{\beta}}_{(i)}^b - \hat{\boldsymbol{\beta}})]'\Bigg\{[\boldsymbol{I}_q \otimes \hat{\boldsymbol{\Upsilon}}]'[T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) \otimes (\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime})][\boldsymbol{I}_q \otimes \hat{\boldsymbol{\Upsilon}}]\\
&+ [\boldsymbol{I}_q \otimes \hat{\boldsymbol{\Upsilon}}]'[T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) \otimes (\boldsymbol{z}_t^b\bar{\boldsymbol{v}}_t^{b\prime})] + \left[[\boldsymbol{I}_q \otimes \hat{\boldsymbol{\Upsilon}}]'[T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) \otimes (\boldsymbol{z}_t^b\bar{\boldsymbol{v}}_t^{b\prime})]\right]'\\
&+ T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) \otimes (\bar{\boldsymbol{v}}_t^b\bar{\boldsymbol{v}}_t^{b\prime})\Bigg\}[\boldsymbol{I}_q \otimes (\hat{\boldsymbol{\beta}}_{(i)}^b - \hat{\boldsymbol{\beta}})].
\end{aligned} \tag{B.108}$$

Notice that the terms inside the curly brackets in (B.108) are pre/post-multiplied by $\hat{\boldsymbol{\beta}}_{(i)}^b - \hat{\boldsymbol{\beta}}$ which is $O_p^b(T^{-1/2})$. Therefore it becomes evident that it is sufficient to show that the terms in the curly brackets are $o_p^b(T^{\alpha})$, for any (small) $\alpha > 0$. We first show that $T^{-\alpha}\mathcal{Y}_T^b \equiv T^{-1-\alpha}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) \otimes (\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) = o_p^b(1)$. We have:

$$(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) \otimes (\boldsymbol{z}_t^b\boldsymbol{z}_t^{b\prime}) = \begin{bmatrix} 1 & \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b & \boldsymbol{\xi}_{t-1}^{b\prime}\mathcal{S}' \\ \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b & \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b\prime}\mathcal{S}_{\boldsymbol{r}}' & \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_{t-1}^{b\prime}\mathcal{S}' \\ \mathcal{S}\boldsymbol{\xi}_{t-1}^b & \mathcal{S}\boldsymbol{\xi}_{t-1}^b\boldsymbol{\xi}_t^{b\prime}\mathcal{S}_{\boldsymbol{r}}' & \mathcal{S}\boldsymbol{\xi}_{t-1}^b\boldsymbol{\xi}_{t-1}^{b\prime}\mathcal{S}' \end{bmatrix} \otimes \begin{bmatrix} 1 & \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b & \boldsymbol{\xi}_{t-1}^{b\prime}\mathcal{S}' \\ \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b & \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_t^{b\prime}\mathcal{S}_{\boldsymbol{r}}' & \mathcal{S}_{\boldsymbol{r}}\boldsymbol{\xi}_t^b\boldsymbol{\xi}_{t-1}^{b\prime}\mathcal{S}' \\ \mathcal{S}\boldsymbol{\xi}_{t-1}^b & \mathcal{S}\boldsymbol{\xi}_{t-1}^b\boldsymbol{\xi}_t^{b\prime}\mathcal{S}_{\boldsymbol{r}}' & \mathcal{S}\boldsymbol{\xi}_{t-1}^b\boldsymbol{\xi}_{t-1}^{b\prime}\mathcal{S}' \end{bmatrix}, \tag{B.109}$$

where $\xi_t^b = \hat{\boldsymbol{F}}_s^{t-[\tau_{s-1}T]} \xi_{[\tau_{s-1}T]}^b + \left(\sum_{l=0}^{t-[\tau_{s-1}T]-1} \hat{\boldsymbol{F}}_s^l\right) \hat{\boldsymbol{\mu}}_s + \sum_{l=0}^{t-[\tau_{s-1}T]-1} \hat{\boldsymbol{F}}_s^l \boldsymbol{\eta}_{t-l}^b$ (given in (B.44)). To show that $T^{-\alpha}\mathcal{Y}_T^b = o_p^b(1)$ we can see from (B.109) that it is enough to show that $T^{-\alpha}\mathcal{Y}_{1,T}^b \equiv T^{-1-\alpha}\sum_{t\in I_i}(\xi_t^b \xi_t^{b'}) \otimes (\xi_t^b \xi_t^{b'}) = o_p^b(1)$ because the other terms follow by similar reasoning. This follows if we can show:

$$T^{-\alpha}\mathcal{Y}_{2,T}^b \equiv T^{-1-\alpha}\sum_{t\in I_i}\sum_{l,l^*,\kappa,\kappa^*=0}^{\tilde{t}-1}(\hat{\boldsymbol{F}}_s^l \otimes \hat{\boldsymbol{F}}_s^{l^*})((\boldsymbol{\eta}_{t-l}^b \boldsymbol{\eta}_{t-l^*}^{b'}) \otimes (\boldsymbol{\eta}_{t-\kappa}^b \boldsymbol{\eta}_{t-\kappa^*}^{b'}))(\hat{\boldsymbol{F}}_s^\kappa \otimes \hat{\boldsymbol{F}}_s^{\kappa^*})' = o_p^b(1).$$

Since $\hat{\boldsymbol{F}}_s - \boldsymbol{F}_s = o_p(1)$, we can write:

$$\tilde{\mathcal{Y}}_{2,T}^b = \sum_{l,l^*,\kappa,\kappa^*=1}^{\Delta\tau_s T-1}(\boldsymbol{F}_s^l \otimes \boldsymbol{F}_s^{l^*})\left(T^{-1}\sum_{t\in I_i}(\boldsymbol{\eta}_{t-l}^b \boldsymbol{\eta}_{t-l^*}^{b'}) \otimes (\boldsymbol{\eta}_{t-\kappa}^b \boldsymbol{\eta}_{t-\kappa^*}^{b'})\right)(\boldsymbol{F}_s^\kappa \otimes \boldsymbol{F}_s^{\kappa^*})' + o_p^b(1),$$

in probability. We now show that $\tilde{\mathcal{Y}}_{2,T}^b = o_p^b(T^\alpha)$ in probability, for any (small) $\alpha > 0$. Because $\boldsymbol{F}_s^l$ is exponentially decaying, it therefore suffices to show that

$$T^{-1-\alpha}\sum_{t\in I_i}(\boldsymbol{\eta}_{t-l}^b \boldsymbol{\eta}_{t-l^*}^{b'}) \otimes (\boldsymbol{\eta}_{t-\kappa}^b \boldsymbol{\eta}_{t-\kappa^*}^{b'}) = o_p^b(1),$$

in probability, or, because $\boldsymbol{\eta}_t^b = \hat{\boldsymbol{A}}_s^{-1}\boldsymbol{g}_t^b$, that:

$$T^{-1-\alpha}\sum_{t\in I_i}(\boldsymbol{g}_{t-l}^b \boldsymbol{g}_{t-l^*}^{b'}) \otimes (\boldsymbol{g}_{t-\kappa}^b \boldsymbol{g}_{t-\kappa^*}^{b'}) = o_p^b(1).$$

Note $(\boldsymbol{g}_{t-l}^b \boldsymbol{g}_{t-l^*}^{b'}) \otimes (\boldsymbol{g}_{t-\kappa}^b \boldsymbol{g}_{t-\kappa^*}^{b'}) = [(\hat{\boldsymbol{g}}_{t-l}\hat{\boldsymbol{g}}_{t-l^*}') \otimes (\hat{\boldsymbol{g}}_{t-\kappa}\hat{\boldsymbol{g}}_{t-\kappa^*}')] \odot [(\boldsymbol{v}_{t-l}\boldsymbol{v}_{t-l^*}') \otimes (\boldsymbol{v}_{t-\kappa}\boldsymbol{v}_{t-\kappa^*}')]$, and $\mathrm{E}^b[(\boldsymbol{v}_{t-l}\boldsymbol{v}_{t-l^*}') \otimes (\boldsymbol{v}_{t-\kappa}\boldsymbol{v}_{t-\kappa^*}')]$ has elements that are uniformly bounded, where the largest moment involved is the fourth moment of ν_t, also uniformly bounded by Assumption 10. Therefore, we analyze only the element of $T^{-1-\alpha}\sum_{t\in I_i}(\boldsymbol{g}_{t-l}^b \boldsymbol{g}_{t-l^*}^{b'}) \otimes (\boldsymbol{g}_{t-\kappa}^b \boldsymbol{g}_{t-\kappa^*}^{b'})$ pertaining to that moment, namely $T^{-1-\alpha}\sum_{t\in I_i}\hat{u}_{t-l}^4 \nu_{t-l}^4$. Because the distinction between $l = 0$ or $l > 0$ is irrelevant for the end result, we show that $\mathcal{Y}_{u,T}^b \equiv T^{-1-\alpha}\sum_{t\in I_i}\hat{u}_t^4\nu_t^4 = o_p^b(1)$ in probability, which then completes the proof, as the rest of the terms can be analyzed in a similar fashion. Letting $\mathcal{Y}_{u,T} = T^{-1-\alpha}\sum_{t\in I_i}\hat{u}_t^4$ and $\mathrm{E}^b(\nu_t^4) = c \le \bar{c}$ (by Assumption 10(ii)), we have, by Markov's inequality, conditional on the data:

$$P^b(\mathcal{Y}_{u,T}^b > \eta) = P^b(T^{-1-\alpha}\sum_{t\in I_i}\hat{u}_t^4\nu_t^4 > \eta) \le \eta^{-1}T^{-1-\alpha}\sum_{t\in I_i}\hat{u}_t^4\,\mathrm{E}^b(\nu_t^4) = \eta^{-1}cT^{-1-\alpha}\sum_{t\in I_i}\hat{u}_t^4 = \eta^{-1}c\mathcal{Y}_{u,T}.$$

We now show that $\mathcal{Y}_{u,T} = o_p(1)$ (uniformly over I_i), which from above implies that $P^b(\mathcal{Y}_{u,T}^b > \eta) \xrightarrow{p} 0$ or, equivalently, that $\mathcal{Y}_{u,T}^b = o_p^b(1)$ in probability, therefore completing the proof. Note that $\hat{u}_t = u_t + \boldsymbol{w}_t'\hat{\boldsymbol{b}}$, where $\hat{\boldsymbol{b}} = \hat{\boldsymbol{\beta}}_{(i)} - \boldsymbol{\beta}^0 = O_p(T^{-1/2}) = o_p(1)$. Therefore,

$$\mathcal{Y}_{u,T} = \sum_{j=1}^5 \mathcal{Z}_{j,T}, \quad \text{where}$$

$$\mathcal{Z}_{1,T} = T^{-1-\alpha}\sum_{t\in I_i}u_t^4, \qquad \mathcal{Z}_{2,T} = 4T^{-1-\alpha}\sum_{t\in I_i}u_t^3\boldsymbol{w}_t'\hat{\boldsymbol{b}}, \qquad \mathcal{Z}_{3,T} = 6T^{-1-\alpha}\sum_{t\in I_i}u_t^2(\boldsymbol{w}_t'\hat{\boldsymbol{b}})^2$$

$$\mathcal{Z}_{4,T} = 4T^{-1-\alpha}\sum_{t\in I_i}u_t(\boldsymbol{w}_t'\hat{\boldsymbol{b}})^3, \qquad \mathcal{Z}_{5,T} = T^{-1-\alpha}\sum_{t\in I_i}(\boldsymbol{w}_t'\hat{\boldsymbol{b}})^4.$$

We now show that each of these terms is $o_p(1)$. First, $\sup_t \mathrm{E}(u_t^4) < M$ by Assumption 9. Therefore, by the Markov inequality, for some $\eta > 0$,

$$P(\mathcal{Z}_{1,T} > \eta) < \eta^{-1}\,\mathrm{E}(\mathcal{Z}_{1,T}) = \eta^{-1}T^{-1-\alpha}\sum_{t\in I_i}\mathrm{E}(u_t^4) \le \eta^{-1}T^{-\alpha}\sup_t \mathrm{E}(u_t^4) < \eta^{-1}T^{-\alpha}M \to 0,$$

so $\mathcal{Z}_{1,T} = o_p(1)$.

Next, consider $\mathcal{Z}_{2,T}$. First, $\|\mathcal{Z}_{2,T}\| \le 4\left(T^{-1-\alpha}\sum_{t\in I_i}\|u_t^3\boldsymbol{w}_t\|\right)\|\hat{\boldsymbol{b}}\| = \left(T^{-1-\alpha}\sum_{t\in I_i}\|u_t^3\boldsymbol{w}_t\|\right)o_p(1)$. So to show $\mathcal{Z}_{2,T} = o_p(1)$, it suffices to show that $T^{-1-\alpha}\sum_{t\in I_i}\|u_t^3\boldsymbol{w}_t\| = o_p(1)$. By Markov's inequality,

$$P\left(T^{-1-\alpha}\sum_{t\in I_i}\|u_t^3\boldsymbol{w}_t\| > \eta\right) < \eta^{-1}T^{-\alpha}\sup_t \mathrm{E}\,\|u_t^3\boldsymbol{w}_t\|. \tag{B.110}$$

Hölder inequality states that $\mathrm{E}\,\|\boldsymbol{ab}\| \leq (\mathrm{E}\,\|\boldsymbol{a}\|^p)^{1/p}(\mathrm{E}\,\|\boldsymbol{b}\|^q)^{1/q}$, where $p, q \geq 1$ and $1/p + 1/q = 1$. Let $\boldsymbol{a} = u_t^3$, $\boldsymbol{b} = \boldsymbol{w}_t$, $p = 4/3$ and $q = 4$. Then: $\mathrm{E}\,\|u_t^3\boldsymbol{w}_t\| \leq (\mathrm{E}|u_t|^4)^{3/4}(\mathrm{E}\,\|\boldsymbol{w}_t\|^4)^{1/4}$. Let M be a universal constant. Then, by Assumption 9, $\mathrm{E}|u_t|^4 < M$. Similarly, $\boldsymbol{w}_t = \boldsymbol{\Upsilon}^{0'}\boldsymbol{z}_t + \bar{\boldsymbol{v}}_t$, where $\bar{\boldsymbol{v}}_t = \mathbf{vect}(\boldsymbol{v}_t, \mathbf{0}_{q_1})$, and by the triangle inequality, $\mathrm{E}(\|\boldsymbol{w}_t\|^4)^{1/4} \leq (\mathrm{E}\,\|\boldsymbol{\Upsilon}^{0'}\boldsymbol{z}_t\|^4)^{1/4} + (\mathrm{E}\,\|\bar{\boldsymbol{v}}_t\|^4)^{1/4} \leq (\|\boldsymbol{\Upsilon}^0\|^4\,\mathrm{E}\,\|\boldsymbol{z}_t\|^4)^{1/4} + (\mathrm{E}\,\|\boldsymbol{v}_t\|^4)^{1/4} < M$, where the latter follows by Assumption 9 and by the proof of Theorem C1 in the Supplementary Appendix (Appendix C), where we showed that $\sup_t \mathrm{E}\,\|\boldsymbol{z}_t\|^4 < M$. It follows that $\sup_t \mathrm{E}\,\|u_t^3\boldsymbol{w}_t\| < M$, and substituting into (B.110), it follows that $T^{-1-\alpha}\sum_{t\in I_i}\|u_t^3\boldsymbol{w}_t\| = o_p(1)$, implying that $\mathcal{Z}_{2,T} = o_p(1)$.

Next, $\|\mathcal{Z}_{3,T}\| \leq 6\left(T^{-1-\alpha}\sum_{t\in I_i}\|u_t\boldsymbol{w}_t\|^2\right)\|\hat{\boldsymbol{b}}\|^2 = 6\left(T^{-1-\alpha}\sum_{t\in I_i}\|u_t\boldsymbol{w}_t\|^2\right)o_p(1)$. So it suffices to show $T^{-1-\alpha}\sum_{t\in I_i}\|u_t\boldsymbol{w}_t\|^2 = o_p(1)$. By Hölder's inequality with $p = q = 2$, we have $\mathrm{E}\,\|u_t\boldsymbol{w}_t\|^2 = \mathrm{E}|u_t|^2\|\boldsymbol{w}_t\|^2 \leq (\mathrm{E}|u_t|^4)^{1/2}(\mathrm{E}\,\|\boldsymbol{w}_t\|^4)^{1/2} < \infty$, which by Markov's inequality applied as in (B.110), shows that $T^{-1-\alpha}\sum_{t\in I_i}\|u_t\boldsymbol{w}_t\|^2 = o_p(1)$ and therefore that $\mathcal{Z}_{3,T} = o_p(1)$.

Now consider $\mathcal{Z}_{4,T}$. By similar arguments as above, we have $\|\mathcal{Z}_{4,T}\| \leq 4\left(T^{-1-\alpha}\sum_{t\in I_i}|u_t|\,\|\boldsymbol{w}_t\|^3\right)\|\hat{\boldsymbol{b}}\|^3 = 4\left(T^{-1-\alpha}\sum_{t\in I_i}|u_t|\,\|\boldsymbol{w}_t\|^3\right)o_p(1)$. By Hölder's inequality with $p = 4$ and $q = 4/3$, we have: $\mathrm{E}|u_t|\,\|\boldsymbol{w}_t\|^3 \leq (\mathrm{E}|u_t|^4)^{1/4}(\mathrm{E}\,\|\boldsymbol{w}_t\|^4)^{3/4}$. Since $\mathrm{E}\,\|\boldsymbol{w}_t\|^4 < M$ as shown above, and $\mathrm{E}|u_t|^4 < M$, it follows by Markov's inequality applied as in (B.110) that $T^{-1-\alpha}\sum_{t\in I_i}|u_t|\|\boldsymbol{w}_t\|^3 = o_p(1)$ and therefore that $\mathcal{Z}_{4,T} = o_p(1)$.

By similar arguments as above, $\|\mathcal{Z}_{5,T}\| \leq \left(T^{-1-\alpha}\sum_{t\in I_i}\|\boldsymbol{w}_t\|^4\right)\|\hat{\boldsymbol{b}}\|^4 = \left(T^{-1-\alpha}\sum_{t\in I_i}|\|\boldsymbol{w}_t\|^4\right)o_p(1)$. Since $\sup_t \mathrm{E}\,\|\boldsymbol{w}_t\|^4 < \infty$ as shown above, by Markov's inequality applied as in (B.110), it follows that $T^{-1-\alpha}\sum_{t\in I_i}\|\boldsymbol{w}_t\|^4 = o_p(1)$ and therefore that $\mathcal{Z}_{5,T} = o_p(1)$.

Putting the results for $\mathcal{Z}_{j,T}$ together, $j = 1, \ldots, 5$, it follows that $\mathcal{Y}_{u,T} = o_p(1)$, completing the proof. Therefore it follows that $T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}) \otimes (\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}) = o_p^b(T^\alpha)$ for any $\alpha > 0$.

Similarly we can show that $T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}) \otimes (\boldsymbol{z}_t^b\bar{\boldsymbol{v}}_t^{b'}) = o_p^b(T^\alpha)$ and $T^{-1}\sum_{t\in I_i}(\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}) \otimes (\boldsymbol{v}_t^b\bar{\boldsymbol{v}}_t^{b'}) = o_p^b(T^\alpha)$. It follows that $S_{1,2}^b = o_p^b(1)$, and similarly $S_{1,3}^b = o_p^b(1)$ implying $S_1^b = T^{-1}\sum_{t\in I_i}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}(u_t^b)^2 + o_p^b(1)$. Along the same lines, it can be shown that $\boldsymbol{S}_2^b = T^{-1}\sum_{I_i}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}(\boldsymbol{v}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}^b)^2 + o_p^b(1)$ and that $\boldsymbol{S}_3^b = T^{-1}\sum_{I_i}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}u_t^b\boldsymbol{v}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}^b + o_p^b(1)$. Hence, $\hat{\boldsymbol{M}}_{(i)}^b = T^{-1}\sum_{I_i}\hat{\boldsymbol{\Upsilon}}^{b'}\boldsymbol{z}_t^b\boldsymbol{z}_t^{b'}\hat{\boldsymbol{\Upsilon}}^b(u_t^b + \boldsymbol{v}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}^b)^2 + o_p^b(1)$ (uniformly in $\boldsymbol{\lambda}_k$). By similar arguments, it can also be shown for the WF bootstrap that $\hat{\boldsymbol{M}}_{(i)}^b = T^{-1}\sum_{I_i}\hat{\boldsymbol{\Upsilon}}^{b'}\boldsymbol{z}_t\boldsymbol{z}_t'\hat{\boldsymbol{\Upsilon}}^b(u_t^b + \boldsymbol{v}_t^{b'}\hat{\boldsymbol{\beta}}_{\boldsymbol{x}}^b)^2 + o_p^b(1)$ (uniformly in $\boldsymbol{\lambda}_k$).

Because $\hat{\boldsymbol{M}}_{(i)}$ and $\hat{\boldsymbol{M}}_{(i)}^b$ estimate the same part of the variance of $T^{1/2}(\hat{\boldsymbol{\beta}}_{i,\lambda_k} - \boldsymbol{\beta}_{\boldsymbol{x}}^0)$, and $T^{1/2}(\hat{\boldsymbol{\beta}}_{i,\lambda_k}^b - \boldsymbol{\beta}_{\boldsymbol{x}}^0)$ respectively, from Lemmas 2, 6, 9 and 10, it follows that $\hat{\boldsymbol{M}}_{(i)}^b - \hat{\boldsymbol{M}}_{(i)} = o_p^b(1)$ for the WR and WF bootstraps. Putting these results together, $\sup_{c\in\mathbb{R}}\left|P^b\left(\sup\text{-}Wald_T^b \leq c\right) - P(\sup\text{-}Wald_T \leq c)\right| \xrightarrow{p} 0$ as $T \to \infty$. □

Proof of Theorem 2. Inspecting the alternative representation of the $\sup\text{-}Wald_T(\ell+1|\ell)$ in the proof of Theorem C2 in the Supplementary Appendix (Appendix C), and defining the same representation for $\sup\text{-}Wald_T^b(\ell+1|\ell)$, the desired result follows using the same steps as in the proof of Theorem 1. □

Appendix C. Supplementary Appendix

Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.05.019.

References

Andrews, D.W.K., 1988. Laws of large numbers for dependent non-identically distributed random variables. Econometric Theory 4, 458–467.
Andrews, D.W.K., 1993. Tests for parameter instability and structural change with unknown change point. Econometrica 61, 821–856.
Bai, J., Perron, P., 1998. Estimating and testing linear models with multiple structural changes. Econometrica 66, 47–78.
Beran, R., 1986. Discussion: Jackknife, bootstrap and other resampling methods in regression analysis. Ann. Statist. 14, 1295–1298.
Boswijk, H.P., Cavaliere, G., Rahbek, A., Taylor, A.M.R., 2016. Inference on co-integration parameters in heteroskedastic vector autoregressions. J. Econometrics 192, 64–85.
Brady, R.R., 2008. Structural breaks and consumer credit: Is consumption smoothing finally a reality? J. Macroeconom. 30, 1246–1268.
Brown, B.M., 1971. Martingale central limit theorems. Ann. Math. Stat. 42, 59–66.
Bunzel, H., Enders, W., 2010. The taylor rule and 'opportunistic' monetary policy. J. Money Credit Bank. 42, 931–949.
Chang, S.Y., Perron, P., 2018. A comparison of alternative methods to construct confidence intervals for the estimate of a break date in linear regression models. Econometric Rev. 37, 577–601.
Davidson, R., 2016. Computing, the bootstrap and economics. Canad. J. Econ. 48, 1195–1214.
Davidson, R., Flachaire, E., 2008. The wild bootstrap, tamed at last. J. Econometrics 146, 162–169.
Davidson, R., MacKinnon, J.G., 1993. Estimation and Inference in Econometrics. Oxford University Press.
Davidson, R., MacKinnon, J., 1998. Graphical methods for investigating the size and power of hypothesis tests. Manch. Sch. 66, 1–26.
Davidson, R., MacKinnon, J.G., 2010. Wild bootstrap tests for IV regression. J. Bus. Econom. Statist. 28, 128–144.
Eicker, F., 1967. Limit theorems for regressions with unequal and dependent errors. In: Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, vol. 1. University of California Press, Berkeley, pp. 59–82.

Georgiev, I., Harvey, D.I., Leybourne, S.J., Taylor, A.M.R., 2018. Testing for parameter instability in predictive regression models. J. Econometrics 204, 101–118.
Gine, E., Zinn, J., 1990. Bootstrapping general empirical measures. Ann. Probab. 18, 851–869.
Gonçalves, S., Kilian, L., 2004. Bootstrapping autoregressions with conditional heteroskedasticity of unknown form. J. Econometrics 123, 89–120.
Hall, A.R., Han, S., Boldea, O., 2012. Inference regarding multiple structural changes in linear models with endogenous regressors. J. Econometrics 170, 281–302.
Hall, A.R., Osborn, D.R., Sakkas, N.D., 2013. Inference about structural breaks using information criteria. Manch. Sch. 81, 54–81.
Hall, A.R., Osborn, D.R., Sakkas, N.D., 2015. Structural break inference using information criteria in models estimated by two stage least squares. J. Time Series Anal. 36, 741–762.
Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press, Princeton, NJ, USA.
Hansen, B.E., 1992. Convergence to stochastic integrals for dependent heterogeneous processes. Econometric Theory 8, 489–500.
Hansen, B.E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64, 413–430.
Hansen, B.E., 2000. Testing for structural change in conditional models. J. Econometrics 97, 93–115.
Kim, C.-J., Manopimoke, P., Nelson, C.R., 2014. Trend inflation and the nature of structural breaks in the New Keynesian Phillips Curve. J. Money Credit Bank. 46, 253–266.
Kleibergen, F., Mavroeidis, S., 2009. Weak instrument robust tests in GMM and the New Keynesian Phillips Curve. J. Bus. Econom. Statist. 27, 293–311.
Liu, R.Y., 1988. Bootstrap procedure under some non-i.i.d. models. Ann. Statist. 16, 1696–1708.
Mammen, E., 1993. Bootstrap and wild bootstrap for high dimensional linear models. Ann. Statist. 21, 255–285.
Perron, P., Yamamoto, Y., 2014. A note on estimating and testing for multiple structural changes in models with endogenous regressors via 2SLS. Econometric Theory 30, 491–507.
Perron, P., Yamamoto, Y., 2015. Using OLS to estimate and test for structural changes in models with endogenous regressors. J. Appl. Econometrics 30, 119–144.
Qian, J., Su, L., 2014. Structural change estimation in time series with endogenous regressors. Economics Letters 125 (415–421).
Qu, Z., Perron, P., 2007. Estimating and testing structural changes in multivariate regressions. Econometrica 75, 459–502.
White, H., 1980. A heteroskedasticity-consistent covariance matrix estimator and a direct test of heteroskedasticity. Econometrica 48, 817–838.
Wooldridge, J.M., 1994. Introductory Econometrics: A Modern Approach. Thomson-Southwestern Press, Mason, OH, USA.
Wu, C.F.J., 1986. Jackknife, bootstrap and the other resampling methods in regression analysis. Ann. Statist. 14, 1261–1295.
Zhang, C., Osborn, D., Kim, D., 2008. The New Keynesian Phillips Curve: from sticky inflation to sticky prices. J. Money Credit Bank. 40, 667–699.

Please cite this article as: O. Boldea, A. Cornea-Madeira and A.R. Hall, Bootstrapping structural change tests. Journal of Econometrics (2019), https://doi.org/10.1016/j.jeconom.2019.05.019.